UNIVERSITÀ DEGLI STUDI DI MILANO

Scienze biomediche cliniche e sperimentali

Dipartimento di Biotecnologie Mediche e Medicina Traslazionale

Dottorato in Biotecnologie Applicate alle Scienze Mediche (XXV ciclo)

Coordinatore Prof. A. M. Gianni

WNT-dependent regenerative function is induced in leukemia-initiating AC133^{bright} cells

Francesca Lazzaroni

Prof. Alessandro Beghini

Anno Accademico 2011-2012

UNIVERSITÀ DEGLI STUDI DI MILANO

Scienze biomediche cliniche e sperimentali

Dipartimento di Biotecnologie Mediche e Medicina Traslazionale

Dottorato in Biotecnologie Applicate alle Scienze Mediche (XXV ciclo)

Coordinatore Prof. A. M. Gianni

WNT-dependent regenerative function is induced in leukemia-initiating AC133^{bright} cells

BIO/13

Francesca Lazzaroni

Prof. Alessandro Beghini

Prof. Enrico Ginelli

Anno Accademico 2011-2012

To my family.

ABSTRACT

The Cancer Stem Cell model supported the notion that leukemia was initiated and maintained in vivo by a small fraction of leukemia-initiating cells (LICs). Previous studies have suggested the involvement of Wnt signaling pathway in Acute Myeloid Leukemia (AML) by the ability to sustain the development of LICs. A novel hematopoietic stem and progenitor cell marker, monoclonal antibody AC133, recognizes the $CD34^{bright}$ $CD38^{-}$ subset of human acute myeloid leukemia cells, suggesting that it may be an early marker for the LICs. During the first part of my phD program we previously evaluated the ability of leukemic $AC133^{+}$ fraction, to perform engraftment following to xenotransplantation in immunodeficient mouse model $Rag2^{-/-}\gamma c^{-/-}$. The results showed that the surface marker AC133 is able to enrich for the cell fraction that contains the LICs. In consideration of our previously reported data, derived from the expression profiling analysis performed in normal (n=10) and leukemic (n=33) human long-term reconstituting AC133+ cells, we revealed that the ligand-dependent Wnt signaling is induced in AML through a diffuse expression and release of WNT10B, a hematopoietic stem cells regenerative-associated molecule. In situ detection performed on bone marrow biopsies of AML patients, showed the activation of the Wnt pathway, through the concomitant presence of the ligand WNT10B and of the active dephosphorylated β-catenin form, suggesting an autocrine / paracrine-type ligand-dependent activation mechanism. In consideration of the link between hematopoietic regeneration and developmental signaling, we transplanted primary AC133+ AML A46 cells into developing zebrafish. This biosensor model revealed the formation of ectopic structures by activation of dorsal organizer markers

that act downstream of the Wnt pathway. These results suggested that the misappropriating Wnt associated functions can promote pathological stem cell-like regeneration responsiveness. The analyses performed in situ retained information on the cellular localization, enabling determination of the activity status of individual cells and allowing the tumor environment view. Taking this issue into consideration, during the second part of my phD program, I set up the application of a new in situ method for localized detection and genotyping of individual transcripts directly in cells and tissues. The mRNA *in situ* detection technique is based on padlock probes ligation and target priming rolling circle amplification allowing the single nucleotide resolution in heterogenous tissues. The mRNA *in situ* detection performed on bone marrow biopsies derived from AML patients, showed a diffuse localization pattern of WNT10B molecule in the tissue. Conversely, only the $AC133^{bright}$ cell population shows the Wnt signaling activation signature represented by the cytoplasmatic accumulation and nuclear translocation of the active form of β-catenin. In spite of this, we previously evidenced that the regenerative function of WNT signaling pathway is defined by the up-regulation of WNT10B, WNT10A, WNT2B and WNT6 loci, we identified the WNT10B as a major locus associated with the regenerative function and over-expressed by all AML patients. By the molecular evaluation of the WNT10B transcript, we isolated the $WNT10B^{IVS1}$ aberrant splicing variant, that identify Non Core-Binding Factor Leukemia (NCBFL) class and whose potential role is discussed. Moreover, we demonstrate that the function of "leukemia stem cell", present in the cell population enriched for the marker $AC133^{bright}$, is strictly related to regenerative function associated with WNT signaling, defining the key role of WNT10B ligand as a specific molecular marker for leuchemogenesis. This thesis defines the new suitable approaches to characterize the leukemia-initiating cells

(LICs) and suggests the role of WNT10B as a new possible target for AML.

LIST OF PAPERS

1. Beghini A, Corlazzoli F, Del Giacco L, Re M, Lazzaroni F, Brioschi M, Valentini G, Ferrazzi F, Ghilardi A, Righi M, Turrini M, Mignardi M , Cesana C, Bronte V, Nilsson M, Morra E and Cairoli R (2012). Regeneration-associated WNT Signaling Is Activated in Long-term Reconstituting AC133^{bright} Acute Myeloid Leukemia Cells. *Neoplasia,* **14** (12), 1236-1248.

2. Cairoli R, Beghini A, Turrini M, Bertani G, Nadali G, Rodeghiero F, Castagnola F, Lazzaroni F, Nichelatti M, Ferrara F, Pizzolo G, Pogliani E, Rossi G, Martinelli G, and Morra E. Old and new prognostic factors in acute myeloid leukemia with deranged core-binding factor beta. (Submitted).

3. Lazzaroni F and Beghini A. PTPN6 protein tyrosine phosphatase, non-receptor type 6[Homo sapiens]. *Atlas Genet Cytogenet Oncol Haematol*. (Accepted).

Uppsala University- Rudbeck Laboratory

I set up and performed the mRNA *in situ* detection method at the Department of Immunology, Genetics and Pathology, Molecular tools, Rudbecklaboratoriet, Uppsala University (Uppsala, Sweden) in collaboration with *Ulf Landegren, Mats Nilsson, Ola Soderberg* and *Marco Mignardi*.

At the Uppsala University I carried out the image analysis, using CellProfiler and CellProfiler Analyst.

LIST OF ABBREVIATIONS

ABC	Active β- catenin
ACTB	β-actin mRNA
ALL	Acute Limphocytic leukemia
AML	Acute Myeloid Leukemia
APC	Adenomatous Polyposis Coli
BM	Bone Marrow
BP	Base Pair
CBF	Core Binding Factor
cDNA	Complementary DNA
CK1	Casein Kinase 1
CLL	Chronic Lymphocytic Leukemia
CML	Chronic Myelogenous Leukemia
CR	Complete Remission
CRD	Cysteine Rich Domain
CSC	Cancer stem cell
CTNNB1	β -catenin
CV	Coefficient of variation
DAPI	4,6-diamidino-2-phenylindole

DKK1	Dickkopf-1
DNA	Deoxyribonucleic Acid
DVL/DSH	Dishevelled
FAB	French-American-British
FLT3	FMS-related tyrosine kinase 3
FFPE	Formalin-Fixed Paraffin-Embedded
FZD	Frizzled
GAPDH	Glyceraldehyde-3-Phosphate Dehydrogenase
GMP	Granulocyte-Macrophage Progenitor
GSK3	Glycogen Synthase Kinase 3
HeLa	Human epithelial Cervical Cancer Cell Line
HSC	Hematopoietic Stem Cell
ITD	Internal Tandem Duplications
LCM	Laser-Capture Microdissection
LEF1	Lymphoid Enhancer-binding factor 1
LIC	Leukemia-Initiating Cell
LNA	Locked Nucleic Acid
LOD	Limit of Detection
LRP	Low-Density Lipoprotein Receptor-Related protein
LSC	Leukemic Stem Cell

LSC	Leukemic Stem Cell
LT-HSC	Long Term Hematopoietic Stem Cell
MAb	Monoclonal Antibody
MIP	Maximum Intensity Projection
mRNA	Messenger RNA
NES	Nuclear Export Signal
NF-κB	Nuclear factor kappa B
NLS	Nuclear Localization Signal
NOD/SCID	Non-Obese Diabetic mice with Severe Combined Immunodeficiency Disease
NPM1	Nucleophosmin
OLA	Oligonucleotide Ligation Assay
OS	Overall Survival
PB	Peripheral Blood
PROM1	Prominin-1
PSF	Point Spread Function
PTKs	Protein Tyrosine Kinases
RCA	Rolling Circle Amplification

RCP	Rolling Circle Product
RHD	Runt-Homology Domain
RNA	Ribonucleic Acid
RT-qPCR	Reverse Transcription Quantitative Polymerase Chain Reaction
RTK	Receptor Tyrosine Kinases
SFRP	Secreted Frizzled Related Protein
ST-HSC	Short-Term Hematopoietic Stem Cell
TCF	Transcription factor
Tm	Melting Temperature
WHO	World Health Organization
WIF1	Wnt inhibitory factor 1
WNT	Wingless-type

TABLE OF CONTENTS

3. Results

1. INTRODUCTION

1.1 ACUTE MYELOID LEUKEMIA “MOLECULAR”; THE STATE OF THE ART

The term leukemia (from ancient greek: *leukós haima*, “white blood”) was originally described by Rudolf Virchow (Virchow, 1856). The first case to be described in medical literature dates to 1827 when a French Doctor Velpeau described a patient who developed the symptoms such as fever, weakness, urinary stones, enlargement of the liver and spleen. He noticed that the blood of this patient had a consistency like “gruel”. Rudolf Virchow was the first to describe the abnormal amount of white blood cells in patients and called the disease “leukemia” in 1856. In 1889, Wilhelm Ebstein presented the name “acute leukemia” to the leukemias that are rapidly progressive and fatal. Finally, in 1900 Naegeli divided leukemias into myeloid and lymphocytic based on cell specificity. Nowadays leukemias are divided into several major groups: acute myeloid leukemia (AML), chronic myelogenous leukemia (CML), acute lymphocytic leukemia (ALL) and chronic lymphocytic leukemia (CLL). Molecularly, leukemias are a heterogeneous disease entity with different rearrangements and dysregulations of genes with important functions in cellular growth, differentiation, and death (apoptosis). At the cellular level, acute leukemias are characterized by an expansion of immature white blood cells (blasts) in the bone marrow and blood, where a lack of mature blood cells together with a suppression of normal residual hematopoiesis, eventually leads toanemia, thromobocytopenia, and leukopenia, which result in fatigue, bleeding, and infections. Deregulated hematopoiesis ultimately leading to leukemia is a consequence of acquisition of genetic and epigenetic changes in blood stem and progenitor cells, which otherwise produce the huge numbers of mature, red and white blood cells. These

so-called blast cells have lost their ability to differentiate and response to normal regulation of proliferation and survival, and progressively displace normal blood cells in the bone marrow (BM) with the result of fatal infections, bleedings, and infiltration of other organs *(Tenen, 2003).* Depending on which blood cell lineage is affected, leukemia is subdivided into lymphoid or myeloid leukemia. Furthermore, chronic and acute leukemia are distinguished by the amount of blast cells in the (BM). The occurrence of more than 20 percent blasts in the BM is considered as acute leukemia. AML is diagnosed as a result of a complete blood count showing decreased numbers of red blood cells (anemia), platelets (thrombocytopenia) and neutrophil granulocytes (neutropenia) whereas the total white blood count are most commonly increased (leukocytosis) due to an accumulation of leukemic blast cells. For identification of blasts an examination of peripheral blood analysis can be carried out, but definitive diagnosis requires a BM sample, which is analyzed morphologically by microscopy and flow cytometry to diagnose the presence of leukemic blasts and to differentiate from other types of leukemia. In addition, cytogenetic analysis is carried out to identify chromosomal abnormalities and translocations which can have prognostic significance. The classification of AML is complex due to the diversity of cytology, clinical prognosis, and genetic diversity. Two main classifications for AML have been defined over the years. The earliest was designed by a French-American-British cooperative group (FAB) in 1976, and was based on cytological and cytochemical criteria assessing lineage of origin of leukemic cells and stage of differentiation block. The AMLs were divided into 8 groups from M0 to M7 with later groups presenting more differentiated cells than the earlier ones. A new leukemia classification was introduced by the World Health Organization (WHO) *(Jaffe, 2009; Vardiman, 2010; Vardiman et al., 2002; Vardiman et al., 2009)*. This

classification is not limited to cytological or cytochemical characteristics of the cells, but was designed to include all currently available information, including morphology, cytochemistry, immunophenotype, genetics, and clinical features in order to define clinically relevant diseases *(Vardiman et al., 2009).* The latest WHO classification includes a number of entities that are defined by the common genomic rearrangements rather than other cellular characteristics. Two provisional entities were included that are defined based on mutations in CCAAT/enhancer binding protein α (CEBPA) and nucleophosmin (NPM1) genes. Although FLT3 mutations did not define a separate category, screening was recommended due to their profound effect on prognosis. Separate entities include myeloid neoplasms with therapy-related changes, as they have significantly different complete response, overall survival and response to treatment. Another interesting change was the introduction of a separate group for myeloid proliferations related to Down syndrome that are often associated with mutated GATA-1. As for not otherwise specified AMLs (AML, NOS), they generally incorporate FAB classification based on stage of maturation of myeloid leukemia cells.

The molecular pathogenesis of AML

A key characteristic of hematopoietic stem cells (HSC) is the ability to self-renew. According to Rizo et al., the processes of self-renewal and differentiation, are based on the relationship between stem cell and microenvironment *(Rizo et al., 2006)*. The characteristics of quiescence, self-renewal and differentiation of hematopoietic stem cell (HSC) are regulated by intrinsic mechanisms, such as chromatin remodeling and extrinsic mechanisms, mainly driven by the microenvironment of the niche, which

guides the fairies of the stem cell. Several genes and signaling pathways control the fine balance between self-renewal and differentiation in HSC and potentially also in leukemic stem cells. Besides pathways such as Wnt signaling, Hedgehog signaling and Notch signaling, transcription factors (FoxOs) and cell fate determinants may also play a role in stem cells. While some of these pathways seem to be dispensable for maintenance of adult HSC, there may be a distinct requirement in leukemia stem cells for leukemic self-renewal. Genetic changes involving tyrosine kinases, important regulators in basic cellular processes such as proliferation, survival and differentiation, are frequent in leukemia. Deregulation of signaling pathways, including those involving tyrosine kinase signaling, have been connected to leukemogenesis and progression of leukemic disease, thereby making them attractive targets in antileukemic drug discovery *(Chase & Cross, 2006; Stapnes et al., 2009).* The molecular pathogenesis of acute myeloid leukemia (AML) has not yet been completely defined. Recurrent chromosomal structural variations (e.g., t(15;17), t(8;21), inv(16), t(9;21), t(9;11), del5, del7) are established diagnostic and prognostic markers, suggesting that acquired genetic abnormalities play an essential role in leukemogenesis *(Betz and Hess, 2010).* However, nearly 50% of AML cases have a normal karyotype (NK), and many of these lack recurrent structural abnormalities, even with high density comparative genomic hybridization (CGH) or SNP arrays *(Bullinger et al., 2010; Suela et al., 2007; Walter et al., 2009)*. Targeted sequencing efforts have identified several mutations that carry diagnostic and prognostic information, including mutations in FLT3, NPM1, KIT, CEBPA, and TET2 *(Bacher et al., 2009; Stirewalt & Radich, 2003).* The advent of massively parallel sequencing enabled the discovery of recurrent mutations in DNMT3A *(Ley et al., 2010; Yamashita et al., 2010)* and IDH1 *(Mardis et al., 2009).* Despite these efforts, more than 25% of AML patients carry no mutations in the

known leukemia-associated genes *(Shen et al., 2011).* Furthermore, defining the molecular consequences of recurring mutations (e.g., whether a mutation is an initiating or a cooperating event) has been challenging.

1.1.1 DANCING TO THE TUNE OF DANGER SIGNALING

Historically, the signaling cascades have been described as pathways, based on biochemical and genetic evidence placing some signaling intermediates "downstream" of others. One of the first pathways to be described from the activation of RTKs to the alteration of transcription in the nucleus was the Ras-MAPK pathway, where the small GTPase Ras is activated via several intermediate steps. Activated, GTP-bound Ras in turn activates a cascade of dual specificity and serine/threonine kinases that finally activate a group of mitogen activated protein kinases, the MAPK. These kinases phosphorylate important transcriptional regulators of cell cycle progression and thus induce a proliferative response of the target cell. Another signaling pathway that recently received much attention is the PI3- Kinase/Akt pathway that can be activated directly by binding and phosphorylation of a regulatory subunit of PI3-Kinase to an activated RTK or indirectly by several mechanisms involving the activation of Src-family kinases or the binding of adaptor proteins to activated kinases. PI3-Kinase generates a small lipid second messenger that activates several serin/threonine-kinases like Akt or the mammalian target of Rapamycin mTOR. These proteins are involved in the regulation of apoptosis, proliferation and, through a direct effect on protein translation, in the control of cell size. Alterations of signal transduction increase proliferation and survival potential of hematopoietic stem and progenitor cells, often through constitutive

activation of protein tyrosine kinases (PTKs) but normally do not affect cellular differentiation. Examples include gain-of-function mutations of ABL, JAK2, FLT3, PDGFR, KIT, activating mutation of the RAS family members, as well as loss-of-function alterations of NF1 or PTPN11. Activating JAK2 mutations are found in chronic myeloproliferative neoplasms and in more than 95% of patients with polycythemia vera (PV). Activating mutations of FLT3, KIT and RAS are present in more than 50% of AML patients. FLT3 is the most commonly mutated gene in approximately one third of AML. In 20%–25% of cases of AML, internal tandem duplications (ITD) in the juxtamembrane domain of FLT3 results in loss of an autoinhibitory domain leading to constitutive activation. Other FLT3 mutations consist of substitutions, small deletions, or insertions within the activation loop of the second kinase domain found in 5% to 10% of AML patients. The overall consequence of these FLT3 mutations is ligand- independent receptor dimerization and/or constitutive activation of its tyrosine kinase activity, leading to uncontrolled activation of several downstream signaling pathways, such as RAS, MAPK, and STAT5 pathways. The RAS/MAPK signal transduction pathway is a critical regulator of proliferation and survival of hematopoietic progenitors. Leukemic blasts from about 40% of AML patients showed constitutive activation of RAS signaling. However, expression of most class I mutations in murine bone marrow generally leads to a lethal myeloproliferative disease (MPD), but not acute leukemia. In addition, most of these disorders such as the FLT3-ITD induced MPD are not transplantable into secondary recipient mice suggesting that the class I mutations do not confer self-renewal potential to the transformed cells. JAK/STAT pathway is also involved in oncogenesis, as JAK2 mutations are commonly found in CML and MDS *(Reuther, 2008).* JAK is an intracellular kinase that is activated by different receptors and in turn

phosphorylates STAT proteins. The phosphorylated STAT TFs dimerize, are transported into nucleus and become active as regulators of transcription. Among different genes regulated by STAT is anti-apoptotic BCL family member BCL-xL. STAT5 also promotes cell survival by inducing PIM1 expression *(Steelman et al., 2008).* Additionally, it is shown that activation of STAT5 by mutant FLT3 increases proliferation and alters differentiation of hematopoietic stem cells *(Moore et al., 2007).* Activation of STAT5 may be enhanced by ERK phosphorylation, indicating a possible cross-talk between JAK/STAT and MAPK pathways *(David et al., 1995; Winston & Hunter, 1995).* The most commonly mutated genes found in human AMLs, as well as many of the genes constantly upregulated in AMLs, are involved in the described pathways. C-KIT RTK, which is almost ubiquitously expressed on AML cells, results in direct activation of PI3K/AKT pathway and activates STAT3 *(Ning et al., 2001a).* RAS GTPases are also directly activated by c-KIT receptor, leading to upregulated RAF/MEK/ERK cascade *(Scholl et al., 2008a).* FLT3-ITD and other activating mutations of FLT3 activate AKT phosphorylation, RAS/RAF/MEK/ERK and STAT5 in AML cells *(Brandts et al., 2005; Hayakawa et al., 2000; Kim et al., 2005; Mizuki et al., 2000; Spiekermann et al., 2003).* JAK2, as described above, is a member of JAK/STAT oncogenic pathway. Commonly mutated in AML RAS proteins are downstream targets of type III RTKs (including c-KIT and FLT3) and activators of RAF/MEK/ERF pathway. Finally, a major player in PI3K/AKT pathway AKT is overexpressed in a large human AML *(Tamburini et al., 2007).* Many of these players of leukemic transformation are targets for specific inhibitors therapy, although cross-talks between the pathways and autoregulatory feedback loops make a task of cancer prevention by signaling inhibition challenging *(Breitenbuecher et al., 2009; Chu & Small, 2009; Huang et al., 2003; Huang & Houghton, 2001).* Mutations in components of the RTK

signaling pathway are the most common mutations found in CBF leukemias *(Haferlach et al., 2010).* The KIT and FLT3 genes play crucial roles in proliferation and survival of hematopoietic progenitors *(Ikeda et al., 1991; Pollard et al., 2010; Wang et al., 1989; Yokota et al., 1997).* The oncogenic mutations that activate the FLT3 receptor are found in 5% to 7% of CBF AML cases, including the ITD and TKD mutations *(Boissel et al., 2006b; Care et al., 2003; Schnittger et al., 2002; Shih et al., 2008),* and have an unfavorable prognosis *(Döhner et al., 2010; Mrózek et al., 2007).* The activating mutations in KIT (the most common are D816V and N822K) are found in approximately 25% of CBF AML and are rare in other AMLs *(Beghini et al., 2000a; Cairoli et al., 2006; Boissel et al., 2006b; Care et al., 2003; Goemans et al., 2005; Paschka, 2008; Shih et al., 2008).* It was also proven by using mouse models that both FLT3-ITD *(Kim et al., 2008; Schessl et al., 2005b)* and mutant KIT *(Wang et al., 2011a)* cooperate with CBF fusions in leukemia development in mice. The CBF AMLs can also present oncogenic mutations in NRAS and KRAS, but rarely in HRAS *(Neubauer et al., 2008).* Rat sarcoma (RAS) genes encode membrane anchored GTPases that are involved in multiple signaling pathways and were shown to have oncogenic activity *(McCubrey et al., 2007).* In inv(16) AML cases around 30% harbored constitutively active mutated RAS genes with the most commonly mutated NRAS. The incidence of RAS mutations in t(8;21) AML cases is around 10% *(Boissel et al., 2006b; Care et al., 2003; Goemans et al., 2005; Shih et al., 2008).* It is noteworthy that KIT, FLT3 and RAS mutations in inv(16) and t(8;21) human AML cases were mutually exclusive, thus supporting the idea that they act as class I mutations in the "two hit" model for leukemia development *(Care et al., 2003; Shih et al., 2008)*. Approximately 8 to 10% of t(8;21) AML cases show oncogenic mutations in the tyrosine kinase Janus kinase 2 (JAK2) gene, frequently at residue V617F *(Dohner et al., 2006; Schnittger et al., 2007a),*

but this occurs less frequently in inv(16) AML. Interestingly, these mutations in JAK2 are often found in CML and MDS, but are rare in non-CBF AMLs *(Illmer et al., 2007).*

1.1.2 MAKING THE CUT: THE CORE BINDING FACTOR STORY

The Core Binding Factor (CBF) is a master regulator of hematopoietic development, maintenance and differentiation, in embryonic and adult hematopoiesis. CBF is a heterodimeric TF that consists of subunits called α and β *(Kamachi et al., 1990).* In mammals, the α-subunit is encoded by three alternative genes with high level of homology (over 90% identity at the amino acid level) called Runt-related transcription factors RUNX1, RUNX2 and RUNX3. The nomenclature of the CBF factors has been confusing due to its history until the year 2004 when researchers from these fields working with the CBF proteins proposed a unified nomenclature in which the mammalian homologs of the Drosophila gene runt would be officially called RUNX1, RUNX2, and RUNX3, while the cofactor would remain with the name given in the developmental field as CBFB *(van Wijnen et al., 2004).* Interaction between RUNX1 and CBFβ subunits is critical for CBF function *(Wang et al., 1996b).* The RUNX protein binds to the DNA consensus sequence YGYGGTY in promoters, enhancers and silencer regions *(Wang & Speck, 1992).* This binding is dependent on heterodimerization with CBFβ, as CBFβ increases RUNX affinity to DNA over five-fold *(Golling et al., 1996; Wang et al., 1993)*, and protects it from proteasomal degradation *(Huang et al., 2001).* All three RUNX proteins share similar conserved domains with defined functions. The highly conserved Runt-homology domain (RHD) between amino acids 50 and 177 is responsible for

binding to DNA and CBFβ *(Golling et al., 1996).* The RHD also participates in RUNX binding a variety of cofactors. For example, the v-Ets avian erythroblastosis virus E26 oncogene homolog 1 (ETS1) TF binds to the RHD as RUNX1 recruits it to the T-cell receptor *a* (TCRa) enhancer in T cells *(Giese et al., 1995).*

Core Binding Factor Leukemias

Approximately 20-25% of AML cases are associated with genomic alterations involving CBFB or RUNX1 genes *(Look, 1997).* Specifically, 12% of cases of AML have chromosomal inversion inv(16)(p13q22), disrupting the CBFβ gene and fusing it inframe with the second half of the Myosin heavy chain 11(MYH11) gene that encodes the Smooth muscle myosin heavy chain (SMMHC) protein. The resulting gene is called CBFβ -MYH11, and its encoded fusion protein is called CBFβ-SMMHC *(Liu et al., 1993).* In 12% of AML cases the translocation t(8;2l)(q22;q22) results in the creation of the RUNX1-ETO fusion gene that fuses RUNX1 in exon 5 with Eight Twenty One (ETO, also known as RUNX1T1) gene starting from exon 2 *(Rowley, 1973).* In addition to these rearrangements, AML samples can also harbor point mutations in RUNX1 that create a protein with dominant negative or hypomorphic function *(Schnittger et al., 2011; Silva et al., 2009).* By FAB classification, 68% of cases associated with the CBFβ-MYH11 fusion gene fall into M4Eo group, while 20% are M4 and 12% are M2 *(Marcucci et al., 2005; Schnittger et al., 2007a).* There are rare cases of myelodysplastic syndrome and chronic myeloid leukemia crisis also associated with CBFβ-MYH11 expression *(Tirado et al., 2010),* but these are temporary conditions that lead to development of AML. More than 90% of the M4Eo cases carry an

abnormality that leads to CBFB-MYH11 expression, whether it is inv(16), t(16;16) or del(16)(q22) *(Hernandez et al., 2000; Pirc-Danoewinata et al., 2000).* However, the fusion transcript CBFβ-MYH11 has been detected in all M4Eo cases tested, suggesting the presence of micro-rearrangements in some cases. Since the presence of the CBFβ-MYH11 fusion in AML samples is unique at the molecular and clinical levels, the WHO classification defined it as a separate AML subtype *(Vardiman, 2010).* The CBFβ-MYH11-expressing AML subtype is considered to have a favorable prognosis, with CR rate around 90% and OS rate around 50% *(Appelbaum et al., 2006; Delaunay et al., 2003; Marcucci et al., 2005).* The CR and OS rates were lower in older patients, patients with complex genotype and high white blood cell (WBC) counts *(Appelbaum et al., 2006).* AML cases in M4Eo group are characterized by myeloblastic and/or monoblastic infiltration into BM, elevated monocytic counts in peripheral blood (PB), and the presence of atypical eosinophils in both BM and PB. The eosinophil population usually accounts for 5% or more of white blood cells and carries the inv (16), being therefore part of the leukemic population and not a secondary change in AML (*Haferlach et al., 1996).* Taken together, the data from human patients suggest that CBFβ- MYH11 causes development of AML with partial abnormal differentiation into monocytic/myeloid lineage with small populations keeping expression of early hematopoietic stem/progenitor markers. A majority (91%) of AML cases associated with the fusion gene RUNX1-ETO are in the FAB subtype M2, while only 6% are in the less differentiated subtype M1 and 3% in the more differentiated subtype M4 *(Marcucci et al., 2005).* The WHO classification defines a separate subtype based on the presence of t(8;21) translocation expressing RUNX1-ETO. The AML subtype is considered to have a favorable prognosis, but the OS rate is lower in the RUNX1-ETO-related AMLs than in CBFβ-MYH11-related AMLs *(Appelbaum et al.,*

2006; Marcucci et al., 2005), being around 45%. The risk factors were similar to those for CBFβ-MYH11 subtype: age, WBC count, complex genotype with additional genomic alterations. Race was an additional factor that was not found significant in CBFβ- MYH11 AML subtype, but was significant for OS of patients with RUNX1-ETO AMLs *(Marcucci et al., 2005)* being around45%.

1.2 LEUKEMIA INITIATING CELL (LIC): GETTING THE STEM OF AML

The "stem cell" concept was first proposed by Till and McCulloch following their pioneering studies of the blood system regeneration in vivo. Ten days after transplanting limiting number of syngenic bone marrow (BM) cells into recipient mice, they observed cellular colonies that formed in the spleens of recipient mice. Analysis of these colonies revealed that a very small sub-population of the donor BM cells possessed two remarkable properties: (1) the ability to generate multiple types of myeloerythroid cells, and (2) the ability to self- replicate *(Becker et al., 1963; Siminovitch et al., 1963; Till & McCulloch, 1961).* These findings introduced the two defining criteria of stem cells i.e. multi-potency and self-renewal. Hematopoietic Stem Cells (HSCs) are the only cells within the hematopoietic system that possess the potential for both multi-potency and self-renewal. In the case of HSC, multi-potency is the ability to differentiate into all functional blood cells, while self-renewal is the ability to give rise to identical daughter HSCs without differentiation. In 1988, the initial prospective purification of hematopoietic stem cells from mouse BM was achieved utilizing the relatively new technologies of multi-color fluorescence-activated cell sorting and monoclonal antibodies *(Spangrude et al., 1988).* The resultant population of enriched mouse HSCs had a surface marker phenotype of Thy-1^{low} Lin^{-} Sca-1+, and represented approximately 0.05% of the mouse adult BM cells. Spangrude and colleagues demonstrated that these

were the only cells in mouse BM capable of transferring long-term reconstitution of the entire hematopoietic system when transplanted into lethally irradiated mice *(Spangrude et al., 1988).* A reductionist approach by Uchida et al showed that Thy1.1low, but not Thy1.1high or Thy1.1– cells could give rise to donor-derived long-term multilineage reconstitution of irradiated hosts; this was true of Sca-1+, but not Sca-1– cells, and of Lin–, but not Lin+ cells *(Uchida & Weissman, 1992).* Since these initial studies, mouse HSCs have been more extensively purified by identifying and then utilizing additional cell-surface markers to distinguish them from other cells in BM; these included, but were not exclusively single cells that could self-renew and give long-term multilineage maturation *(Uchida & Weissman, 1992; Ikuta and Weissman, 1992)*. In 1994 the population isolated by Spangrude et al was shown to include at least three multipotent populations: Long-Term (LT)-HSC, Short-Term (ST)-HSC, and Multi-Potent Progenitor (MPP), a cell population that has lost the self-renewal capacity of HSC *(Morrison & Weissman, 1994).* In 1996, HSCs from adult mouse BM were sufficiently enriched to conduct single-cell transplantation experiments, and these studies revealed that one in three CD34–/low c-Kit+ Sca-1+ Lin– cells showed myelolymphoid long-term reconstitution in lethally irradiated recipients after a single cell transplant *(Osawa et al., 1996).* Despite the fact that hematopoietic tissues contain both stem and progenitor cells, rapid and sustained engraftment of syngenic and even of H2 incompatible allogenic hosts can only be achieved with HSC, the time to engraftment depending on the number of HSC transplanted *(Uchida et al., 1997).* While the phenotypic and functional properties of HSCs have been extensively characterized *(Morrison et al., 1995; Weissman, 2000),* a fundamental question that remains is how self- renewal is regulated. In most cases, combinations of growth factors that can induce extensive proliferation are unable to

prevent differentiation of HSCs in long-term cultures. Although progress has been made in identifying conditions that maintain HSC activity in culture for a brief period of time *(Miller & Eaves, 1997),* it has proven exceedingly difficult to identify combinations of growth factors that cause significant expansion in culture in the number of progenitors with transplantable HSC activity. The first one to identify the leukemic stem cell (LSC) were Bonnet and Dick in 1997 *(Bonnet & Dick, 1997).* Until then, there was much conflict in reasoning about the identity of the initiating cell in leukemia *(Wang & Dick, 2005).* It was not clear whether the stochastic model or the hierarchical model was applicable, and within the hierarchical model it is still not clear whether the normal SC is targeted and transformed and/or a progenitor cell initiates leukemia. By using non-obese diabetic mice with severe combined immunodeficiency disease (NOD/SCID mice) it was shown that cells that are able to initiate leukemia (the SCID leukemia-initiating cells or SL-ICs) have the ability to proliferate, self-renew, and differentiate via asymmetrical division. The cells identified by Bonnet as the leukemia initiating cells reside in the CD34+CD38- immunophenotypic compartment *(Bonnet & Dick, 1997).* Cancer stem cells (CSCs) constitute a subpopulation of malignant cells capable of self renewal and differentiation *(Al-Hajj et al., 2004; Jordan et al., 2007; Jordan, 2006; Pardal et al., 2003; Reya et al., 2001; Rossi et al., 2008; Wang & Dick, 2005).* It is now half a century since bone-marrow reconstitution experiments, following lethal irradiation in mice, first indicated the existence of the haematopoietic stem cell (HSC) a cell first postulated to exist by Artur Pappenheim as early as 1917. Although this cell population is still not fully characterized, its discovery awakened the field of stem-cell biology. Elegant experiments by Philip Fialkow and colleagues, in which they used patterns of inactivation in X-linked genes, had previously shown that leukaemias such as chronic myelogenous leukaemia

(CML) and AML *(Huntly & Gilliland, 2005),* together with preleukaemic diseases such as the myeloproliferative disorders, are clonal in origin. In a series of seminal experiments in 1997 by investigators based at the University of Toronto, where James Till and Ernest McCulloch had first shown the radioprotective effects of mouse bone marrow, the LSC for AMLwas first identified. The first descriptions of LSC (Leukemia Stem Cell) in human AML by Lapidot et al. and Bonnet and Dick from John Dick's laboratory identified a subpopulation of CD34+ CD38− human AML cells that were able to serially transplant leukemia in a mouse xenograft model. In contrast, the more committed progenitors expressing CD38, lacked this potential. These reports demonstrated that LSC were rare, however the frequency of these LSC varied greatly between different AML samples, ranging from 1 in 10^4 to 10^7 cells. In this xenograft model, LSC were not limited to their ability to cause leukemia, but also gave rise to progeny that lost leukemia initiating activity, leading to the hypothesis that AML is arranged in a hierarchy with the LSC at the apex and the more "differentiated" blasts representing the bulk, non-transplantable tumor population. This deterministic model differs from the original stochastic model based on observations that only rare cells within tumors randomly possessed or acquired the ability to form colonies and transplant disease. Until then, there was much conflict in reasoning about the identity of the initiating cell in leukemia. It was not clear whether the stochastic model or the hierarchical model was applicable, and within the hierarchical model it is still not clear whether the normal SC is targeted and transformed and/or a progenitor cell initiates leukemia. By using non- obese diabetic mice with severe combined immunodeficiency disease (NOD/SCID mice) it was shown that cells that are able to initiate leukemia (the SCID leukemia-initiating cells or SL-ICs) have the ability to proliferate, self-renew, and differentiate via asymmetrical division. The cells

identified by Bonnet as the leukemia initiating cells reside in the CD34+CD38- immunophenotypic compartment. LSC, like their normal HSC counterparts, possess a range of characteristics that enable their long-term survival and some of these also facilitate their escape from the cytotoxic effects of chemotherapy. For example, LSC express the p- glycoprotein multidrug resistance efflux pump, ABCB1 (also known as MDR1) that can remove potentially cytotoxic chemotherapeutic agents from the cell. By reducing cytotoxic stress, LSC may become a reservoir for the selection of mutants that are resistant to targeted or conventional therapy. LSC are characterized by limitless self-renewal and experimental evidence implicates the primacy of key, developmentally conserved self- renewal pathways such as Bmi-1, Wnt/β-catenin and Hedgehog in this process. Increased expression of Hox genes, such as HoxA9, has been implicated in the pathogenesis of MLL- AF9-induced AML. However, not surprisingly, this dependence may be context and oncogene dependent as evidenced by the importance of Hedgehog pathway signaling in CML LSC but not in MLL-AF9-induced AML. LSC may also evade apoptosis by up-regulation of the pro-survival factor NF-kB or evasion of programmed cell death mediated by Fas/CD95 interactions. The development of the NOD/SCID-leukaemia model allowed the separation of leukaemia cells into subpopulations that could be evaluated for engraftment and serial-transplantation potential. Cells derived from seven patients representing each subtype of AML (according to French-American-British classification standards), were separated into CD34+CD38+ and CD34+CD38- fractions. These subpopulations were then injected intravenously into sublethally irradiated NOD/SCID mice that received regular injections of human cytokines. The mice were assessed at 4–8 weeks for engraftment, based on human-specific DNA sequences. Human cells from the bone marrow of transplant recipients were then isolated, based on

the expression of human form of CD45 and transplanted into secondary recipients. These experiments showed that the capacity to transfer human AML to recipient mice resided exclusively within the CD34+CD38- cell fraction. Furthermore, these cells had the same capacity to induce all subtypes of AML except for M3 acute promyelocytic leukaemia. The leukaemias that developed in the secondary recipients closely resembled the human cancer, demonstrating that LSCs have long-term self-renewal capabilities and also determine the stage of the differentiation block during leukaemogenesis. Based on these findings, the authors proposed a hierarchical organization for AML that is similar to normal haematopoiesis. In this model, the LSC (or SL-IC, as the authors named it) is responsible for both self-renewal and the production of clonogenic leukaemic progenitors that have proliferative capacity, but not the capacity of self-renewal. Based on similarities in the organization of the respective systems and phenotypic similarities, they also proposed that the HSC was the most likely target for transformation into an LSC *(Bonnet & Dick, 1997).* Experiments performed by Craig Jordan and Donna Hogge have refined the immunophenotype of the LSC in AML to be CD34+CD38-CD90-interleukin 3 receptor (IL-3R)+CD71- human leukocyte antigen (HLA)-DR-CD117-. In order to identify the LICs by immunophenotype, Gibbs et al. transplanted into irradiated wild-type mice the purified hematopoietic stem/progenitor cells, which express c-kit (Lin-kit+), kit+ cells coexpressing lymphoid markers (Lyn+kit+) or myeloid markers (Gr1+kit+), and mature myeloid cells (Gr1+kit^{low}) from HoxA9-Meis1 (H9M) mice. All transplanted cell fractions generated leukemia in the recipients except the Gr1+kit cells. These findings show that the LICs represent malignant hematopoietic progenitors that are immunophenotypically heterogeneous. Furthermore, each immunophenotype of LIC from primary leukemias could recapitulate all the LIC immunophenotypes in secondary

recipients. Because the three distinct LIC populations seemed to share developmental lineage potential, Gibbs et al. suggested that they may possess common cell signaling and genetic properties. In human leukemias, it has also been difficult to find exact immunophenotypic markers that characterize all LICs. For example, Goardon et al. and Sarry et al. found that LICs represented hematopoietic progenitors with several different immunophenotypes *(Goardon et al., 2011; Sarry et al., 2011).* Goardon et al. examined a large group of primary AML patients and found two different progenitor-like LICs: one multipotent population and another representing granulocyte-macrophage progenitors (GMPs). Their relationship was hierarchical because the multipotent population could generate the GMP-like LICs, but not the other way around. Another recent study also reported progenitor-like LICs in primary AML with different immunophenotypes, some without expression of lineage markers (Lin-CD38-) and others expressing lineage markers CD38 or CD45RA *(Sarry et al., 2011)* (Figure 1.1).

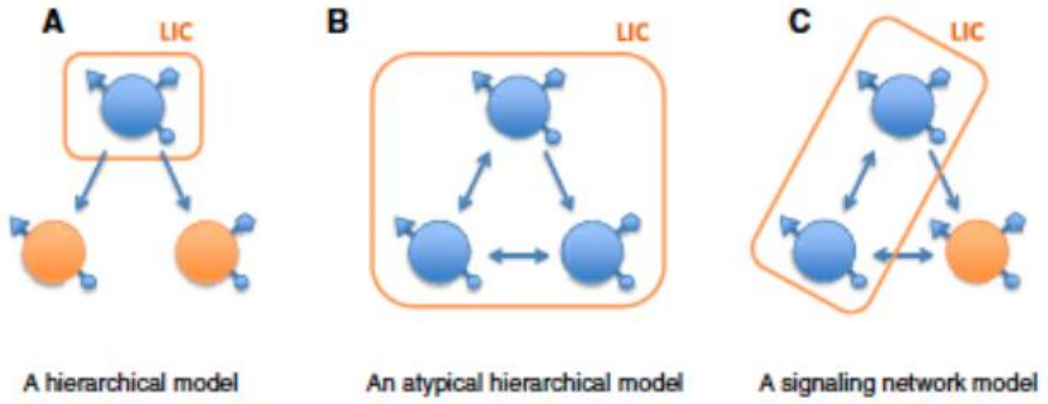


Figure 1.1. Different models for Leukemia-Initiating Cells. (A) A classical hierarchical model: leukemia-initiating ability is restricted to a small fraction of hematopoietic-stem-cell-like cells, that generate progeny cells that lack tumor-initiating ability. (B) An atypical hierarchical model: various types of hematopoietic-stem-cell-like or progenitor-like cells comprise leukemia-initiating cells (LICs), and these cells can independently generate other LICs with distinct immunophenotypes and therefore exhibit an atypical hierarchy. Hence it is not possible to identify LICs with a single immunophenotype. (C) A signaling network model: LICs with different immunophenotypes exhibit activation of common signaling pathways. Cell bodies in blue are LICs, while the orange cells do not have leukemia-initiating activity. Figure from *Miharada & Karlsson, 2012*.

In recent study, the capacity of various strains of immunodeficient mice [NOD/it-SCID and NOD/Shi-SCID, NOD/it-SCID/IL-2Rγnull (NSG), and NOD/Shi-SCID/IL-2Rγnull (NOG)] to act as recipients for human HSCs was compared, showing that both NSG and NOG improved the engraftment in peripheral tissues compared to the parental strains of NOD/SCID mice. NSG mice provide greater human engraftment in bone marrow *(McDermott et al., 2010)*. The NOD/SCID repopulating assay demonstrated that human

HSCs are present in the CD34+CD38− fraction of human hematopoietic cells *(Bhatia et al., 1997; Civin et al., 1996)*. Using NOD/SCID mice strains with enhanced immunosuppression as recipients, it was shown that the CD34+/CD38+ cell fraction also possesses some repopulating activity *(Hogan et al., 2002; McKenzie et al., 2006)*. However, CD34+/CD38+ cells possess only a short-term SCID-repopulating activity, while the long-term repopulating activity is limited to the CD34+/CD38− cell population *(Hogan et al., 2002)*. It is important to mention that several studies have characterized a rare SCID-repopulating population observed at the level of CD34−Lin− cells: these cells, like CD34+/CD38− cells, possess a long-term repopulating capacity *(Wang et al., 2003; Kitamura et al., 2007; Kitamura et al. 2010)*. Importantly, CD34− HSCs are able to generate in vivo CD34+ HSCs *(Kitamura et al., 2010)*. According to their capacity of repopulating hematopoiesis, the hematopoietic stem cell pool can be subdivided into three groups: short-term HSCs, capable of generating clones of differentiating cells for only 4–6 weeks; intermediate-term HSCs, capable of sustaining a differentiating cell progeny for 6–8 months before becoming exinct; and long-term HSCs, capable of maintaining hematopoiesis undefinitely *(Benviste et al., 2010)*. Initial studies have shown that the leukemia-initiating cells, as assayed in the NOD/SCID assay model, are detected only within the CD34+/CD38− fraction of the majority of AML samples, but none were found in any other fraction, including CD34+/CD38+ fraction *(Lapidot et al., 1994; Bonnet & Dick, 1997; Ailles et al., 1999)*. Since AML-CFUs are contained only in the CD34+/CD38+ fraction, these studies provided a direct experimental proof that the AML clone is organized as a hierarchy that originates from leukemia-initiating cells, which generate a cell progeny initially composed by AML-CFU and then by leukemic blasts arrested at various stages of differentiation. Serial transplantation studies in NOD/SCID

mice provided evidence that the pool of AML-initiating cells, like normal HSCs, is composed of distinct leukemia stem cells that differ in their repopulating leukemia capacity and self-renewal capacity *(Hope et al., 2004)*. However, some recent studies have shown that leukemia stem cells are present also in the CD34+/CD38+ fraction. In fact, it was shown that, in a significant proportion of AMLs, cells contained in the CD34+/CD38+ fraction are capable of initiating and maintaining the leukemic process when grafted to NOD/SCID mice *(Taussig et al., 2008)*. Interestingly, in some AMLs in which the CD34+/CD38− fraction was unable to initiate leukemia after injection in NOD/SCID mice, the CD34+/CD38+ fraction was able to do it *(Taussig et al., 2008)*. The discrepancy between these observations and previous studies relies in the observation that the anti-CD38 monoclonal antibody used for cell fractionation studies has a marked negative effect on the engraftment of AML repopulating cells in NOD/SCID mice *(Taussig et al., 2008)*. Very recently, it was shown that, in a significant proportion of AMLs, leukemia-initiating cells are observed within the CD34− fraction. These studies were based on the analysis of SCID leukemia-initiating cells in a group of patients bearing nucleophosmin mutations using the most immunodeficient SCID mice available. These AMLs are classified as a separate entity and are characterized by a low CD34 expression. In half of these AMLs, the CD34− fraction contained leukemia-initiating cells, while the CD34+ fraction gave rise to normal multilineage hematopoiesis. In the remaining half of the patients, leukemia-initiating cells are observed among both CD34+ and CD34− AML cells *(Taussig et al., 2010)*. The AMLs with exclusively CD34− leukemia initiating cells may have arisen either from CD34+ stem/ progenitor cells that have lost CD34 expression through the leukemia-transforming events or from CD34− stem cells. These observations further reinforce the concept that the membrane phenotype of leukemia-

initiating cells is heterogeneous in various AMLs. While cell surface characteristics may not define LICs, recent studies have identified several cell surface markers that predict the ability of LICs to grow and metastasize. For example, antibodies against CD44 expressed on LICs could markedly deplete human LIC engrafted in NOD/SCID mice. Similarly, in HOXA10-Meis1 induced murine leukemia, prevention of secondary leukemias was achieved upon treatment of the recipients with antibodies against CD44 *(Jin et al., 2006; Quere´ et al., 2011).* Increased protein levels of CD44 were discovered by screening the LICs by proteomics, indicating that expression profile analysis may not always be sufficient to find key molecules that are dysregulated in LICs *(Quere´ et al., 2011).* Interestingly, Hertweck et al., evidenced that CD44 plays two pivotal roles in early hematopoiesis: (1) mediation of the interaction of the progenitor cells with their respective niche in the bone marrow (BM), (2) stimulation of cell proliferation and differentiation by regulation of local cytokine secretion *(Hertweck et al., 2011).* CD44 signaling plays a pivotal role in acute myeloid leukemia (AML), depicting three different putative points of attack: differentiation arrest, bone marrow niche dependency of leukemia-initiating cells (LIC), and acquired therapy resistance. The authors focused on hematological neoplasias where CD44 has three main functions: first, its role as prognostic marker; second, its potential role for diagnosis; and third, its role as a promising therapeutic target. Distinguishing between different prognostic subtypes of one neoplastic disease entity is of great advantage. For preventing secondary malignancies induced by too aggressive treatment regimes a deliberate risk stratification based on the expected tumor prognosis is imperative. Another example gives the observed treatment resistance to several standard therapeutics associated with CD44 expression in a limited group of patients with DLBCL (diffuse large B-cell lymphoma).

This provides the possibility of improved individual treatment decisions, which have been a big aim in cancer therapy in the last years. More and more, CD44 is considered being of diagnostic use.

1.3 THE AC133 ANTIGEN: OLD DOG NEW TRICKS?

The pentaspan membrane protein CD133 (Prominin-1) was first identified by Huttner and colleagues in embryonic and adult mouse epithelial cells *(Weigmann et al., 1997),* and it was charachterized in human hematopoietic stem and progenitor cells by Miraglia and colleagues *(Miraglia S et al., 1997).* CD133was found to be enriched at subdomains of the cell surface, such as microvilli of neuroepithelial cells and in cell protrusions, like filopodia, lamellipodia and microspikes in non-epithelial cells. Due to its specific location on the cell surface, this protein was termed´Prominin`, from the Latin word "prominere", which means to stand out, to be prominent. In the same year, the homolog of mouse CD133 was detected in human CD34-positive hematopoietic stem cells, by using an antibody against the surface antigen AC133 *(Yin et al., 1997).* In the mouse, CD133 is encoded by the Prom1 gene on chromosome 5 (location 5 B3). In 1997, Yin et al., produced a novel monoclonal antibody (MAb) that recognized the AC133 antigen, a glycosilation-dependent epitope of CD133. Prominin-1 is a product of a single-copy gene on chromosome 4 (4p15.33) in humans or chromosome 5 (5 B3) in mice. Both the mouse and human genes have similar genomic organizations, consisting of at least 37 (human) and 34 (mouse) exons spanning approximately 160 kb. The transcript size is about 4.4 kb in both humans and mice. Human Prominin-1 is a transmembrane glycoprotein of 865 amino acids (aa) with a total molecular weight of 120 kDa (858 aa,

115 kDa in mouse). Prominin-1 has a unique structure consisting of an N-terminal extracellular domain, five transmembrane domains with two large extracellular loops, and a 59 aa cytoplasmic tail (Figure 1.2).

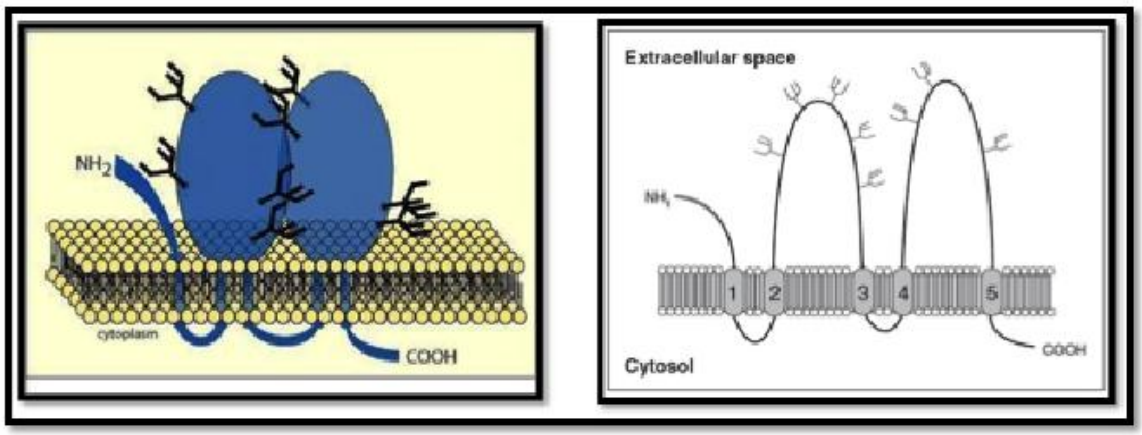


Figure 1.2 Structure of AC133 antigen. Adapted from *Fargeas et al., 2004.*

Human and mouse Prominin-1 share roughly 60% homology. Phylogenetically conserved, Prominin-1 can also be found in *C. elegans*, Drosophila and Zebrafish. Prominin-2, a recently discovered second member of the Prominin family, shares about 26% and 29% homology with human and mouse Prominin-1, respectively *(Fargeas, et al., 2003).* A number of alternatively spliced isoforms of Prominin-1 have been identified in mice and humans, with tissue-specific distribution *(Corbeil, et al., 2001; Shmelkov et al., 2004; Yu, et al., 2002).* CD133 has five putative transmembrane domains with two extracellular loops which contain more than 250 aminoacids each, an extracellular N-terminal and a cytoplasmic C-terminal domain. Eight potential N-gycosylation sites are

located at the extracellular loops *(Fargeas et al., 2003)*. Several splice variants are identified so far (in the mouse variant s1-s8) and their expression seems to be tissue-specific and developmentally regulated *(Fargeas et al., 2007)* (Figure 1.3). Murine prominin-1 was identified as a novel marker of neuroepithelial cells, primary progenitor cells of the mammalian central nervous system, whereas its human counter part constituted a new hematopoietic stem and progenitor cell (HSPC) marker (initially referred to as AC133 antigen).

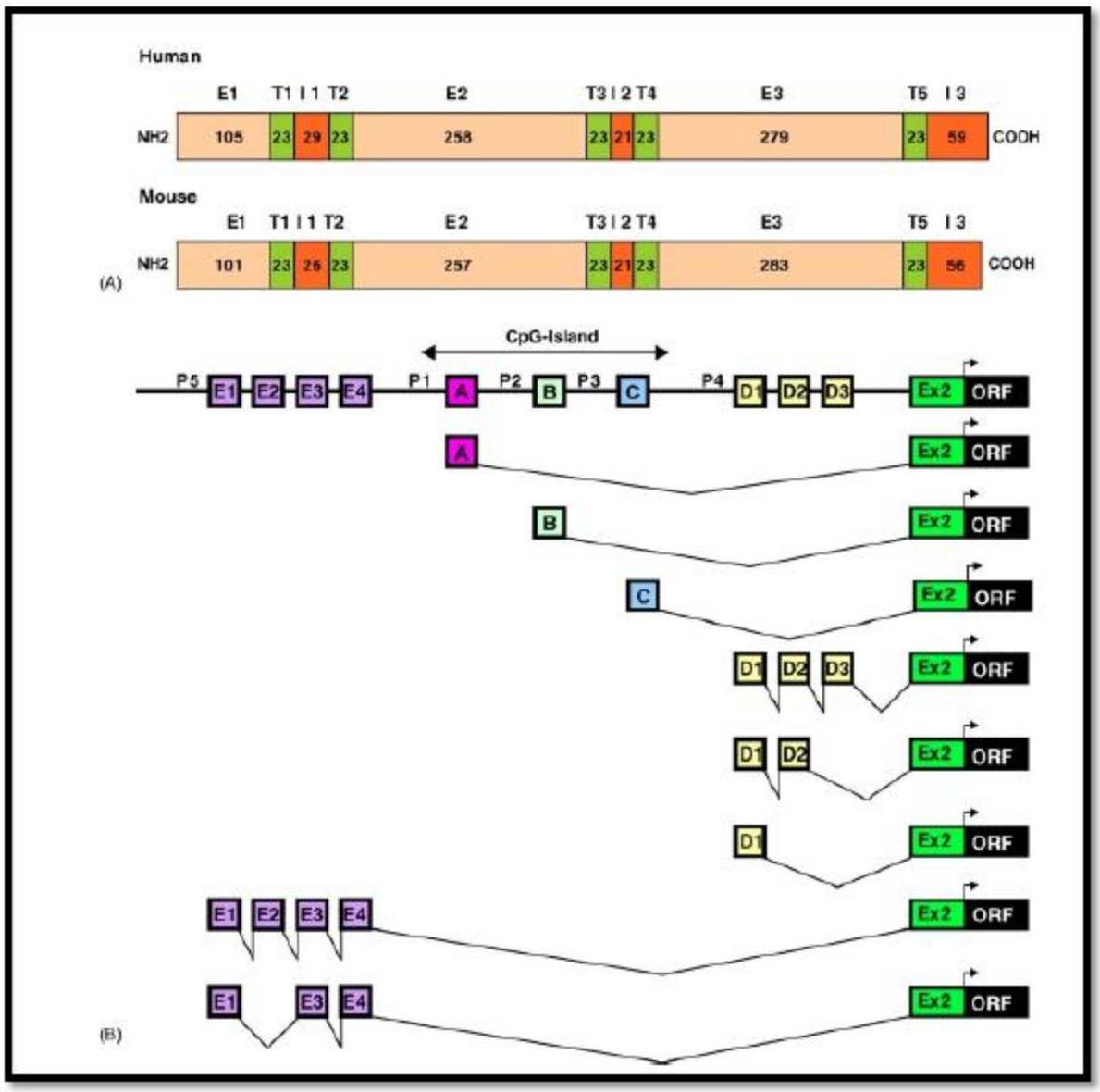


Figure 1.3 Schematic representation of the 5′ region of CD133. CD133 has five alternative promoters, each containing one or multiple corresponding exons. Adapted from *Kemper et al., 2010; Xin et al., 2008*.

As a cell surface marker, prominin-1 is now used for somatic stem cell isolation. CD133+ stem and progenitor cells might become clinically important, particularly with regard to brain injury/disease and bone marrow transplantation. It is important to note that,

although various stem and progenitor cells express prominin-1, its expression is not limited to primitive cells. For instance, prominin-1 is detected in several epithelia in adult mice and humans where it appears to be restricted to the apical (luminal) side. Additionally, prominin-1/ CD133 is often used for the identification and isolation of putative cancer stem cell populations from malignant tumors of brain, prostate, liver, pancreas, lung, and colon *(Yin et al., 1997; Kania et al., 2005; Miraglia et al., 1997; Weigmann et al., 1997)*. Its expression in cancer stem cells has broadened its clinical value, as it might be useful to outline new prospects for more effective cancer therapies by targeting tumor initiating cells. Cell biological studies of this molecule have demonstrated that it is specifically concentrated in various membrane structures that protrude from the planar areas of the plasmalemma. Prominin-1 binds to the plasma membrane cholesterol and is associated with a particular membrane microdomain in a cholesterol-dependent manner. Although its physiological function is not yet fully determined, a recent finding has shown that transgenic mice carrying human mutation for the prominin-1 gene (PROM1) undergo progressive photoreceptor degeneration in the retina consistent with that found in human patients, suggesting a functional role for prominin-1. Prominin-1 is expressed on a subpopulation of CD34+ HSPCs derived from various sources including fetal liver and bone marrow, adult bone marrow, cord blood and mobilized peripheral blood. Interestingly, an immunomagnetic selection of CD133+ HSPCs allowed the enrichment of a sufficient amount of cells to perform hematopoietic stem cell transplantation, and pilot trials with leukemic children have proven the feasibility of CD133+ selection for allogeneic transplantation. Other studies have shown the successful transplantation of haploidentically mismatched peripheral blood stem cells using CD133+ purified stem cells. Thus, the immunomagnetic isolation procedure

of HSPCs based on prominin-1 appears to be an interesting alternative to CD34. Moreover, accumulating studies illustrate that prominin-1/CD133+ progenitor cells can exert beneficial effects in treating of different pathological disorders, including cardiac and hepatic malignancies Although its physiological function in stem cells remains elusive, Bauer et al., observed that during the cell division of neural and HSPCs prominin-1 was either symmetrically or asymmetrically distributed between the two nascent daughter cells *(Bauer et al., 2008; Fargeas et al., 2006; Kosodo et al., 2004).* Prominin-1 containing lipid rafts might host key determinants or players necessary to maintain stem cell properties, their quantitati\ve reduction or loss might result in differentiation (*Fargeas et al., 2006; Marzesco et al., 2005)*. In agreement with this, Kosodo et al could demonstrate that neurogenic cell divisions of neural progenitors, but not proliferative ones, involve such an asymmetric distribution of prominin-1 *(Kosodo et al., 2004).* Additionally, in the developing embryonic mouse brain, prominin-1 is released from neural progenitor cells into the lumen of the neural tube during the early phase of neurogenesis *(Marzesco et al., 2005).* Therein, prominin-1 is associated with small (50–80 nm) membrane vesicles that were distinct from exosomes and appeared to bud from microvilli and/or primary cilium *(Dubreuil et al., 2007; Marzesco et al., 2005).* Such prominin-1 containing membrane vesicles (prominin-1-CMV) were also found in human cerebral fluids, and remarkably appear to be up-regulated in glioblastoma patients suggesting that putative cancer stem cells might release them as well *(Huttner et al, 2008)*. Intriguingly, we demonstrated that, in contrast to other prominin-expressing cell types studied so far, e.g. epithelial cells *(Karbanova´ et al, 2008; Weigmann et al, 1997),* HSPCs contain an important intracellular pool of prominin-1 in addition to the cell surface one. Inside the cell, prominin-1 is located primarily in membrane vesicles within

multivesicular bodies. Taken together, the release of prominin-1-containing exosomes concomitant with cellular differentiation may represent an alternative mechanism of externalization of stem/progenitor properties-containing lipid rafts, in addition to the budding mechanism underlying the release of prominin-1 from plasma membrane protrusions (i.e. microvillus and primary cilium) of neural progenitors *(Ettinger et al., 2011; Corbeil et al., 2010; Dubreuil et al., 2007; Marzesco et al., 2005, 2009).* Thus, the concept of ‘stem cell-specific lipid rafts’ holding molecular determinants necessary to maintain the stem cell properties is attractive in this context *(Fargeas et al., 2006; Marzesco et al., 2005)* and the characterization of those containing prominin-1 including their proteome and lipidome may reveal new aspects of stem cell biology. Along the same line, it will be of interest to determine whether similar phenomena occur in cancerous tissues other than the brain *(Huttner et al., 2008)* given that the expression of prominin-1 is often associated with transformed, treatment-resistant cells exhibiting tumour initiating properties *(Bao et al., 2006; Liu et al., 2006).*

1.4 WNT RECEPTOR SIGNALING: PROCESSING, SECRETION AND RECEPTION

The Wnt pathway, like several other signaling systems, is a major molecular mechanism that controls animal development. Moreover, deregulation of Wnt signaling is tightly linked to human disease, such as multiple forms of cancer and bone malformation *(Clevers, 2006; Klaus & Birchmeier, 2008).* In sporadic colorectal cancer, the mostcommon form of intestinal cancer, mutations in multiple Wnt signaling components have been found, ectopically activating the Wnt pathway. This is best illustrated by the gene adenomatous polyposis coli (APC) which is mutated in more than two-thirds of the cases *(Segditsas & Tomlinson, 2006).* More recently, Wnt signaling received additional attention for its important role in specification and maintenance of stem cells in various tissues and organs *(Wend et al., 2010).* Therefore, it is of critical importance to have a good understanding of this pathway. Over the last few decades much work has focused on the different signal transduction mechanisms initiated by binding of the Wnt ligand to receptors of Wnt responsive cells. More recently, also the Wnt posttranslational modifications and the secretion mechanismin the Wnt producing cells have received much attention. Wnt proteins were originally identified in Drosophila *(Nusslein-Volhard & Wieschaus, 1980)* and mice *(Nusse & Varmus, 1982),* which were called Wingless(Wg) and Int1, hence the name Wnt. Wnt proteins are secreted lipid-modified glycosylated signaling molecules that are essential in various developmental processes. Wnt proteins contain an N-terminal signal sequence and are palmitoylated on a conserved cysteine residue. The palmitate is added in the endoplasmatic reticulum by the protein Porcupine

(Porc) and is essential for signaling. These proteins are characterized by a high number of conserved cysteine residues and are glycosylated and lipid modified at two conserved residues, which makes Wnt proteins highly hydrophobic. Surprisingly, however, Wnt proteins have been shown to diffuse over several cell diameters in the extracellular space. Additional factors in the extracellular space are able to regulate the activity of the Wnt morphogenetic gradient. Proteins of the secreted frizzled related protein (SFRP) family and the Wnt interacting factor 1(Wif1) can bind Wnts and are therefore seen as Wnt inhibitors. Other factors act at the level of the Wnt receptors such as the Wnt signaling inhibitor Dickkopf (Dkk) and Wnt activator R-spondin. Wnt receiving cells can induce several signaling cascades in response to the morphogenetic Wnt gradient. The best studied signal transduction mechanism is the Wnt/β-catenin pathway, which is also known as the canonical Wnt pathway *(Clevers, 2006).* Here, binding of Wnt to its receptors induces the stabilization of β-catenin, which is otherwise targeted for proteasomal degradation. Stabilized β-catenin translocates into the nucleus and activates TCF-LEF depended transcription of Wnt target genes (Figure 1.4).

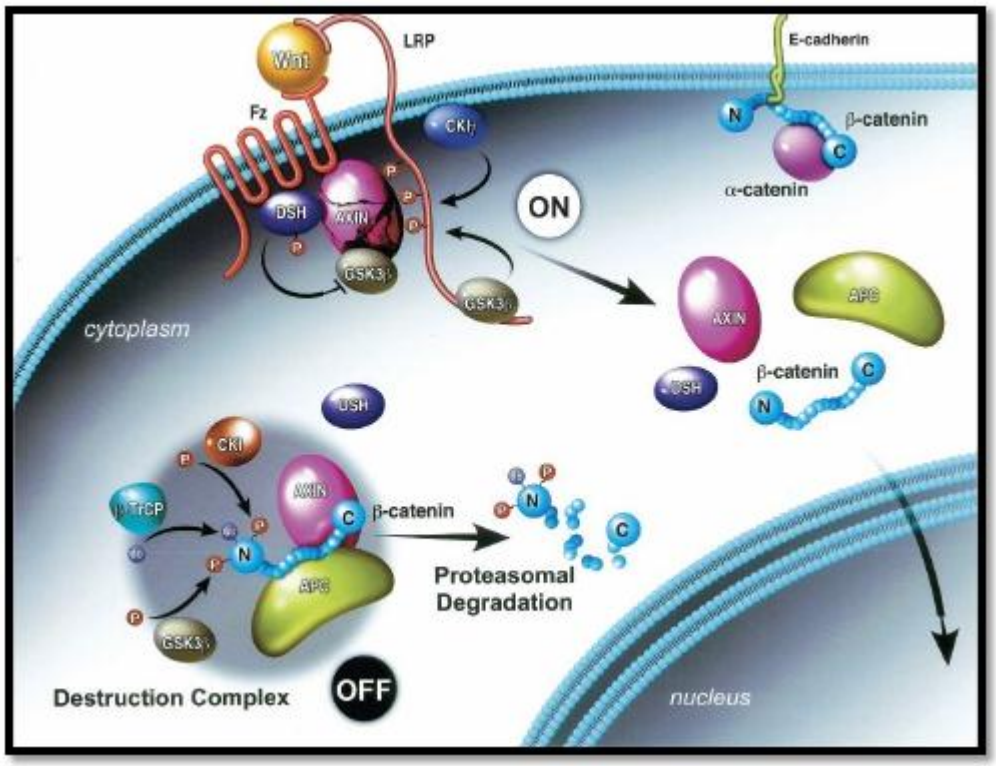


Figure 1.4 Activation of Wnt signaling pathway. In the absence of Wnt ligands, β-catenin binds to a destruction complex containing APC, Axin, and the CK1 and GSK3 kinases and is marked for proteolytic destruction ("off"). Wnt signaling promotes CK1γ and GSK3β mediated hyperphosphorylation of LRP5/6 and enhances Dsh phosphorylation, which jointly recruit Axin to the receptor complex at the plasma membrane ("on"), where it undergoes proteolytic degradation. Unphosphorylated β-catenin is no longer rapidly degraded and enters the nucleus. Many of the β-catenin destruction complex components are shuttling proteins that distribute both in the cytoplasm and the nucleus, and some of these, such as APC, Axin, and Bcl-9/Lgs, are required for the accumulation or retention of β-catenin in the nucleus. Adapted from *Willert and Jones, 2006.*

Wnt- signalling participates in embryogenesis, stem cell biology, and human cancer. At present there are 19 Wnt genes identified encoding unique ligands *(Cadigan & Nusse, 1997.)*. Wnt- β-signalling has been classified in two major pathways, canonical and non-

canonical. Wnts induce intracellular signalling in both pathways by binding to Frizzled receptors *(Bhanot et al., 1996; He et al., 1997; Yang-Snyder et al., 1996; Strutt, D. 2003.).* Six Frizzled genes have been identified so far. The Wnt- non-canonical pathway can be further divided in two different signal transduction pathways. The Wnt/PCP pathway, which affects planar cell polarity through rearrangements of the cytoskeleton involving JNK and ROCK, and the Wnt/Ca2+ pathway affecting NF-AT transcription activity, through calcineurin, PKC and PLC. In the canonical pathway, the Wnt co-receptors LRP5/6 are also required for signal induction *(Pinson et al., 2000).* Canonical Wnt- β signalling involves activation of β -catenin, the key effector of this pathway, which is a ubiquitously expressed multifunctional protein. β-catenin links E-cadherin to the actin cytoskeleton and thus has an important function in cell-cell adhesion. In the nucleus, β-catenin functions as a transcription cofactor regulating cell proliferation and differentiation *(Gumbiner et al., 1995).*

1.4.1 CANONICAL WNT SIGNALING: COLLABORATION AND CONNIVANCE

Signaling pathways are a fundamental aspect of how cells communicate with one another and respond to their environment, influencing cell growth, cellular differentiation and apoptosis *(Clevers, 2006; Logan & Nusse, 2004).* In the simplest form, a signaling pathway functions through the binding of a ligand to its specific receptor which in turn activates the receptor in order to elicit an intracellular response (Figure

1.5).

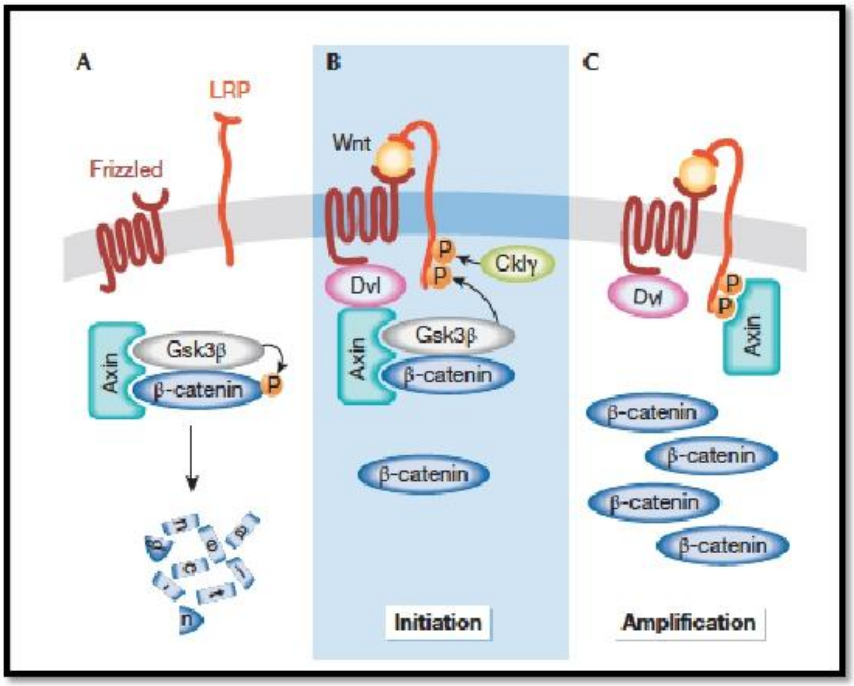


Figure 1.5 Model for the activation of the Wnt/ β -catenin pathway. (A) In absence od WNT ligand. (B) On binding of Wnt to the receptors, FZD and LRP, Dvl binds to FZD and recruits the destruction complex through interaction with axin. Subsequently, LRP is phosphorylated and acts as docking site for axin. (C) Binding of axin to LRP leads to inhibition of the destruction complex and stabilization of β -catenin. *(Fuerer and Nusse, 2008).*

The Wnt family of secreted glycoproteins are the ligands which promote cell proliferation, cell polarity, neural differentiation, and cell fate determination during embryonic development and tissue homeostasis *(Clevers 2006; MacDonald et al., 2009).*

β-catenin is a fundamental player in the canonical Wnt pathway. In the presence of ligand, β-catenin is bound to the destruction complex which is composed of several proteins including a scaffolding protein Axin, adenomatous polyposis coli (APC), casein kinase 1α (CK1), and glycogen synthase kinase 3β (GSK3β). When bound, β-catenin is phosphorylated on Serine (Ser) 45 by CK1α which primes β-catenin for the sequential phosphorylation of Threonine (Thr) 42, Ser 39, and Ser 37 by GSK3β. This phosphorylation of β-catenin promotes the recognition by β-TRCP, an E3 ubiquitin ligase, which leads to the ubiquitination of β-catenin and proteasomal degradation. This continuous degradation prevents cytoplasmic β-catenin from translocating into the nucleus, binding to transcription factors T‐cell factor and lymphoid enhancer factor 1 (TCF/LEF-1), and stimulating Wnt responsive genes. In the presence of Wnt glycoproteins, Wnts bind simultanously to the seven-pass transmembrane Frizzled receptors (Fzd) and the lipoprotein receptor-related protein 5 or 6 (LRP5/6) co-receptor. Proteoglycans, such as Dally *(Lin and Perrimon, 1999)* or Syndecan 1, concentrate Wnt ligands at cell surfaces where they can bind to LRP5/6 and Fz receptors to mediate their interaction. β-catenin is the key player in this pathway and is degraded by the proteasome in the absence of an activating Wnt signal. The E3-ubiquitin-ligase β-transducin-repeat-containing protein (β-TRCP) targets β-catenin for proteasomal destruction, but only recognizes its substrate in a phosphorylated state. β-catenin is phosphorylated by the destruction complex, which is composed of at least the Axis inhibition protein 1 (Axin1), adenomatous polyposis coli (Apc), casein kinase 1 (CK1) and glycogen synthase kinase 3β (GSK-3β). Axin1 interacts directly with all the other members of the destruction complex and functions as a scaffold protein. The two

serine/threonine kinases CK1 and GSK3 β phosphorylate β-catenin, which is then ubiquitinylated by βTRCP and degraded by the proteasome. Signals induced by Wnt proteins interrupt the formation of the degradation complex, there by preventing the phosphorylation and destruction of β-catenin .

WNT proteins

Wnt proteins bind to the extracellular N-terminal cysteine-rich domain of the Frizzled (Fz) receptor, which is in a complex with the low density lipoprotein receptor-related protein 5 or 6 (LRP5/6). After an activating Wnt signal the protein Dishevelled (Dvl) is recruited to the receptor complex and the cytoplasmic tail of LRP5/6 is phosphorylated by CK1 and GSK3β. This provides a docking site for Axin1, which is then recruited to the receptor complex. Axin1 is sequestered and assembly of the destruction complex is disrupted. β-catenin will accumulate in the cytoplasm and translocate to the nucleus, where it initiates transcription by activating T cell factor/lymphoid enhancer factor (TCF/LEF) transcription factors. In the absence of β-catenin TCF forms a transcriptional repressor complex with Groucho. Groucho is physically displaced by β-catenin and Pygopus and Legless are recruited to assemble a transcriptional activator complex *(Clevers, 2006; Logan & Nusse, 2004; MacDonald et al., 2009; Mosimann et al., 2009; Staal & Clevers, 2005).* β-catenin is rapidly turned over by ubiquitination and degradation by the proteasome pathway under unstimulated conditions. This requires phosphorylation of β-catenin by a "degradation complex" consisting of APC, Axin, GSK3, and CK1, followed by binding of β-Trcp *(Rubinfeld et al., 1996; Munemitsu et al.,*

1995). Mutations in components of the "degradation complex" lead to constitutive accumulation of β-catenin in a number of cancers. APC mutations are common in colorectal cancers *(Kinzler & Vogelstein, 1997).* Axin1 mutations have been reported in hepatocellular carcinomas *(Satoh et al., 2000)* while β-Trcp mutations have been reported in prostate cancers *(Gerstein et al., 2002).*

1.4.2 THE MANY FACES OF THE CANONICAL WNT PATHWAY

Wnt proteins are characterized by a high number of conserved cysteine residues *(Miller, 2002)* and are post-translationally glycosylated and lipid modified. Secreted Wnt proteins form concentration gradients in the extracellular space, to which cells expressing the appropriate receptors respond in a concentration dependent manner. Although much research has focused on the signaling cascades that are triggered by Wnt proteins, the importance of Wnt producing cells in the Wnt signaling pathway is only recently becoming apparent. With the identification of several components required in the Wnt producing cells for Wnt signaling, the idea evolved that the production and secretion of Wnt molecules requires a specialized secretory machinery that offers an additional layer of control to the Wnt signaling pathway. When Wnt proteins were first isolated, a surprising finding was that they are highly hydrophobic *(Bradley & Brown, 1990)*. This was explained by the identification of two lipid modifications; a saturated acyl chain, palmitate, attached to a conserved cysteine residue (C77 in mouse Wnt3a) *(Willert et al., 2003)* and a mono-unsaturated acyl chain, palmitoleate, on a conserved serine residue (S209 in mouse Wnt3a) *(Takada et al., 2006).* Mutating the C77 position in

Wnt5a also led to a loss of hydrophobicity, suggesting that the lipid modification on this residue is conserved *(Kurayoshi et al., 2007),* although it is not known whether the same type of acyl chain is used. Mutational analysis has shown that lipid modifi cation is required for the secretion as well as the signaling activity of Wnt proteins. In mammalian Wnt proteins, mutating C77 to alanine strongly interferes with signaling activity, while secretion is largely unaffected *(Galli et al., 2007; Komekado et al., 2007; Kurayoshi et al., 2007; Willert et al., 2003)*. Mutating the S209 residue, on the other hand, has a strong effect on secretion, resulting in accumulation of the mutant Wnt protein in the endoplasmic reticulum (ER) *(Takada et al., 2006).* The effect of mutating the lipid modified residues is however less clear in Drosophila. In S2 tissue culture cells, both C77 and S209 mutants of the Wnt protein Wg are secreted, while in transgenic animals, there is accumulation of the C77A mutant Wg protein in the ER *(Franch-Marro et al., 2008a).* Despite these discrepancies, the current consensus is that lipidation is important for the exit of Wnt from the ER. The observed accumulation of Wnt proteins in the ER might be the result of defects in the folding of the mutated protein. Alternatively, membrane tethering might be required for the interaction of Wnt with the ER exit machinery. It cannot be excluded, however, that the ER accumulation is caused by overexpression of the mutated protein. Therefore, to better understand these results, it will be important to analyze these mutant Wnt proteins under physiological conditions. In *C. elegans* as well as in Drosophila , alleles of Wnt were isolated in which the C77 residue is mutated. This leads to a strong loss of Wnt signaling, although not as severe as observed in Wnt null mutants *(Coudreuse et al., 2006; Couso & Martinez Arias, 1994; Willert et al., 2003),* suggesting that C77 mutated Wnts still have residual signaling activity. The subcellular distribution of the mutated Wnt proteins has however not been studied. Interestingly,

genetic screens have not recovered mutants with mutations in the S209 residue. The only functional data to compare the two lipid modification mutants is the above mentioned S2 cell secretion assay *(Franch-Marro et al., 2008a).* Here it was found that, in agreement with the C. elegans and Drosophila mutants, the C77 mutant has only residual activity, while the S209 mutant has retained more activity. Another interesting question is the biological significance of the S209 acylation with a mono-unsaturated fatty acid. The resulting double bond will induce a kink in the acyl chain, which may negatively influence the interaction with lipid ordered membrane domains *(Moffett et al., 2000).* It was shown for proteins such as Fyn, Annexin II and Gai that acylation with unsaturated fatty acids results in displacement of proteins from membrane domains with ordered lipid structure *(Liang et al., 2001; Moffett et al., 2000; Zhao & Hardy, 2004).* Therefore, this unsaturated lipid modification could play an important role in targeting of Wnt to specific membrane domains during secretion. Furthermore, modification with an unsaturated acyl chain will slightly decrease membrane affinity compared to a saturated acyl chain, which may be important for the extracellular spreading of the secreted Wnt protein. A good candidate for mediating the lipid modification of Wnt is Porcupine (Porc), which was originally identified as a segment polarity gene in Drosophila *(van den Heuvel et al., 1993)* and encodes a member of the membrane-bound O-acyltransferase (MBOAT) family. Porc localizes to the ER, where Wnts are thought to be lipid modifi ed *(Zhai et al., 2004),* and interacts with Wnts in a region that includes the C77 residue *(Tanaka et al., 2002; Tanaka et al., 2000).* Depletion of Porc leads to a complete block in Wnt secretion and accumulation of Wnt in the ER *(van den Heuvel et al., 1993),* similar as observed with the murine Wnt3a (S209) mutant. Furthermore, Porc depletion strongly reduces the hydrophobicity of Wnts *(Zhai et al., 2004),* whereas Porc overexpression

increases hydrophobicity *(Galli et al., 2007).* Given the similarity in phenotype, it is likely that Porc is responsible for the S209 O-esterification of Wnt with palmitoleic acid. Whether Porc is also responsible for the C77 lipid modification is less clear, since this might be masked by the S209 modification and subsequent ER retention. Furthermore, one might question whether a single enzyme can specifically catalyze both oxyester (S209) as well as thioester (C77) formation using two different substrates (palmitoyl and palmitoyleoyl CoA) *(Jing and Trowbridge 1987; Rose et al., 1984).* A possible explanation could be that Porc is responsible for the S209 acylation of Wnt and that subsequent C77 acylation is dependent on a different enzyme. Another explanation could be that both C77 as well as S209 get palmitoylated by Porc and that the S209 palmitate gets de-saturated to palmitoleate by another enzyme. Experimental evidence for this possibility is however lacking. In addition to a role in Wnt maturation and secretion, lipidation is also important for the signaling activity of Wnt proteins, as illustrated by the S2 reporter assays and genetic mutants discussed above and by the observation that enzymatic de-lipidation of purified secreted Wnt strongly inhibits signaling activity. Lipid modification is thought to restrict the spreading of Wnt in the extracellular environment and to concentrates it at the membrane for signaling. In addition, the lipid modifications are also required for the interaction with its receptors, as de-lipidation or mutation of C77 also results in a strong reduction in the affinity of Wnt for the Wnt binding domains of its receptors LRP and Frizzled *(Cong et al., 2004; Franch-Marro et al., 2008a; Komekado et al., 2007; Kurayoshi et al., 2007).* Wnts are modified by multiple N-linked glycosylations, which, like the lipid modifications, are not well understood and are likely involved in regulating both secretion and signaling. For example, upon enzymatic de-glycosylation of secreted Wnt, signaling activity is strongly

reduced, even though it still interacts with the Frizzled receptor *(Komekado et al., 2007).* Moreover, mutating the presumed modified residues or inhibiting the modification by a glycosylation inhibitor strongly affects Wnt secretion (although this may not be the case for all Wnts) *(Komekado et al., 2007; Kurayoshi et al., 2007).* These effects on secretion could be caused by folding defects, as one of the functions of glycosylation is to facilitate the folding of proteins in the ER *(Caramelo & Parodi, 2007).* Surprisingly, overexpression of Porc stimulates both lipidation and glycosylation and Porc depletion was reported to decrease glycosylation *(Galli et al., 2007; Tanaka et al., 2002).* It seems unlikely that Porc is directly responsible for the glycosylation of Wnt. It may, however, be involved in the recruitment of the oligosaccharide transferase (OST) complex. Alternatively, Porc overexpression may facilitate glycosylation by promoting membrane tethering of Wnt, making it better accessible to the OST complex. Three independent groups identified the multi-pass transmembrane protein Wls (also known as Evenness interrupted/Evi or Sprinter) as a critical component of the Wnt secretion machinery *(Banziger et al., 2006; Bartscherer et al., 2006; Goodman et al., 2006).* Wls is specific for Wnt secretion, since its depletion does not affect other secreted proteins. It is a highly conserved multi-pass transmembrane protein that binds Wnt in co-immunoprecipitation experiments *(Banziger et al., 2006; Coombs et al., 2010; Fu et al., 2009)*. Wls localizes most prominently to the Golgi, the plasma membrane and endosomes *(Banziger et al., 2006; Bartscherer et al., 2006; Belenkaya et al., 2008; Franch-Marro et al., 2008b; Port et al., 2010; Yang et al., 2008),* indicating that it functions downstream of Porc in the secretory pathway. Analysis of Wls mutants in Drosophila showed that in the absence of Wls, Wnt accumulates in the Golgi, suggesting that Wls functions as a sorting receptor that transports Wnt from the Golgi to the cell surface for release. It was found that upon

inhibition of clathrin mediated endocytosis (by depletion of the AP2 μ subunit DPY-23 in C. elegans or dynamin in Drosophila) Wls accumulates at the plasma membrane *(Belenkaya et al., 2008; Pan et al., 2008; Yang et al., 2008).* As this leads to a strong defect in Wnt signaling, it was proposed that Wls is endocytosed and is recycled to take part in multiple rounds of Wnt secretion (through a pathway that will be discussed below). In this model, Wls plays a central role in the Wnt secretion pathway. Regulation of the trafficking of Wls therefore represents a mechanism to closely control Wnt secretion. A key question that remains to be addressed is how Wnt is released from Wls. It has recently been shown that endosomal acidification is essential for the dissociation of Wnt form Wls, as inhibition of acidification by treatment with bafilomycin, a V-ATPase inhibitor, interferes with Wnt secretion *(Coombs et al., 2010).* Intracellular and plasma membrane levels of both Wnt and Wls were increased and importantly, Wnt and Wls remained in complex together. This suggests that vesicular acidification is somehow involved in the release of Wnt from Wls. However, a decrease in pH was not enough to dissociate a purified Wls-Wnt complex in vitro. The Wnt binding domain of Wls was mapped to the first intraluminal loop and modeling of the structure of this domain revealed that it may fold into a lipocalin-like structure *(Coombs et al., 2010).* Lipocalins are a family of secreted proteins that can bind a range of hydrophobic molecules, including palmitate *(Flower, 1996),* indicating that the lipocalin-like domain of Wls might bind to the Wnt lipid modifications. In support of this possibility is the recent observation from the Nusse laboratory that soluble Wnt isolated from tissue culture medium is in complex with the lipocalin family member Swim *(Nusse et al., 2008)*. It was shown that the S209 lipid modification is essential for the binding of Wnt to Wls, whereas the palmitoylated C77 residue is not *(Coombs et al., 2010),* indicating that in the

case of Wls the lipocalin domain may bind the mono-unsaturated palmitoleic acid residue instead of the palmitate residue. Interestingly, it has been shown that a plant lipocalin family member dimerizes upon vesicular acidification *(Arnoux et al., 2009).* This suggests the intriguing possibility that vesicular acidification en route to the plasma membrane induces the release of Wnt from Wls by triggering dimerization of the Wls Wnt binding domains. A recent study provides an example where Wnt may be required to traffic through the endocytic pathway for release. The idea that the Wnt producing cell itself might promote Wnt spreading by regulating its own plasma membrane composition is appealing. Another mechanism that may assist the release and spreading of Wnt depends on Reggie-1/flotillin, which is a major component of membrane microdomains *(Katanaev et al., 2008).* Overexpression of Reggie-1 strongly expands the signaling range of Wnt by facilitating the capacity of Wnt to diffuse in the tissue. The mechanism by which Reggie-1 achieves this increase in Wnt spreading is however still unknown. Since Wnt signaling is such a key mechanism of development and adult tissue homeostasis, close regulation of Wnt signaling is essential. It is clear that there are several layers of regulation of the Wnt secretion process. First, there is transcriptional control of Wnt, but also of Wnt secretion factors. This is illustrated by Wls in the mouse, which is a target of the canonical Wnt/β-cateninpathway. This suggests a positive feedback mechanism in which Wnt stimulates Wnt secretion through an autocrine or paracrine mechanism *(Fu et al., 2009).* The induction of Wls expression by canonical Wnt signaling is however not evolutionarily conserved, as constitutive activation of canonical Wnt/ β-catenin signaling in the Drosophila wing disc did not alter Wls levels *(Port et al., 2008)*.

Frizzled (FZD) receptors

Wnt ligands interact with the cell surface receptor, Frizzled (FZD). The initial connection between seven-transmembrane-span proteins of the Fz family and Wnt proteins came from studies in Drosophila cell culture. Transfection of Drosophila frizzled 2 (fz2) confers the ability to bind Wingless (Wg; a fly Wnt) and stabilize Armadillo (Arm; the fly β-catenin) upon cells that do not express Fz2 and are unresponsive to Wg *(Bhanot et al., 1996).* Additional evidence implicating Fzs in Wnt signaling came from mis-expression studies in Xenopus *(Yang-Snyder et al., 1996; He et al., 1997)* and the finding that a mutation in lin-17, which encodes a Fz, affects T-cell polarity in Caenorhabditis elegans *(Sawa et al., 1996),* which is controlled by LIN-44, a Wnt *(Herman et al., 1995).* FZD receptors are seven-pass transmembrane receptors which have cycteine-rich domains (CRD) in their N-terminus. Through the CRD, FZD receptor binds Wnt ligands. In general, it is thought that a monomeric FZD receptor transmit signals downstream upon binding with Wnt ligand, however, the crystallographic resolution of the structure of the mouse FZD8 and sFRP3 CRD domains suggested that CRDs might be able to homodimerise or heterodimerise. Furthermore, there are reports showing that dimerisation of FZD receptor activates the Wnt/β-catenin pathway and that FZD form specific homo-and hetero-oligomers. These reports suggest the wide possibility of the signal transmission mechanism downstream of FZD receptor. Upon the binding of Wnt to FZD receptor, the intracellular amino sequences, K-T-X-X-X-W directly binds to Dishevelled proteins. There are 10 reported human frizzled receptors. Phylogenetically, the Frizzled receptors fall into four groups. Frizzled-1, 2 and 7, and Frizzled-3 and 6 make up two related groups, while Frizzled-5 and 8 comprise a third group, and Frizzled-4, 9 and 10 generate a distant fourth group *(Carron et al. 2003).*

LRP5, LRP6

There also exist co-receptors of FZD receptor. A genetic study using flies showed that a single-pass trans-membrane receptor, Arrow, is required to establish a segment polarity triggered by Wg signaling. Arrow is homologous to two members of the mammalian low-density lipoprotein receptor (LDLR)-related protein (LRP) family, LRP5 and LRP6. LRP5/6 function as co-receptors of FZD receptor and binding of Wnt ligand to both FRZ receptor and LRP5/6 co-receptor activates Wnt/β-catenin pathdway. LRP5 was first cloned as an apolipoprotein E binding receptor in hepatocytes and adrenal cortex *(Kim et al., 1998; Hey et al., 1998; Brown et al., 1998)*. The gene is highly conserved among different species *(Houston & Wylie et al., 2002)* and is designated "arrow" in invertebrates *(Wehrli et al., 2000).* LRP5 was later associated to type 1 diabetes *(Figueroa et al., 2000; Twells et al., 2003),* and being essential in cholesterol metabolism and glucose-induced insulin secretion *(Fujno et al., 2003: Magoori et al., 2003).* LRP5 was shortly thereafter identified as an essential component in Wnt signalling *(Tamai et al., 2000; Wehrli et al., 2000).* LRP5 functions as a coreceptor and binds Wnt ligands together with Frizzled receptors, with subsequent activation of the β-catenin dependent Wnt signalling pathway *(Schweizer & Varmus, 2003; Holmen et al., 2002)*. Phosphorylation of LRP5/6 by CK1 gamma and GSK-3 transduce activating signals *(Davidson et al., 2005),* while CK1 epsilon phosphorylation has an inhibitory effect *(Swiatek et al., 2006).* Receptor phosphorylation leads to Axin binding and subsequent stabilisation of β-catenin *(Tamai et al., 2004).* LRP5 activity is inhibited by DKK1 through binding to kremen *(Mao et al., 2002).* Using live imaging of vertebrate cells expressing fluorescently tagged Axin and LRP6, *(Bilic et al. 2007)* demonstrated that Wnt signaling induces plasma membrane-associated LRP6 aggregates. In unstimulated cells, Axin localizes to

intracellular punctate while LRP6 uniformly stains at the cell membrane. Wnt stimulation results in the rapid formation of LRP6 punctate, referred to as "LRP6 signalosomes." This event is followed by Axin recruitment to the aggregates. The Tp1479 antibody was used to show that the punctae are enriched for CKI γ-mediated phosphorylated LRP6, and that these structures also contain Fz8, Dvl2, GSK, and Axin *(Bilic et al., 2007).*

Kremen, Ror2 and Ryk

A single-pass trans-membrane receptor, Kremen, was initially identified as a binding partner of a negative regulator of Wnt/ β-catenin signaling, Dkk1. Upon binding to Dkk1, Kremen is internalized by endocytosis with LRP5/6, leading to a suppression of Wnt/ β-catenin pathway *(Mao et al., 2002).* The Ror family of receptor tyrosine kinases (RTK) consists of two structurally related proteins, Ror1 and Ror2. They have an extracellular CRD, a membrane proximal kringle (KR) domain, and intracellular cytoplasmic tyrosine kinase domain and a proline-rich domain near the c-terminus *(Forrester er al., 1999).* Ror2 has been shown to act as an alternative receptor or co-receptor for Wnt5a *(Mikels & Nusse, 2006).* In addition to its ability to bind Wnt5a, Oishi and colleagues reported the ability of Ror2 to bind some FZD receptors as well *(Oishi et al., 2003),* suggesting that Ror2 might play a role as a co-receptor. The extracellular domain of Ror2 associates with Wnt5a but not with Wnt3a. Furthermore, Ror2 mediates Wnt5a signaling by activating the Wnt/JNK pathway and/or inhibiting the β-catenin/Tcf pathway. It has also been shown that Ror2 interacts with filamin A and that

it mediates Wnt5a-dependent cell migration *(Nishita et al., 2006).* Ryk is a single-pass transmembrane RTK and Ryk can interact at least with Wnt1 and Wnt3a *(Lu et al., 2004).* Ryk family members have been shown to be required for Wnt signaling in several contexts. For example, knockdown of Ryk reduces the Wnt1-dependent TCF activation in HEK-293 cells *(Lu et al., 2004).* However, whether Ryk mediates Wnt signaling in concert with Fz-LRP5/6 or independently is not clear and also how Ryk activates the intracellular signaling cascade after binding to Wnt ligands has not yet been uncovered.

WNT negative regulators sFRPs (Secreted Frizzled-Related Proteins)

The sFRPs are a group of Wnt-binding glycoproteins that resemble the transmembrane receptor FZD. Their actions are mainly considered to be inhibitory to Wnt activity, however, there are also some reports showing their actions stimulatory to Wnt activity at low concentrations *(Uren et al., 2000).* There are presently eight known members of the family, sFRP1 to sFRP5, Sizzled, Sizzled2 and Crescent. On the basis of sequence homology, sFRP1, sFRP2 and sFRP5 form a subgroup, as do sFRP3 and sFRP4, which are quite distantly related to the other sFRPs. Sizzled, Sizzled2 and Crescent form a third group, but they have not been identified in mammals and Drosophila *(Kawano & Kypta, 2003).* All sFRPs are secreted and derived from unique genes and none are alternate splice forms of the FZD family *(Jones & Jomery, 2002).* They share sequence similarity with the Frizzled receptor CRD (cysteine rich domain), but lack the transmembrane and intracellular domains *(Leyns at al., 1997; Wang et al., 1997; Finch et al., 1997).* Through its CRD, sFRP exhibits the ability to bind Wnt. Furthermore, the CRD of sFRP1 also

appears to interact with itself to make dimmers or multimers and with FZD *(Bafico et al., 1999).* Thus, sFRPs may block Wnt signaling either by interacting with Wnt proteins to prevent them from binding to FZD receptors or by forming nonfunctional complexes with FZD receptors *(Rattner et al., 1997; Lin et al., 1997; Bafico et al., 1999).* Human sFRP1 is also known as SARP2 (secreted apoptosis-related protein 2) and FrzA (Frizzled in aorta). sFRP1 has been reported to bind to Wnt1 *(Dennis et al., 1999; Bafico et al., 1999),* Wnt2 *(Bafico et al., 1999; Xu et al., 1998),* Wnt8 *(Jaspard et al., 2000),* Wnt4 and Wnt3a *(Hussain et al., 2004).* However, it does not bind to Wnt5a *(Dennis et al., 1999; Xu et al., 1998)*. In any event, in binding to Wnts, sFRP1 would seem to act primarily as an inhibitor of Wnt signaling *(Dennis et al., 1999; Uren et al., 2000)*. sFRP1 binding to Wnt1 is reported to be antagonistic to Wnt activity *(Bafico et al., 1999).* sFRP1 also has been reported to protect cells from apoptosis, but this may be context dependent *(Barandon et al., 2003; Hussain et al., 2004; Roth et al., 2000; Dufourcq et al., 2002).* Other functions associated with sFRP1 include endothelial cell migration and capillary tube formation *(Dufourcq et al., 2002),* myofibroblast recruitment and collagen deposition, and a sFRP-induced decrease in MMP-9 activity *(Barandon et al., 2003).*

WIF1 (Wnt-inhibitory factor 1)

WIF1 is a Wnt binding protein secreted by variety of tumors and embryonic tissues. WIF1 has an N-terminal signal sequence, a unique WIF domain (WD) that is highly conserved across species, and five EGF-like repeats. Although WIF1 does not share any sequence similarity with the CRD domain of FZD or sFRPs, it can bind to Wnt ligands

(Kawano & Kypta, 2003). It apparently does so by forming a non-covalent complex with Wnt8 and Wnt1 *(Hsieh et al., 1999).*

Dkks (Dickkopfs)

The Dkk family comprises four structurally-related members (Dkk1 to Dkk4) and a unique Dkk3-related protein named Soggy (Sgy), which possesses homology to Dkk3. Dkks contain two characteristic cysteine-rich domains separated by a linker region of variable length *(Glinka et al., 1998; Krupnik et al., 1999)*. Dkk1 is a negative regulator of Wnt-mediated LRP signaling. Dkk1 interacts with LRP5/6 and a single-pass transmembrane proteins Kremen1 (Krm1) and Kremen2 (Krm2), which are endocytosable molecules *(Mao et al., 2002).* Using these interactions, Dkk1 can form a "bridge" between LRP and Kremen leading to the endocytosis of Kremen accompanied by internalization of Dkk/LRP. This internalization blocks LRP deactivation/destabilization of axin and results in the phosphorylation / degradation of β-catenin *(Kawano & Kypta, 2003; Caricasole et al., 1999; Nakamura et al., 2001).* Thus, Dkk1 acts as a negative regulator of Wnt/ β-catenin signaling, but not PCP signaling or calcium signaling.

β-catenin (CTNNB1)

In the late 1980s, β-catenin was independently discovered twice, on the basis of its different functions: structural and signalling. The group of Rolf Kemler isolated β-catenin, together with two other molecules (α-catenin and γ-catenin/plakoglobin), as proteins associated with E-cadherin, the key molecule of Ca2fl-dependent cell adhesion. These proteins were named catenins (in Latin catena means chain) to reflect their linking of E-cadherin to cytoskeletal structures *(Ozawa et al, 1989).* The signalling potential of β-catenin was exposed through its Drosophila orthologue Armadillo: the armadillo gene was discovered in the seminal screens for mutations affecting segmentation of the Drosophila embryo, performed by Eric Wieschaus, Christiane Nusslein-Volhard, and Gerd Jurgens *(Wieschaus et al, 1984).* Epistatic analysis determined that the armadillo segmentation function is regulated by Wingless *(Riggleman et al, 1990).* This finding was a key step in the subsequent characterization of the Wnt/β-catenin (or Wingless/ Armadillo, respectively) signalling cascade, and of the functions and mutual interactions of its individual components. Another important part of this mosaic was revealed by the description of the basic pathway leading from the Wingless ligand through Dishevelled to regulation of Armadillo stability by Shaggy/Zeste-white-3 (GSK3 in vertebrates) *(Siegfried et al, 1994).* Finally in the mid-1990's several groups independently found that the signalling function of β-catenin/Armadillo in the nucleus is mediated via T-cell factor (TCF)/Lymphoid enhancer-binding factor (Lef) transcription factors, which in association with β-catenin trigger Wnt-mediated transcription *(Behrens et al., 1996; Huber et al, 1996; Molenaar et*

al, 1996; Brunner et al, 1997; van de Wetering et al, 1997). The β-catenin protein (781 aa residues in humans) consists of a central region (residues 141–664) made up of 12 imperfect Armadillo repeats (R1-12) that are flanked by distinct N- and C-terminal domains, NTD and CTD, respectively. A specific conserved helix (Helix-C) is located proximally to the CTD, adjacent to the last ARM repeat (residues 667–683) *(Xing et al, 2008).* The NTD and the CTD may be structurally flexible, whereas the central region forms a relatively rigid scaffold. This scaffold serves as an interaction platform for many β-catenin binding partners, at the membrane, in cytosol, and in the nucleus *(Huber et al, 1997).* β-Catenin is a founding member of the Armadillo (ARM) repeat protein superfamily. Each ARM repeat of its central region comprises B42 residues, forming three helices arranged in triangular shape. Together, all ARM repeats form a superhelix that features a long, positively charged groove. Biochemical and crystal structure analyses revealed that many of β-catenin's binding partners share overlapping binding sites in the groove of the central β-catenin region: consequently, these partners cannot bind to β-catenin simultaneously. This mutual exclusivity is certainly valid for the key β-catenin interacting molecules: E-cadherin (the main partner in adherens junctions), APC (the main partner in the destruction complex), and TCF/Lef (the main partner in the nucleus). All these β-catenin interactors bind to the core binding site comprising ARM repeats R3-R9, where they form salt bridges with two key amino-acid residues, Lys312 and Lys435. Other ARM repeats are also involved, at least in strengthening the interaction *(Graham et al, 2000; Huber & Weis, 2001; Poy et al, 2001).* Free β-catenin is recognized by the key scaffold molecules Axin and APC, both of which can directly interact with β-catenin and also inter se. The scaffold establishes a platform for

associated kinases to phosphorylate β-catenin *(Kimelman & Xu, 2006; Roberts et al, 2011).* CK1a phosphorylates β-catenin at Ser45, priming the sequential phosphorylation of Thr41, Ser37, and Ser33 by GSK3 (preferentially by GSK3β) *(Liu et al, 2002; Xing et al, 2003).* As a next step, β-catenin bound to APC leaves the destruction complex and interacts with the ubiquitin machinery. Ser33 and Ser37 phosphorylated β-catenin is recognized by β-TrCP and recruited to the Skp1/Cul1/Fbox/ β-TrCP (SCF β-TrCP) E3 ubiquitin ligase complex. Ubiquitin-conjugated β-catenin is subsequently degraded by the 26S proteasome *(Hart et al, 1999)*. Alternatively, phospho-β-catenin can be ubiquitinylated by the single unit E-3 ligase Jade 1 *(Chitalia et al, 2008).* The scaffolding proteins Axin and APC are essential for the GSK3-mediated phosphorylation of β-catenin: although GSK3 can modify a plethora of different proteins within a cell as a free molecule, it modifies β-catenin only if it is associated with Axin and APC *(Hur & Zhou, 2010; Wu & Pan, 2010).* The rate limiting protein Axin greatly enhances the activity of GSK3 on β-catenin *(Dajani et al, 2003).* APC contributes to the establishment of the destruction complex, and stabilizes β-catenin's phosphorylation status. If N-terminally phosphorylated β-catenin is not associated with APC, after leaving the destruction complex, then it is immediately dephosphorylated by PP2A *(Su et al, 2008).* Activation of Wnt signalling leads to the disassembly of the β-catenin destruction complex and GSK3 activity is blocked. The ubiquitously expressed multi-functional protein and proto-oncogene β-catenin fulfils important functions in cell-cell adhesion by linking E-cadherin to the actin cytoskeleton, and in the canonical Wnt-signalling pathway by regulating cell proliferation and differentiation *(Funayama et al., 1995; Aberle et al., 1994).* β-catenin also plays an important role in interactions between

cadherins and transmembrane proteins, such as the epidermal growth factor receptor. In the absence of growth or differentiation signals free cytoplasmic β - catenin is rapidly turned over, initiated by phosphorylation of its amino terminus (Ser-33, Ser-37, Thr-41, Ser-45). A multiprotein complex consisting of GSK-3/APC/Axin and other components regulates this phosphorylation and promotes subsequent binding of β-Trcp, ubiquitination and degradation of β-catenin by the proteasome pathway *(Rubinfeld et al., 1996; Munemitsu et al., 1996).* In melanoma cells, mutations of the β-catenin phosphorylation sites in exon 3 (Ser-33, Ser-37, Thr-41, Ser-45) resulted in stabilisation of the protein, cytoplasmic/nuclear accumulation and activation of transcription *(Rubinfeld et al., 1996).* Aberrant activation of the Wnt-signalling pathway, by stabilising β-catenin mutations in exon 3 was first described in sporadic colorectal cancers and melanomas at a frequency of 6-10%.

The nuclear complex

β-Catenin can dynamically shuttle between the cytoplasm and nucleus. Surprisingly, it does not contain any classical nuclear localization signal (NLS) or nuclear export signal (NES) within its polypeptide sequence. Indeed nuclear import of β-catenin was shown to occur importin-karyopterin independently *(Fagotto et al, 1998).* Recently, β-catenin was shown to directly interact with different nuclear pore complex components (NPCs; *Shitashige et al, 2008; Sharma et al, 2012).* By transiently and sequentially binding to different NPCs, β-catenin could pass through the nuclear pores. Of the various NPCs

interacting with β-catenin, Nup358 seems to be important for docking/undocking of β-catenin to nuclear pores during the process of nuclear translocation *(Sharma et al, 2012).* ARM repeats R10-12 were revealed as crucial for β-catenin nuclear import (and export). Moreover, Tyr654 in the last ARM repeat was supposed to affect β-catenin import, as a phospho-mimicking mutation strongly enhances nuclear import, indicating yet another way in which (p-Tyr654 can promote the signalling activity of β-catenin at the expense of its adhesive function *(Sharma et al, 2012).* There are a few other molecular mechanisms facilitating nuclear translocation of β-catenin. The Forkhead-box transcription factor FoxM1 was shown to directly interact with the ARM repeats R11-12 and thereby promotes β-catenin nuclear import in mammalian cells: immortalized neural stem cells or MEF's lacking FoxM1 display a strong reduction of nuclear β-catenin and thus reduced β-catenin signalling activity *(Zhang et al, 2011).* FoxM1 might provide its NLS to β-catenin and bipartite β-catenin-FoxM1 complexes could be imported to the nucleus. Alternatively, or additionally, FoxM1 could act as a nuclear anchor preventing export of β-catenin out of the nucleus (Figure 1.6).

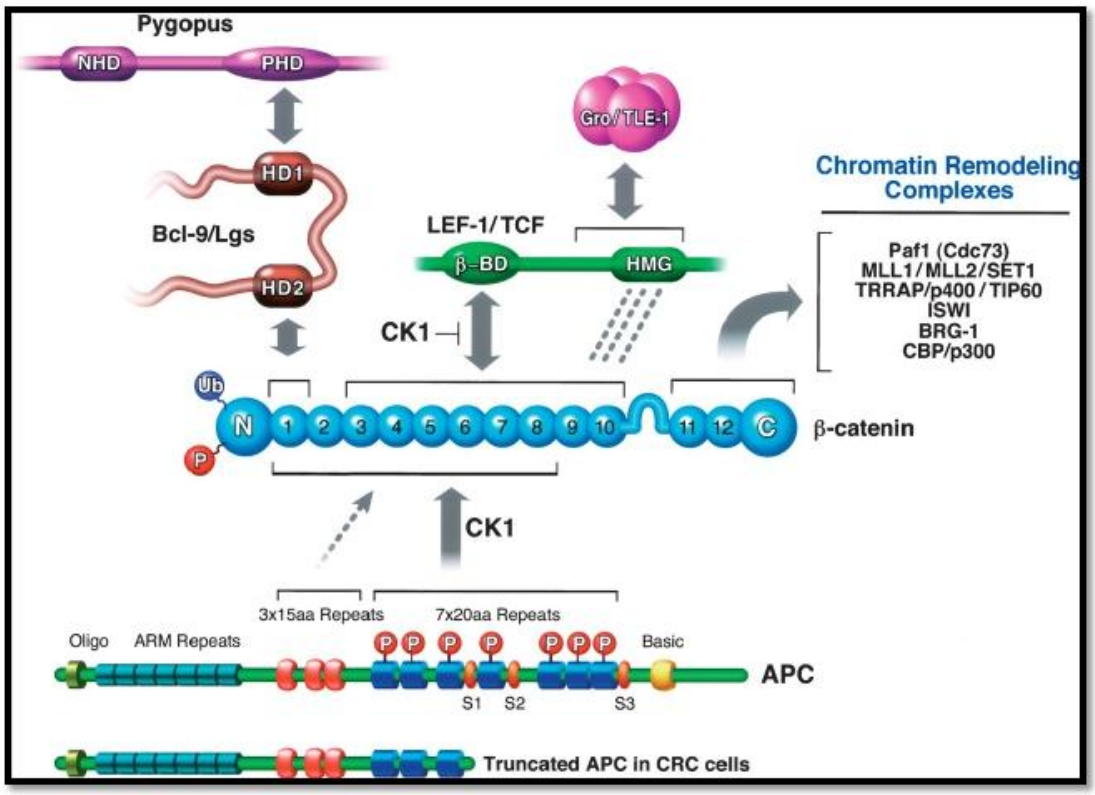


Figure 1.6 Schematic rapresentation of interactions between different Wnt transcriptional regulators. The schematic diagrams indicate the overlapping sites within the β-catenin armadillo (ARM) repeats for LEF-1/TCF and APC. Phosphorylation of APC by CK1 dramatically enhances its affinity for β-catenin (Ha et al. 2004; Xing et al. 2004), which would disfavor binding to LEF-1/TCF. Similarly, CK1 phosphorylation of LEF-1 has been reported to disrupt binding to β-catenin (Hammerlein et al. 2005). Other post-translational protein modifications, potentially including β-TrCP-mediated ubiquitination of β-catenin, are likely to regulate the protein interactions within the complex; however, the precise mechanism is unknown. The C terminus of β-catenin interacts, directly or indirectly, with several distinct chromatin remodeling complexes. Sequential interactions with these various complexes is likely to govern transcription initiation and elongation at Wnt target genes*. (Willert and Jones, 2006).*

Once in the nucleus β-catenin can activate transcription of Wnt/β-catenin target genes. Since β-catenin does not possess a DNA binding domain it needs DNA binding partners tobring it to the promoters of its target genes *(Huber et al, 1997; Huber et al., 2001; Xing et al, 2008).* Hence, β-catenin initiates transcription onlyas a member of bipartite or multimeric complexes wherein one partner provides association with specific responseelements on target genes (e.g., Wnt response elements, WREs) and β-catenin acts as the central transcriptional activator. TCF/Lef transcription factors serve as the main nuclear partners of β catenin guiding it to specific DNA loci. Within the coactivator complex, β-catenin functions as a scaffold to link the LEF-1/TCF proteins to specific chromatin remodeling complexes, as well as to the Wnt coactivators, Bcl-9/Lgs and Pygopus. The N-terminal armadillo (ARM) repeat of β-catenin interacts directly with Bcl-9/Lgs, which forms part of a "chain of adaptors" *(Stadeli et al., 2006)* that connects LEF-1 to the Pygopus (Pygo) PHD finger protein *(Belenkaya et al. 2002; Kramps et al. 2002; Parker et al. 2002; Thompson et al. 2002).* Bcl-9/Lgs and Pygopus are implicated in nuclear localization of β-catenin *(Townsley et al. 2004)* as well as transcription *(Thompson 2002; Hoffmans et al. 2005)*. LEF-1/TCF proteins bind to the β-catenin central ARM repeats in a region that largely overlaps the binding sites for APC or E-cadherin. Other β-catenin- interacting proteins include the DNA ATPase/helicase TIP49a/Pontin52 *(Bauer et al. 2000),* which is present in mammalian TRRAP/p400/TIP60, INO80, and SWRCAP/SWR1 chromatin remodeling complexes, Brg-1, and CBP/p300 (Figure 1.7).

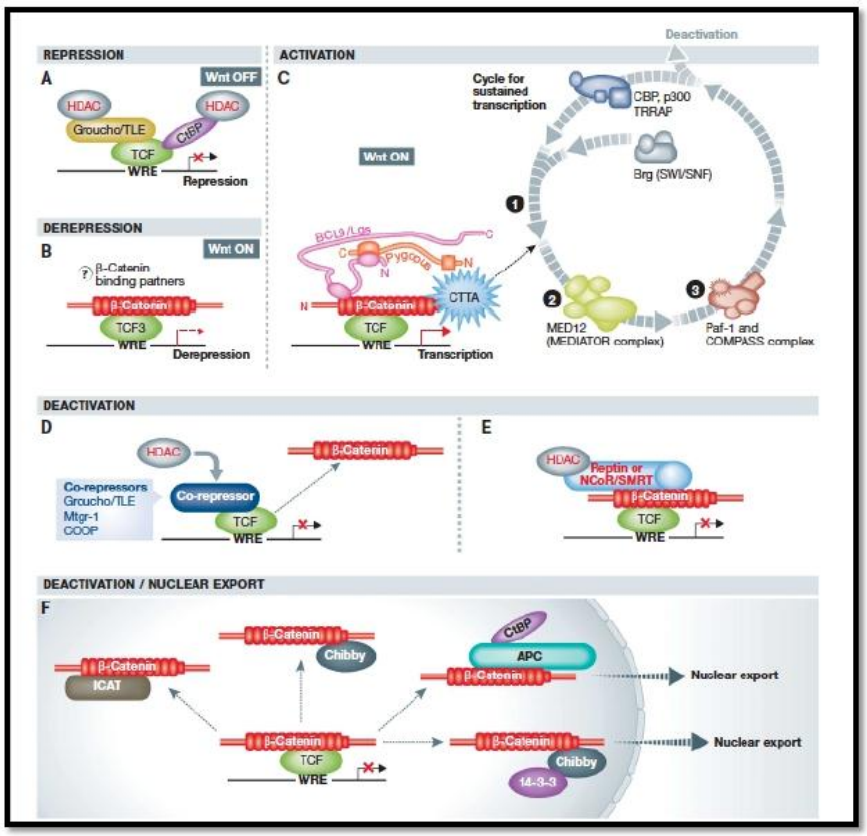


Figure 1.7 The TCF/b-catenin transcriptional switch. (A) In an unstimulated (Wnt OFF) situation, TCF/Lef transcription factors associated with Groucho and/or CtBP co-repressors act as transcriptional repressors. (B) The binding of b-catenin displaces co-repressors, thereby promoting transcriptional derepression *(Wray et al., 2011).* (C) b-Catenin converts TCF/Lef into transcriptional activators providing an interacting platform for the multitude of dynamically cycling transcriptional co-activators. (D) TCF/b-catenin-mediated transcription may also be deactivated by transcriptional corepressors, such as Mtgr-1, COOP, or Groucho/TLE kicking off b-catenin. (E) TCF/b-catenin-mediated transcription may also be deactivated by direct interaction with transcriptional co-repressors (Reptin or NCoR/ SMRT) recruiting histone deacetylases. (F) Chibby with 14-3-3 protein or APC with CtBP can sequester b-catenin away from target promoters and export it out of the nucleus. Binding of ICAT or Chibby blocks the interaction between TCFs and b-catenin *(Valenta et al., 2012).*

1.5 HEMATOPOIETIC REGENERATION: TOO MUCH OF A WNT THING

Canonical Wnt signaling has been reported to play a major role in the modulation of the delicate balance between stemness and differentiationin several adult stem cell niches, including hair follicles in the skin, mammary gland, and intestinal crypt. It was reported that β-catenin activation is crucial for the development of cancer stem cells in various HMs, including AML *(Wang et al., 2010)*, mixed-lineage leukemia (MLL) *(Yeung et al., 2010)*, and CML *(Heidel et al., 2012*. Moreover, canonical Wnt signaling plays a crucial role in the maintenance and establishment of fetal HSCs, further indicating that the signaling pathway sustains LSC development. Evidence from fetal HSC indicates that the deletion of β-catenin during the development of these cells results in the impairment of self-renewal *(Heidel et al., 2012)*, and very recent studies have revealed that β-catenin is also crucial in adult HSC maintenance *(Ruiz-Herguido et al., 2012)*.Along with maintaining a steady state of hematopoietic cell production for the lifetime of an organism, hematopoietic stem cells(HSCs) also have the ability to proliferate rapidly and repair loss of the hematopoietic compartment in response to injury. Loss of cells due to events such as chemotherapy, infection, and radiation exposure ablates rapidly cycling cells, which includes much of the developing hematopoietic system. In response to such injury, HSCs initiate a program of rapid proliferation in order to regenerate the lost hematopoietic compartment. Presently, neither the changes in the extracellular microenvironment nor the intracellular signals activated by HSCs during the proliferative phase of regeneration are known. Furthermore, it is also unknown whether

the signals that modulate regeneration after injury are the same as or distinct from those implicated in the homeostatic maintenance and growth of HSCs and progenitors. Regeneration requires the rapid expansion of HSCs followed by differentiation of these cells into mature lineages. This process is mediated by crosstalk between cell intrinsic factors and external cues from the microenvironment. Any injury to the hematopoietic system results in a significant loss of cells and consequent changes in the microenvironmental milieu of HSCs. It is likely that HSCs sense these changes and in turn activate intracellular signals to initiate the proliferation process that ultimately leads to successful regeneration of the hematopoietic compartment. Studies examining the bone marrow environment following injury have revealed changes in mRNA levels of multiple growth factors, including stem cell factor, stromal cell derived factor 1, and transforming growth factor β1. Congdon et al., demonstrate that following injury the soluble fraction of the bone marrow microenvironment develops an enhanced ability to support HSCs and examining of the injured bone marrow microenvironment they revealed that both hematopoietic and stromal cells specifically upregulate their expression of WNT10B. Furthermore, regenerating HSCs show increased activation of the Wnt signaling pathway. In consideration that WNT10B functions as an HSC growth factor, Congdon et al., proposed that elevated levels of WNT10B after injury may serve to extrinsically activate the Wnt pathway in regenerating HSCs. These findings are the first to demonstrate that the canonical Wnt cascade is activated by regenerating HSCs and that this activation coincides with an increase in the microenvironmental availability of a specific Wnt ligand, thus providing novel insight into both the microenvironmental changes that occur after injury and how these changes are integrated with the regeneration of hematopoietic cells. It's interesting to note that that renewal signals are

reactivated during tissue injury lends support to the proposal that recurrent activation of renewal signals during tissue injury may form the basis of oncogenic transformation. Wnts are defined as a specific cue that is upregulated early after injury in the bone marrow suggests that Wnts and other, similar developmental pathways should be investigated as possible players in repair after injury in other organs and tissues as well. The Congdon's et al work, defined that after injury, the microenvironment changes to be more supportive of HSCs. According to Bowman's idea, developmental signal transduction pathways, such as BMP and Wnt, are often reactivated during regeneration, suggesting that link between the hematopoietic regeneration and developmental pathway. In addition, Wnt10b ligand an HSC growth factor, is increased by both stromal and hematopoietic cell subsets after an injury induced by Cy/G treatment. Concomitant with the increased expression of Wnt ligand, regenerating stem and progenitor cells demonstrate increased activation of the Wnt signaling pathway. According to Bowman et al., developmental signal transduction pathways, such as BMP and Wnt, are often reactivated during regeneration *(Bowman et al., 2012)*.

1.6 mRNA *IN SITU* DETECTION: LOCK AND ROLL

Many of the decisions that cells take concerning survival, growth and differentiation are reflected in altered patterns of gene expression. Messenger RNA transcripts, mRNA, play an important role in living cells as an intermediate step during the protein synthesis process. RNA detection methods span from the specific quantitative detection of a few transcripts of interest to the large-scale parallel detection of thousands of different RNA molecules. Methods detecting specific RNA transcripts such as northern blot, real-time PCR or the invader technique can be used to detect and quantify transcripts present in different tissues or cell types. mRNA expression can be studied globally with techniques such as reverse transcription quantitative polymerase chain reaction (RT-qPCR) *(Bustin et al., 2005),* microarrays or more recently sequencing technologies *(ten Bosch & Grody, 2008; Wang et al., 2009).* In the RT-qPCR the RNA is first reverse transcribed into its complementary DNA (cDNA) using a reverse transcriptase enzyme *(Chelly et al., 1990).* The RT-qPCR technology is easy to perform, capable of high-throughput analysis of up to hundreds of known transcripts at a time and can combine low LOD with reliable specificity *(Bustin, 2002).* Real-time detection of the ongoing PCR via the TaqMan assay *(Heid et al., 1996; Gibson et al., 1996)* exploits the nuclease activity of the Taq-polymerase *(Holland et al., 1991).* A probe complementary to the amplified target sequence, labeled with a reporter fluorophore at the 5’- end and a quencher fluorophore at the 3’- end *(Lee et al., 1993; Livak et al., 1995),* is added to the PCR. During the extension step Taq-polymerase degrades hybridized probes, releasing the 5’ fluorophore. The fluorescence from released fluorophores is proportional to the amount of degraded probes, and hence

the amplification cycle at which the fluorescence exceeds a threshold value can be taken as a measure of the amount of target present in the reaction. Alternative probe designs have been developed for real-time detection, such as molecular beacons *(Tyagi & Kramer, 1996)* and scorpion probes *(Whitcombe et al., 1999).* These probes contain reporter fluorophores and quencher molecules as in the TaqMan probes previously described, but are constructed so that degradation of the probes is not necessary. Furthermore, they can be used to distinguish transcript variants that differ in a single-nucleotide position *(Heid et al., 1996; Gibson et al., 1996).* The PCR step renders the Taqman method sensitive to false positive results due to contamination of the samples, and multiplexing is difficult, demanding labor-intensive optimizations. Biased results may arise from the reverse transcription step. This is avoided in the invader technique, since RNA is used as the target in the reaction. These assays can also be used for quantification of mRNA targets. To achieve proper quantification of mRNA expression by RT-qPCR the RNA should be of good quality, internal controls as well as standard curves should be included and samples should be normalized against relatively constantly expressed genes, so called housekeeping genes *(Bustin et al., 2005).* RT-qPCR experiments that rely on RNA extraction of sometimes complex tissue samples will give an average value from numerous variable subpopulations of cells of different lineage at diverse stages of differentiation. This average value can be misleading in attempts to compare mRNA expression levels between different individuals *(Bustin et al., 2005).* Moreover, comparison of gene expression patterns of housekeeping genes, in subpopulations of cells derived from the same individual, revealed differences in mRNA levels. This provides evidence that cellular subpopulations of the same origin are highly heterogeneous *(Goidin et al., 2001).* The microarray technology allows simultaneous

characterization of expression levels on a genome-wide scale and has been applied for not only detection of mature mRNA, but also non-coding RNAs *(Pozhitkov et al., 2007).* For gene expression profiling by microarray technology, transcripts are isolated, labeled and hybridized to thousands of probes that are attached to a solid surface. Unreacted targets (that are not bound to the DNA probes) are washed away and the remaining signals are detected and measured. However, array-based assays suffer from certain limitations such as unreliable detection of low abundant genes and cross hybridization which gives rise to unspecific signals *(Pozhitkov et al., 2007; Draghici et al., 2006).* This type of hybridization-based approach rely on prior knowledge about the genome sequence in contrast to sequence-based approaches that directly can determine the expression of novel transcripts, thereby allowing identification of previously uncharacterized genes. The limitations associated with traditional sequencing, i.e. Sanger sequencing, such as relatively low throughput and high costs, resulted in the development of tag-based methods, e.g. Serial Analysis of Gene Expression (SAGE) *(Velculescu et al., 1995)* that offer higher throughput and precise digital gene expression levels. However, disadvantages with this approach limit the use of it, such as laborious technical procedure and difficulties in resolving similar transcripts *(Velculescu et al., 1995; Morozova et al., 2009).* The development of a panel of next-generation sequencing technologies (e.g. 454/Roche, Illumina, SOLiD and Helicos) provided new transcriptome studies for gene expression profiling as well as for identification of genetic variants such as mutations, splice variants and fusion genes and was termed RNA-Seq (RNA sequencing) *(Wang et al., 2009; Morozova et al., 2009; Edgren et al., 2011).* Although it is still in the early stage of use, RNA-Seq is believed to have many advantages over previously described methods, such as the deep coverage and base level resolution.

However, the newly described sequencing technique is associated with some limitations or difficulties, such as nonuniformity of transcript coverage and transcript-length bias, which will be important to further advance RNA-Seq in becoming an invaluable tool for the characterization and quantification of the transcriptome *(Wang et al., 2009; Ozsolak & Milos, 2011).*

Single-cell studies of gene expression

Expression analysis of single genes in single cells is important, for example, for finding rare events in a sample *(Levsky & Singer, 2003)*, e.g. cancer cells that are hidden in a group of normal cells. Moreover, studies reveal that gene expression can be highly diverged, even within a clonal population of cells *(Kaufmann & van Oudenaarden, 2007),* and that genes are transcribed in bursts with long periods of expression inactivity *(Raj et al., 2006).* In these cases, single cell detection techniques that can identify cell-to-cell differences within a population become a preferable method of choice. Furthermore, multiplex in situ analysis is required to appreciate the interplay between different cells in a heterogeneous tissue and the respective transcript expression profiles. A plausible risk with bulk measurements is the limitation to see differences at the inter- and intracellular level and instead end up with false positives or negatives that represent the average value in that sample. Thus, the advantage of studying single molecules in individual cells is that it gives the correct frequency distribution of expressed molecules for single genes, yielding much more detailed information than can be gleaned from the mean value alone. The in vitro techniques described in the previous section use isolated

mRNAs from cells or tissues to determine the expression levels. However, the precise dissection of tissues might be difficult to attain without inadvertently including some irrelevant surrounding cells which can lead to false results. Exact sampling is especially important for diagnostic analysis where samples need to be as pure as possible, with no contamination from normal cells, to prevent uncertain or incorrect results. One method that offers mRNA as well as protein expression analysis is laser-capture microdissection (LCM) of single-cells isolated from a certain location in a heterogeneous tissue *(Emmert-Buck et al., 1996).* TMA's are produced by punching out samples from selected tissue-regions and distributing them on a single slide *(Kononen et al., 1998).* The TMA technology has been widely applied within the field of cancer research for diagnostic and drug target discovery *(Kallioniemi et al., 2001; Sugimura et al., 2010).* However, LCM has a number of potential drawbacks. The procedure is expensive, time-consuming and limited to amplification-based techniques *(Curran et al., 2000).* Moreover, another major limitation is the need to identify the cells of interest based on morphologic characteristics, which in turn, requires a trained histologist or pathologist *(Liu, 2010).* In situ analyses can achieve precise and spatial localization within morphological preserved cells or tissues as they occur in their natural situation. Studying tissues can also give comprehensive information of the origins of the different cell types and find regions containing cells of similar characteristics. A technique for in situ hybridization was first described in 1969 for detection of ribosomal DNA. At first radio-labeled probes were used, however many non-isotopic labeling variations have been developed. Non-radioactive hybridization methods can be divided into two groups: direct and indirect.

- The **direct method**: the probe is bound directly to the target molecule so the resulting hybrid can be visualized in a microscope immediately after

hybridization. This can be accomplished by introducing labeled nucleotides (fluorophores) to the probes.

- The **indirect method**: the labeled probe is not visualized directly. Instead, a reporter molecule is bound to the label after hybridization which enables the visualization of the target in a microscope. Commonly used labels for indirect approaches are streptavidin and digoxigenin.

The direct method is simple and fast and best suited for detection of repetitive sequences and multicopy genes, whereas the indirect method is more labor intensive and instead suitable for low-copy target sequences.

Performing an experimental assay

In order to perform an essay, there are features, that must be considered:

- **Sensitivity** is the capability of a method to discriminate between small differences in concentration of target molecule in a sample. The assay sensitivity is measured by the ability of the assay to recognize all positive cases as such. Extending this definition for methods used to detect certain molecules, the sensitivity of a method is measured by its ability to detect every molecule present in a reaction.
- **Limit of detection (LOD)** describes the lowest detectable concentration of analyte in a sample. Limit of detection (LOD) is the minimum amount of molecules that an assay can detect significantly above background signal. Depending on the level of significance that is required, the LOD can be calculated as, for example, the concentration of a certain analyte that corresponds to a signal

that is two or three standard deviations above background signal.

- **Specificity** describes how efficient the assay targets the correct analytes in a sample, i.e. that it only detects the molecules of interest. The assay specificity is determined by the ability of the assay to recognize all negative cases as such. As with sensitivity, specificity in the context of detection methods refers to the ability of the method to detect only the correct molecules.
- **Selectivity,** defines the method's ability to distinguish between closely related targets.
- **Dynamic range** is the ratio between the highest and lowest measureable amount of a changeable quantity. The dynamic range of an assay (or the linear dynamic range) is the range of protein concentrations that lies between the LOD and the point of saturation, at which point the greatest possible amount of protein has been detected.
- **Coefficient of variation (CV)** is defined as the ratio of the standard deviation(σ) to the mean (μ). It applies only for non-zero means.

mRNA in situ detection

To increase the sensitivity of traditional in situ hybridization, target amplification strategies involving in situ polymerization have been developed. Over two decades ago, in situ PCR was developed for detection of DNA molecules. The method conducts PCR directly on cells and tissues with elongation of sequence specific primers and amplification in a conventional thermal cycler. The amplified targets can be detected

directly with labeled nucleotides or indirectly via in situ hybridization of labeled target specific-probes, which is more specific and therefore preferred. The ligase-based analysis, is characterized by:

- **High specificity:** pairwise hybridization to targets
- **High selectivity:** enzymatic substrate recognition
- **Efficient linking:** affinity and covalent interaction
- **Amplifiable product:** formation of a new strand of DNA
- **Carries information:** DNA tags included

Further modifications of the technique lead to in situ RT-PCR for detection of RNA (Figure 1.8). In practice, however, the method is associated with many problems, such as low amplification efficiency, poor reproducibility, sensitivity and specificity as well as problems with high background, which makes the practical application limited. A similar technique, also based on in situ polymerization, is the primed in situ labeling (PRINS) procedure in which an unlabeled, target-specific probe is hybridized and used as primer for chain elongation in situ using Taq polymerase and labeled nucleotides. Although PRINS has the specificity to discriminate between single nucleotide differences it cannot detect low copy-number sequences due to low LOD. In situ PCR and PRINS are target amplification-based techniques that were developed to address the need to amplify and detect targets in situ and represent rapid and relatively inexpensive alternatives to some in situ hybridization applications. However, due to the limitations described above none of these methods is suitable for detection and visualization of low-copy single nucleotide variants in situ. The first application of padlock probes in the in situ setting was for detection and genotyping of centromeric sequences in chromosomes 13 and 21 *(Lizardi et al., 1998).* In this study padlock probes labelled with biotin and digoxigenin were

detected using labelled streptavidin and antibodies. Detection of ligated padlock probes using antibodies suffered from a lack of sensitivity, preventing studies of rare targets due to background generated by unspecific binding of the detection reagents. Introduction of the rolling circle amplification mechanism increased the specificity of detection by increasing the signal strength and decreasing the unspecific background *(Nilsson et al., 1997)*.

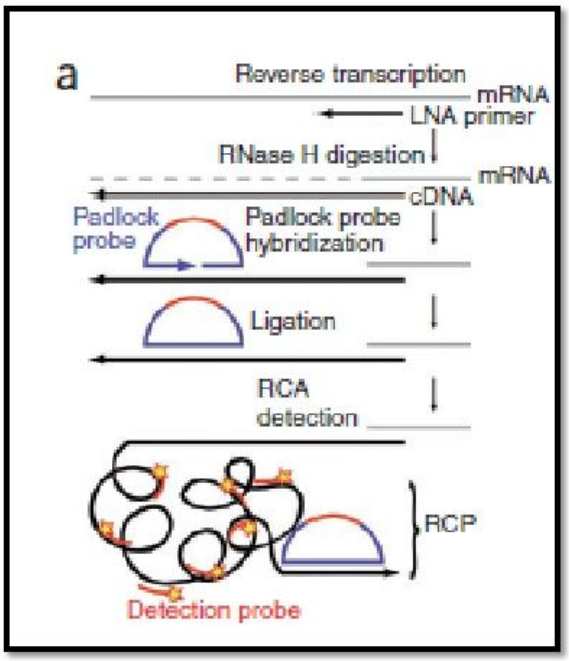


Figure 1. 8 Schematic representation of the mRNA in situ detection. cDNA is created using locked nucleic acid (LNA)-modified primers and is probed after degradation of mRNA by RNase H. RCPs are identified through hybridization of fluorescent detection probes *(Larsson et al., 2010)*.

Locked nucleic acids (LNA)

Locked nucleic acids (LNA) are another type of nucleic acid analogue that has exceptional hybridization affinity towards complementary DNA and RNA molecules. The synthetic LNA molecule contains a methylene bridge on the ribose ring between the 2´-oxygen and the 4´-carbon thereby locking the structure into a high binding-affinity with reduced conformational flexibility (Figure 1.9).

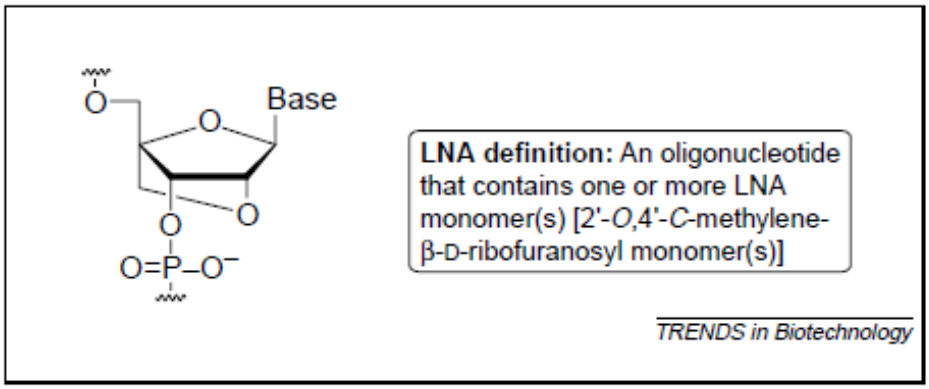


Figure 1.9 The chemical structure and definition of locked nucleic acid (LNA). *(Petersen & Wengel, 2003).*

An LNA-DNA duplex provides a substantial increase in thermal stability with the ability to increase the Tm of an oligonucleotide with +1 to +8 °C for DNA and +1 to +10 °C for RNA per LNA monomer introduced. Furthermore, LNA probes have high discriminatory power between matched and mismatched sequences which make them well suited for sensitive nucleic acid detection. LNA-modified oligonucleotides are fully soluble in water, which simplifies experimental implementation. LNA- modified probes have been used in

various applications including in vitro discrimination of single-base mismatches, FISH-studies for detection of repeated genomic sequences and miRNAs and for whole mount in situ hybridization detection of mRNAs. In the Nilsson's work, the effect of LNA base incorpaporation in the primer for cDNA synthesis in situ, was evaluated. Primers with every second base at the 5'-end substituted with LNA performed better than primers with substitutions of every third base. Primers with five, seven or nine LNA bases in total were also investigated and it was found that adding nine LNA bases resulted in a small decrease in the amount of signals in situ. To ensure that the LNA would not interfere with the ability of the reverse transcriptase to synthesize cDNA from the primer, LNA bases were placed on the 5' -side of the primers, leaving the 3'-end unmodified.

The Padlock probe

The concept with padlock probes was invented two decades ago and is an extension of the oligonucleotide ligation assay (OLA). Padlock probes have many advantageous characteristics and offer highly selective detection of DNA and RNA in solution and in situ. Padlock probes can be synthesized using standard solid-phase chemical synthesis or via a novel PCR-based method. Padlock probes are linear oligonucleotides of approximately 70 to 100 nucleotides in length with target-complementary 5 ´ - and 3'-ends which constitute dual target recognition when both probe arms must hybridize correctly to the target. This property allows for highly multiplex assays with limited cross-reactivity between probes. When the padlock probes hybridize to their correct target the ends of the padlock probe are brought together in a head to tail orientation,

with only a nick in between. The nicks can be sealed by a DNA ligase creating circles that are locked onto the target strands as padlocks. This nick ligation will only occur if there is a perfect match between probe and target at the ligation junction, leaving allelic probes linear and unamplified. Padlock probes are excellent tools to detect single-nucleotide variants in RNA and DNA. This is because the simultaneous hybridization of the two target-complementary segments ensures a high specificity in recognizing a target sequence. The discriminating power of the padlock probes is increased by the fact that the ligation reaction is strongly inhibited by any mismatches at the ligation junction, especially at the 3' ends of the hybridized probes *(Larsson et al., 2004; Nilsson et al., 1994; Larsson et al., 2010)* (Figure 1.10). Distinction of single-nucleotide sequence variants is therefore possible with padlock probes. The catenation of the probes to the correct target sequence after hybridization and ligation renders the probe resistant to stringent washes if the probes are bound between two points of the target molecule that are attached to a solid phase, reducing the background signal from non-specifically hybridized probes. After the circularization, the probes can act as template for amplification using PCR or rolling-circle amplification (RCA). Moreover, the arising probe/target duplex becomes topologically locked and will thereby resist extreme washes, which reduce the amount of non-specific signals.

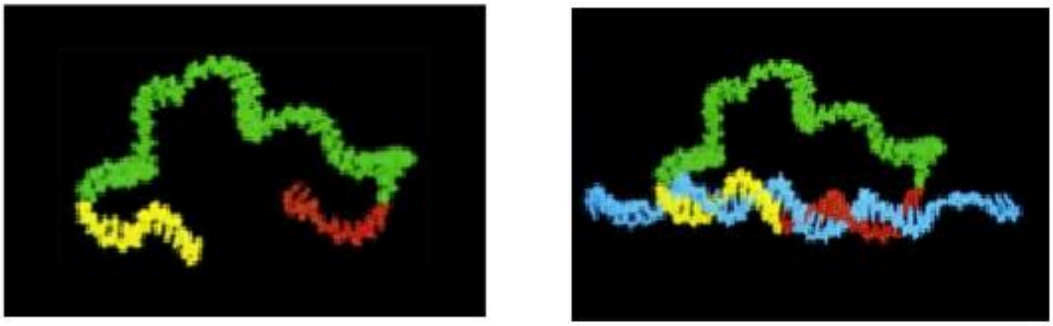


Figure 1.10 Schematic rapresentation of padlock probe ligation. Adapted from *Nilsson et al., 2004.*

DNA ligases: Ampligase *vs* T4 DNA ligase

DNA ligases were first isolated in the late 1960s *(Weiss et al., 1968; Zimmerman et al., 1967; Gellert, 1967)* and applied for molecular cloning *(Jackson et al., 1972)* and synthesis of long oligonucleotides *(Agarwal et al., 1970).* In their cellular environment DNA ligases catalyze the joining of Okazaki fragments produced during the replication of genomic DNA, forming the lagging strand, and repair nicks that arise during DNA damage repair and DNA recombination. DNA ligases have been isolated from a number of organisms, and share general reaction mechanism, but differ in substrate requirements and temperature optima. All ligases contain a lysine residue in their active site. Depending on the origin of the enzyme, this lysine is adenylated by reaction with either ATP or NAD+. The ligation reaction begins when the adenylated ligase enzyme

binds to a single-stranded break (with a phosphorylated 5' end) in double-stranded DNA. The adenylyl group is transferred to the 5' phosphate, which is in turn attacked by the 3' hydroxyl group, releasing the adenylate. The enzyme is released and is ready for another ligation after acquiring a new adenlyate charge *(Lehman, 1975).* Ligases isolated from thermophilic organisms have higher optimal reaction temperatures, and can be used for thermocycled ligation reactions *(Abravaya et al., 1995).* Higher reaction temperatures also destabilize transiently hybridized imperfectly matched DNA complexes, further decreasing the risk of chance ligation events. The substrate specificity of DNA ligases ensures that only nicks with correctly matched base pairs are closed. This mismatch discrimination is due to the structure of DNA ligases, whose footprint covers several bases upstream and downstream of the nick position *(Doherty & Dafforn, 2000).* DNA ligases catalyze the formation of phosphodiester bonds between the adjacent 3’-hydroxyl and 5’-phosphate termini at breaks in one or both strands of a DNA duplex. DNA ligases can be divided according to the cofactor requirements of the enzymes. The known eukaryotic, archebacterial and viral ligases all require ATP as a cofactor and are monomeric proteins ranging from 30 to >100 kDa. The eubacterial ligases are a more homogenous group of proteins in the range of 70-80 kDa, and they require NAD+ as a cofactor *(Higgins & Cozzarelli, 1979; Engler & Richardson, 1982; Doherty & Suh, 2000).* ATP- and NAD+- dependent DNA ligases share six conserved motifs involved in building up the active site of the enzymes *(Doherty & Dafforn, 2000).* The ligation reaction mechanism can be divided into three steps. First, the ligase is activated through the formation of a covalent protein-AMP intermediate with the concomitant release of PPi or NMN, depending on the cofactor. The AMP molecule is bound to the ε-amino group on the lysine residue in the conserved KXDG motif. After

the recognition of and binding to a nick, the AMP molecule is transferred to the phosphorylated 5' end of the nick. In the third and final step, the enzyme catalyzes a nucleophilic attack of the adjacent 3' hydroxyl group on the pyrophosphate bond between the AMP and the 5' phosphate. The subsequent formation of a phosphodiester bond seals the nick, and releases the AMP and the ligase *(Higgins & Cozzarelli, 1979; Doherty & Suh, 2000; Rossi et al., 1997).* There are many commercially available ligases for molecular biology, but mainly two DNA ligases are used for padlock probe ligation. The first is the bacteriophage T4 DNA ligase, which is an ATP-dependent ligase most active around 37 °C. The second ligase is the thermally stable Thermus thermophilus (Tth) ligase, also known as Ampligase, which joins DNA nicks in double stranded DNA through an NAD+dependent reaction. T4 DNA ligase is derived from bacteriophage T4. Tth ligase originates from the eubacterium Thermus thermophilus. T4 DNA ligase has two temperature optima for ligation, 28°C and 37°C, while Tth ligase is thermostable and has an optimal temperature range between 65-72° C for nick closure. The optimal pH for nick closure ranges between 7.2-7.8 for T4 DNA ligase, and 8.5 for Tth ligases *(Wu & Wallace, 1989; Engler & Richardson, 1982; Tong et al., 1999).* Both of these enzymes require divalent cations present in the active site during the ligation reaction, probably in a similar configuration that has been proposed for T7 and Tfi ligase *(Doherty & Suh, 2000).* In general, Mg2+ is used as the divalent metal ion for ATP-dependent and NAD+-dependent ligases. Slightly increased ligation rates have been observed with the T4 DNA ligase when Mn2+ was used *(Engler & Richardson, 1982).* Tth ligase exhibited about 6 fold higher mismatch ligation rate with Mn2+ compared to Mg2+ *(Tong et al., 1999).* DNA ligases need a certain length of double-stranded DNA, referred to as a footprint, for nick ligation. For the T7 ligase, the footprint has been shown to be asymmetrical, spanning 7-

9 and 3-5 bases at the 5’ and 3’ ends of the nick, respectively *(Doherty & Dafforn, 2000).* T4 DNA ligase and Tth ligase are more discriminating toward mismatches at the 3’ end than the 5’ end of the ligation junction *(Landegren et al., 1988; Wu & Wallace, 1989; Pritchard & Southern, 1997).* T4 DNA ligase mismatch discrimination can be enhanced further by a NaCl concentration of 200 mM in the ligation reaction *(Landegren et al., 1988; Wu & Wallace, 1989).* Some ATP-dependent ligases, such as T4 DNA ligase, also have the ability to ligate DNA hybridized to RNA *(Kleppe et al., 1970; Fareed et al., 1971).* Ampligase is active at high temperatures, enabling stringent hybridization of oligonucleotides, and has been found to have significantly higher ligation specificity than T4 DNA ligase *(Luo et al., 1996).* Although specificity is generally higher with Ampligase, T4 DNA ligase has the advantage of working at low temperatures which makes it suitable for some padlock probe applications. Ampligase® Thermostable DNA Ligase catalyzes the NAD-dependent ligation of adjacent 3’-hydroxyl and 5’- phosphate termini in duplex DNA structures. The half-life of Ampligase DNA Ligase is 48 hours at 65°C and more than 1 hour at 95°C. Ampligase DNA Ligase has no detectable activity on blunt ends or RNA substrates. Ampligase DNA Ligase does not replace T4 DNA Ligase in most conventional cloning applications because it has: i) no activity on blunt ends; and ii) low activity at temperatures where 2- and 4-base cohesive ends form stable duplexes.

Rolling-circle amplification (RCA)

The rolling circle replication mechanism is used by many plasmids and bacteriophages for replication of their circular genomes. The replication process is isothermal and usually requires DNA polymerase, a primer to initiate the replication, DNA nucleotides and also DNA binding and unwinding proteins. The rolling circle replication (RCR) mechanism, or also called rolling circle amplification (RCA), can also function well with small DNA or RNA circles as templates in in vitro reactions, producing many copies in tandem of a sequence complementary to the initial circle *(Fire & Xu, 1995; Liu et al., 1996; Daubendick et al., 1995).* Fire & Xu used several commercially available DNA polymerases and DNA circles as small as 26 nucleotides and generated RCR products consisting of many copies of the complementary circular template. RCR products longer than 12000 nucleotides were obtained, which means that the enzyme traveled at least 280 times around the circle before dissociating. This rolling circle replication does not require DNA binding or helix-unwinding proteins in the polymerization process. Since the contiguous rolling circle products (RCPs) will by nature collapse into micrometer-sized DNA-bundles, RCA is highly suitable for localized detection. The RCPs become detectable in a fluorescence microscope by the local enrichment of short fluorescent probes that hybridize to the detection sites of the coiled RCPs. The enzyme used in RCA, Φ29 DNA polymerase *(Bacillus subtilis)*, possesses several important features which make it most suitable for the efficient amplification of circular DNA molecules *(Blanco & Salas, 1985; Esteban et al., 1993).* The polymerase carries a 3´ to 5´ exonuclease activity that enables proofreading of the newly synthesized DNA. This feature also becomes handy for initiation of polymerization in cases when no complementary primer is added

and should instead be primed by the target *(Blanco & Salas, 1985).* The Φ29 DNA polymerase is a DNA dependent polymerase which synthesizes DNA with high fidelity due to efficient proofreading hoice for performing RCA in situ. During polymerization, one polymerase molecule can synthesize more than 70 kb DNA in length. The polymerase has a strand displacement activity which could be of importance when replicating a circular molecule. It has been argued however that it is more likely the bending force associated with replication of small circular molecules such as padlock probes that cause the strand displacement needed for efficient replication of these molecules with RCA. The Φ 29 DNA polymerase can further also use RNA as primer for DNA synthesis, which is of special interest in RNA analysis with padlock probes. In order for the RCA to begin, the target strand must have a free end close to where the padlock probe hybridizes. This is because the reaction is inhibited by the topological link formed between the padlock probe and its target sequence. If a nearby free 3´ end is introduced close to the padlock probe site at the target DNA strand, this inhibition is circumvented and the reaction can proceed efficiently. This variant of RCA has been named target-primed RCA. Once the appropriate primer is created and in place the Φ29 DNA polymerase can then switch its activity and starts incorporating nucleotides in 5´ to 3´ direction. After replicating the entire circular padlock probe sequence, the polymerase will reach the priming site again. The polymerase then displaces the newly synthesized strand and goes on for another round. This strand-displacement activity will keep the amplification continuously going, creating longer and longer RCPs, until inactivation of the polymerase or change of temperature. After approximately one hour of amplification the resulting concatemeric product is roughly one μm in diameter and contains about 1,000 copies of the original padlock probe (Figure 1.11). This RCP can then be labeled by

hybridization of complementary fluorescent detection oligonucleotides and is easily detected and visualized as a bright spot in a fluorescence microscope.

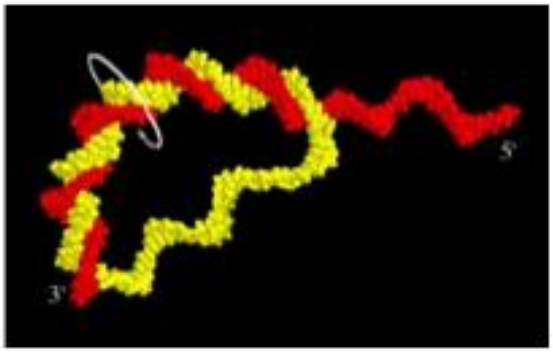


Figure 1.11 A rolling circle product (RCP). Padlock probe amplification with RCA creates a long linear molecule that spontaneously coils up to a DNA-bundle. The RCP becomes detectable by hybridization of fluorescence labeled probes that is visualized as a bright spot (square) in a cell.

1.7 IMAGE ANALYSIS TAKES OFF: EXTRACTING RICH INFORMATION FROM DIGITAL IMAGES

The rapid growth in digital imaging techniques associated with light microscopy allows researchers from the fields of biology and medicine to produce large amounts of image data in a variety of experiments. This sometimes overwhelming amount of image data needs to be handled carefully to allow the extraction of the required information in a resourceful manner. Thus the role of image analysis is not limited only to the analysis of the acquired image. It is always better for the image analysis to start with "good quality images" rather than trying to "make them good" for processing later. While numerous commercial and free software packages exist for image analysis, many of these packages are designed for a very specific purpose, such as cell counting. Other packages are sold with accompanying hardware for image acquisition (e.g., yeast colony counters), but these are expensive and do not allow measurement of features beyond those that are already built-in. Most commercial software is proprietary, meaning that the underlying methods of analysis are hidden from the researcher. At the other end of the continuum, some software packages are very flexible, especially for interactive analysis of individual images (e.g., Image-Pro Plus, MetaMorph®, and the open-source ImageJ/National Institutes of Health (NIH) Image). While users can program custom algorithms or record macros, these customized routines are challenging to adapt without knowing a programminglanguage or interacting directly with the macro code. The CellProfiler™ project was developed to address these software challenges by providing the scientific community with an easy-to-use opensource platform for automated image analysis. The compiled software is freely available for Macintosh®, PC, and Unix platforms at

www.cellprofiler. org. It can accommodate adaptation to many biological objects and assays without requiring programming, due to its modular design and graphical user interface. There are many existing software packages available for specific applications in biology, but CellProfiler accomplishes many of the same goals in one open-source program *(Carpenter et al., 2006)*.

Fluorescence Microscopy

Cells and the internal structures of the cells can be observed using manydifferent forms of light microscopy ranging from the normal bright fieldmicroscopy to advanced systems like the Stimulated Emission Depletionmicroscopy (STED) *(Davidson & Abramowitz, 2008; Klar et al., 2001*). With the advances of the Green Fluorescent Proteins (GFP) *(Tsien, 1998; Cantrill, 2008)* a whole new color palette becomes available for fluorescence microscopy enabling the study of protein dynamics and functioning in livingcells. The applicability of this wide range of techniques in a particularsituation differs depending on the questions asked and also the availability ofthe techniques. Fluorescence microscopy has become an indispensable technique forperforming precisely localized detection of interactions within the cells. The microscopy and imaging need to be associated with specially designed techniques and the advances in both have enabled the study of dynamic processes in living cells *(Stephens & Allen, 2003).* In fluorescence microscopy a sample is irradiated with a specific band of wavelengths. These wavelengths are absorbed by fluorophores and light of longer wavelengths are emitted. Through optical filters only specific wavelengths of the emitted light reach the

detector. In fluorescence microscopy, the use of a fluorophore capable of emitting light in the detectable visible range, defined by the filters, is required to visualize the sample *(Rittscher et al., 2008).* In contrast to bright field microscopy, where the sample is observed together with the incident light, fluorescence microscopy makes use of the difference in excitation and emission wavelengths to block the incident light. This results in an image with high contrast between sample and background. Fluorescent molecules (fluorochromes) absorb short wavelength light (high energy) and are excited to a higher electronic energy state. The duration of the unstable high energy state (fluorescence lifetime) before the molecule relaxes by photon emission to the ground state is in theorder of nanoseconds. The fluorescence phenomenon is depicted in the Jablonski diagram. In ordinary fluorescence microscopy setups the number of photonsreaching the detector is very low. This is because the ratio between the number of energy quanta emitted by the tissue sample compared to the number of absorbed quanta is very low (also known as quantum yield). The fluorescent light is emitted in all directions where only a fraction reaches the objective lens. The light also has to pass through a setup of filters and dichroic mirrors (beam splitters) that divert the emitted light from the optical section, and filters the fluorescence before it reaches the imaging device. To have a large number of photons reaching the imaging device, high energy mercury (short wavelengths) or xenon (long wavelengths) arc lamps are used. Laser (acronym for ”light amplification by the stimulated emission of radiation”) sources are also commonly used to achieve the high-intensity illumination needed toimage weak fluorescence signals *(Denk et al., 1990; Huisken et al., 2004).* The type of light source to use dependson the wavelengths needed to excite the fluorochromes. A limiting property in using fluorescence microscopy is the phenomenonof photobleaching (commonly known as

fading). Photobleaching occurs when fluorochromes permanently lose their ability tofluoresce. The number of fluorescence cycles that the fluorochromes can perform before photobleaching occurs is limited (from a few to millions of cycles), and dependent on the molecule and its environment. During the excitation stage the fluorochrome may interact with a nearby molecule, and thus form a new configuration for which fluorescenceis no longer possible. Increasing the intensity of the exciting light will speed up the process of photobleaching. Instead of increasing the laser intensity for imaging thick tissue, or having prolonged exposure due to high resolution demands, one may have to use lower resolution orswitch imaging technique to more advanced optical path setups, single (or selective) plane illumination microscopy (SPIM), or 2photon microscopy *(Denk et al., 1990; Huisken et al., 2004).* The dichroic mirror reflects light below a given wavelength while transmitting the light of longer wavelengths. Together withthe excitation and emission filters, it is possible to expose the sample to only the light of the absorption wavelength and to only let the emitted light of the specific wavelength to be captured at the detector.

Fluorochromes

The fluorescence phenomena is inherent to many plants and animal tissues, if illuminated by short wavelength light. This is known as primary fluorescence (autofluorescence), and has been used in research hand industry. In the study of animal tissue the primary fluorescence is usually very faint, and does not always appear where wanted. Instead fluorochromes are introduced, allowing for target specificity, and

significantly better quantum yield. This is know n as secondary fluorescence, and there are several methods to stain a specimen with a fluorescent dye. The use of secondary fluorescence has increased due to the development of hundreds of fluorochromes with well known excitation (absorption) and emission spectra, and with new techniques to increase the specificity for a given biological target. One of the most common fluorochromes is 4,6-diamidino-2-phenylindole (DAPI). DAPI is a nucleic acid dye, with two highly nucleophilic parts *(Kapuscinski, 1995)*. The dye binds to the adenosine-thymidine (A-T) base pairs inDNA, and fluoresces in the blue region of visible light when excited byultraviolet light. Another common fluorochrome is Rhodamine. The decision of which fluorochrome to use for fluorescence microscopy has to be based on that the quantum yield should be sufficient given the light conditions, and that the fluorochrome can stay attached to the target given the treatment of the specimen and the environment. Immunofluorescence is a very important application of fluorescence microscopy based on mainly using antibodies specifically targeting an antigen (protein) of interest.

- **Direct immunofluorescence:** by chemically attaching a fluorochrome to an antibody (also known as a conjugate) and adding many of themin the presence of the antigen of interest, they bind to the antigen increasing the local concentration of the fluorochrome. The antibody remains bound after the specimen is washed, and the presence of the antigen is detected after excitation with specific wavelengths.
- **Indirect immunofluorescence:** when unstained antibodies are incubated together with its related antigen to form an antibody-antigen complex. The conjugates attach to the complex, and the complex is then detected by the fluorescence after excitation. This usually produces fluorescence with higher

signal to noise ratio, since several conjugates are likely to react with the same primary antibody.

Genetic information contained in the DNA (deoxyribonucleic acid) can be specifically stained by fluorochrome-conjugated oligonucleotides (short segments of DNA, or RNA) that bind to particular DNA sequences.

Point Spread Function

The resolution, i.e., how close two objects can be within an image and still be resolved as two distinct objects, depends on the imaging system properties, particularly on the point spread function (PSF). The point spread function (PSF) describes the relationship between the point object and the blurred response produced in the microscope. An image from a microscope consists of a sum of all PSF from the point objects in the scene. A wider PSF will decrease the resolution in the acquired image. The PSF differs between different microscopes and microscopy techniques. More blurring is usually seen in the z--direction (axial) than in the x–y direction (lateral). Confocal microscopes decrease the size of the PSF in all directions, i.e., improving the resolution in all directions. Even though the axial resolution is improved, it is still lower than the lateral *(Bolte & Cordelieres, 2006; Wallace et al., 2001).* If the PSF of the microscope is known, deconvolution methods can be used to reduce the blurring effect caused by the PSFor a given microscopy system the PSF is usually modeled with a Gaussian function *(Zhang et al., 2007)* and the full width at half maximum (FWHM) of the PSF is used to measure resolution. The determined PSF can be used to reassign the out-of-focus light mathematically *(Bolte & Cordelieres, 2006).* When the PSFs of wide-field microscopy and

confocal microscopy are compared, confocal systems show improved resolution in both axial and lateral directions. The z resolution is still poor compared to the xy resolution. This has to be taken into account when performing image analysis.

Digital image analysis

The image or image volume representations are definied by the spatial domain and by the frequency domain where the image content is represented as frequencies. In general, slowly varying intensity components in the spatial representation are represented as low frequency components while sharp transition in intensity are represented as high frequency components. The Fourier transform maps the spatial domain representation of an image into the frequency domain representation, where N is the number of samples.

$$F(\omega) = \int_{-\infty}^{\infty} f(t) e^{-i\omega t}\, dt$$

$$F(\omega) = \sum_{t=0}^{N-1} f(t)[cos(2\pi\omega t/N) - sin(2\pi\omega t/N)]$$

A digital image is an image represented in a computer, which differs in many aspects from the continuous image we perceive through our visual system. Scientific visualization is the process of communicating a message based on the images from image processing or image analysis/computer graphics *(Sonka et al., 2007).* The origin of digital image analysis dates back to the 1950s and1960s with the establishment of the artificial intelligence and robotics branches of the growing field of computer science. The precursor for the computer science field is the development of general purpose computersduring the 1940s. Pioneers and early contributions to image analysis include, e.g., Azriel Rosenfeld (http://www.cfar.umd.edu/AR), who wrote the first textbook *(Rosenfeld, 1969)* on the subject in 1969 and was a founding member of the IEEE Computersociety's Technical Committee on Pattern Analysis and Machine Intelligencein 1965, and King Sun Fu *(Kashyap et al., 1986),* who was a founding member of the International Association for Pattern Recognition (IAPR) in 1978. A digital image is generally represented with a square grid consisting of picture elements (pixels) in 2D and in 3D the elements are referred to as voxels. Each element has a value that describes the content of the position of the imaged object that it represents. In a binary image the value is either 1 (part of an object) or zero (part of the non-object or background regions). In a gray valued or gray scale image the range of values change depending on how the structure that stores the information is defined. If an 8-bit representation is used 28 is the upper limit and each element can have a value ranging between 0 and 255. This is the most common form of representation but it is not so uncommon to use other representations such as the 16-bit representation that gives a value between 0 and 65535. A color image follows the same structure with the addition that, instead of having a single number representing a gray value, each element has a value for each color

component. In an RGB (Red, Green, Blue) representation three values represent how much red, green and blue are contained in each element. In essence this means that instead of one matrix of values (in 2D) we have three matrices of gray values, one for each color component. A set of images collected over time will have an additional dimension, the time. An important issue in acquiring these time-lapse sequences is the temporal resolution that defines the rate at which the images are acquired. A suitable rate should be based on the viability of the live cells imaged *(Meijering et al., 2008).* Putting things together more formally, a digital image can be represented as a discrete integer-valued function:

$$f(x);\ x = (x, y, z,)$$

x, y,z: spatial coordinates; t: spatial coordinates; λ: the color component.

Pre-processing

The samples may need to be prepared differently depending on the modality that is used for imaging. If the specimen is to be imaged several times, possibly in different systems, some landmarks may need to be introduced at this stage to facilitate registration of images. Image acquisition; the all important imaging step that often makes or brakes a project. Great care should be taken to generate the best images possible from an analysis point of view (not necessarily the prettiest pictures) at this stage, since the image quality reflects the output of the whole project. If noise or errors were introduced in the imaging step, they should be taken care of now. Registration of images from different times or

modalities can be considered as pre-processing. Each image volume consisted of layers that were merged along the z-axis using a maximum intensity projection (MIP). However, the MIP is somewhat sensitive to noise, as it will act as maximum filter for each 1 × 1 × 16 pixel array. To reduce the amount of non-point-like signal before the MIP, the background was removed in each focus layer using top-hat filtering *(Haralick & Shapiro, 1992)*. The top-hat filtering may be described using a rolling ball analogy. Draw a squiggly line on a piece of paper, and let an imaginary ball (a structuring element) with a certain radius roll along the curve. Every point that the ball manages to touch, being small enough to fit in to large holes and grooves, is set to zero. Points that the ball does not touch, due to being too large, are given values according to the distance between the ball and the line.

Segmentation

The aim of segmentation is to divide the image into different regions that are homogeneous with respect to certain criteria. Thresholding is the process of separating the image into foreground (objects), and background based on selecting a threshold value using global, or local properties of the image intensity distribution. The threshold value divides the image such that pixels with gray levels below the threshold belong to the background, and pixels with higher, or equal gray level belong to the foreground (or the opposite if the image contains dark objects on a bright background). The output from the thresholding is a binary image g (x, y) (an image consisting of 0 and 1). A threshold value T is defined and pixels with values above T are considered to be part of the object

in the image f (x, y):

$$g(x,y) = \begin{cases} 1 & \text{if } f(x,y) \geq T \\ 0 & \text{if } f(x,y) < T \end{cases}.$$

The value T can be set manually or selected automatically from certain criteria. The image histogram p(f) of the image f (x, y) is the probability density function which gives the frequency of the different pixel values in f (x, y). A popular thresholding method based on the histogram is Otsu's *(Otsu, 1979)* where the automatic selection of an optimal threshold is performed based on the class seperability of the histogram. If an image contains objects with fairly similar gray values that are significantly different from those of the background, the histogram will have two distinct peaks with a valley in between. A gray value close to the valley can then successfully be used to separate the objects from the background. However, if the objects and background have overlapping intensities then the histogram will not have distinct peaks and finding the optimal value for thresholding is difficult. Image pre-processing can be useful in reducing the variations in the gray values within objects and also within the background to increase the between class separability. The watershed algorithm was originally presented by Beucher and Lentuéjoul (1979) *(Lantuejoul & Beucher, 1981)* and later refined in a more efficient implementation by Vincent and Soille (1991) *(Vincent & Soille, 1991).* The

watershed segmentation is a region-based segmentation method and has been extensively used in many areas of image analysis, e.g., cell segmentation *(Carpenter et al., 2006; Malpica et al., 1997; Wahlby et al., 2004).* The watershed segmentation can be understood by seeing the image as a landscape. The watershed algorithm is a region growing method commonly explained by a "rain falling on a mountain landscape" analogy. The gray-level intensity represents the elevation in this landscape. The landscape image can be, e.g., the original gray-level image, a distance transformed image (an image containing a distance value to the nearest object pixel or nearest background pixel) *(Maurer et al., 2003),* or a gradient magnitude image. The algorithm can be illustrated by letting water enter through the local minima and start to rise. A lake around a local minimum is created and referred to as catchment basin. When the water fronts from different catchment basins meet they form a dam or watershed that separates the catchment basins. All that is left after the watershed segmentation are watershed lines separating the objects. If objects of interest are bright rather than dark, the image is inverted before applying watershed segmentation. Beside bilevel thresholding, multilevel thresholding techniques also exist, where the problem lies in selecting two or more thresholds for dividing the image. The probability function gives the likelihood of a certain gray level to occur in the image. In Otsu's method (*Otsu, 1979),* an optimal threshold is reached by minimizing the variance between classes (background and foreground), over the total image variance (three equivalent formulations exist). The threshold that produces the smallest variance ratio is selected. This approach was used in the signal segmentation project. In minimum error thresholding *(Kittler & Illingworth, 1986)* the gray level histogram is viewed as two mixed normally distributed populations. An optimal threshold is the value that

minimizes a criterion function introduced to avoid estimating the mean and variance of the two populations. The method was used with good results in the testicle project, after some interactive treatment of parts of images lacking tissue.

1.7.1 CELLPROFILER: DIGGING DEEP AND WIDE INTO SINGLE CELLS

CellProfiler is freely available, open-source software that enables researchers without training in computer programming to measure biological phenotypes quantitatively and automatically from thousands of images. With an interface designed by biologists and underlying algorithms developed by computer scientists, CellProfiler bridges the gap between advanced image analysis algorithms and scientists who lack computational expertise. *(Carpenter et al., 2006; Lamprecht et al., 2007).* CellProfiler was initially designed for high-throughput image analysis but is often used for small-scale projects. This highlights the trend toward quantifying information in images regardless of experiment size.CellProfiler's interface lets researchers build customized chains of interoperable image analysis modules to identify and measure biological objects and features in images (Figure 1.12).

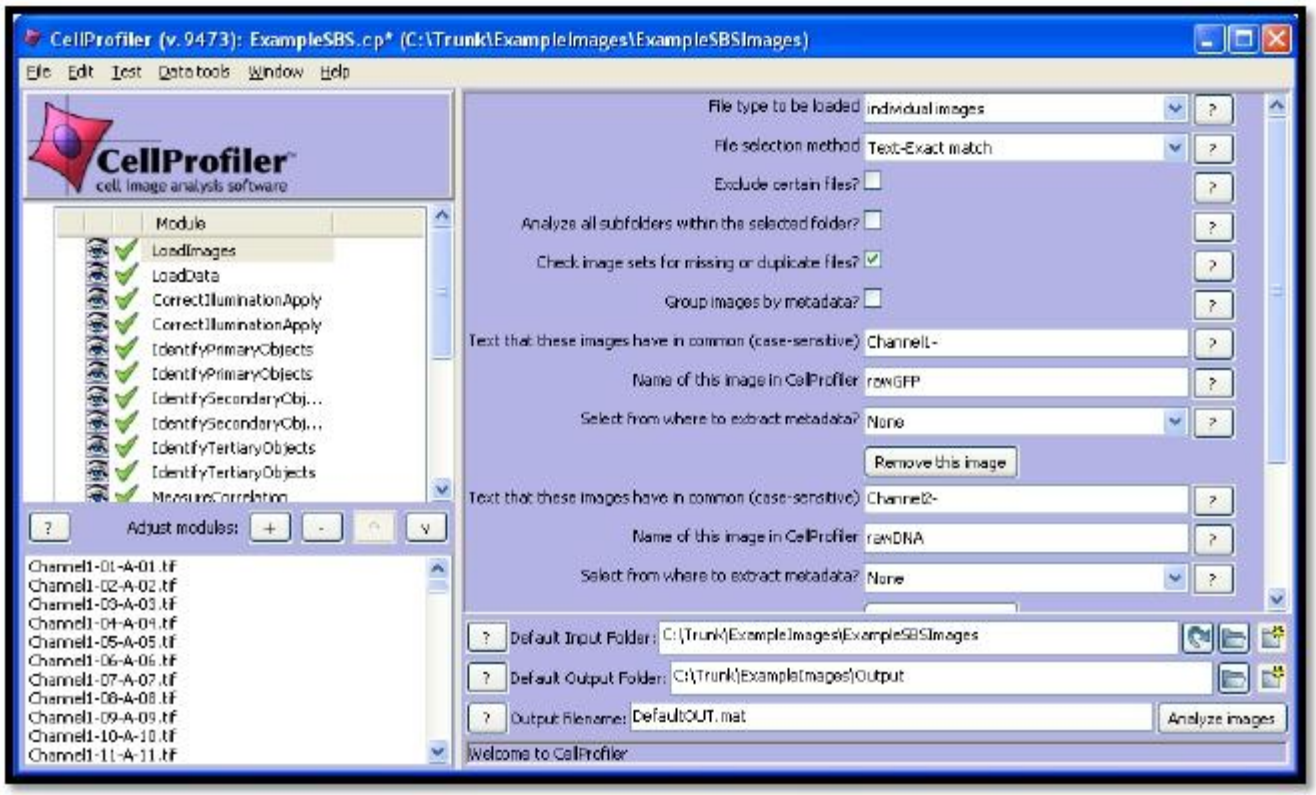


Figure 1.12 An example of CellProfiler's pipeline on web interface.

CellProfiler has been used to measure individual cells, colonies of cells and whole organisms in a wide range of assays (e.g. counting cells, measuring staining intensities and scoring complex phenotypes with machine learning) and at many experimental scales (from a few to hundreds of thousands of images). A variety of cell types have been

analyzed, including budding yeast, Drosophila, mouse, rat and dozens of human cell types. The diverse measurements generated by CellProfiler provide raw material for machine-learning algorithms that can identify challenging phenotypes *(Jones et al., 2009; Misselwitz et al., 2010; Ramo et al., 2009).* The first MATLAB *(MATLAB version 7.5.0 2007)* language was redesegned to the open-source Python language, making use of the high-performance scientific libraries NumPy and SciPy *(Oliphant, 2007).* Cython (http://www.cython.org) is used to implement computationally intensive algorithms, as well as bridge to precompiled libraries including Java via the Java Native Interface. The Java/Python bridge allows CellProfiler 2.0 to load nearly 100 image formats via the Open Microscopy Environment Consortium's Bio-formats library (http://www.loci.wisc.edu/software/bio-formats). A bridge with ImageJ was built in order to run ImageJ (http://rsbweb.nih.gov/ij), macros in the context of a CellProfiler pipeline. CellProfiler can be run in batch mode: sets of images are partitioned between CellProfiler instances running on separate computing cores or cluster nodes in a distributed environment. In CellProfiler 2.0, images can be loaded via HTTP or located based on a comma-delimited text file containing image file locations, which might be generated by automated microscopes or laboratory information systems. Metadata about the images can also be loaded similarly. CellProfiler 2.0 has enhanced database capabilities and is now able to upload directly to MySQL or SQLite databases during image processing. CellProfiler 2.0's FlagImage module can exclude images from analysis based on measurements of image quality, such as blurriness and presence of debris. Images can be grouped for aggregate operations, such as illumination correction of images on a per-plate basis or analysis of multiple time-lapse movies or three-dimensional image stacks. Illumination often varies more than 1.5-fold across the field of

view, even when using fiber optic light sources, and occasionally even when images are thought to be already illumination-corrected by commercial image analysis software packages (TRJ, AEC, DMS, and PG, unpublished data). This adds an unacceptable level of noise, obscures real quantitative differences, and prevents many types of biological experiments that rely on accurate fluorescence intensity measurements (for example, DNA content of a nucleus, which only varies by two-fold during the cell cycle). CellProfiler contains standard methods plus our new methods *(Jones et al., 2006)* to address illumination variation, allowing various methods to be compared side by side and, ultimately, providing less noisy quantitative measures.

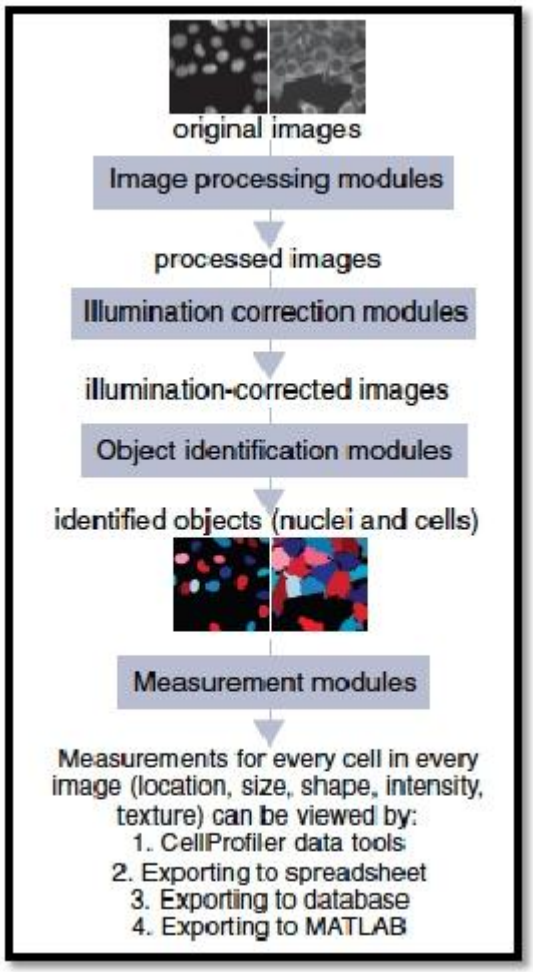


Figure 1.13 Schematic representation of a typical CellProfiler pipeline. The original image is processed, in order to apply the modules of illumination correction. Then, the subsequent step is the identification of primary and secondary objects (nuclei and cells).

The final step is the analysis of data (Figure 1.13). Object identification (also called segmentation) is the most challenging step in image analysis and its accuracy determines the accuracy of the resulting cell measurements. In most biological images, cells touch each other, causing the simple, fast algorithms used in some commercial software packages to fail. The first objects identified in an image (called primary objects) are often nuclei identified from DNA-stained images, although primary objects can also be whole cells, beads, speckles, tumors, and so on. After primary objects (often nuclei) are identified, the edges of secondary objects that surround each primary object (often

cell edges) can be found more easily. Measuring cell size in Drosophila was not previously feasible because the commonly used watershed method *(Meyer & Beucher, 1990)* often fails to find the borders between clumped cells *(Jones et al., 2005).* Other subcellular compartments can also be identified, including the cytoplasm (the part of each cell excluding the nucleus) and the cell or nuclear membrane (the edge of the cell or nucleus).

2. MATERIALS AND METHODS

2.1 SAMPLE COLLECTION

In Beghini et al., work, bone marrow (BM) mononuclear cells were collected from 33 newly diagnosed, unselected non-promyelocytic AML patients according to standard procedures, after obtaining the informed consent, according to Niguarda Hospital's Ethical Board approved protocols. AML samples were classified according to the French-American-British (FAB). According to the revised Medical Research Council risk group stratification, based on cytogenetic and molecular markers, samples included 14 adverse, 13 intermediate and 6 favorable risk patients. Human adult BM cells obtained from 10 healthy donors were derived from the posterior iliac crest, after obtaining informed consent, according to the Niguarda Hospital's institutional review board guidelines (Table 2.1). The subsequent gene expression analysis, was performed on n=112 AML patients, 1 idiopathic myelofibrisis and n=3 healthy donors, including some samples recruited in Beghini et al., work (Table 2.2). Bone marrow tissue sections of AML patients, were obtained from Niguarda Hospital.

TABLE 2.1 Clinical characteristics of AML patients

AML	FAB	CYTOGENETICS
AML 1	M2	45, XY, -7
AML 2	M1	Complex karyotype
AML 4	M4	46, XY, del(20)(q11;q13)
AML 5	M4	46, XX
AML 6	M4	46, XX, +11
AML 9	M0	46, XY
AML 10	M5a	46, XX
AML 13	M1	46, XY, t(6;9)(p23;q34)
AML 14	M2	46, XX
AML 16	M2	46, XX
AML 17	M1	46, XY
AML 19	M1	46, XX, del (11)(q23)
AML 21	M1	46, XY
AML 23	M1	46, XY
AML 24	M1	46, XY
AML 25	M1	46, XX, del (11)(q23)
AML 30	M5a	46, XY
AML 32	M2	46, XX
AML 34	M0	46, XY, del (11)(q13;q23)
AML 38	M1	46, XX

AML 39	N.A	Complex karyotype
AML 40	M2	Complex karyotype
AML 41	M4	46, XX
AML 42	M4	46, XX
AML 44	M2	46, XX,t(8;21)(q22;q22)
AML 46	M2	46, XY
AML 47	M2	46, xx, +21
AML 48	M5a	46, XY
AML 49	Biphen	45, XX, -7,t(9;22)(q34;q11)
AML 50	M4	Complex karyotype
AML 51	M5b	46, XX
AML 52	M2	46, XY
AML 53	M1	46, XX

TABLE 2.2 Clinical characteristics of AML patients.

AML	FAB	CYTOGENETICS
AML 1	M2	45, XY, -7
AML 2	M1	Complex karyotype
AML 4	M4	46, XY, del(20)(q11;q13)
AML 9	M0	46, XY
AML 10	M5a	46, XX
AML 11	AREB-T	46, XY, +8
AML 12	M2	46, XX
AML 13	M1	46, XY, t(6;9)(p23;q34)
AML 14	M2	46, XX
AML 16	M2	46, XX
AML 17	M1	46, XY
AML 18	M5a	46, XY
AML 19	M1	46, XX, del (11)(q23)
AML 21	M1	46, XY
AML 25	M1	46, XX, del (11)(q23)
AML 30	M5a	46, XY
AML 39	N.A.	Complex karyotype
AML 40	M2	Complex karyotype
AML 42	M4	46, XX
AML 43	M0	46, XY, t(4;12)(q21;p13)

AML 44	M2	46, XX,t(8;21)(q22;q22)
AML 46	M2	46, XY
AML 49	Biphen.	45, XX, -7,t(9;22)(q34;q11)
AML 53	M1	46, XX
AML 57	M1	47, XX, +4
AML 58	M1	47, XX, +4
AML 59	N.A	47, XX, +8
AML 61	M1	46, XY
AML 62	M5a	46, XX
AML 63	M5b	46, XX
AML 64	M2	46, XX
AML 65	M2	46, XX,t(8;21)(q22;q22)
AML 67	M1	47, XX, +8
AML 200	M2	45,X,-Y,t(8;21)(q22;q22)
AML 201	M4	48,XY,inv(16),+22,+9
AML 202	M4	46,XY,inv(16)
AML 203	M4	46,XY,inv(16)
AML 204	M4	46,XX,inv(16)
AML 205	M4	46,XY,inv(16)
AML 206	M4	46,XX,inv(16)
AML 207	M4	46,XX,inv(16)
AML 208	M4	45,X0,inv(16)
AML 209	M4	46,XY,inv(16)
AML 210	M4	46,XY,inv(16)

AML 211	M4	46,XY,inv(16)
AML 212	M2	45,X, -Y,t(8;21)(q22;q22)
AML 213	M2	49,XY,t(8;21)(q22;q22),+4,+6,+19
AML 214	M4	46,XX,inv(16
AML 215	M2	45,X,-Y,t(8;21)(q22;q22)
AML 216	M4	46,XY,inv(16)
AML 217	M2	46,XY,t(8;21)(q22;q22)
AML 218	M2	46, XX,t(8;21)(q22;q22)
AML 219	M2	47,XY,t(8;21)(q22;q22),+13
AML 220	M2	46,XX
AML 221	M2	45,X, -Y,t(8;21)(q22;q22)
AML 222	M2	46, XX,t(8;21)(q22;q22)
AML 223	M2	46, XY,t(8;21)(q22;q22)
AML 224	M2	46, XX,t(8;21)(q22;q22)
AML 225	M1	45,X,-Y
AML 226	M2	47,XY,+11
AML 227	M2	46,XX
AML 228	M2	46,XY
AML 229	M4	46,XX
AML 230	M1	46,XY
AML 231	M1	46,XY
AML 232	M1	46,XX
AML 233	M1	46,XX
AML 234	M4	46,XX,inv(16)

AML 235	M4	46,XX,inv(16)
AML 236	M1	46, XX
AML 237	M2	46,XY,inv(16)
AML 238	M4	46,XX,inv(16)
AML 239	M2	45,X, -Y,t(8;21)(q22;q22)
AML 240	M2	46,XY,t(8;21)(q22;q22),+4
AML 241	M2	45,X,-X,t(8;21)(q22;q22),add(4)(p16),-9,+mar
AML 242	M4	46,XY,inv(16)
AML 243	M4	46,XY,inv(16)
AML 244	M4	47,XY,inv(16),+6
AML 245	M1	47, XX, +4
AML 246	M4	46, XY
AML 247	M2	Complex karyotype
AML 248	M1	46, XY
AML 249	M4	46, XX
AML 250	M0	46, XY
AML 251	M1	46, XX
AML 252	M1	46, XY
AML 253	M4	46, XY
AML 254	M1	46, XY
AML 255	M4	46, XY
AML 256	M4	46,XX,inv(16)
AML 257	M1	46, XX
AML 258	M1	46, XY

AML 259	M4	46, XY
AML 260	M1	46, XX
AML 261	M4	46, XY
AML 262	M1	46, XX
AML 263	M4	46, XY
AML 264	M0	46, XX
AML 265	M1	46, XY
AML 266	M4	46, XX
AML 267	M2	46, XX,t(8;21)(q22;q22)
AML 268	M2	46, XY,t(8;21)(q22;q22)
AML 269	M1	46, XX
AML 270	M1	46, XX
AML 271	M4	46,XX,inv(16)
AML 272	M4	46,XY,inv(16
AML 273	M4	46,XX,inv(16)
AML 274	M1	46, XX
AML 275	AREB-T	46, XX
AML 276	M4	46, XY,t(8;21)(q22;q22)
AML 277	M5	46,XX
AML 278	M1	46, XX
I.M. 66	-	46, XX

2.2 CELL SORTING AND FLOW CYTOMETRY

The mononuclear cells were obtained by a density gradient centrifugation using "Ficoll-Hypaque" (Lymphoprep, AXIS-SHIELD, Oslo, Norway d = 1.077 g/m), a mixture of polysaccharides of high molecular weight, which allows the separation of different hematopoietic components. This procedure allows the separation of two cellular fractions: I) erythrocytes and polymorphonuclear granulocytes, which precipitate to the bottom of the tube, II) mononuclear cells (MNC) which are collected at the interface between Ficoll and serum.

The $AC133^+$ cell fraction was isolated by immunomagnetic separation after labeling with CD133/1 (AC133)-biotin antibody and anti-biotin MicroBeads on LS columns and Midi MACS separator (Miltenyi Biotec, Bergisch Gladbach, Germany). The purity of the $AC133^+$ fraction, evaluated by flow cytometry analysis, was greater than 97%.

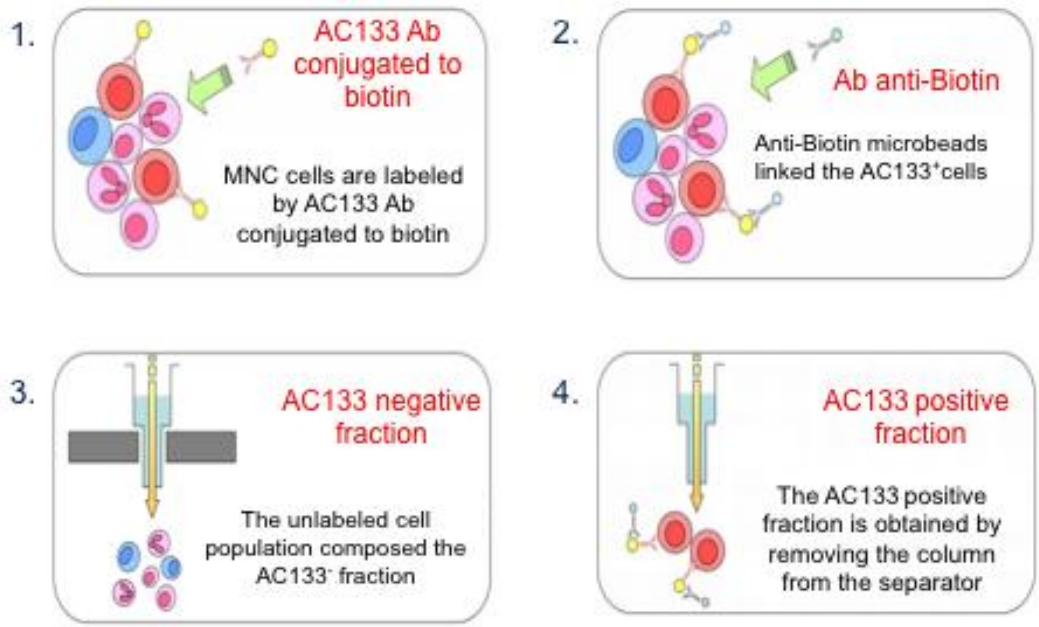


Figure 2.1 AC133 positive selection.

2.3 MICROARRAY EXPRESSION ANALYSIS

Microarrays have evolved from Southern Blotting, a technique which is used to identify a specific DNA sequence in DNA samples *(Southern, 1975)*. The first studies that involved microarrays were published in 1980s, but the real microarrays take-off began with a publication by Fodor et al. *(Fodor et al., 1991)* from Affymax, a company which later changed its name to Affymetrix and became the market leader for microarray technology.

Typically, a microarray consists of a large number microscopic DNA spots attached to a solid surface. Each of the spots contains many short DNA sequences called *probes*. All of the probes inside one spot have the same sequence. Also, all of the probe sequences are complementary to a part of the gene from an organism which is being analyzed.

The HG-U133 arrays use as few as 11 probes in a probeset *(Bolstad, et al., 2003).* The size of a standard GeneChip is 1.28 cm x 1.28 cm; and over 6.5 million squares, or features are present on each chip. In each feature, there are millions of identical probes. The design of Affymetrix probes is not usually in the hands of the researchers. A probe consists of a short oligonucleotide sequence containing 25 nucleotides, called a 25-mer; and all the probes are synthesised on the chip one base at a time, and in parallel at all locations. A paired probe is composed of: a) a perfect match (PM), which is the exact sequence of the chosen fragment of the gene, b) a mismatch (MM), which is same as PM but contains a mismatch nucleotide in the middle of the fragment. The mismatch probe contains a single mismatch located directly at the 13th position in the 25-mer probe sequence . This mismatch probe is used as a background control and also to overcome

the low specificity of the short oligonucleotide used. While the perfect match probe provides measurable fluorescence when the sample binds to it, the paired mismatch probe is used to detect and eliminate any false or contaminating fluorescence within that measurement.

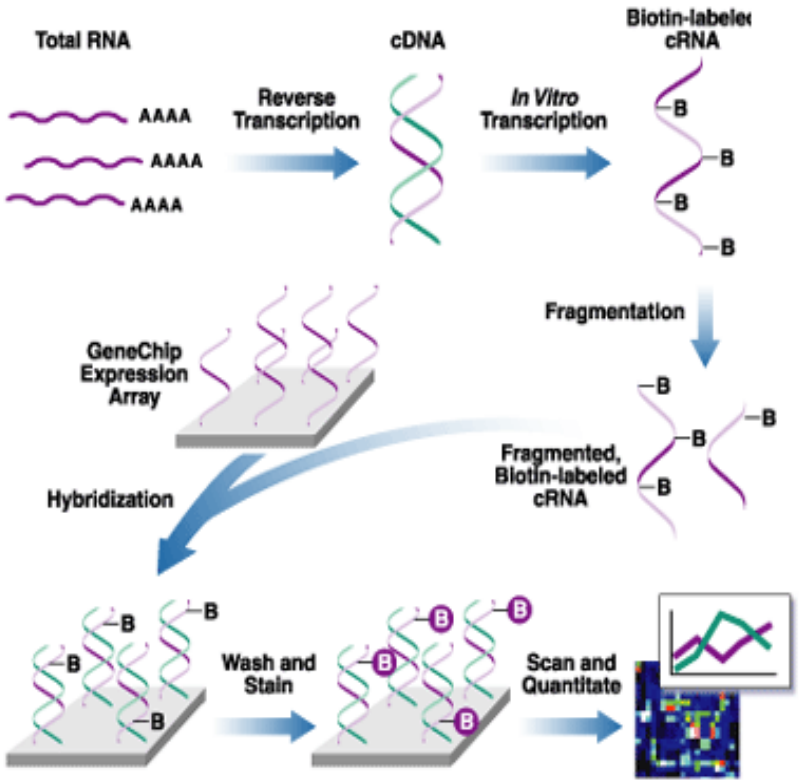


Figure 2.2 Microarray technology overview.

The mismatch probe serves as an internal control for its perfect match partner because it hybridizes to nonspecific sequences about as effectively as its counterpart, allowing misleading signals, from cross hybridization for example, to be efficiently quantified and subtracted from a gene expression measurement or genotype call. Affymetrix anticipates that the MM probe does not hybridize well to the target transcript, but hybridizes to

many transcripts to which the PM probe cross-hybridizes *(Simon et al., 2004).* Therefore, the intensity difference between PM and MM paired probe is considered to be a better estimate of the hybridization intensity to the true target transcript.

The value that is usually taken as representative for each gene‘s expression level is the average difference between PM and MM. Ideally, this average value is expected to be positive because the hybridization of the PM is expected to be stronger than the hybridization of the MM. However, many factors, including non-specific hybridizations and a less than optimal choice of the oligonucleotide sequences representative of the gene, might result in an MM hybridization stronger than the PM hybridization for certain probes. The calculated average difference might be negative in such cases, and these negative values introduce noise into the dataset.

The outputs of microarray experiments require processing before they can be used for extracting meaningful information. Image processing and normalization are the two preliminary microarray data processing stage. Regardless of the technology, the arrays are scanned after hybridization and independent, 16 bit, digital, grey-scale TIFF images are generated for query and control samples *(Causton et al., 2003).* Analysis of a cDNA image seeks to extract intensity for each spot or feature on the array, and it involves various image processing stages that can be carried out through different microarray image analysis software. Affymetrix has integrated its image processing algorithms into the experimental process of GeneChip software, and thus, there are no decisions to make for the end users *(Stekel, 2006).*

Affymetrix GeneChip experiments are managed with the Affymetrix GeneChip Operating Software (GCOS). Once the fluorescent-tagged nucleic acid sample is injected into the hybridization chamber, and hybridization takes place to the complementary

oglionucleotides on the chip, the hybridized chip is scanned and the laser excited fluorescence across the chip is converted to a 2D image. This image data file (.DAT) can be exported as a .TIFF image. The image data file is used by the software to generate a .CEL file that gives the position and intensity information of each probe for one GeneChip, in addition to the position of masks and outliers. The Affymetrix output result file is the .CHP file, where the average signal intensities are linked to gene identities. The report file (.RPT) is generated from the .CHIP file, and it summarizes the quality control information about expression analysis settings and probe set hybridization intensity data. Besides, there are two more files that are used in the actual analysis process - Experiment File (.EXP) and Chip Description file (.CDF). The former contains parameters of the experiment such as probe array type, experiment name, equipment parameters and sample description. The .CDF file is provided by Affymetrix and describes the layout of the chip.

4 staged Affymetrix scanning output files

- **Experiment File *.EXP**: This file contains the parameters of the experiment such as Probe Array Type, Experiment Name, Equipment parameters, Sample Description, and others. This file is not used for analysis, but is required to open other GCOS* files for the designated chip experiment.
- **Image Data File *.DAT:** This file is the image file generated by the scanner from the Probe Array after processing on the Fluidics Station. This file can be viewed in GCOS or exported as a *.TIFF image. This file is used in GCOS to generate the *.CEL file.
- **Cell Intensity File *.CEL**: The cell file contains the processed cell intensities from the primary image in the *.DAT file. The cell file is used by GCOS to generate the *.CHP file, which contains the numerical data from the *.DAT, and *.CEL files.
- **Probe Array Results File *.CHP:** The chip file is the output file from the GCOS expression analysis of the Probe Array. The chip file contains the data that will be used for statistical analysis and data mining analysis.

* GCOS is the Affymetrix software suite, which controls the hybridization and fluidics station as well as the scanner. GCOS regulates the final laboratory processing producing the specified files as well as having the option to do statistical pre-processing within its environment. Alternatives to pre-processing in this environment are described below.

Data normalisation is an important aspect, and plays an important role in the early stage of microarray data analysis as the subsequent analytical results are very much dependent on it. The normalization methods rely on the fact that gene expression data can follow a normal distribution, and the entire distribution can be transformed about the population mean and median without affecting the standard deviation. The objective of normalization is to eliminate the measurement variations and measurement errors, and to allow appropriate comparison of data obtained from the expression levels of genes so that the genes that are not really differentially expressed have similar values

across the arrays. Normalization is also used to identify and eliminate questionable and low quality data. Normalization approaches typically use either a control set of genes or the entire genes from an array. The use of a control set requires only one assumption, i.e., the control genes are detected at constant levels in all of the samples being compared. For GeneChips, Affymetrix Inc. claims to have integrated the housekeeping genes and spiked-ins, or spiked controls in the chips after supposedly testing them on a large number of various tissue types with the resultant low variability in those samples *(Wang et al. 2002).* For cDNA microarrays, normalization involves determining the amount by which the genes of the red channel are over- or under expressed relative to the green channel. Approaches to calculate the normalization factor can be divided into three categories: global normalization, intensity-based normalization and location-based normalization as well as a hybrid of intensity- and location-based normalization. The Affymetrix data normalization was performed using the Robust Multi-Array Average or RMA *(Irizarry et al., 2003)* based on quantile normalization, and is used for normalizing the chips. RMA is largely the work of Terry Speed‘s group at University of California at Berkeley, and only uses PM probes as the method assumes that including the MM probes introduces more variability than the correction is worth. In RMA, the expression measure is obtained using three steps: convolution background correction, quantile normalization, and a summarization method based on a multi-array model fit that uses the median polish algorithm *(Tukey, 1977).* Starting with the raw probe-level data from a set of GeneChips, the perfect-match (PM) values are background-corrected, quantile normalized, and then finally the linear model is fit to the normalized data to obtain an expression measure for each probe set on each array.

PROCEDURE

RNA isolation

Total RNA for expression profiling analysis was extracted using RNAqueus 4PCR kit (Ambion, Austin TX) from AC133 selected cells. 500 µL of Lysis/Binding Solution was added to a sample and vortexed vigorously; an equal volume of 64% Ethanol was added to the lysate and the tube was inverted several times. The samples were applied to columns, spun for 30 seconds at 10,000 RCF, and the flow through was discarded. 700µL of Wash Solution #1 was added, centrifuged, the flow through was then discarded. Then 500µL of Wash Solution #2/3, was added and centrifuged and the flow thought was subsequently discarded. This process was repeated with another 500 µL of Wash Solution #2/3. RNA was eluted with 50µL of preheated Elution Solution, and centrifuged at 10,000 RCF for 30 seconds, and eluted again with 15 µL of Elution Solution. Solution was then treated with 7.5µL of 10x DNase1 buffer and 1.0 µL of DNase1 enzyme to destroy residual DNA. The product was then incubated for 30 minutes at 37ºC before 8.0 µL of DNase1 inactivation reagent was added and incubated for 2 minutes at room temperature. The tube was then spun down at 10,000 RCF for 1 minute to pellet the inactivation reagent.

RNA quality evaluation

Integrity of RNA samples extracted from dried blood spots was checked using Experion™ (Bio-Rad, USA). Experion system was included with automated electrophoresis station, priming station, vortex station for RNA analysis and RNA std sens analysis kit which included with chips and reagents for standard-sensitivity RNA. Following procedure performed for RNA analysis using the Experion system. In order to avoid any contamination during RNA integrity analysis, electrodes of the Experion system were cleaned using Experion electrode cleaner (800 µl) in the first step. After repeating this step for one more time, electrodes were rinsed with DEPC treated water (500 µl) for 5 minutes using electrode cleaning chip. At the end lid was kept open for 60 second to evaporate remaining water on electrodes. RNA stain, RNA loading buffer and RNA gel from the RNA std sens kit were removed from 4˚ C and equilibrated at room temperature for 20 minutes. RNA stain was wrapped in aluminum foil to avoid its light sensitive degradation. RNA gel was filtered from filter tube at 2000 RPM for 10 minutes. Filtered gel (65 µl) was taken into RNase-free microfuge tube and mixed with RNA stain (1 µl). RNA ladder was removed from -20˚ C and thawed it on ice for 10 minutes. RNA ladder (1 µl) and RNA samples (3 µl) was taken into RNase-free microfuge tube. RNA ladder and RNA samples were denatured at 70˚ C for 2 minutes. Ladder and samples were immediately placed on ice for 5 minutes, spun down for 2-5 seconds and stored on ice until needed. Gel-stain solution (9 µl) was taken in well labeled as GS on RNA std sens chip without forming any air bubble. Chip was primed by setting appropriate pressure for sufficient time on priming station. Chip was inspected for any air bubbles in micro

channels and for incomplete priming. Gel-stain solution (9 µl) was taken other well labeled GS. Filtered gel (9 µl) was taken to well labeled as G. Loading buffer (5 µl) was taken to each sample well 1-12 including ladder well. RNA ladder (1 µl) was taken to the well labeled as L. RNA samples were taken to all wells numbered as 1-12. Chip was placed tightly and vortexed for 60 seconds on vortex station. Primed chip loaded with RNA samples and ladder was then kept on electrophoresis station for 5 minutes and electrophoresis run was started. The use of a RNA ladder as a mass and size standard during electrophoresis allows the estimation of the RNA band sizes. After completion of the run, electrodes were cleaned using DEPC water (800 µl) filled in a cleaning chip. Electropherograms generated were analyzed by Experion software version 3.2. Integrity of the RNA may be assessed by visualization of the 18S and 28S ribosomal RNA bands *(Mueller et al., 2004).* The intact RNA preparation shows high 18S and 28S rRNA peaks as well as a small amount of 5S RNA.

Microarray

The samples are hybridized to GeneChip HGU133plus 2.0, using the Two Cycle Target Labeling protocol, according to the manufacturer's procedures.

TOTAL RNA	PROTOCOL
1 µg- 15 µg	One-cycle target labeling
10 ng-100 ng	Two-cycle target labeling

For using as target, the total mature, spliced, poly-A tail added RNA isolated from the cell being studied is turned into a double stranded cDNA through reverse transcription. At the time of running the array, the cDNA is allowed to go through in vitro transcription back to RNA (now known as *cRNA*), and labelled with *biotin*. The labelled cRNA is then randomly fragmented in to pieces anywhere from 20 to 400 nucleotides in length, and the cRNA fragments are added to GeneChip for hybridization. The hybridization occurs at a 45°C, 60 rpm for 16 hours. After hybridization, performed in fluidic station 400, the difference in hybridization signals between PM and MM, as well as their intensity ratios, was detected by scanning the array with a laser serves as indicators of specific target abundance. After hybridization, the chips were scanned by a scanner directly connected to the computer, which also manages the fluid through the GCOS software. The image captured by the scanner is a .DAT file, where the pixels in a feature are then grouped into a single value. CEL. The file .CEL is created by the algorithm Cell Analysis Algoritm of GCOS and contains the light intensity measured for each probe cell. The GCOS software then transformed the .CEL files into .CHP file, in order to obtain qualitative and quantitative analysis.The file. CHP can then be exported to a file. TXT. In our study we used the method of the logarithmic transformation, used in the pre-processing and normalization of gene expression data. For the pre-processing of the data was used package Affy, defined as Affymetrix oligonucleotide Methods for Arrays. The system used

in our study provides data normalization using RMA (Robust Multy-Array Average) in R environment, a programming language for bioinformatics derived from S, a program developed by John Chambers and colleagues in the 80s at the Bell Labs. R is an Open-Source Bioconductor initially developed by Ross Ihaka and Robert Gentleman at University of Auckland (New Zealand), which operates through the web: www.rproject.org. Background correction used in RMA is aimed at correcting only PM values, and is a nonlinear correction using a probabilistic model, done on a per-chip basis. It involves a convolution of an exponentially distributed (with mean, α) signal, X and normally distributed (with mean, μ and standard deviation, σ) noise, Y caused by optical noise and non-specific binding. Therefore, the observed PM intensity, S = X + Y.

To evaluate the effect of normalization on the Affymetrix arrays, an assessment is carried out evaluating MA plots, Array intensity distribution and Nuse plot.

Data mining

Hierarchical clustering *(Johnson, 1967)* is useful to find the closest associations among gene profiles under evaluation where it seeks unsupervisedly to build a hierarchy of clusters based on relatedness. Whether any unwanted change has been caused to the microarray data through the process of ratio-transformation can be evaluated through hierarchical clustering. The method when applied to the pre- and post- transformed microarray data would highlight if any change has occurred to the overall state of the data.

In our study, In order to evaluate the genes that are differentially expressed between all patients and healthy subjects, we set up parametric and non-parametric analysis. The statistics used are as follows:

- **t-Test:** standardized difference in the mean absolute or relative gene expression (fold changes). They are constructed by dividing the difference between two parameters (usually two means) to the standard error of their difference (usually estimated by the variances of the two samples from which the averages are calculated). Under the assumption that the two samples are drawn randomly from the same distribution eindipendente and that this is Gaussian (normal), the test follows the Student's t distribution.

$$t = \frac{\log I_G - \log I_B}{\sqrt{\frac{\sigma_G^2}{n_1} + \frac{\sigma_B^2}{n_2}}}$$

- **Mann-Whitney U, Wilcoxon W-test.** The nonparametric tests have the advantage of not requiring assumptions about the distribution of the parameter of interest and dedicated software

Microarray data have been deposited in ArrayExpress (http://www.ebi.ac.uk/arraexpress/), with accession number E-MTAB-220.

2.4 BIOINFORMATIC ANALYSIS

The Gene Ontology (GO) *(Ashburner et al., 2000)* was created in 1998 by the Gene Ontology Consortium in an effort to address the need for a controlled, structured and unified vocabulary for genome annotation. The Gene Ontology Consortium is a collaborative project whose founding members are the model organism databases Flybase *(Tweedie et al., 2009),* Mouse Genome Informatics (MGI) *(Blake et al., 2011)* and the Saccharomyces Genome Database (SGD) *(Cherry et al., 1998)*. In the last ten years, the list of member projects has more than quintupled and now includes, among others, dictyBase *(Fey et al., 2009),* Gene Ontology Annotation @ EBI (GOA) *(Barrell et al., 2009),* Gramene *(Jaiswal et al., 2006),* Rat Genome Database (RGD) *(Twigger et al., 2006),* Reactome *(Croft et al., 2011),* The Arabidopsis Information Resource (TAIR) *(Swarbreck et al., 2008),* WormBase *(Harris et al., 2010)* and Zebrafish Information Network (ZFIN) *(Bradford et al., 2011).*

The GO consists of three orthogonal structured vocabularies or "sub-ontologies", namely:

• **Molecular function (MF),** i.e. the activity, at molecular level, of a gene product;

• **Biological process (BP**), i.e. the larger overall process that a gene product is involved in;

• **Cellular component** i.e. the component of the cell that a gene product acts in.

Genome wide analysis

We performed a genome-wide analysis in order to select genes differentially expressed between AML AC133+ patients and AC133+ healthy donors (Welch t-test, 0.05 significance level). The resulting set of differentially expressed genes has been analyzed for functional enrichment with respect to the terms of the Biological process (BP) branch of the Gene Ontology (GO) and the pathways of the KEGG database. We performed the functional enrichment analysis using:

- **GOStats**: version 2.12.0 of the Bioconductor package (http://www.biocunductor.org/packages/release/bioc/html/GOstats.html);
- **DAVID (Database for Annotation, Visualization and Integrated Discovery)**: http://david.abcc.ncifcrf.gov/home.jsp ;
- **Dysregulated pathway analysis:** iterative procedure, based on non parametric test proposed by Majeti et al.

The first two tests are based on the hypergeometric distribution, whereas the last one is based on a non-parametric test and on iterative procedure. All tests were applied to all genes of the Hgu133plus2 gene chip, considering all the genes included in the Affymetrix platform as Universe with p-value < 0.05.

Generation of an evolutionarily conserved human transcriptional regulators dataset

The entry point of our selection procedure was the extraction from Pfam (http://pfam.sanger.ac.uk/) of all the domains involved in transcription regulation. Next, we used the set of 400 Pfam domains to query the entire set of ensemble human genes using the BioMART web interface (http://www.ebi.ac.uk/biomart/). This produced an initial set of 2,971 non redundant genes associated to a unique HGNC symbol. In order to extract all the one-to-one orthology relationships , we thus adopted a direct SQL access to the ensembl_compara_51 database (hosted at ensembldb.ensembl.org) occurring between the genes in the initial set and the complete gene set of the following organisms: *Bos taurus, Canis familiaris, Danio rerio, Drosophila melanogaster, Equus caballus, Gallus gallus, Loxodonta africana, Mus musculus, Saccharomyces cerevisiae, Takifugu rubripes and Xenopus tropicalis*. We then compiled an evolutionarily conserved set of TRs by arbitrarily requiring the presence of at least 5 one-to-one orthology relationships disregarding the involved species. This produced a set composed by 1,989 non redundant genes. We finally tested each of the 1,989 gene symbols obtained through the orthology filtering for the association with at least 1 Affymetrix HGU 133 Plus 2 microarray platform (according to the content of BioMART). This produced a final set of conserved putative transcriptional regulators composed by 1,919 genes. We thus check our conserved TRs set for the overlap with annotated NCBI's UniGene clusters. Again, the required annotations were obtained using the BioMART web interface using the whole human ensembl known gene set (version 51) and limiting the extraction to the genes set produced by the orthology 1-to-1 filter. 1,620 conserved TRs genes were found

to be overlapped with at least 1 NCBI UniGene cluster. We thus removed from our transcriptional regulator set all the gene symbols lacking an annotation in the Bioconductor annotation package associated with the Affymetrix HGU 133 Plus 2 platform. This filtering procedure produced the reference conserved TRs set composed by 1,611 evolutionarily conserved genes. The genes comprised in the conserved TRs set were tested for differential expression in AC133+AML cells by mean of a standard t-test. In order to identify a set of genes suitable for ontology driven functional enrichment investigations we applied a standard hypergeometric test. We followed this approach and we selected a set of overexpressed conserved TRs by mean of a standard Welch t-test with statistical cutoff of 0.05 (upper tail). This procedure identified a set of 734 (out of 1,611 TR genes) conserved putative TRs overexpressed in AC133+AML cells. The TR overexpressed genes were then tested for functional enrichment using the Bioconductor GOstats package (function: hyperGTest).

2.5 CELL COLTURE

Selected AC133+ cells from A46 BM at AML diagnosis were cultured for 16 weeks, using synthetic medium StemSpam H3000 (StemCell Technologies, Vancouver, Canada) in the absence of serum and cytokines.

2.6 IMMUNOSTAINING

Indirect immunostaining

Indirect immunostaining was performed following standard procedures. Bone marrow biopsies of AML patients, previously embedded in paraffin blocks were cut in 5 μm thick sections and mounted on slides. Slides were loaded into glass slide-holders and dewaxed as follows: twice in 100% xylene (Carlo Erba Reagents, Rodano, Italy) 15min and 10min, twice in 100% EtOH (Panreac, Quimica Sau, Castellar del Vallès, Spain) for 5min, once in 95% EtOH for 5min, once in 85% EtOH for 5min, once in 65% EtOH for 5min, once in 30% EtOH for 5min and three times in water for 5min each. Epitope retrieval was performed using the boiling method with buffer citrate (10mM, pH6). Slides were dipped into already boiling buffer citrate for 20min and then brought to room temperature in water. The blocking procedure was performed overnight using Phosphate Buffered Saline (PBS) plus 5% Bovine Serum Albumin (BSA) (Sigma Aldrich, St. Louis, US) and 0.05% Tween 20 (Roche, Mannheim, Germany). The day after, slides

were incubated with primary antibodies such as mouse anti-Active-β-Catenin (1:100, Millipore, Billerica, MA, US), rabbit anti-WNT10B (H70) (1:100, Santa Cruz Biotechnology, Santa Cruz, CA, US), for 5h and then washed three times with PBS plus 5% BSA and 0.05% Tween 20.

PRIMARY ANTIBODIES	DILUITION	SECONDARY ANTIBODIES	DILUITION
Mouse anti-Active-β-Catenin	1:100	Donkey anti-mouse Alexa Fluor 488	1:500
Rabbit anti-WNT10B	1:100	Donkey anti-rabbit Alexa Fluor 568	1:500

Samples were incubated with the secondary antibodies donkey anti-mouse Alexa Fluor 488 (1:500, Life Technologies, Carlsband, CA, US), and donkey anti-rabbit Alexa Fluor 568 (1:500, Life technologies). Exposure times for slides images were 600-620 ms for DAPI; 450-480 ms for Alexa Fluor 568 and 370-400 ms for Alexa Fluor 488. Nuclei were counterstained with 100ng/ml DAPI (Sigma-Aldrich). Cells were analyzed using the upright microscope (Leica, DM 4000B).

FLUOROPHORES	λ ABSORPTION	λ EMISSION
DAPI	360/40 nm	470/40 nm
488 nm	495 nm	519 nm
568 nm	578 nm	603 nm

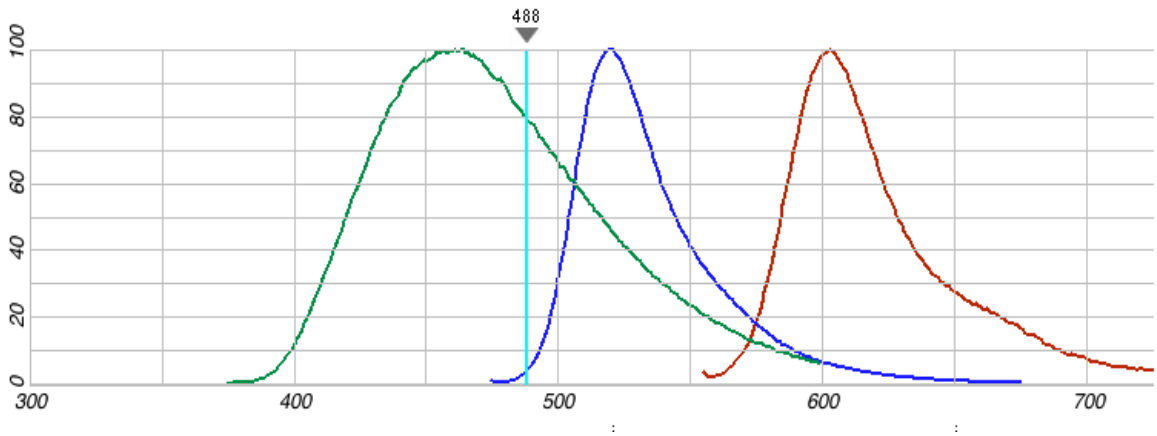


Figure 2.3 Image from Fluorescence Spectra Viewer (Life Technologies). DAPI emission is shown in green, 488 nm emission is shown in blue and 568 nm emission is shown in red.

Direct immunostaining

Direct immunostaining was performed following standard procedures. Bone marrow biopsies of AML patients, previously embedded in paraffin blocks were cut in 5 μm thick sections and mounted on slides. Slides were loaded into glass slide-holders and dewaxed as follows: twice in 100% xylene (Carlo Erba Reagents, Rodano, Italy) 15min and 10min, twice in 100% EtOH (Panreac, Quimica Sau, Castellar del Vallès, Spain) for 5min, once in 95% EtOH for 5min, once in 85% EtOH for 5min, once in 65% EtOH for 5min, once in 30% EtOH for 5min and three times in water for 5min each. Epitope retrieval was performed using EDTA (0.5M, pH 8.0) at 37°C for 30 min. The blocking procedure was performed overnight using Phosphate Buffered Saline (PBS) plus 5% Bovine Serum Albumin (BSA) (Sigma Aldrich, St. Louis, US) and 0.05% Tween 20 (Roche, Mannheim, Germany). Primary antibodies mouse anti- CD133.1 (AC133) (Milteny Biotech, Bergisch Gladbach, Germany) and mouse anti-Active-β-Catenin (Millipore, Billerica, MA, US), were direct labeled using and Mix-n-Stain™ CF™ 647A Antibody Labeling Kit (5-20μg Sigma Aldrich, St. Louis, US). The day after, slides were incubated with primary antibodies for 3h and then washed three times with PBS plus 5% BSA and 0.05% Tween 20. Exposure times for slides images were 300-350 ms for DAPI; 550-660 ms for 488 nm dye and 470-490 ms for 647 nm dye. Nuclei were counterstained with 100ng/ml DAPI (Sigma Aldrich, St. Louis, US). Cells were analyzed using the upright microscope (Leica, DM 4000B).

FLUOROPHORES	λ ABSORPTION	λ EMISSION
DAPI	360/40 nm	470/40 nm
488 nm	495 nm	519 nm
647 nm	650 nm	668 nm

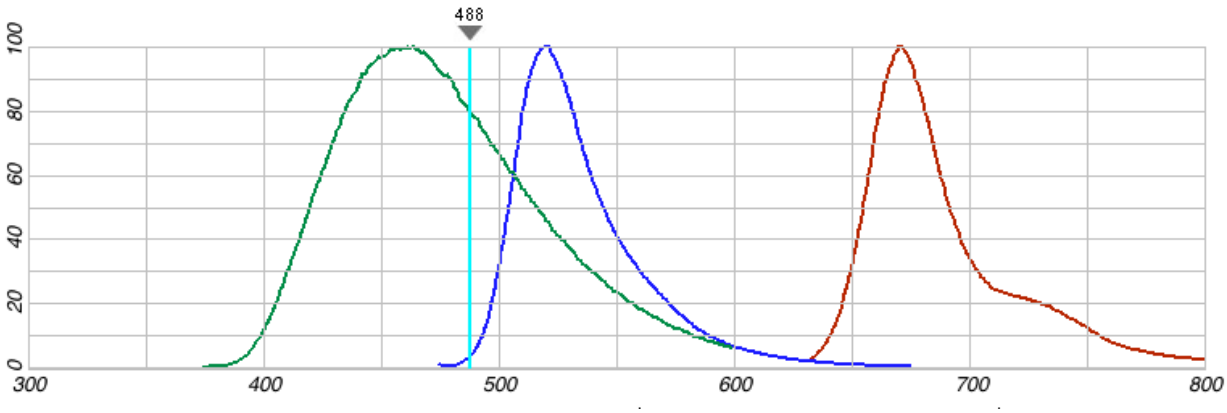


Figure 2.4 Image from Fluorescence Spectra Viewer (Life Technologies). DAPI emission is shown in green, 488 nm emission is shown in blue and 647 nm emission is shown in red.

2.7 PROBE DESIGN: MFOLD

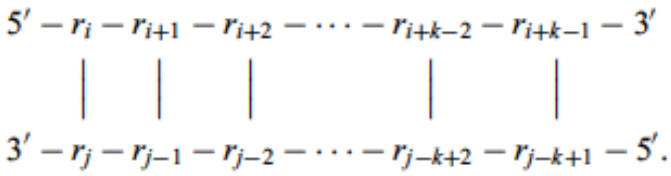
The concept of RNA secondary structure began with the work of Doty and Fresco *(Doty et al., 1959; Fresco et al., 1960)*. The prediction of RNA secondary structure (folding) by energy minimization using nearest neighbor energy parameters began with Tinoco and colleagues *(Borer et al., 1974; Tinoco & Uhlenbeck, 1971; Tinoco et al., 1973; Uhlenbeck et al., 1973)* and also with Delisi and Crothers *(Delisi & Crothers, 1971)*. Efficient algorithms for RNA secondary structure prediction using dynamic programming methods borrowed from sequence alignment were developed independently by a number of people *(Waterman & Smith, 1978; Waterman, 1978; Nussinov et al., 1978; Nussinov & Jacobson, 1980; Zuker & Stiegler, 1981; Sankoff et al., 1983; Zuker, 1989; Devereux et al., 1984)*. The 'mfold' software for RNA folding was developed in the late 1980s *(Salser, 1977; Freier et al., 1986; Zuker, 1989)*. The 'm' simply refers to 'multiple'. The core algorithm predicts a minimum free energy, ΔG, as well as minimum free energies for foldings that must contain any particular base pair. Any base pair, *ri-rj* , between the *i* th nucleotide and the *j* the nucleotide that is contained in a folding no more than δδG from the minimum, is plotted in a triangular plot called the 'energy dot plot'. The base pair *ri-rj* is plotted in row *i* and column *j* of this matrix. The free energy increment, δδG , is chosen a priori by the user, who selects a 'percent suboptimality', P . From this, δδG is computed to be P / 100 |ΔG| . Base pairs within this free energy increment are chosen either automatically, or else by the user, and foldings that contain the chosen base pair are computed. They have minimum free energy conditional on containing the chosen base pair. The description and use of the mfold package has appeared in a number of articles *(Jaeger et al., 1989; Jaeger et al., 1990; Zuker, 1994; Zuker, 1999)*. The closely related 'RNAstructure'

program has also been described (*Mathews et al., 1998; Mathews et al., 2000)*. The Turner group has published numerous articles over the years that detail the development of the RNA folding parameters. A subset of these articles are what I would call 'major works' that summarize the current state of the art. Version 1 of the mfold package used free energies that were described by Freier et al.. Versions 2.1 to 2.3 used the parameters from Walter et al. *(Walter et al., 1994)*, although the incorporation of coaxial stacking parameters into the minimization algorithm has not been accomplished. The current version 3 software uses free energy data from Mathews et al. *(Mathews et al, 1999).* DNA folding prediction with the mfold software began in 1996, when DNA specific parameters were added to the mfold package through a collaboration with the SantaLucia group (John SantaLuica Jr., Department of Chemistry, Wayne State University, Detroit, MI).

The mfold web server comprises a number of separate applications that predict nucleic acid folding, hybridization and melting temperatures (Tm s).

- Sequence name. A sequence name may be typed or pasted (entered) within the 'Enter a name for your sequence:' text field. Long names are truncated to 40 characters.
- Sequence. A sequence must be entered into the sequence text area box. All characters except for 'A–Z' and 'a–z' are removed. Lower case characters are converted to upper case. For RNA folding, 'T' or 't' are converted to 'U', while 'U' or 'u' are converted to 'T' in DNA folding. The letter ' N' should be used for an unspecified base.

- Multiple constraints of any form are allowed in any order. Force a specifi c base pair or helix to form.
- will force the formation of the helix (single base pair if k=1).



The triple (*i , j , k*) refers to k consecutive base pairs, where *ri -rj* is the exterior closing base pair. If any of these base pairs cannot exist, then an error will be generated and the job will fail.

RNA and DNA sequences may be linear or circular.

- The folding temperature is fixed at 37°C for RNA folding using version 3.0 energy rules. For RNA folding with the version 2.3 parameters, or for DNA folding, any integral temperature between 0 and 100°C may be chosen.
- Ionic conditions may be altered for DNA folding only. For RNA, the ionic conditions are fi xed at [Na^+]= 1M and [Mg^{++}]= 0 M. For folding, these are equivalent to physiological conditions. The following constraints apply: [Na^+]= 0.01 M, [Mg^{++}]= 0.1 M, and [Na^+]= 0.3M if [Mg^{++}]> 0M. For the purposes of folding, Na^+ may be considered equivalent to Li^+, K^+ and $NH4^+$, while Mg^{++} is equivalent to Ca^{++}.

PROCEDURE

We evaluate our three padlock probes, using MFold:

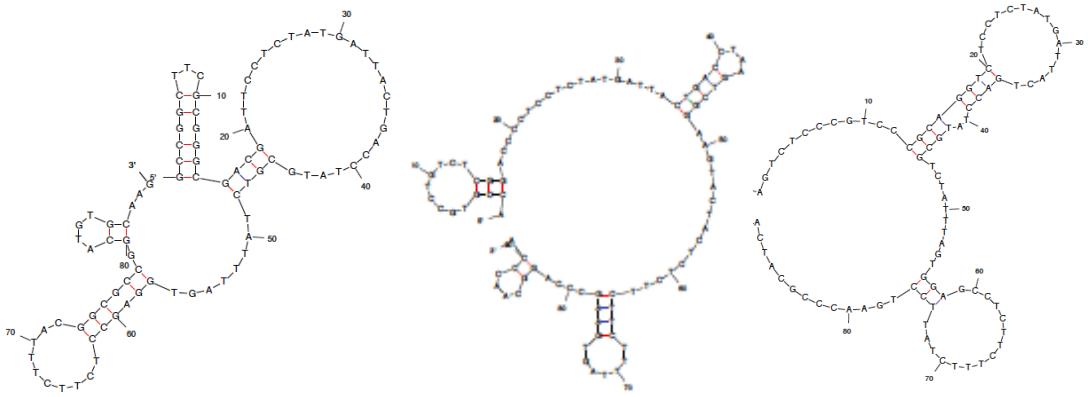


Figure 2.5 MFold output for β-actin, WNT10B, WNT10B^{IVS1} padlock probes.
The mRNA *in situ* detection is performed in denaturing conditions, without secondary structures of probes.

2.8 *IN SITU* mRNA DETECTION

In situ detection of individual mRNA molecules was performed as described. Bone marrow biopsies of AML patients, previously embedded in paraffin blocks, were cut in 5

µm thick sections and mounted on slides. Slides were dewaxed as follows: twice in 100% xylene for 15 min and 10 min, twice in 100% EtOH for 2 min, twice in 95% EtOH for 2 min, twice in 70 % EtOH for 2 min, and washed in DEPC-H_2O for 5 min and in DEPC-PBS for 2 min. Tissue fixation was performed in 3.7% (w/v) paraformaldehyde in PBS for 10 min at RT. After a wash in DEPC-PBS for 2 min, the tissue sections were then permeabilized with 2 mg/ml pepsin (Sigma Aldrich, St. Louis, US) in 0.1 M HCl at 37 °C for 2 min. Slides were washed in DEPC-H_2O for 5 min, in DEPC-PBS for 2 min and then fixed in 3.7% (w/v) paraformaldehyde in PBS for 10 min at RT. Tissue sections were then dehydrated through a series of 70%, 85% and 100% ethanol for 1 min each. Molecular reactions were carried out with a reaction volume of 100 µl in secure-seals (13 mm in diameter, 0.8 mm deep; Grace Bio-Labs) mounted over the tissue. One µM of locked nucleic acid (LNA)-modified cDNA primer was added to the slide with 10 U/µl of M-MULV reverse transcriptase (Fermentas), 500 nM dNTPs (Invitrogen), 0.2 µg/µl BSA (New England Biolabs, NEB) and 1 U/ µl RiboLock RNase Inhibitor (Fermentas) in the M-MULV reaction buffer. Slides were incubated for 3h at 37°C. After incubation, slides were washed in PBS-T (DEPC-PBS with 0.05% Tween20), followed by a post-fixation step in 3.7% (w/v) paraformaldehyde in DEPC-PBS for 30 min at RT. After post-fixation the sample were washed twice in DEPC PBS-T. To make the target cDNA strands available for padlock probe hybridization, the RNA portion of the created RNA-DNA hybrids was degradated with RNaseH (Fermentas). Ligation was then carried out with 0.1 µM of the WNT10b padlock probe and β-actin padlock probe in a mix of 0.5 U/µl Ampligase (Epicentre), 0.4 U/µl RNase H (Fermentas), 1 U/µl RiboLock RNase Inhibitor (Fermentas), Ampligase buffer, 50 mM KCl and 20% formamide. Incubation was performed first at 37 °C for 30 min, followed by 45 min at 45 °C. After ligation reaction,

the slides were washed twice in DEPC-PBS with 0.05% Tween20. Rolling Circle Amplification (RCA) was performed with 1 U/µl ϕ29 DNA Polymerase (Fermentas) in the supplied reaction buffer, 1 U/µl RNase Inhibitor (Fermentas), 250 µM dNTPS (Invitrogen), 0.2 µg/µl BSA (NEB) and 5% glycerol. Incubation was carried out for 5h at 37°C, and it was followed by a twice wash in PBS-T. Rolling Circle Particles (RCPs) were visualized using 100 nM of detection probe in 2X SSC and 20% formamide at 37°C for 20 min. Slides were then washed in DEPC-PBS. . Nuclei were counterstained with 100ng/ml Hoechst 33258 (Sigma-Aldrich). The Secure-seals were removed and the slides were dehydrated using a series of 70%, 85% and 99.5% ethanol for 3 min each. The dry slides were mounted with Invitrogen Slowfade. Images of bone marrow tissue slides were acquired using an Axioplan II epifluorescence microscope (Zeiss) equipped with a 100 W mercury lamp, a CCD camera (HRM, Zeiss), and a computer-controlled filter wheel with excitation and emission filters for visualization of DAPI, Cy3, and Cy5. A ×20 (Plan-Apocromat, Zeiss) and ×40 (Plan-Neofluar, Zeiss) objective were used for capturing the images. Images were collected using the Axiovision software (release 4.3, Zeiss). Exposure times for slides images were 520–680 ms (at ×20 magnification), 320–480 ms (×40) for DAPI; 300 ms (x20), 650 ms (×40) for Cy3; 250 ms (×20), 580 ms (×40) for Cy5. Images were collected as *z*-stacks to ensure that all RCPs were acquired, with a maximum intensity project created in Axiovision. For quantification, the numbers of RCPs and cell nuclei in images were counted digitally using CellProfiler (www.cellprofiler.org) on three x20 microscope images . The total number of RCPs was divided by the number of nuclei for each image. The average for each sample was then calculated from the result of the five images and is reported as RCPs per cell. The threshold for different color channels was set with ImageJ 1.41 .

β- ACTIN LNA PRIMER

FLUOROPHORES	λ ABSORPTION	λ EMISSION
HOECHST 33258	346 nm	460 nm
CY3	550 nm	570 nm
CY5	622/36 nm	667/30nm

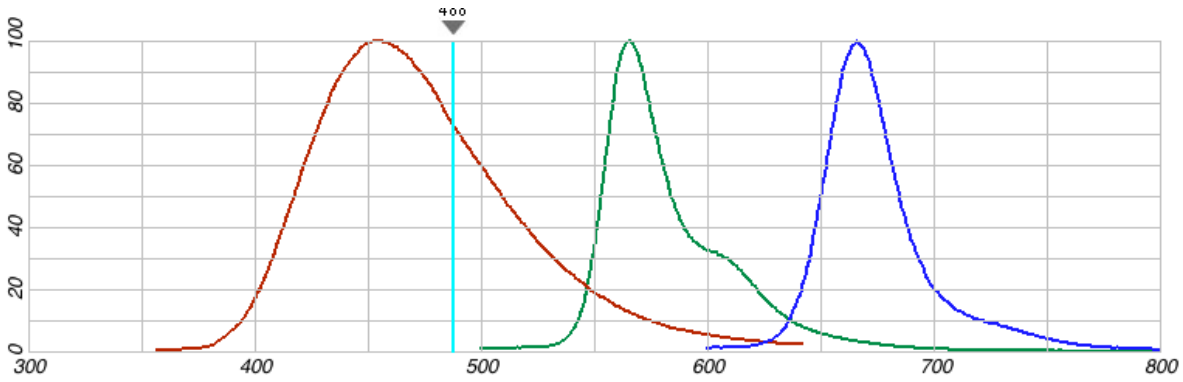


Figure 2.6 Image from Fluorescence Spectra Viewer (Life Technologies). Hoechst emission is shown in red, CY3 emission is shown in green and CY5 emission is shown in blue.

5’-C+TG+AC+CC+AT+GCCCACCATCACGCCC-3’
WNT10B LNA PRIMER
5’-C+A+G+G+C+CGGACAGCGTCAAGCACACG-3’

β- ACTIN PADLOCK PROBE
5’-[Phos]GCCGGCTTCGCGGGCGACGATTCCTCTATGATTACTGACCTATGCGTCTATTTAGTGGAGCCTCTTCTTTACGGCGCCGGCATGTGCAAG-3’
WNT10B PADLOCK PROBE
5’-[Phos]-ACCGTGCCTGTCTCGGACCCCTCCTCTATGATTACTGACCTAAGTCGGAAGTACTACTCTCTTCTTCTTTTAGTGAAGCCCAGGCAACCCA3’
WNT10B^{IVS1} PADLOCK PROBE
5’-[Phos]AGTCTCCCGTCCCGCAGGTCTCCTCTATGATTACTGACCTATGCGTCTATTTAGTGGAGCCTCTTCTTTCTATTCCTGAACCCGCATCA-3’

CY3 DETECTION PROBE
5’-[Cy3]TGCGTCTATTTTAGTGGAGCC -3’
CY5 DETECTION PROBE
5’-[Cy5]AGTCGGAAGTACTACTCTCT-3’

2.9 MICE AND TRANSPLANTATION PROCEDURE

The RAG genes encode two recombination activating proteins, RAG-1 and RAG-2, whose cellular expression is restricted to lymphocytes during their developmental stages and is fundamental for the development of mature B and T lymphocytes. In 1992, Shinkai and colleagues developed a RAG-2-deficient murine model characterized by the absence of functional B and T cells, similarly to *scid* mice *(Shinkai et al., 1992).* Unlike *scid* mutation, due to the specificity of RAG function and its involvement in the initial phases of VDJ rearrangement, the resulting phenotype is completely stable and affects only lymphocyte maturation. The deletion of RAG-2 activity caused several alterations in the physiology of the immune system: mice showed decreased or occasionally absent thymus, and a reduction of the cellularity of other lymphoid organs as spleens and lymph nodes. Crossing of RAG-2- and γ- common chain-deleted mice led to the generation of a completely genotipically defined immunodepressed murine model named $Rag2^{-/-}\gamma c^{-/-}$ mice. As previously reported, the common cytokine receptor γ- chain is a high and median affinity component of the receptors for several cytokines, such as IL-2, involved in the activation of T and NK cells and in control of the peripheral tolerance, IL-7 responsible for the development and functionality of lymphocytes, IL-9 that acts as a growth factor for mast cells, IL-15 involved in the regulation of the development and maturation of NK cells and the homeostasis of memory lymphocytes, IL-4 and IL-21 that act on immunoglobulin production and regulate the isotypic switch. $Rag2^{-/-}\gamma c^{-/-}$ mice show a combined immunodeficiency characterized by the complete absence of B, T and NK cells and share all the properties described for NOG mice, with

the exception that they are completely resistant to radiations *(Shultz et al., 2007).*

PROCEDURE

$Rag2^{-/-}\gamma c^{-/-}$ BALB/c mice were bred and maintened under specific pathogen-free conditions in the mouse facility of Istituto Oncologico Veneto. $Rag2^{-/-}\gamma c^{-/-}$ mice at 6 weeks were sublethally irradiated with 5Gy and transplanted via the tail vein with $1x10^6$ of A46 AC133+ AML cells. Three weeks after transplantation, recipient mice were sacrificed and BM cells were harvested by flushing femurs and tibias. BM engrafment was evaluated using human antibodies CD34, CD38, AC133 and CD45 (BD Biosciences Bedford, MA and Miltenyi Biotec GmBH). The cell engraftment was evaluated by the human panleukocyte CD45 marker expression.

2.10 ZEBRAFISH MODEL AND TRANSPLANTATION

PROCEDURE

Zebrafish (Danio rerio) was first pioneered as an animal model system by George Streisinger at the University of Oregon at Eugene in the 1970s *(Stahl, 1985; Westerfield, 1995)*. In recent years, the zebrafish has become one of the most important vertebrate model organisms to study biological processes in vivo. This is due to a combination of advantages making it an ideal organism for the researchers in many aspects of embryonic development, physiology and disease.

Zebrafish is a teleost fish of the cyprinid family in the class of ray-finned fishes. The lineage leading to the cyprinds and mammals split about 450 million years ago *(Nusslein-Volhard, 2002).* The teleosts include other model organisms such as medaka (Oryzias latipes), the pufferfish (Takifugu ruripes, Tetraodon nigrovirdis) and the three spined stickleback (Gasterosteus aculeatus) *(Nusslein-Volhard, 2002).* The zebrafish genome is about 1.7 gigabases in size, which is just more than half of tetrapods, such as human and mice and is divided into 25 chromosomes *(Nusslein-Volhard, 2002).* The absolute number of genes in zebrafish is currently unknown, however, a large number of gene has been cloned and analyzed to date.

As a vertebrate, it has many of the strengths of invertebrate model systems, such as small size, easy maintenance, a large number of offspring and a short generation time. Its transparent, easily accessible embryos, as well as their large size make it ideally suited to micromanipulation and in vitro observations *(Nusslein-Volhard, 2002).* Moreover, its very rapid and synchronous embryonic development greatly facilitates phenotypic analysis and large-scale experimental approaches. During the first 24 hours

of development, the embryos are completely transparent, allowing the visualization of developing organs. After about 2 days all common vertebrate specific body features can be seen, including compartmentalized brain, eyes, ears and all internal organs *(Westerfield, 1995).* In comparison to higher vertebrates, the organs in zebrafish are like a minimalist version, using far fewer cells to fulfill the equivalent function in the organism. For instance, the kidney of the larval zebrafish consist of a single glomerulus which runs most of the body length, in comparison to the tens of thousands of glomeruli in the mammalian kidney *(Nusslein-Volhard, 2002).* The zebrafish larvae have hatched and are able to swim and search food as soon as 5 days after fertilization. The zebrafish reaches sexual maturity in approximately 2-3 months *(Nusslein-Volhard, 2002).*

With the community of zebrafish researchers growing worldwide, new techniques and methodologies are becoming available at an ever-increasing rate and adaptation of techniques from other model organisms. In addition, the zebrafish genome has been sequenced to facilitate disease and orthologous gene identification (http://www.sanger.ac.uk/modelorgs/zebrafish.html).

PROCEDURE

Embryos were handled according to relevant guidelines. Fish of the AB strain were maintained at 28°C on a 14-hr light/10-hr dark cycle. Embryos were collected by natural spawning and staged according to Kimmel work. Zebrafish embryo cells transplantation was performed as previously reported in Hatta et al. work. Briefly, fluorescently labeled A46 cells were resuspended in 1x PBS and injected into zebrafish blastulae (between

100 and 200 cells per injection) at 3 hpf. Injected live embryos were immediately observed under a fluorescent microscope to ensure for the presence of labeled A46 cells. Embryos were collected at the desired developmental stages and immediately fixed and processed for whole-mount *in-situ* hybridization (WISH) according to Thisse and colleagues using *gsc*, *ntl*, and *pax2a* DIG-labeled riboprobes.

2.11 RAPID AMPLIFICATION OF cDNA (RACE PCR)

The Rapid Amplification of cDNA (RACE PCR) is a procedure for amplification of nucleic acid sequences from a messenger RNA template between a defined internal site and unknown sequences at either the 3' or the 5' -end of the mRNA. This method is used to extend partial cDNA clones by amplifying the 5′ sequences of the corresponding mRNAs. The technique requires knowledge of only a small region of sequence within the partial cDNA clone. During PCR, the thermostable DNA polymerase is directed to the appropriate target RNA by a single primer derived from the region of known sequence; the second primer required for PCR is complementary to a general feature of the target, in the case of 5′ RACE, to a homopolymeric tail added (via terminal transferase) to the 3′ termini of cDNAs transcribed from a preparation of mRNA. This synthetic tail provides a primer-binding site upstream of the unknown 5′ sequence of the target mRNA. The products of the amplification reaction are cloned into a plasmid vector for sequencing and subsequent manipulation.

In general, PCR amplification of relatively few target molecules in a complex mixture requires two sequence-specific primers that flank the region of sequence to be amplified. In this procedure, mRNAs are converted into cDNA using reverse transcriptase (RT) and an oligo-dT adapter primer. Specific cDNA is then directly amplified by PCR using a gene-specific primer (GSP) that anneals to a region of known exon sequences and an adapter primer that targets the poly(A) tail region. This permits the capture of unknown 3'mRNA sequences that lie between the exon and the poly(A) tail. 5' RACE, or “anchored” PCR, is a technique that facilitates the isolation and characterization of 5' ends from low-copy messages. The method has been reviewed by both Frohman *(Frohman, 1990*) and Loh *(Loh, 1991).*

PROCEDURE

1µg of total RNA derived from AML46 sample, was used to prepare 5'RACE ready cDNA using 5´ RACE System for Rapid Amplification of cDNA Ends, Version 2.0 amplification kit (Invitrogen) and following the protocol supplied.

First strand cDNA is synthesized from total or poly(A)+ RNA using a gene-specific primer (GSP1) that the user provides and SuperScript™ II, a derivative of Moloney Murine Leukemia Virus Reverse Transcriptase (M-MLV RT) with reduced RNase H activity. This permits cDNA conversion of specific mRNA, or related families of mRNAs, and maximizes the potential for complete extension to the 5' -end of the message. Following cDNA synthesis, the first strand product is purified from unincorporated dNTPs and GSP1. TdT (Terminal deoxynucleotidyl transferase) is used to add homopolymeric tails to the 3' ends of the cDNA. After first strand cDNA synthesis, the original mRNA template is removed by treatment with the RNase Mix (mixture of RNase H, which is specific for RNA:DNA heteroduplex molecules, and RNase T1). Unincorporated dNTPs, GSP1, and proteins are separated from cDNA using a S.N.A.P. Column. A homopolymeric tail is then added to the 3'-end of the cDNA using TdT and dCTP. Since the tailing reaction is performed in a PCR-compatible buffer, the entire contents of the reaction may be directly amplified by PCR without intermediate organic extractions, ethanol precipitations, or dilutions. PCR amplification is accomplished using Taq DNA polymerase, a nested, gene-specific primer (GSP2) that anneals to a site located within the cDNA molecule, and a novel deoxyinosine-containing anchor primer

provided with the system.

PRIMER GSP1	5'-CCCAGAATCTCATTGCTTAGAG-3'
PRIMER GSP2	5'-CTCCTCCAGCATGTCGAAGC-3'

Conventional agarose gel electrophoresis was used to analyze DNA fragments. The electrophoretic mobility of DNA fragments mainly depends on the fragment size and to a lesser extent on the conformation of the DNA, type and concentration of agarose used as well as applied voltage and electrophoresis buffer used. Agarose gels have greater range of separation and can resolve DNAs from 50 bp to 20 kbp in length. For separating smaller sized PCR fragments 2% and 3% gels were used. Agarose gels were prepared by dissolving agarose powder in Tris-Boric Acid-EDTA buffer (TBE). 1-2 % (m/v) agarose gels were used depending on the size of the PCR amplicon.

PCR products were cloned into pCR®II-TOPO® plasmid using the TOPO TA Cloning® kit (Invitrogen). 2µl PCR Product was mixed with 1µl salt solution (1.2 M NaCl; 0.06 M MgCl2), 1µl sterile H2O, and 1µl TOPO®vector. For each transformation, 1-50 µg of DNA was added to 25-50 ml of chemically competent cells. The reaction was incubated at

room temperature for 5-30 minutes, and subsequently chilled on ice for 5 minutes. 0.5-2µl of the ligation reaction was transformed into Escherichia Coli strains Top10F (Invitrogen). Incubated on ice for 30 min, the cells were next subjected to heat shock at 37 or 42 °C for 1 min and next incubated on ice for 2 min. The cells were recovered in SOC (2% bacto-tryptone, 0.5% bacto-yeast extract, 10mM NaCl, 2.5mM KCl, 10mM MgCl2, 20mM glucose) broth and then incubated for 1 h at 37°C with shaking (200-250 rpm). Cells were plated on LB-agar containing appropriate antibiotics (ampicillin 100µg/ml) with an incubation at 37°C overnight.

To allow blue/white screening of bacterial colonies 40µl of 40mg/ml X-gal (Genespin), and 40µl 100µg/µl IPTG (Genespin) were applied to plates, and left to dry before plating. White colonies were selected for restriction analysis of miniprep DNA. Miniprep DNA was performed using a boiling method.

DNA sequencing was carried out using the dideoxy nucleotide chain-termination method *(Sanger et al., 1977)* with dye terminator labelling of purified plasmids or PCR products. Specific forward and reverse primer for PCR products and universal M13 forward and reverse primer for purified plasmids were used. The DNA samples were precipitated using ethanol and the pellets were washed twice with 75% ethanol. The purified templates were dissolved in sterile water and the Applied Biosystems Big-Dye ver 3.1 chemistry kit was used to set up the sequencing reactions. Sequences were analysed on ABI 3130XL automatic sequencers (Applied Biosystems).

PRIMER M13 FW	5’-TTGTAAAACGACGGCCAGT-3’
PRIMER M13 REV	5’-CAGGAAACAGCTATGACC-3’

2.12 RT-PCR

First strand cDNA was synthesised from total RNA using IMPROM™ Reverse Transcription (Promega) system following the manufactures procedure.

For primer design, genomic sequences were retrieved from the University of California

Santa Cruz (UCSC) Genome Browser (http://genome.ucsc.edu/cgi-bin/hgGateway, version February 2009) and primers were designed by the use of the publicly available software Primer3 (http://frodo.wi.mit.edu/primer3/input.htm).

RNA was denatured with GAPDH and WNT10B primer for the gene-specific retranscription and random primers at 70 ˚C for 5 min. Then single strand cDNA was synthesised for 1hr in a mix containing 1.0 ul of reverse transcriptase, 0.5 mM dNTPs, 20U of RNase inhibitor and 5X reverse transcription buffer, 1.5 mM $MgCl_2$. The reaction was performed at 42˚C for 1 h, followed by heating at 70°C for 15 min.

The WNT10B gene-specific reaction was puriefied by using 2.5M of Sodium Acetate (pH 5.2) and 2 volumes of EtOh 100%; 30 min at -20°C followed by 30 min in a centrifuge at 4°C.

Nucleic acid sequences of interest were amplified by RT-PCR using cDNA as the template. The standard mixtures (20.0 µl) for the regular PCR reaction in this work were prepared by the following recipe:

1 µl cDNA (RT-PCR), 1.25 U of EuroTAq DNA polymerase (EuroClone), 2.5 µl of the supplied 10 x buffer (EuroClone), 1.50 µl 50 mM $MgCl_2$ (EuroClone), 0.5 mM dNTPs ,0.5 µM of forward and reverse primer.

Reactions were carried out in microcentrifuge tubes in a themocycler MJ Mini™ Personal Thermal Cycler (Bio-Rad).

with a range of cycling conditions, specific to each primer set.

The reactions were performed by the following cycling parameters: initial denaturation, amplification, elongation and final extension.

- GAPDH: 94°C 1 min, 94°C 20 sec, 60°C 15 sec, 72 °C 12 sec (30 cycles), 72°C, 5

min, 4°C forever.

- WNT10B: 94°C 2 min, 94°C 20 sec, 61°C 20 sec, 72 °C 30 sec (30 cycles), 72°C, 5 min, 4°C forever.
- $WNT10B^{IVS1}$: 94°C 2 min, 94°C 20 sec, 63°C 20 sec, 72 °C 30 sec (30 cycles), 72°C, 5 min, 4°C forever.
- WNT10A: 94°C 2 min, 94°C 20 sec, 58°C 20 sec, 72 °C 30 sec (30 cycles), 72°C, 5 min, 4°C forever.
- WNT2B: 94°C 2 min, 94°C 20 sec, 57°C 20 sec, 72 °C 30 sec (30 cycles), 72°C, 5 min, 4°C forever.
- WNT6: 94°C 2 min, 94°C 20 sec, 59°C 20 sec, 72 °C 30 sec (30 cycles), 72°C, 5 min, 4°C forever.

PRIMER RT WNT10B	5'-ATCTCATTGCTTAGAGCCCGAC-3'
PRIMER WNT10B EX1 FW	5'-GCAGCACTAGTGAAGCCCAG-3'
PRIMER WNT10B IVS1 FW	5'-CCTGAACCCGCATCAAGTCTC-3'
PRIMER GSP2	5'-CTCCTCCAGCATGTCGAAGC-3'

PRIMER WNT10A FW	5’-GTGCTCCTGTTCTTCCTACTG-3’
PRIMER WNT10A REV	5’-GATCTTGTTGCGAGTCTCCA-3’
PRIMER WNT2B FW	5’-GACAACTCTCCAGATTACTGTG-3’
PRIMER WNT2B REV	5’-AATTGGAGGGATGAGTGAGG-3’
PRIMER WNT6 FW	5’-CAACAGGACATTCGGGAGAC-3’
PRIMER WNT6 REV	5’-ATAAGAGCCTCGACTTCTCGT-3’
PRIMER GAPDH FW	5’-ACCTGCCAAATATGATGACATC-3’
PRIMER GAPDH REV	5’-CAGTGTAGCCCAGGATGC-3’

2.13 IMMUNOBLOT

The lysates were applied with the same volume of 12 µl to one slot of a 10% stacking gel. The samples were run for approximately 30 min at 30 V, and subsequently for 45 min at 100 V. The stacking gel was removed from the separating gel. Thereafter, the gel was applied followed by the PDVF membrane (PVDF Transfer Membrane, Thermo Scientific)

which was activated by dipping into methanol. Tha transfering step, was performed in Biorad instruments Trans Blot® Turbo Transfer System, with transfer buffer (192 mM glycine, 25 mM Tris-base, 0.05% SDS, and 20% methanol). The electroblotting was done for 10 min h with a voltage of 25 V. As the protein transfer was finished, the membrane was separated from the sandwich and washed three times with TBST (20 mM Tris-base, 140 mM NaCl, and 0.05% Tween® 20; adjusted to pH 7.6 with NaCl) for 5 min at RT. Afterwards, bound proteins were blocked with 5% BSA (Sigma-Aldrich) powder dissolved in TBST under continuous shaking O/N at 4°C. The incubation with the primary antibody occurred over night at 4 °C. Protein expression was assessed by immunoblot analysis using standard procedures, applying anti–active β -catenin (ABC) monoclonal mouse (anti-ABC clone 8E7; Millipore, Billerica, MA), anti–WNT10B monoclonal mouse (5A7; Abcam), anti– glyceraldehyde-3-phosphate dehydrogenase (GAPDH) polyclonal rabbit (ab97626; Abcam), antibodies. Secondary antibodies used were anti– mouse HRP, and anti– rabbit HRP (Thermo Fisher Scientific, Waltham, MA).
The detection of the HRP conjugated antibodies on the membrane was done after three washing steps with the help of the ECL™ Western Blotting Detection Reagents (EuroClone).

3. Results

3.1 AC133+ CELL FRACTION IS HIGHLY EXPANDED IN AML

According to literature's data, the AC133 antigen expression is restricted to a rare cell population with long-term reconstituting activity, ranging from 20% to 60% of all CD34+ cells, and resulting barely detectable in CD34-Lin- cells. In Brioschi et al., we evaluated the ability clonogenic of AC133 + selected cell in qualitative terms as capacity to produce colonies granulocyte / macrophage (CFU-GM) and / or erythroid (BFU-E Burst-Forming Units-Erythroid) in the presence of appropriate stimulation and in term quantitative by comparing the results with those obtained from the seeding of mononuclear cells in toto. We previously shown in Brioschi et al. work, through the comparative analysis of the data of clonogenicity, that AC133+ LT-HSCs being also highly enriched in colony forming-units (CFU) and have a stronger granulocyte/macrophage (CFU-GM) differentiation potential relative to the unsorted bone marrow mononuclear cells (BM-MNCs). Conversly, their BFU-E forming potential is lower *(Brioschi et al., 2010).*

The first step of our work is the descriptive assessment of AML and control cell populations, represented by mononuclear cells separated by density gradient and the fractions obtained for positive and negative immunomagnetic separation. In order to determine the dynamic range of expansion of AC133+ cell fraction in AML, flow cytometric quantification of CD133.1 (AC133) expression either in single staining or in combination with the pan-hematopoietic marker CD45 was performed on 25 primary AML bone marrow human samples and 10 age-matched healthy volunteer adult donors. We then proceeded to the statistical analysis defining the P value (α=0.001) by the non-

parametric test of Mann-Whitney test.

The resulting CD133.1+ cell fraction was expanded among AML patients by an average of 31.5% (interquartile range (IQR): 16.5%-53.4%) with respect to normal donors (median 0.54%, IQR: 0.17%-1.14%, P<0.0001) (Figure 3.1).

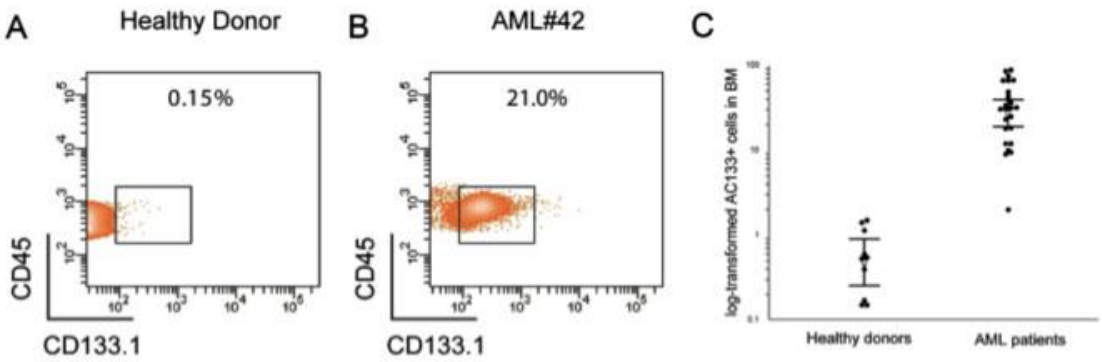


Figure 3.1 Human AC133+ cell fraction is highly expanded in AML.

Dot plots of the immunophenotype analysis from bone marrow of healthy donor (A) and an AML patient (B). The CD45/CD133.1 co-staining was perfomed on BM MNCs and percentages on cellularity are shown for gated normal and AML populations. (C). Flow cytometry analysis of the AC133 antigen in BM MNCs of healthy donors (n=10) and AML patients (n=25).

3.2 THE WNT SIGNALING IS DYSREGULATED IN AC133+ AML CELL FRACTION

In order to highlight the de-regulated pathway that characterize AC133+ AML cell fraction versus normal long-term reconstituting AC133 HSC cells, we performed a genome-wide functional enrichment analysis. The identification of differentially over-represented pathways in AC133+ AML cells is based on the functional enrichment analysis through three computational tools: GOstats (version 2.12.0 of the Bioconductor package), DAVID (Database for Annotation Visualization and Integrated Discovery; http://david.abcc.ncifcrf.gov/home.jsp) and dysregulated pathway analysis, employing the pathway information from GO BP (Gene Ontology Biological Processes) and KEGG databases *(Falcon & Gentleman, 2007; Huang et al., 2009; Majeti et al., 2009)*. We identified:

- 212 GO BP terms functionally enriched with GOStats (p-value <0.01);
- 284 GO BP terms with DAVID (p-value <0.05);
- 616 GO BP terms with the non-parametric test (p-value<0.05).

Moreover, GOStats selected 16 KEGG pathways (p-value < 0.05), DAVID 24 KEGG pathways (p-value < 0.05) and the dysregulated pathway analysis 3 KEGG pathways (p-value < 0.05). Thus, the three computational types of analysis applied to all genes, selected the term *"Wnt receptor signaling pathway"* (p-value= 0.022) as the most specific de-regulated pathway (Figure 3.2).

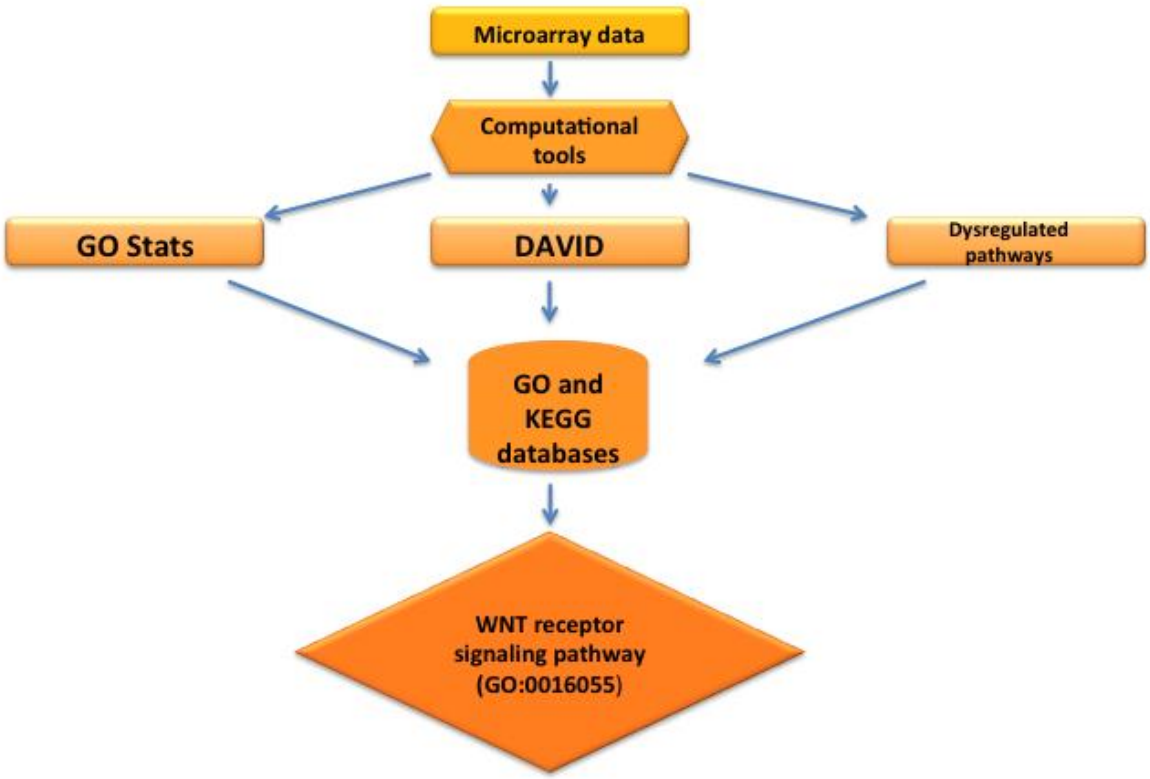


Figure 3.2 Schematic representation of expression profiling analysis and identification of de-regulated pathway. The microarray data, obtained from the gene expression profiling study, are used as query for functional enrichment analysis through three computational tools: GOstats, DAVID and dysregulated pathway analysis, employing GO BP (Gene Ontology Biological Processes) and KEGG databases.

Since the accepted role of transcriptional factors (TFs) in myeloid leukemogenesis, we

performed also the same functional enrichment analysis focusing on genes annotated as transcription regulators (TRs), employing the evolutionarily conserved human transcriptional regulators data set (http://homes.dsi.unimi.it/~re/TRdataset/). The first step of our analysis consisted in the definition of a set of genes annotated as transcriptional regulators (TRs) and conserved in a broad range of organisms, from yeast to human. The entry point of our selection procedure was the extraction from Pfam of all the domains involved in transcription regulation, which resulted in a set of 400 Pfam domains. Next, we used this list to query the entire set of ensemble human genes using the BioMART web interface. This produced an initial set of 2,971 non-redundant genes associated to a unique HGNC symbol. We thus adopted a direct SQL access to the ensembl_compara_51 database (hosted at ensembldb.ensembl.org) in order to extract all the one-to-one orthology relationships occurring between the genes in the initial set and the complete set of genes in the following organisms: *Bos taurus, Canis familiaris, Danio rerio, Drosophila melanogaster, Equus caballus, Gallus gallus, Loxodonta africana, Mus musculus, Saccharomyces cerevisiae, Takifugu rubripes and Xenopus tropicalis*. We then defined an evolutionarily conserved set of TRs by requiring the presence of at least 5 one-to-one orthology relationships disregarding the involved species. This produced a set composed of 1,989 non redundant genes. We finally tested each of the 1,989 gene symbols obtained through the orthology filtering for the association with at least 1 probe set in the Affymetrix hgu133plus2 array. This defined a final set of conserved putative transcriptional regulators composed of 1,919 genes. We can perform a first comparison with the study conducted by Messina et al., in which the analysis was carried out starting from 479 genes, present in the TF TRANSFAC database and integrated with the 852 genes in the InterPro database *(Messina et al., 2004)*. The data obtained,

followed by an orthology test with the FlyBase database of the Gene Ontology, were further analyzed, in manual mode, in order to combine the data present in the literature with the role played by TF and then confirm that each mRNA emerged from the analysis was associated with a regulator of transcription *(Hunter et al., 2008; Levine & Tijan 2003)*. One feature that differentiates the two studies is the analysis of orthology, performed against only one animal species, in the work of Messina et al. and instead made against eleven different species in our study; the test of orthology multi-species executed following analysis of comparative genomics, through the orthology and paralogy proposed by Ensembl (version 51) and by alignments with WU-BLAST, has allowed us to greatly enhance the data obtained, as it allowed the selection of only those proteins that, subject to evolutionary pressure, have retained their biological role in the species compared. Besides, the pipeline adopted by the authors resulted in the definition of a set of 1,468 human TFs. In order to explain the difference in size, we compared our pipeline with the procedure presented in Messina et al., work. After considering the absence of a comparative genomics filtering, the most relevant difference between the compared approaches is the absence in our pipeline of a filtering aimed at providing strong support for the expression of our set of TRs. We thus check our conserved set of TRs for an overlap with annotated NCBI's UniGene clusters. Moreover, the required annotations were obtained using the BioMART web interface and the whole human known Ensembl gene set (version 51) and limiting the extraction to a set of genes produced by the orthology 1-to-1 multi-species filter. One thousand, six hundred and twenty conserved TRs genes were found to overlap with at least 1 NCBI UniGene cluster. Therefore we removed from our set of TRs all the gene symbols lacking an annotation in the Bioconductor annotation package associated with the Affymetrix Hgu133plus2 array.

This filtering procedure produced the reference conserved TRs set composed of 1,611 evolutionarily conserved genes (http://homes.dsi.unimi.it/~re/TRdataset/). We tested genes comprised in the conserved TRs set for differential expression in AC133+ AML cells. In this stage of the analysis our goal was to isolate a set of genes highly or moderately overexpressed in the leukemic condition in order to identify genes suitable for ontology driven functional enrichment investigations using a standard hypergeometric test. We selected a set of overexpressed conserved TRs by means of Welch t-test (α=0.05, one-tailed).

This procedure identified a set of 734 conserved putative TRs overexpressed in AC133+ AML cells (http://homes.dsi.unimi.it/~re/TRdataset/) (Figure 3.3).

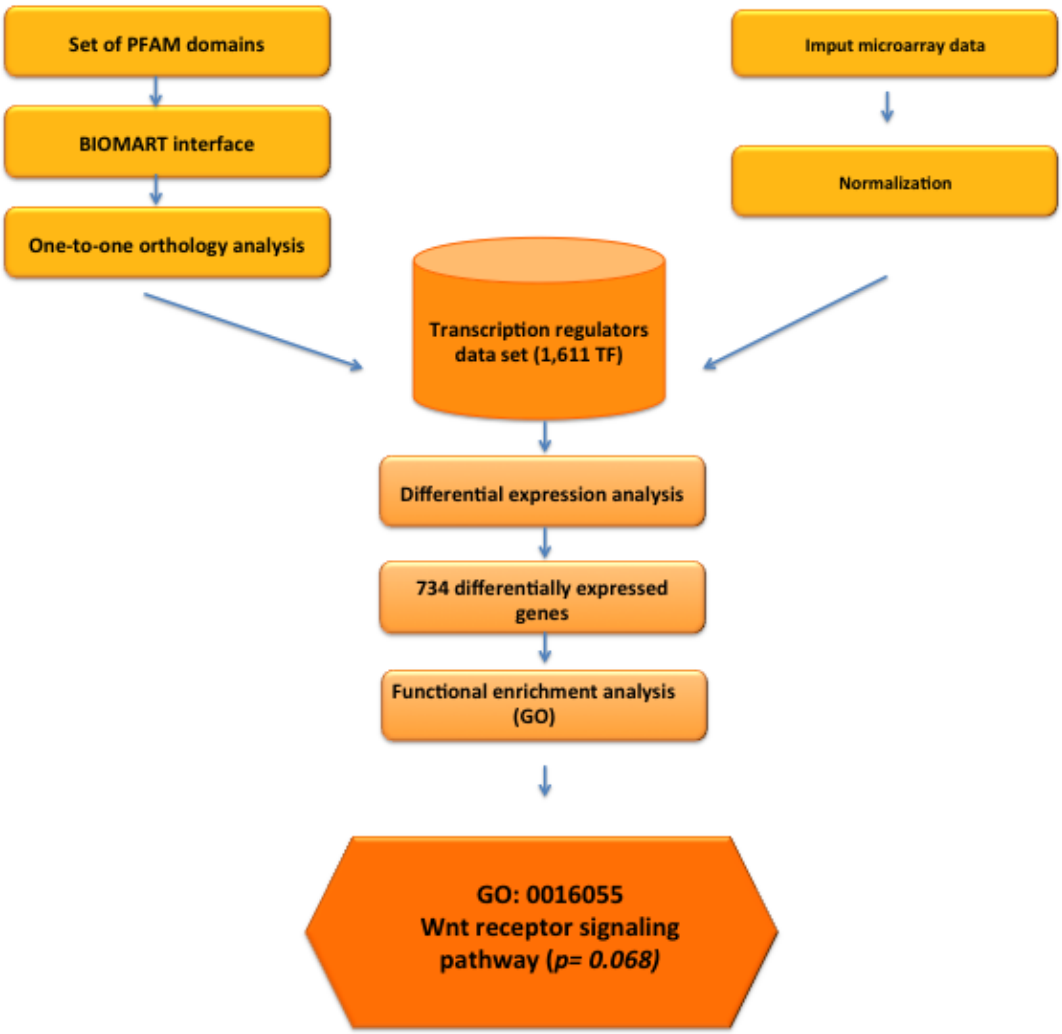


Figure 3.3 Schematic representation of definition of transcription factors dataset and functional enrichment analysis. Construction of conserved human transcriptional regulators (TRs) data set, obtained after the bioinformatic tools interrogation; this dataset was used as a filter for gene expression profiling data in order to define the transcriptional program differentially expressed between AML and healthy donors. The "Wnt signaling pathway" (GO:0016005) term is the most specific de-regulated Go term in AML.

The overexpressed TR genes were then tested for functional enrichment using the Bioconductor GOstats package (www.rproject.org *function: hyperGTest*). All the functional terms retrieved are associated with regulatory processes because of the nature of the universe set, which is entirely composed of TRs. Among the selected functional terms the most specific one, with respect to the GO Biological Process ontology, is the term 'GO:0016055', associated with the "Wnt receptor signaling pathway" and composed of 38 genes, 26 of which resulted significant by the Welch t-test (http://homes.dsi.unimi.it/~re/TRdataset/).

If we consider the study conducted by Majeti et al., our microarray analysis has not been simply adopted to evaluate the differences in gene expression between the normal and pathological hematopoietic stem cell, but the data set of transcription factors has allowed to evaluate the differences between the two types of cells as a function of the transcript *(Majeti et al., 2009).* Besides, unlike the study of Majeti et al., the gene expression analysis was performed on fractions $AC133^+$ obtained and not on the $CD34^+$ cell population, using microarray technology, Affymetrix platform® and followed by analysis of pre-processing for the validation data.

Thus, the two independent types of analysis applied to all genes and to transcription regulators, selected the term "Wnt receptor signaling pathway" as the most specific dysregulated pathway in AML.

The ligand-dependent activation of the regeneration-associated Wnt pathway

In order to look insight into the Wnt signaling in AC133+ leukemia cells, we performed a second step analysis of microarray data by using all the WNT probe sets annotated to GO class 'GO: 0016055' or to any of its children GO terms. The total list of analyzed WNT genes includes 480 probe sets, mapping to 193 different genes (http://homes.dsi.unimi.it/~re/TRdataset/).

In order to assess differential expression of WNT genes we employed Welch *t*-tests (two-tailed) on the 480 probe sets, thus identifying 103 differentially expressed genes (http://homes.dsi.unimi.it/~re/TRdataset/).

The bioinformatics analysis defined the differentially expressed WNT ligands (WNT10B, WNT10A WNT2B and WNT6), the Wnt/β-catenin signaling agonists, the antagonists and deregulated targets within the canonical Wnt/β-catenin pathway with the lower *P* value as evaluated by the Bonferroni correction. Considering this data our results are define the ligand-dependent activation of the regeneration-associated Wnt pathway *(Beghini et al., 2012)* (Figure 3.4).

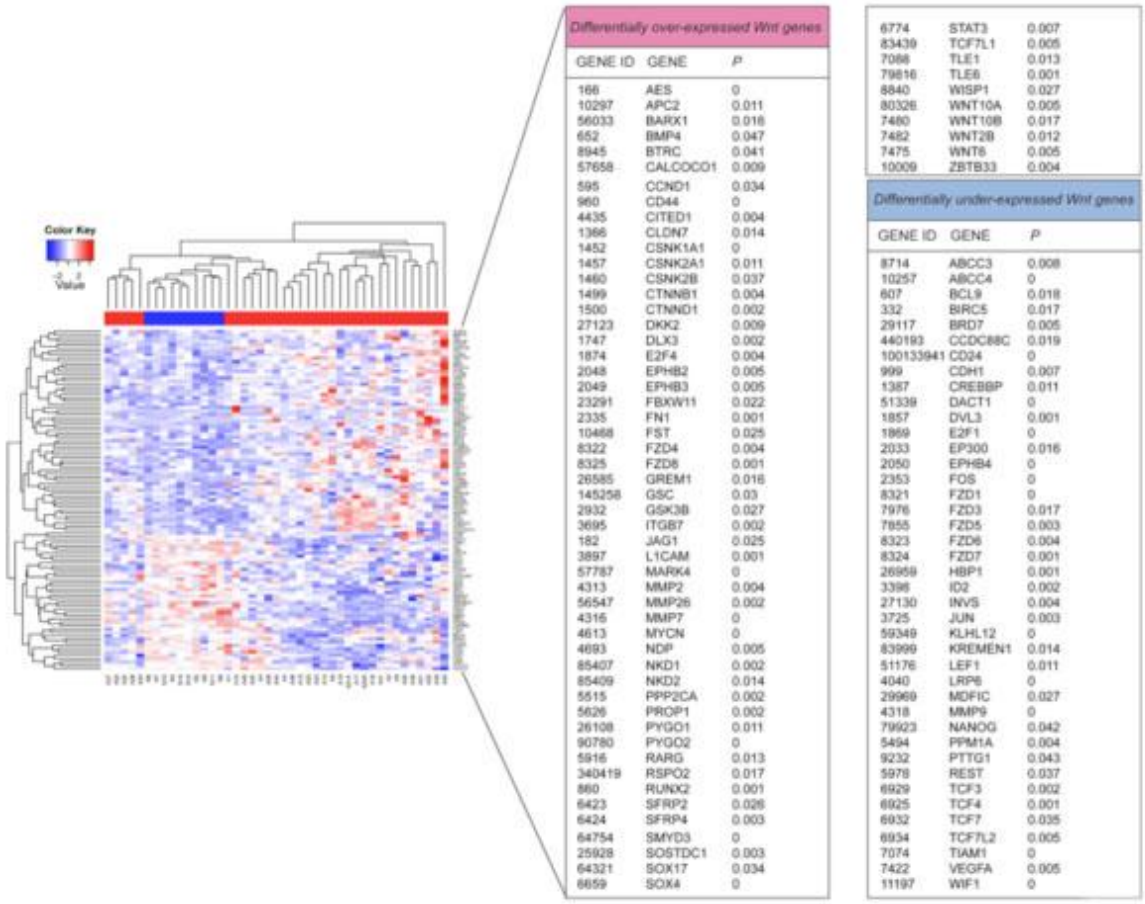


Figure 3.4 Heatmap of the differentially expressed WNT-associated genes. The names, accession numbers as well as P values for the differentially overexpressed and underexpressed genes are shown in the right panels. The double analysis of clustering, that is shown in the heat-map of probe-sets differenzially expressed, makes the distinction between patients and controls.

3.3 THE WNT10B HEMATOPOIETIC REGENERATIVE MOLECULE IS EXPRESSED IN AML

According to our previously data indicating transcriptional activation of canonical WNTs, we investigated how expression of WNT10B is related to AML phenotype through new in situ approaches *(Beghini et al., 2012).*

The ability to study individual transcripts in single cells has been sought after for a long time, but existing methods have limitations in selectivity or efficiency. For single-cell analysis, technologies must be very sensitive. Gene expression is a highly stochastic process, with substantial cell-to-cell variations in mRNA levels. Thus, the concept of the average cell is considered a myth and cellular biomolecules and genetic varations should instead be studied on a single-cell level *(Levsky and Singer, 2003).* The mRNA in situ detection is initiated by in situ reverse transcription using LNA containing primers, followed by digestion of the mRNA part of the mRNA/cDNA duplex for creation of a single-stranded target. Padlock probes are thereafter hybridized to the cDNA transcripts and circularized upon complete match with the correct targets. Finally, RCA is primed from the cDNA creating long concatemeric DNA molecules that will be visualized by labeling with fluorescence tagged oligonucleotides.

Using this approach, we detected WNT10B related transcript in bone marrow sections obtained at diagnosis from two randomly selected patients. In order to set up the experiment, we first tested the method to detect WNT10B transcripts in cultured human cells, in order to evaluate the correct ligation of the specific padlock probe. According to the literature's data *(Kirikoshi & Katoh, 2002),* WNT10B mRNA is relatively highly expressed in HeLa cell line; therefore we tested our padlock probe on Hela cells seeded

and fixed on slides, with β-actin as as a reference transcript. In the figure 3.5, we can observe abundant, bright, spot-like β-actin signals localized to the cytoplasm of cells.

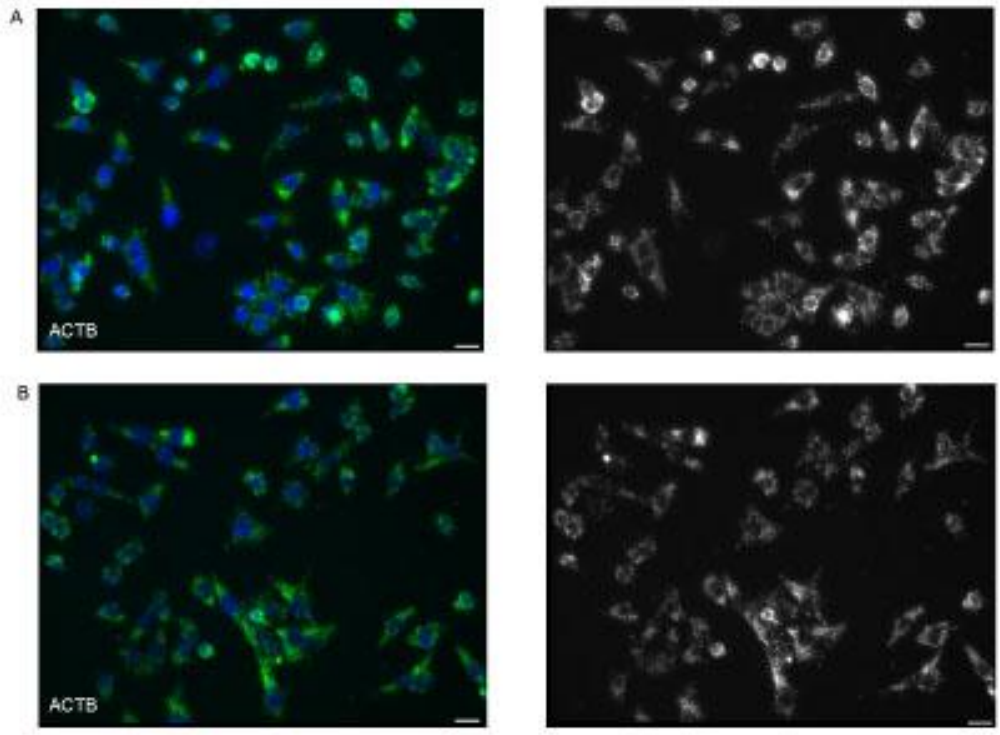


Figure 3.5 In situ detection of β-actin in cultured human HeLa cells. mRNA *in situ* detection of β-actin, visualized by green Rolling Circle Particles (A and B right). In the left panel are shown images acquired in gray scale. Cell nuclei are shown in blue. Scale bar 10 μm.

We acquired the RGB images, but also the gray scale images in order to segment all cell nuclei in the image and count the RCPs for single cell through the CellProfiler pipeline (http://www.cellprofiler.org/). Most gray level image segmentation techniques can be extended to color images as well. However, when a color image is separated to the three color components R, G and B, the color information is scattered and the information that humans can perceive is lost. When compared to gray scale images, color images have additional information that can sometimes be useful or even necessary in image analysis tasks. In the last step of the analysis, through Human Cell CellProfiler's pipeline we counted the detected signals inside each cytoplasm and nucleus. In order to consider all signals, we acquired the images by the Maximum Intensity Projection (MIP), considering x, y and z axes. Through the Human Cells CellProfiler's pipeline, we counted 110 β-actin RCPs for each HeLa cell. We can observe the distribution of RCPs around the nuclei, and we also can observe the presence of the signals on different focal planes. Following the β-actin detection, we performed the same experiment in order to test the WNT10B padlock probe on HeLa cells. As we can see in the figure 3.6 and 3.7, we found expression to vary widely among the cells. Using the same image pipeline of β-actin counting, we counted 120 bright RCPs around the cell nuclei, without any aspecific signal. The CellProfiler output data, were subsequently analyzed through R (http://www.rproject.org/), a language and environment for statistical computing and graphics.

We observed an average ratio close to 1 between the houskeeping gene and the hematopoietic regenerative-associated Wnt ligand (WNT10B) expression, indicating an high expression of WNT10B in a tumor context (Figure 3.8).

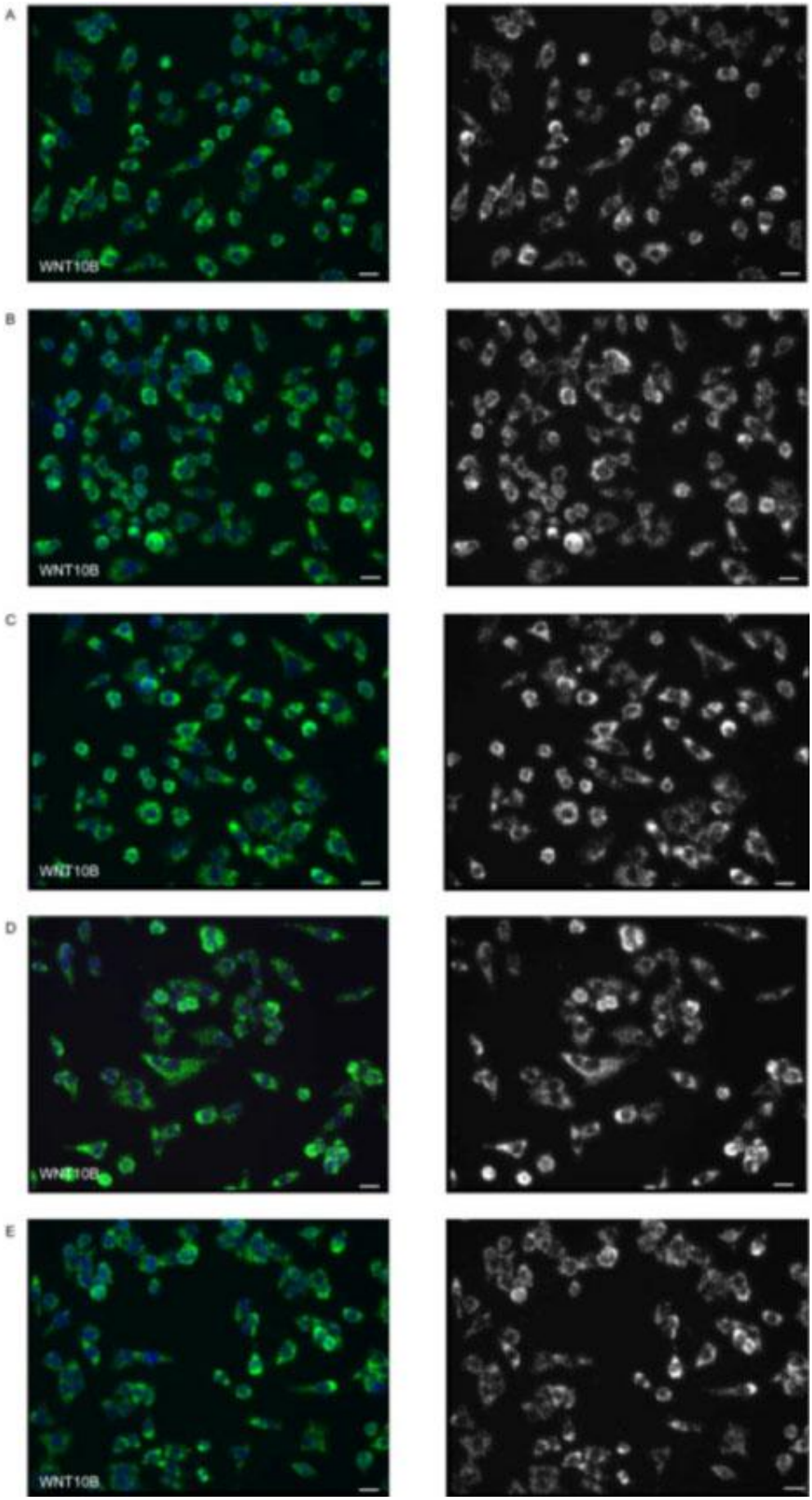

A
WNT10B
B
WNT10B
C
D
E
WNT10B

Figure 3.6 WNT10B mRNA in situ detection in HeLa cells. WNT10B RCPs are shown in green. In the figure are shown five different fields (A,B,C,D,E), acquired at X20. In the left panel are shown the same images acquired in a grey scale. Cell nuclei are shown in blue. Scale bar 10 μm.

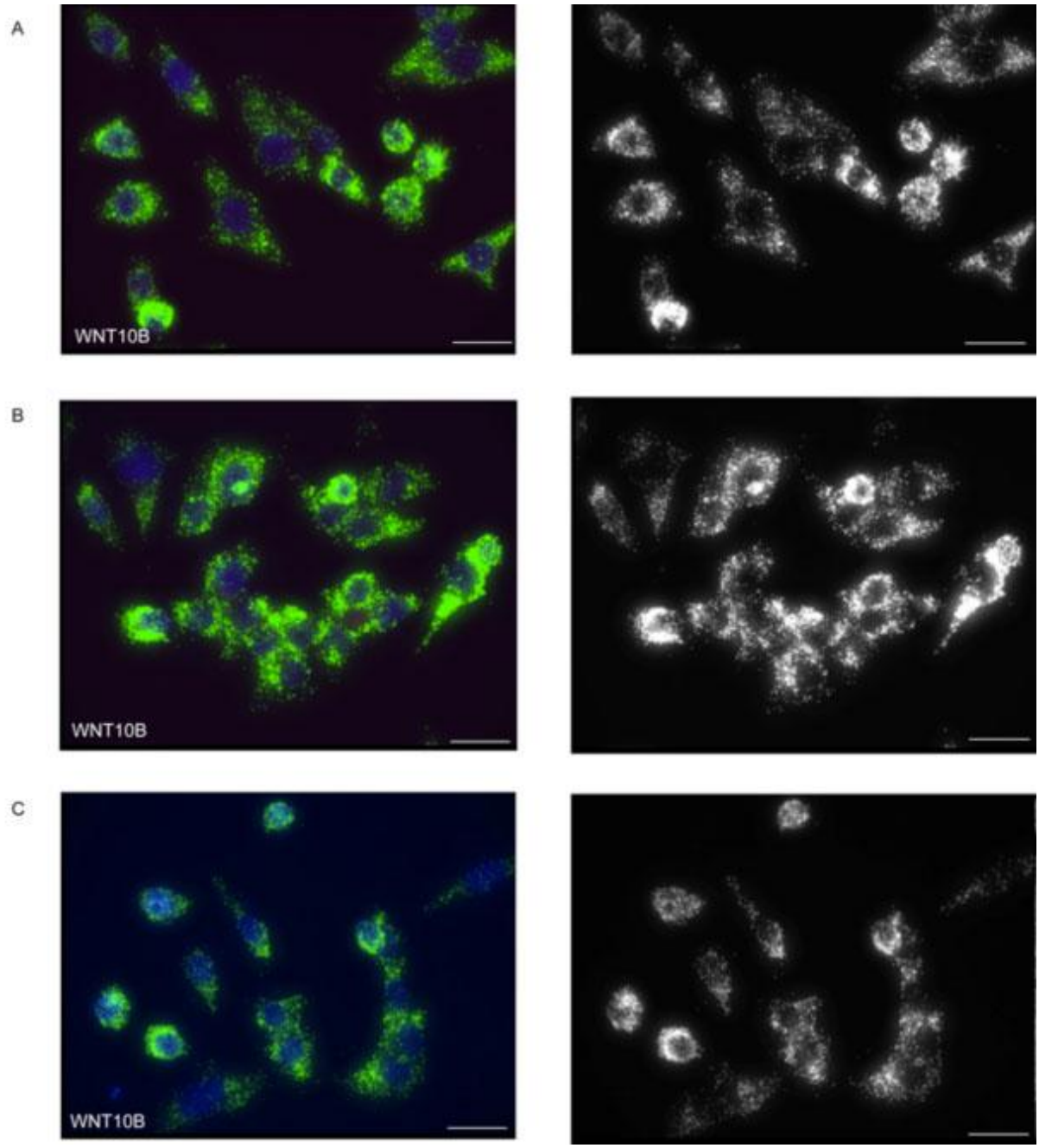


Figure 3.7 Highlits of WNT10B mRNA in situ detection. WNT10B RCPs are shown in green. In the figure are shown three different fields (A,B,C,), acquired at X40. In the left panel are shown images acquired in a grey scale. Cell nuclei are shown in blue. Scale bar 10 µm.

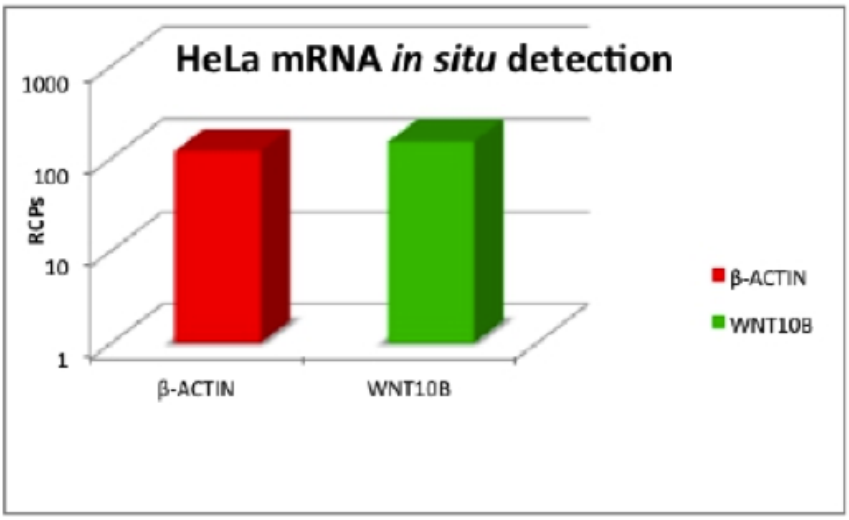


Figure 3.8 β-actin and WNT10B RCPs counting. CellProfiler output, that define the number of β-actin (red) and WNT10B (green) RCPs in each cell. The ratio between β-actin and WNT10B is close to 1.

Therefore, following the experiment's set up, we further investigated WNT10B related transcript in bone marrow sections obtained at diagnosis from two randomly selected patients, with β-actin as reference transcript in consecutive sections (Figure 3.9). Considering the presence of the β-actin mRNA, we acquired 24 fields of the AML63 bone marrow biopsy, in order to visualize the complete distribution of β-actin RCPs (Figure 3.10). The DAPI and Cy3 signals, were acquired with the same parameters in each field.

Then we performed the WNT10B mRNA in situ detection. We can note in the Figure 3.11 that the WNT10B RCPs are spread on the entire area, with an omogeneous distribution. This data suggests that the hematopoietic regenerativeassociated Wnt ligand (WNT10B) is expressed at mRNA level on both leukemic blasts and stromal-like cells, indicating a possible autocrine/paracrine involvement of Wnt in the bone marrow microenvironment.

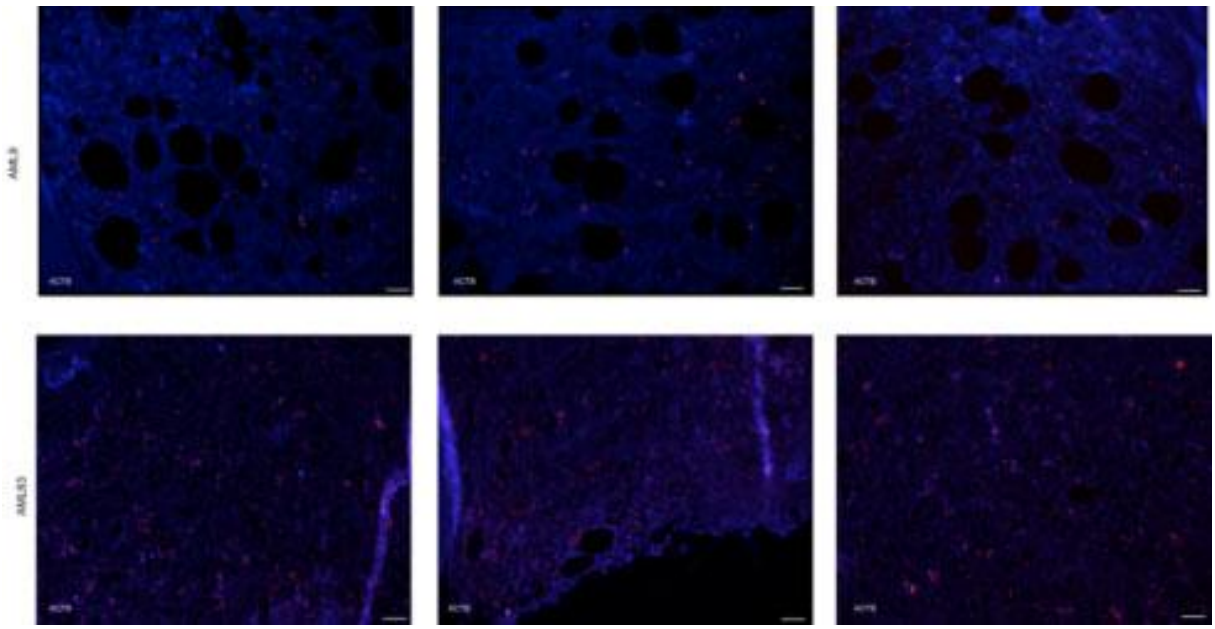


Figure 3.9 β-actin mRNA in situ detection on bone marrow biopsies. β-actin RCPs are shown in red. Cell nuclei are shown in blue. Scale bar 20 μm.

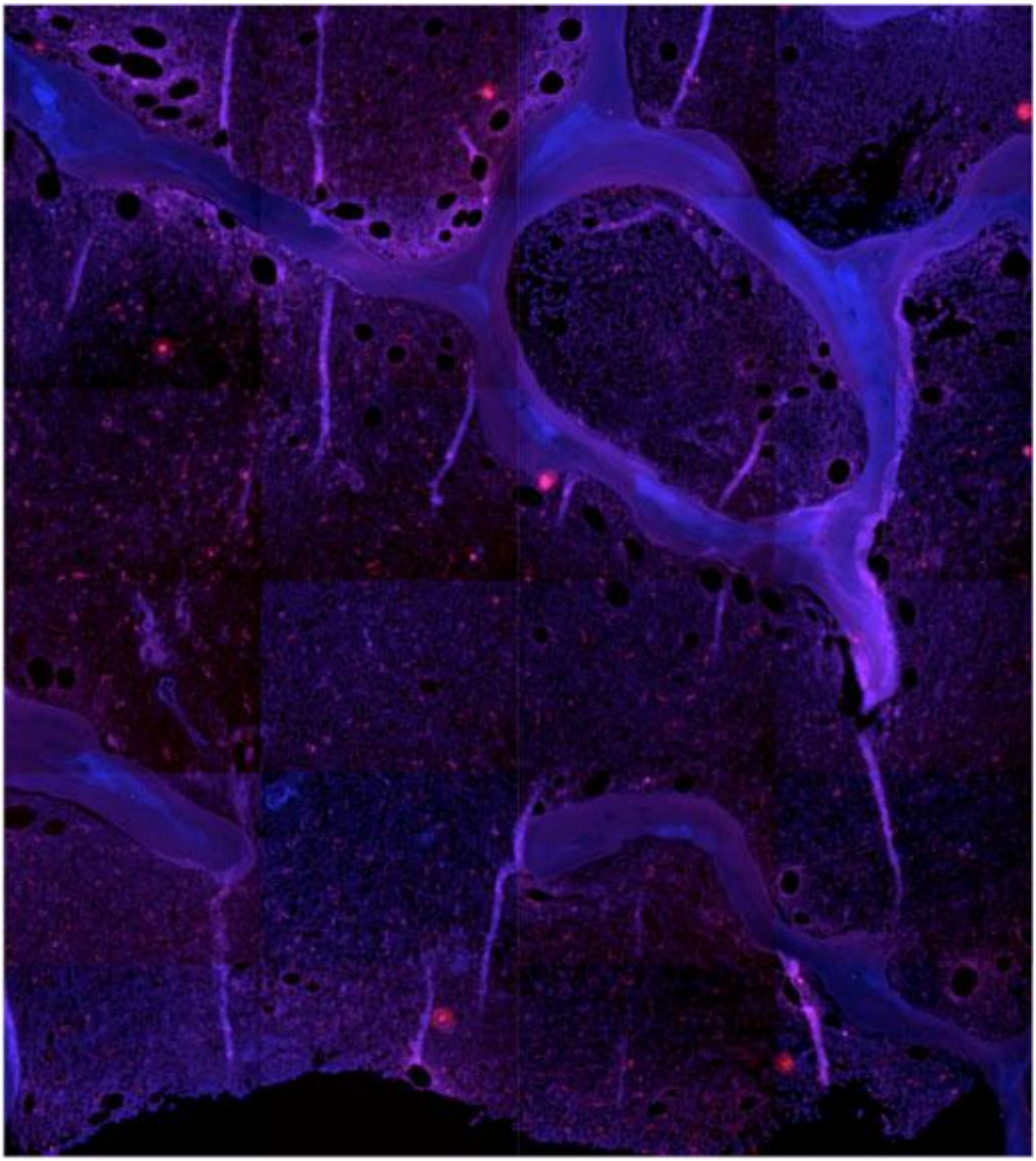


Figure 3.10 A scan of β-actin mRNA in situ detection. β-actin RCPs are shown in red. In the scan, were acquired 24 fields on x and y axes at X20. Cell nuclei are shown in blue.

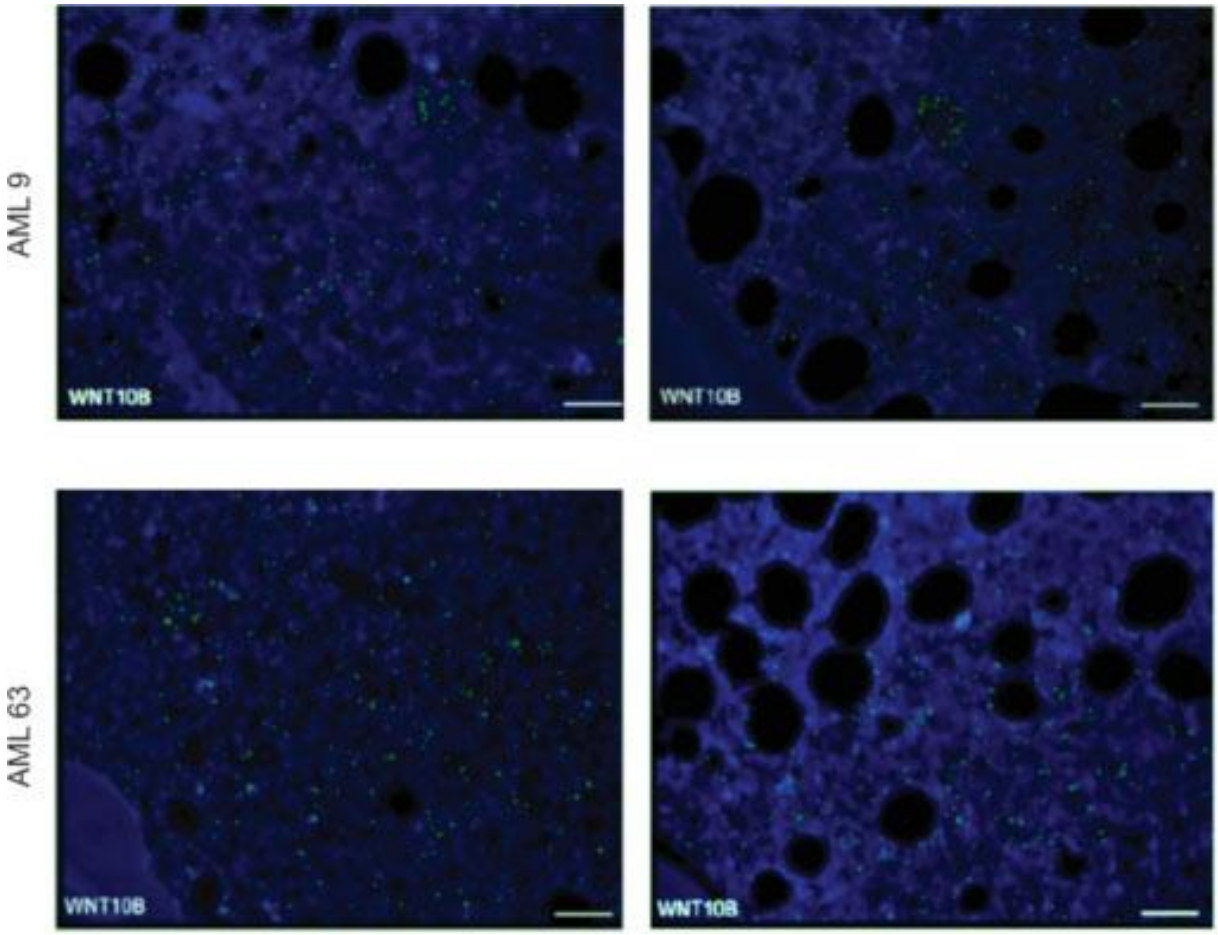
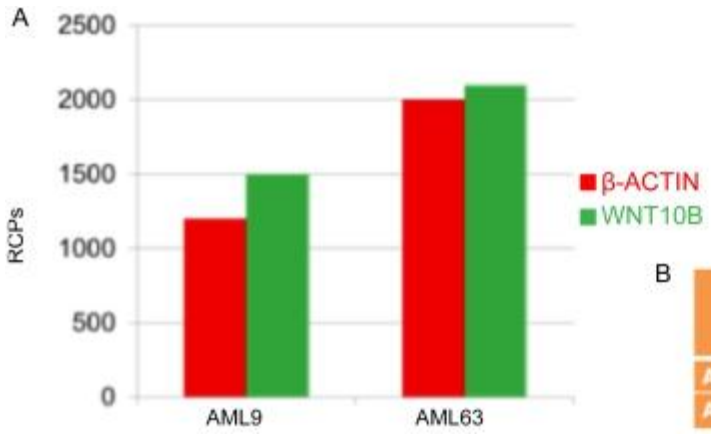


Figure 3.11 WNT10B mRNA in situ detection. WNT10B RCPs are shown in green (Cy5). In the figure are shown mRNA in situ detection performed on two AML patents biopsies, AML9 and AML63. The fields were acquired at X40. Cell nuclei are shown in

blue. Scale bar 20 µm.

If we compare the mRNA in situ detection performed on HeLa cells and on bone marrow biopsies, we can note that there are differences about number and distribution of RCP signals. Infact, working with FPPE sections, we must consider that: first of all, the permeabilization step is a crucial step, becouse allow the contact between target and reagents; second of all, we must consider the thickness of the bone marrow section (5-7 µm). Considering the thickness of the section, we can found RCPs distributed in five focal planes, and then we can't detect every signal in the image. In order to solve this problem, we used CellProfiler software as the automated system of image analysis.



B

	n cells	RCPs β Actin	RCPs WNT10B
AML 63	10300	2000	2100
AML 9	7500	1200	1500

Figure 3.12 Quantification of RCPs. The CellProfiler quantification of β-actin RCPs (red) and WNT10B RCPs (green) for five slide fields. In the panel A, is shown the ratio WNT10B/ β actin RCPs, considering the CellProfiler data output (B).

Finding cells, finding molecules

In the signal segmentation project, the collected image stacks of relatively flat cells contained point like signals in focus in different layers. However, the MIP is somewhat sensitive to noise, as it will act as maximum filter for each 1 X 1 X 16 pixel array. To reduce the amount of non-point-like signal before the MIP, the background was removed in each focus layer using top hat filtering. The top-hat filtering may be described by a rolling ball analogy. Draw a squiggly line on a piece of paper, and let an imaginary ball (a structuring element) with a certain radius roll along the curve. Every point that the ball manages to touch, being small enough to fit in to large holes and grooves, is set to zero. Points that the ball does not touch, due to being too large, are given values according to the distance between the ball and the line. In the top-hat transform the imaginary ball has been replaced with a structuring element with a flat top, like a hat. Performing the MIP of the filtered layers results in a single layer image with high contrast between signals and background. The flattened images contain the fluorescent signals from the fluorochromes and some noise. A Distance Transform (DT) of a binary image results in a gray level image where the value of each object pixel denotes the distance from that object pixel to the background. For a 2D or 3D binary image this can be computed by two passes through the image with a filter mask (typically of size 3X3, or 3X3X3). To separate cell nuclei from image background Otsu's method of thresholding, which minimizes the variance of the foreground and background, is used. The nuclei are often clustered and

have to be separated in order to identify individual cells. A distance transform is applied to the binary image obtained from the previous step. This will create a landscape like image where the intensity represents the height in the landscape. Once the nuclei are separated an area defining the cytoplasm belonging to each cell has to be delineated. Since it often is the case that no cytoplasmic stain is present, cytoplasm delineation is purely based on the distance from the nuclei. A distance transform is performed on the background of the image of the nuclei. A user-defined threshold, representing the maximum distance from cell nucleus to cell border, is thereafter applied, defining the outer border of each cytoplasm. A watershed algorithm is then used to label and separate touching cytoplasms. Nuclei touching the border will be flagged in the results, this way the user can chose to omit these cells in the final results (Figure 3.13).

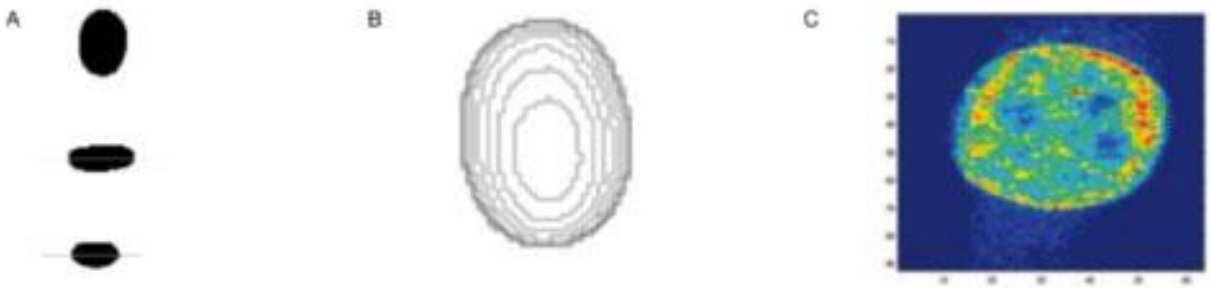


Figure 3.13 The cell nuclei segmentation process. The distance shells for signal concentration measurements. The segmented nucleus in xy, yz and xz views respectively (A). The negative distance shells overlaid on the nucleus (B). The segmented nucleus with pseudo-colors after the segmentation process (C).

When each cell nucleus and cytoplasm is labeled correspondingly we detected and assigned each signal to a particular cell, and thereafter projected into a 2D image by a maximum intensity projection (MIP). In the pre-processing step we set up a filter with a variable sized kernel, defined by us, enhancing local maxima in the image. There are two measurements of signals made; a signal count and an intensity measure. The signal count measures all detected signals as while the intensity measure looks at a 5X5 neighborhood around the center of a detected signal and measures the total intensity in that neighborhood. Working with the formalin-fixed paraffin-embedded (FFPE) tissue sections, we must solve a lot of problems about the background fluorescence, the RCP's counting, and at least the correct segmentation between nuclei and cytoplasm, because the bone marrow section is characterized by a high number of closer blast. Beside, another problem is the thickness of the sample, that determine questions about the z-axes analysis and the Point Spread Function (PSF) definition of RCPs signals. Large scale intensity variations and shading effects in the image caused by uneven illumination and other variations over the field of view are undesirable, as they may introduce problems for the segmentation step as well as position dependence for features such as integrated pixel intensity. Given that the uneven background is due to uneven illumination and the intensity of the fluorescence in each cell is a linear function of the light illuminating the cell, the usually accepted illumination correction is: *(Iraw –Idark)/(Iblank–Idark),* where *Idark* is an image without the illumination and Iblank is an image from a blank part of the slide with illumination. Whether it is feasible to acquire good images of the shading situation directly in conjunction with the imaging procedure, or not, is very much dependent on the imaging environment. Considering these image's feature, we

decreased the background fluorescence, through the illumination correction. A method inspired by the watershed algorithm was used for initial separation of clusters of cells and segmentation of the image into cells and background in one step. Color, as humans perceive it, is the response from the three different spectral sensitivity ranges of the cones in the eye. Therefore most imaging devices are designed to optimize, red, green and blue colors (RGB). In order to localize the point-like signals in relation to the nuclei membranes, a nucleus is considered as a set of shells based on the Eucliedian distance transform. The shells have negative values inside the nuclei, zero at the border and positive values outside. In order to identify the RCPs' signals, using x, y and z axes, we considered the point-like signals using a property of Fourier transform, i.e., computing the coefficients as dot products between the signal and sines and cosines. The idea is based on a method called Stable Wave Detector (SWD) *(Dupaˇc and Hlav.ˇc, 2006)* that is used for landmark detection on 2D images.

We considered the image as a sequence of frames that should overlap more than T/2, the signal width. Sine and cosine Fourier coefficients, b_i and a_i, of the first harmonic wave of the Fourier series are computed using

$$a_i = F_i C$$

$$b_i = F_i S$$

$$C(t) = \cos(2\pi(t/T))$$

$$S(t) = \sin(2\pi(t/T))$$

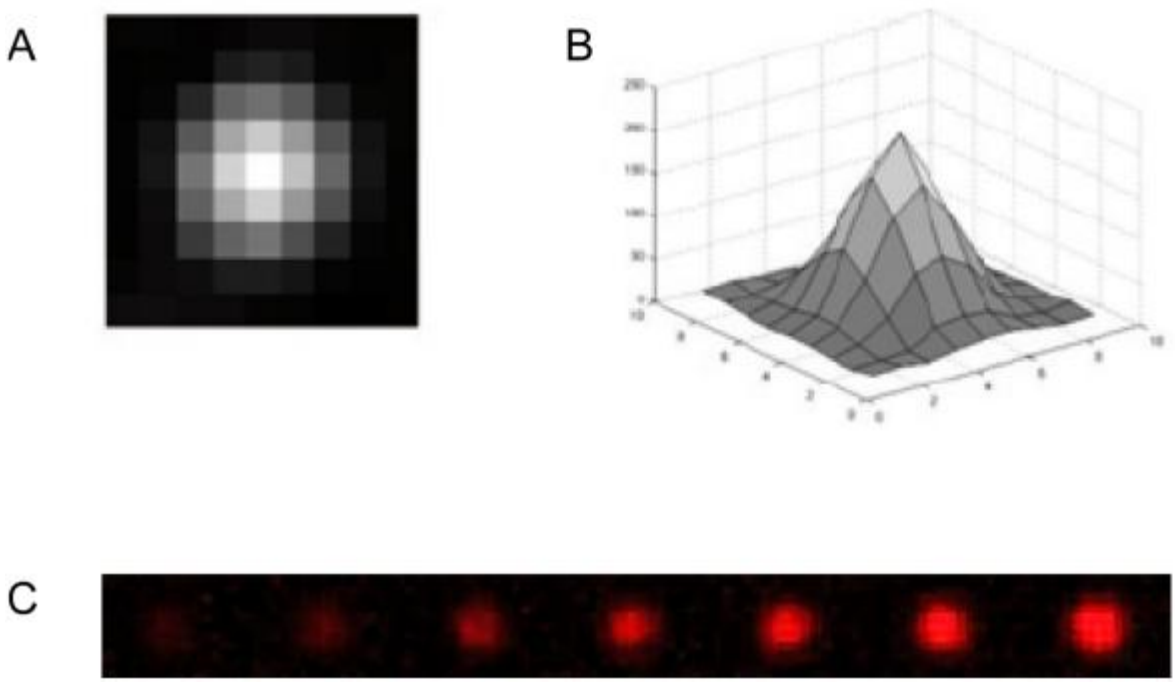


Figure 3.14 Point-like signal detection. A. The mid slice of a simulated point like signal, considering the point spread function of a RCP signal. B. The signal as a surface plot. C. Definition of a single point signal of the RCP, after the image processing.

In order for a pixel to be classified as a true signal, the image is first convolved with the cosine filter C of period T and a threshold is used to identify the potential signals (Figure 3.14). Then, after this pre-processing of the images, we can detect the single RCP signal, in a bone marrow biopsy through CellProfiler pipeline. We applied the Human Cell CellProfiler pipeline, using .tif images acquired at fluorescence microscope and saved in

a folder, as Imput Images. After the images loading, we defined nuclei as Primary Objects and RCPs as Secondary Objects, defined by Otsu's and Waterhshed thresholding methods. The definition of blobs and nuclei, followed the illumination and background correction, is the last step of the image analysis (Figure 3.15).

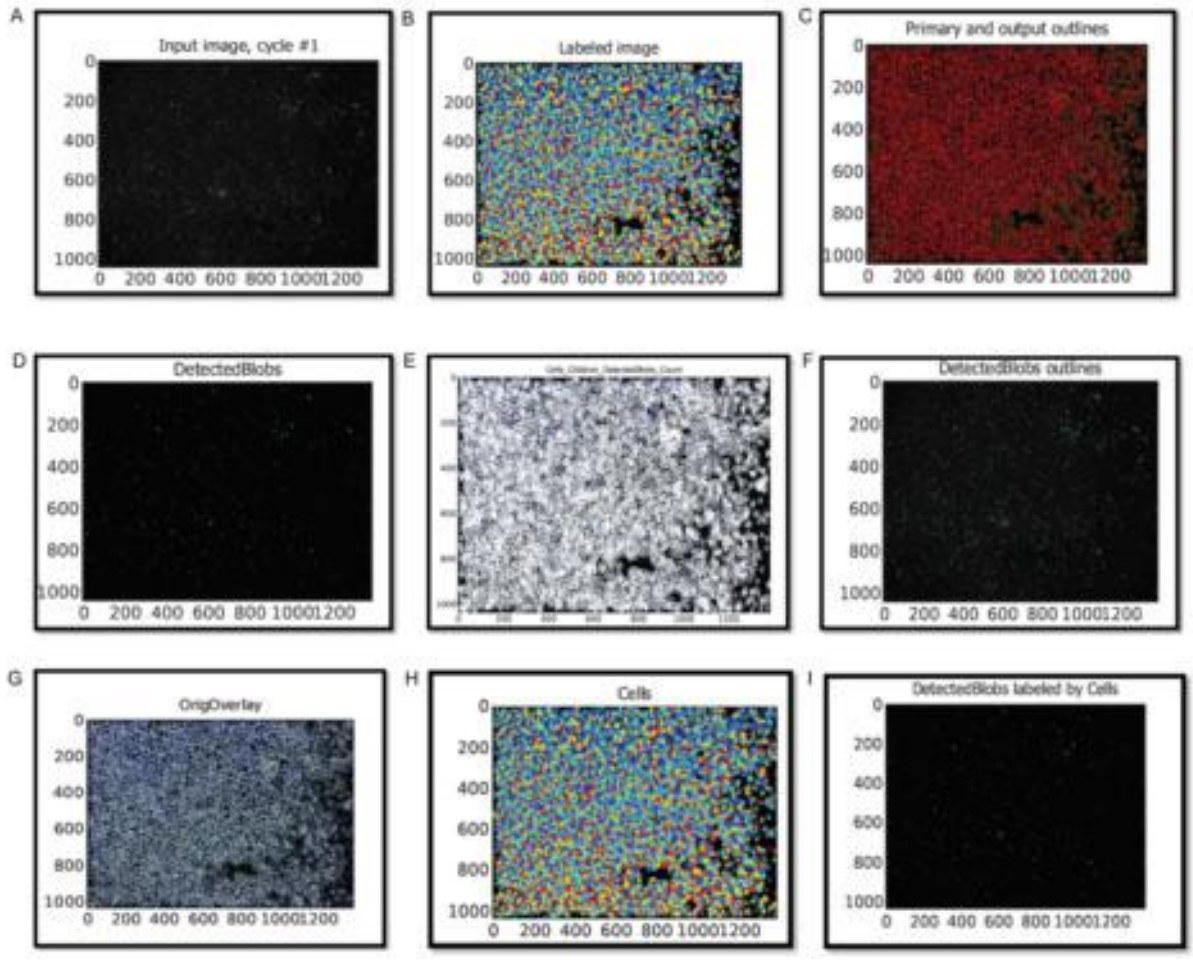


Figure 3.15 CellProfiler analysis. The.tif images were loaded as input image (A), and then the nuclei were labeled (B). In the panel C and D, are shown the Primary and Secondary objects. The overlay between Primary and Secondary Objects (E), is followed by background correction (F, G). The results are shown in panel H and I.

In order to better elucidate the impact of the diffuse WNT10B overexpression on the

leukemic microenvironment, we examined its expression in histological preparations of bone marrow from five randomly selected AML patients at diagnosis. The double immunostaining for WNT10B and ABC, evidenced that WNT10B is expressed by a high proportion of leukemic cells, and its presence correlates with cytoplasmic accumulation of active-β-catenin in a proportion of eight percent of cells. WNT10B antibody staining was also detectable in interstitial spaces, suggesting the secretion and release in the bone marrow microenvironment. Besides, the slides revealed that WNT10B is diffusely expressed but that only a few small cells (8-10 μm diameter of the nuclei), with a clonal appearance and increased N:C ratio, shared the Wnt signaling activation signature represented by cytoplasmic accumulation of the active form of β-catenin, likely induced through an autocrine/paracrine mechanism (Figure 3.16).

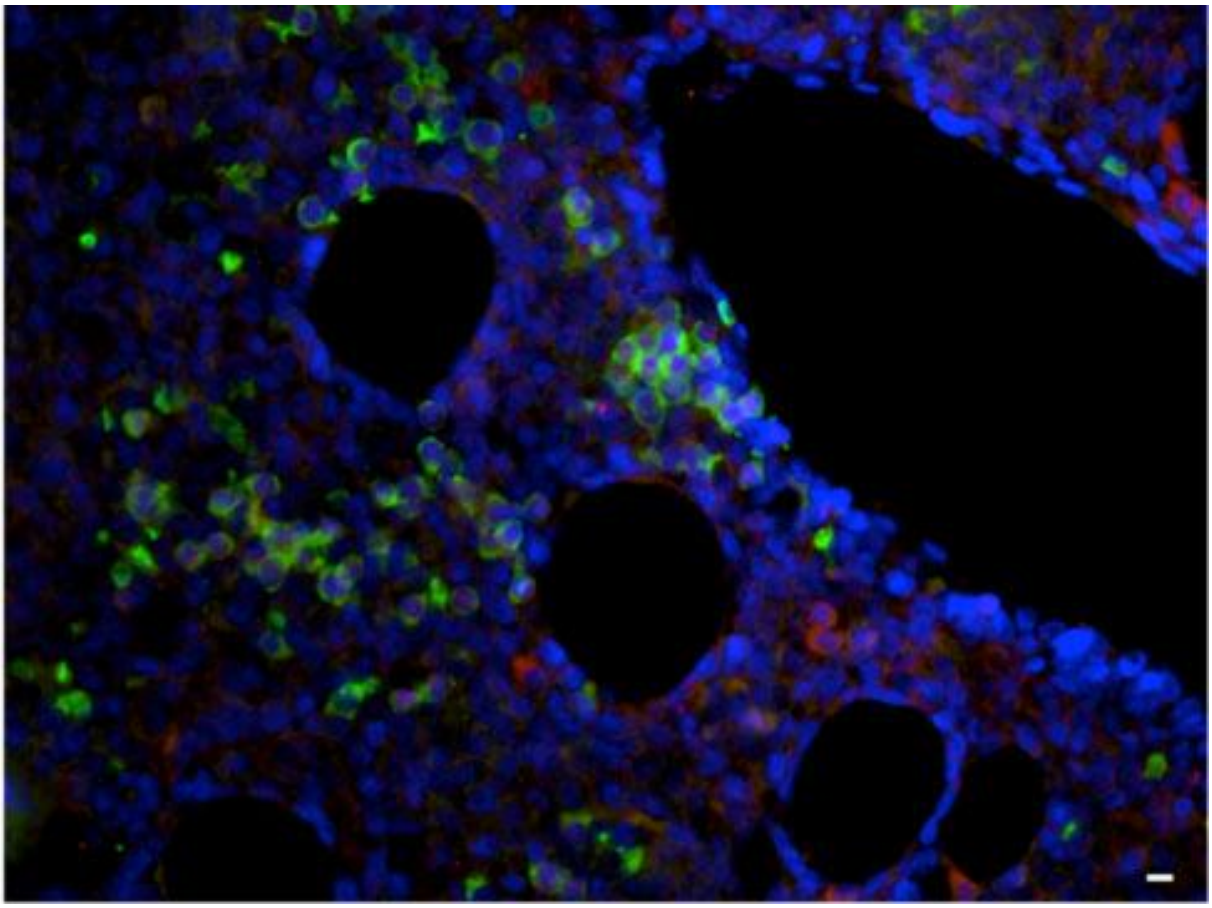


Figure 3.16 β-catenins activation in AML cells expressing WNT10B. Representative bone sections of an AML patient (AML9) were costained for expression of (A) active β-catenin (green), and (B) WNT10B (red). The image was acquired at X40 objective. Scale bar 10 μm.

3.4 AC133 IS EXPRESSED ON A46 AML CELLS SECRETING WNT10B

According to the idea that primary cells cultures closely mimic the *in-vivo* state and generate relevant data, we established a primary AC133+ cell culture (A46). A46 cells, with diploid karyotype, were selected from a 66 years old male at diagnosis of AML-M2. The immunophenotype of unselected cells revealed a dominant CD133.1+CD34+CD38-CD45+CD117+ blast population (59%), representing an optimal source to establish a LICs-enriched primary culture. In order to investigate whether the endogenous WNTs production had any paracrine effect, A46-CM was used to evaluate β-catenin-mediated transcriptional activation. To this aim, HEK293T cells (H293T), transfected with Super8XTOPFlash β-catenin/TCF transcription-based reporter construct, were exposed either to pBA-Wnt10b H293T-Collected Medium as control or A46-Collected Medium. This construct could be efficiently expressed in a dose-dependent manner when transiently exposed to A46-CM for 12h (Figure 3.17).

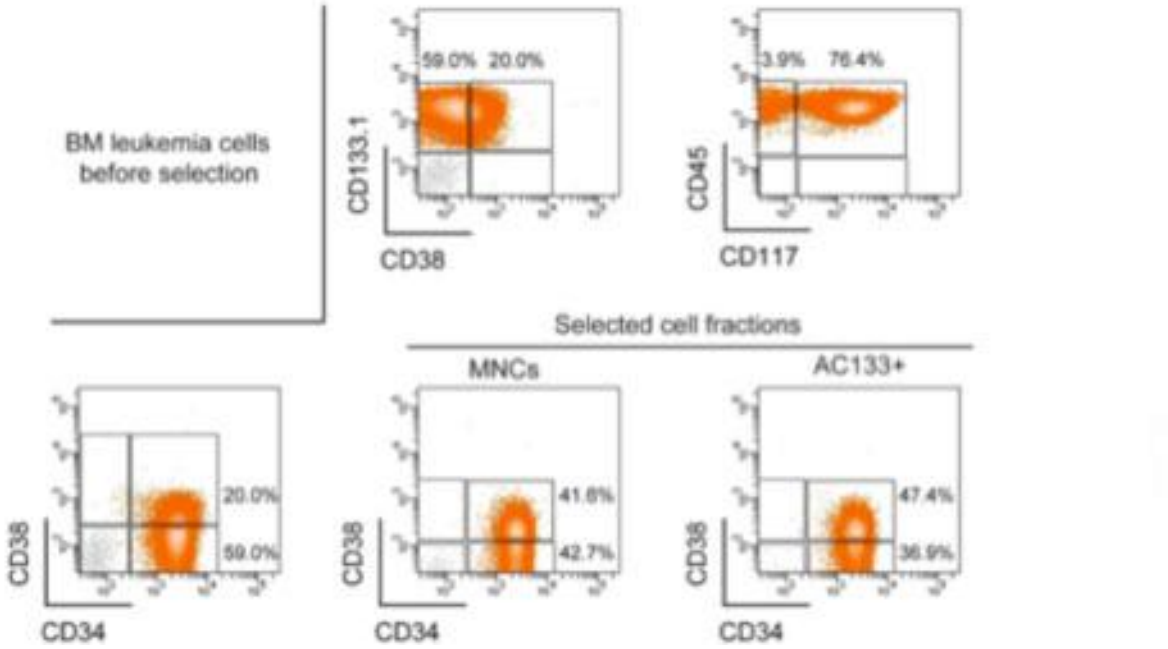


Figure 3.17 Immunoohenotyoe analysis from AML46 BM MNCs at the diagnosis an after selection. (A) Patterns of CD38/CD133.1, CD117/CD45, CD34/CD38 co-staining was gated on BM AML cells before selection. Representative CD34 and CD38 expression on Ficoll™-selected MNCs (below center) and AC133 sorted cells prior to culture is shown in the central part of figure. Percentages on total cellularity are shown for gated AML populations.

In order to clarify the relationship between the abnormal Wnt activation in AC133+ population and the LIC activity we translplanted A46 cells into sublethally irradiated (5 Gy) six weeks old Rag2$^{-/-}\gamma c^{-/-}$ mice via the tail vein. We set up the experiment using Rag2$^{-/-}\gamma c^{-/-}$ mouse, becouse this immunodeficient mouse strain lacks mature B, T and NK cells, supporting efficient engraftment of human AML.

SAMPLE	WEEKS	CD45+	CD133+/ CD34+	CD133+/ CD34-	CD133+/ CD38+	CD133+/ CD38-	Lin-/ Sca1+/Kit+
C#1	3	0.32	0	0	0	0	3.29
C#2	3	0.93	0	0	0	0	4.65
M#1	3	32.17	21.37	67.52	82.67	16.00	1.16
M#2	3	34.02	20.75	70.19	89.09	10.91	0.97
M#3	3	6.96	34.34	59.04	70.18	29.82	0.77
C#3	8	0.16	0	0	0	0	1.20
C#4	8	0.39	0	0	0	0	0.98
C#5	8	0.16	0	0	0	0	0.60
M#4	8	4.39	1.39	0.28	40.00	16.00	0.38
M#5	8	1.92	1.56	1.84	72.73	27.27	0.60
M#6	8	1.60	1.46	1.46	50.00	50.00	0.38

Table 3.1. Immunophenotype analysis of transplanted Rag2$^{-/-}\gamma c^{-/-}$ mice.

Transplanted mice were killed at 3 to 8 weeks after transplantation and analyzed for engraftment of human leukemia cells in BM. The AC133^{+} A46 cells showed engraftment of human CD45 (hCD45)$^{+}$ cells. Then, we evidenced that engrafted hCD45^{+} cells were human myeloid leukemia blasts by measuring CD34/CD38/CD133 expression.

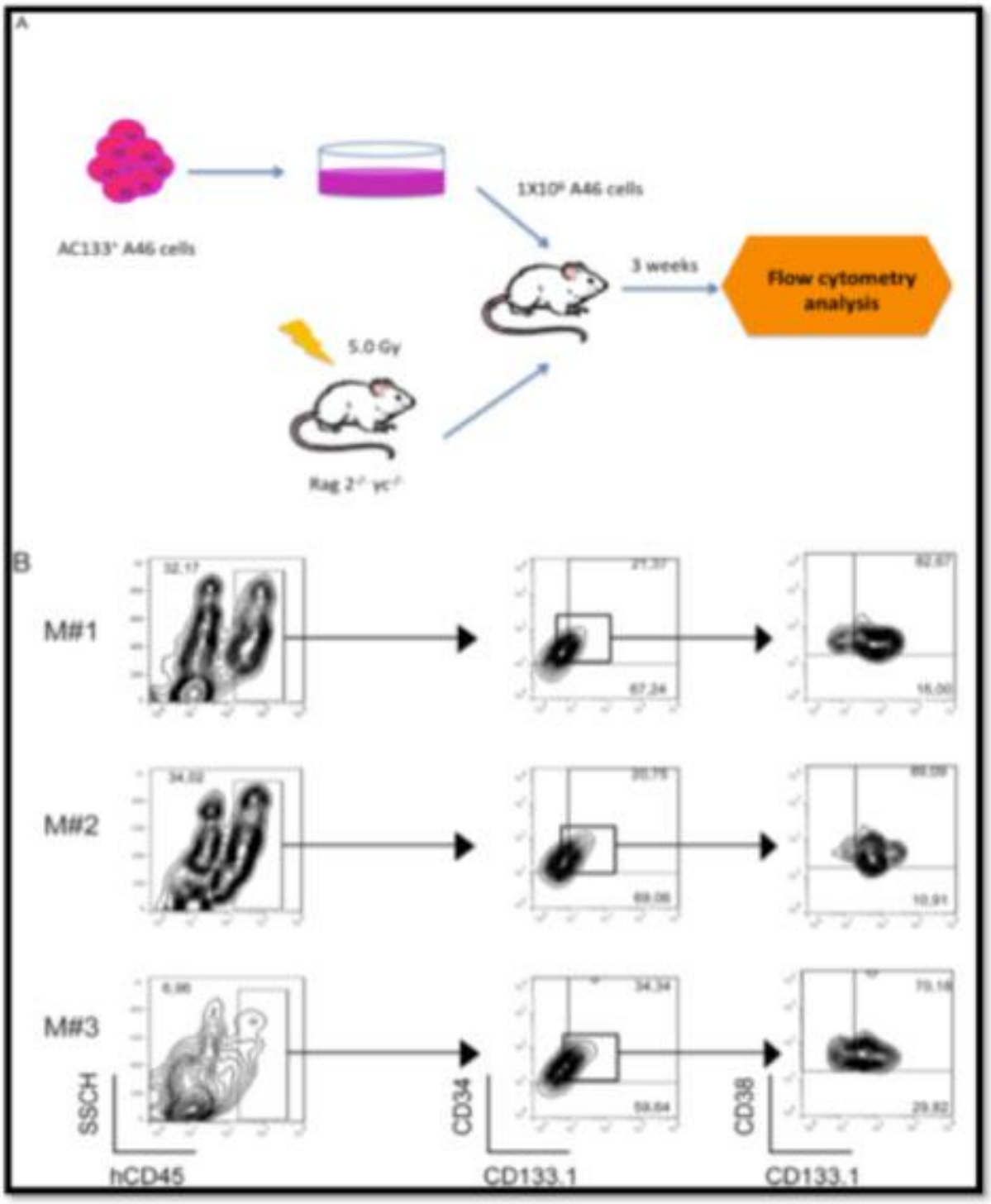


Figure 3.18 AC133+ A46 cells transplantation in Rag2$^{-/-}$γc$^{-/-}$ mice. (A) Represetation of overview of the experimental design: 1*10^6 AC133+ A46 cells were injected into sublethally irradiated Rag2$^{-/-}$γc$^{-/-}$ mice through the tail vein. Three weeks after transplantation BM cells were collected and analyzed by flow cytometry. (B) Expression profiles of three recipient mice (M#1-3) are shown.

3.5 THE TRANSPLANTATION OF A46 CELLS INDUCED ECTOPIC AXIAL STRUCTURES FORMATION IN ZEBRAFISH EMBRYO BY WNT SIGNALING ACTIVATION

In order to define the physiological relevance of WNT factors expression and release and the tumor-linked signals, we used the developing zebrafish model as a biosensor. Wnt pathway is among the most evolutionarily conserved signalings, and Wnt ligands have been strongly implicated in body axis formation in *Xenopus* and *Danio Rerio.* In this models Wnt ligands represents the maternal dorsal determinants necessary to induce the nuclear translocation of β-catenin in the prospective dorsal region of the blastula. Additionally, the Wnt signaling activation in zygotic state of zebrafish development, is required for proper tail development, and its overexpression induces ectopic tail buds. Considering this evidences, we transplanted A46 cells fluorescently labeled with Hoechst 33342, into developing zebrafish embryos (Figure 3.19). We observed that embryos grafted at or before the mid blastula transition stage (~3 hours post fertilization, hpf) A46 cells (Figure A) developed secondary axial structures, ranging from additional tail tissues (Figure E-H) to an almost fully-formed ectopic head (Figure I). Control embryos grafted with normal bone marrow-derived AC133+ cells, did not show any alterations of the normal phenotype. In order to identify the additional tail structures we performed *in-situ* hybridization staining for the notochord and tail bud marker *ntl*. We can observe the emergence of ectopic head structures through the *pax2a*

lstaining that abeling optic stalk, midbrain-hindbrain boundary, and otic vesicles. These results imply that A46 cells might determine the establishment of an additional source of signal, with a Nieuwkoop center-like activity, inducing an extra dorsal organizer. We can underline that this results are similar to the endogenous situation at mid-blastula transition, when maternal Wnt/β-catenin signaling initiates the formation of the dorsal organizer. In consideration of this evidence, we analyzed, in cell-grafted embryos, the expression of the organizer-specific gene *goosecoid (gsc)*, strictly linked with the nuclear translocation of maternal b-catenin triggered by the activation of the Wnt signaling. We can observe that normal AC133+ cells did not alter the normal expression of the gene (Fig B), approximately 30% of the embryos grafted with A46 cells (*n*=208) displayed both the expansion of the *gsc* endogenous domain and the activation of the gene in ectopic positions (Fig. C, D). Besides we can note in the figure that the endogenous (n) and ectopic (*n) *ntl* signals run parallel along the axis of the embryo, indicating the presence of additional axial structures. We also underlined in 24hpf embryo that developed an ectopic head the expression of the brain- marker gene *pax2a*. The optic stalk (OS) in close vicinity to the eye (e), the midbrain-hindbrain boundary (MHB), and the otic vesicles (OV) of the embryo are stained with the *pax2a* riboprobe, as well as several areas of the ectopic head. This data, underlines that A46 cells retain a dorsal-organizer inducing activity possibly correlated with their strong Wnt signaling activation. In our work we also investigated the possible involvement of the Nodal signaling pathway in the dorsal-organizer induction mediated by A46 cells, considering that the secondary axis can be also induced by Nodal, a highly evolutionarily conserved morphogen able to activate *gsc* expression. The microarray data and qualitative RT-PCR in A46 and samples from 13 different AML patients exluded any *Nodal* contribution s to

the formation of the secondary axis and induction of *gsc* expression.

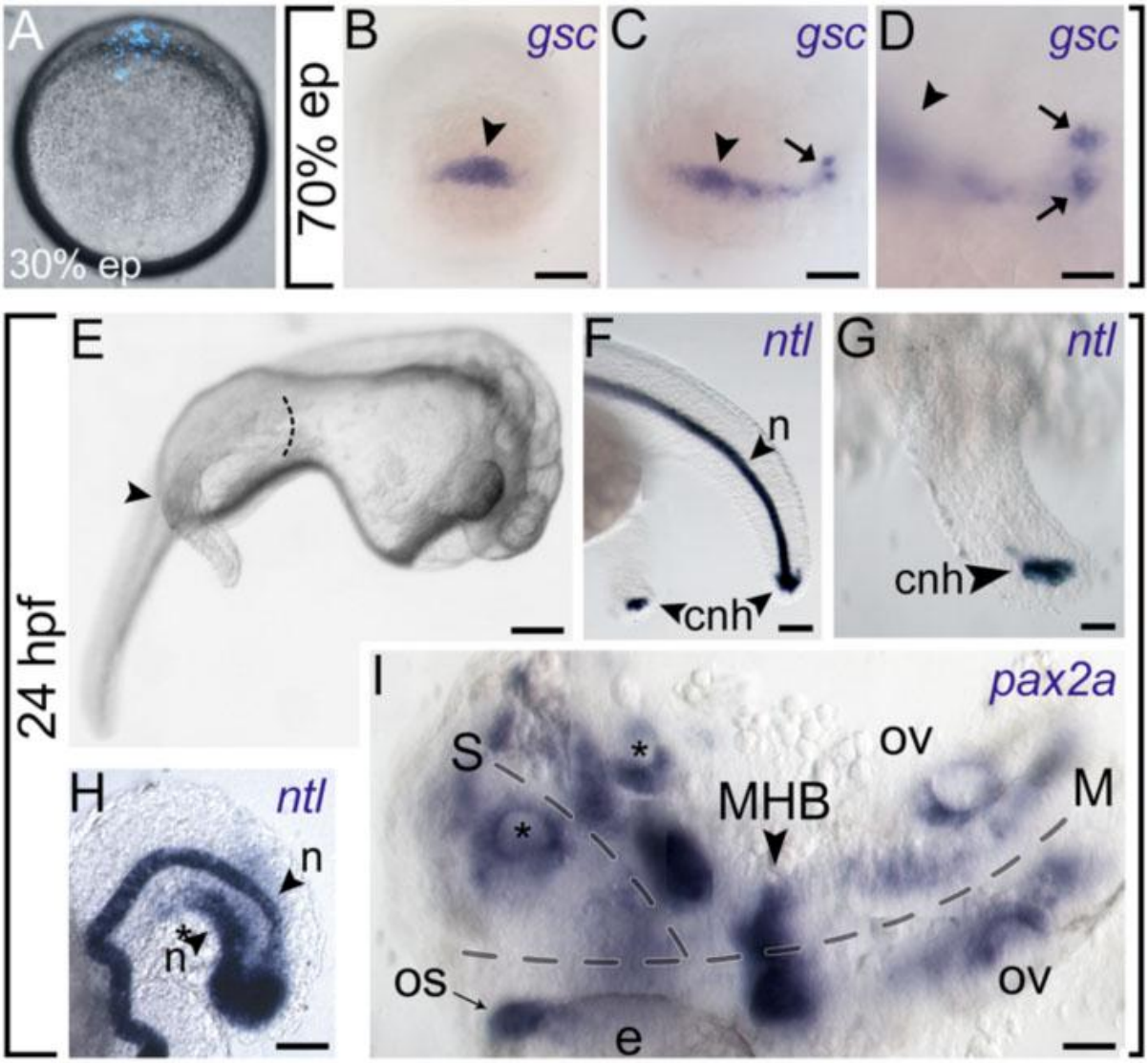


Figure 3.19 A46 AML cells induce ectopic gene expression and secondary body axis formation upon transplantation in zebrafish embryos. (A) Fluorescence microscopy of a live zebrafish embryo at 30% of epiboly (from lateral view) transplanted at 3hpf at

the animal pole region with A46 cells previously blue-stained with Hoechst 33342. (B-D) Dorsal side view of 70% epiboly-stage embryos hybridized with a *gsc*-specific probe. The arrowheads indicate the *gsc*-endogenous signal, and arrows specify the position of zebrafish cells expressing ectopic *gsc*. (E) Brightfield microscopy of a 24hpf zebrafish embryo injected with A46 AML cells (lateral view). The arrowhead and the dotted line indicate the secondary trunk/tail induced by A46 cells. (F,G) The embryo in (E) has been hybridized with a probe specific for the notochord and tail bud marker *ntl*. (F) The probe labels the notochord (n) in the endogenous trunk and the chordoneural hinge (cnh) in both tails (G, higher magnification). (H) Tail of a 24hpf embryo hybridized with *ntl*-specific probe. (I) Dorsal view of a 24hpf embryo that developed an ectopic head on the side of the endogenous one, as indicated by the expression of the brain- marker gene *pax2a*. The dotted lines indicate the main (M) and secondary (S) axes. The image is composed by different pictures corresponding to several focal planes, since the embryo is not flat, and a single focal plane cannot comprise all the labeled structures belonging to the main and secondary axes. Scale bars represent 125 μm (A-D), 150 μm (E), 40 μm (F), 15 μm (G and I), or 25 μm (H).

3.6 THE AC133^{BRIGHT} POPULATION SHARED THE WNT SIGNALING ACTIVATION SIGNATURE

Thirteen years ago, a novel cholesterol (Ch)-interacting, pentaspan membrane glycoprotein, named prominin-1 (CD133) has been identified as a surface marker of both neural *(Weigmann et al, 1997)* and haematopoietic *(Yin et al, 1997)* stem and progenitor cells. Since numerous somatic stem cells and cancer stem cells originating from different organ systems have been demonstrated to express it, prominin-1 might become a biological tool for stem cell. In our work, we noted that the signature of WNT signaling activation, is defined by the cytoplasmatic accumulation of the active form of β-catenin in AML cells. Thus, in order to characterize the activation of the WNT regenerative pathway in AC133+ cellular fractions, we investigated how the expression of WNT10B is related to AML phenotype, through in situ approaches. In order to look and observe the AML context, we performed direct immunostaining in histological preparations of bone marrow from AML patients at diagnosis. Double immunostaining for WNT10B and active β-catenin (ABC), confirmed that WNT10B was expressed by a high proportion of leukemic cells *(Beghini et al., 2012).* We noted that the "responsive" phenotype, defined by expression of the dephosphorylated β-catenin was restricted to only restricted to a rare population of cells, in an estimated proportion of eight percent of cells. Then we performed a direct AC133 immunolabeling on a closer serial section of AML bone marrow biopsy, considering WNT10B/ABC staining, in order to compare the different cell subpopulation distribution *(Beghini et al., 2012).*

In order to clarify the relationship between the responsive phenotype and the AC133 positive cells, we performed series of immunofluorescences followed by image analysis. The first step of our work, was focused on the resolution of the hematopoietic tissue autofluorescence. The formalin-fixed paraffin-embedded (FFPE), show a lot of technical problems as autofluorescent properties of specific tissue constituents, for example lipofuscin granules, collagen fibers and porfirins with a broad-band blue excitation *(Del Castillo et al. 1989; Noonberg et al. 1992; Edwin & Jackman 1981; Verbunt et al. 1992; Belichenko et al. 1996; Banerjee et al. 1999).* On the other hand, autofluorescence, either intrinsic or induced by fixation media and tissue processing may either mask specific fluorescent signals or be mistaken for fluorescent labels *(Del Castillo et al. 1989; Noonberg et al. 1992)*. Besides, using a long-pass green emission filter on FFPE sections, the fluorescence results dull and not bright. When we performed immunofluorescences we also must considered that aldehyde fixatives react with amines and proteins to generate fluorescent products. These tissue-bound free aldehyde groups will combine covalently with any amino group offered to them, including terminal and side-chain (lysine) amino groups of antibodies. Besides, small organic fluorophores are powerful research tools in biological imaging that have enabled unprecedented insights into both cellular and molecular processes. However, their performance can be compromised by undesirable photophysical properties that limit both the fluorescence quantum yield and the total number of photons emitted before photobleaching. Such issues include both transient (blinking) and irreversible (photobleaching) light-induced transitions to dark states. Dark state transitions are particularly limiting in single-molecule studies that demand high illumination intensities. Therefore we performed our in situ studies, by the direct immunolabeling, in order to decrease the autofluorescence, avoiding the aspecific

binding of secondary antibodies. First of all, we set up a series of experiments, in order to define the background of bone marrow sections. We evaluated the fluorescence background after the antigen retrival, followed by the excitation at 460 nm, 488 nm and 647 nm (see Figure 3.1 A, B,C).

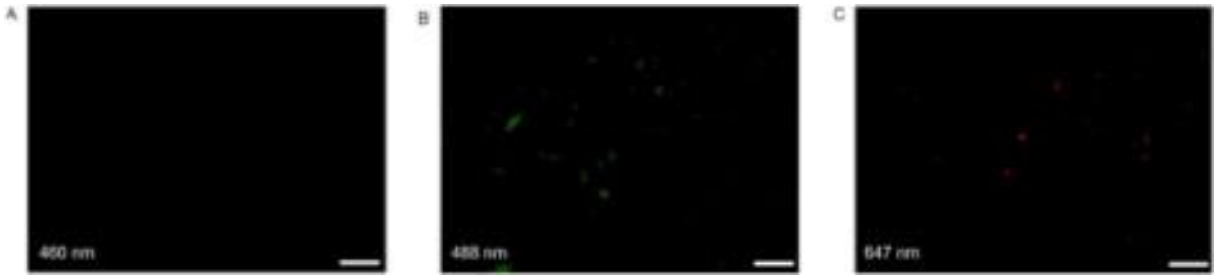
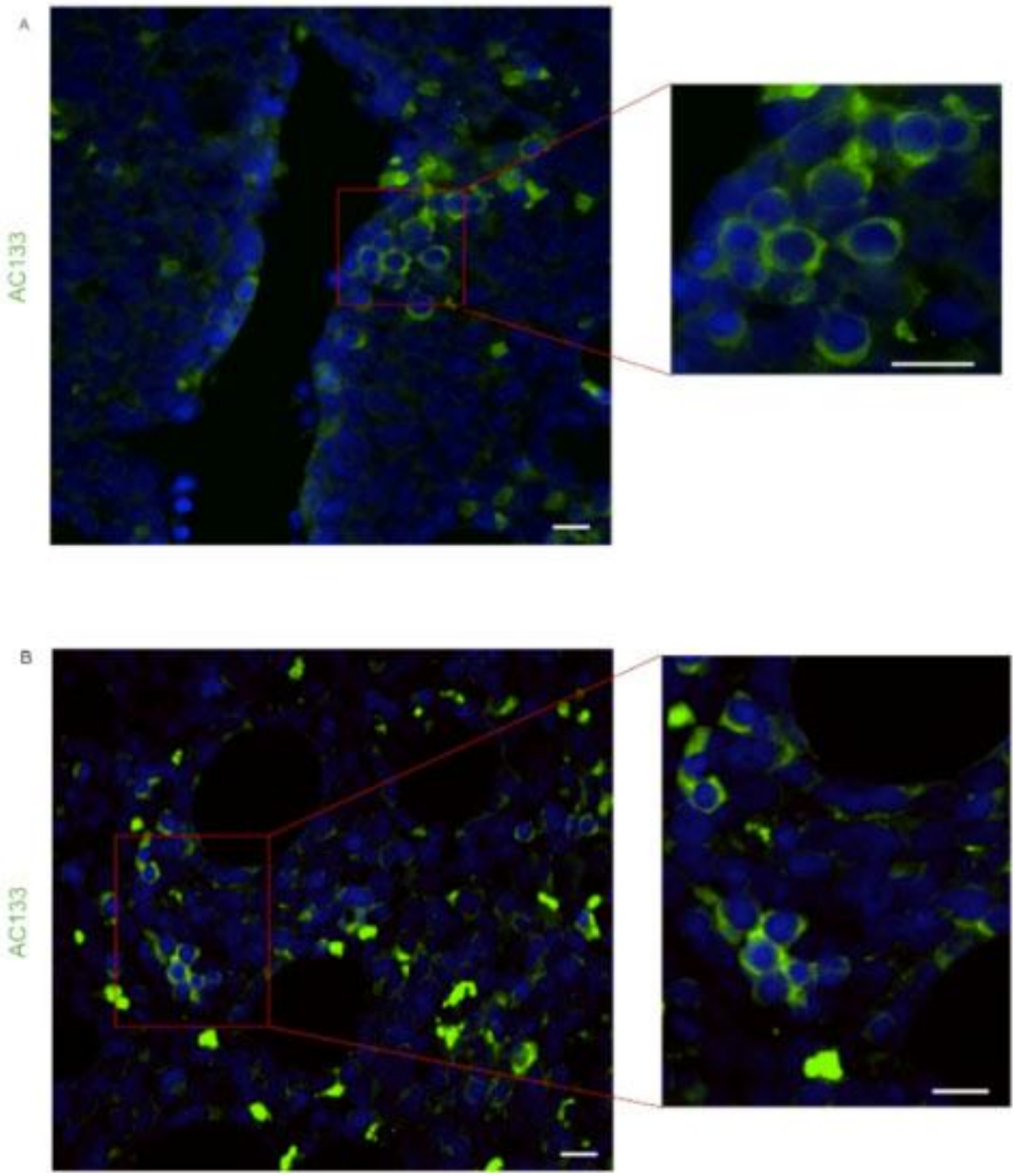


Figure 3.20 Evaluation of fluorescence background of bone marrow section. Following the antigen retrival, evaluation the fluorescence properties of bone marrow section after the excitation at 460 nm (A), 488 nm (B) and 647 (C). Scale bar 10 μm.

In the figure 3.20. we can observe how there are autofluorescent spots, maybe due to granular cells. Besides, at 488 nm emission (see Figure 3.20 B), we can observe, an higher background then the 647 nm emission, own because using a long-pass green emission filter on FFPE sections the hematopoietic tissue has a fluorescent background. The background visualization is useful for the subsequent image analysis, in order to specify the signals and define the parameters for the nuclei segmentation. Then, we examined by immunostaining a number of bone marrow biopsy sections from five randomly selected cases. In all the analyzed samples the AC133 direct immunostaining revealed the islands of highly AC133bright positive cells in an estimated proportion of eight/ten percent of cells, amid AC133dim or negative tumor blasts *(Beghini et al., 2012)*

(Figure 3.21 A and B). The multiple approach to image analysis through ImageJ and CellProfiler Analyst reveals that 70% of cells are AC133 positive cells, but only a 10% of cells present the bright signal positivity in the AML environment. Each cellular image generated in a high-throughput screening experiment contains a tremendous amount of information. In order to analyze the high-content screening (HCS), defined as the high information content inherently present in cell images *(Carpenter, 2009),* we performed the image analysis through ImageJ, CellProfiler, and CellProfilerAnalyst. The cell segmentation is a crucial step in a pre-processing of single cell analysis. The main goal of the segmentation is to, in the end, accurately define signals per cell rather than achieving a perfect delineation of the cytoplasm. In order to define nuclei and cytoplasmatic signals, we performed an image segmentation through two thresholding method: for the nuclei we applied a popular thresholding method based on the histogram named Otsu's *(Otsu, 1979)* where the automatic selection of an optimal threshold is performed based on the class seperability of the histogram. From the histogram, Otsu's algorithm selects a threshold that minimizes a weighted intra-class variance of the background and foreground.



3.21 Detection of AC133^{bright}cells. The AC133 direct immunostaining on two bone marrow sections derived from AML patients, AML9 (A) and AML63 (B). The green signals, obtained with hybridization antibody labeled with dye 488nm define the AC133 positive cells. In both cases, we can observe two types of AC133 postivive cells: a rare group of cells is characterized by the high bright positivity, and other cells showed a dim

cytoplasmatic signal. Cell nuclei are shown in blue. Scale bar 10 μm.

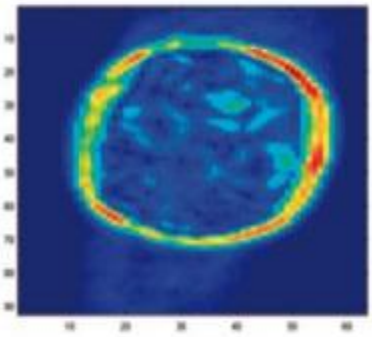


Figure 3.22 The cell nuclei segmentation process. In the figure is shown a cell nucleus MATLAB representation after the Watershed segmentation, through a magnitude gradient representation. In blue is shown the nucleus center, and the AC133 staining is represented as perimeter around the nucleus according to the pixel gradient intensity from low intensity (yellow) to high intensity (red).

In order to define the cytoplasmatic's area and signals, we applied the Watershed segmentation: the general idea is to consider the image as a landscape where the magnitudes of the gray values form the hills and valleys of the landscape. The Watershed segmentation starts by submerging the landscape in water, and water is allowed to rise from each minimum. By these thresholding methods, we divided the image into a set of meaningful regions, nuclei and cytoplasm, reducing the background signal. The digital image processing algorithms, were implemented in MathWorks™ technical computing environment Matlab® (Figure 3.22). This small group of positive cells, are characterized by the nuclei diameter of 6-8 μm and the particular double DAPI emission signal, defined by the integrated intensity of 0.4. Besides, we can note that the green signals are specific and localized around the membrane perimeter. There are rare group defined by bright

signals, amid AC133dim or negative tumor blasts (Figure 3.23).

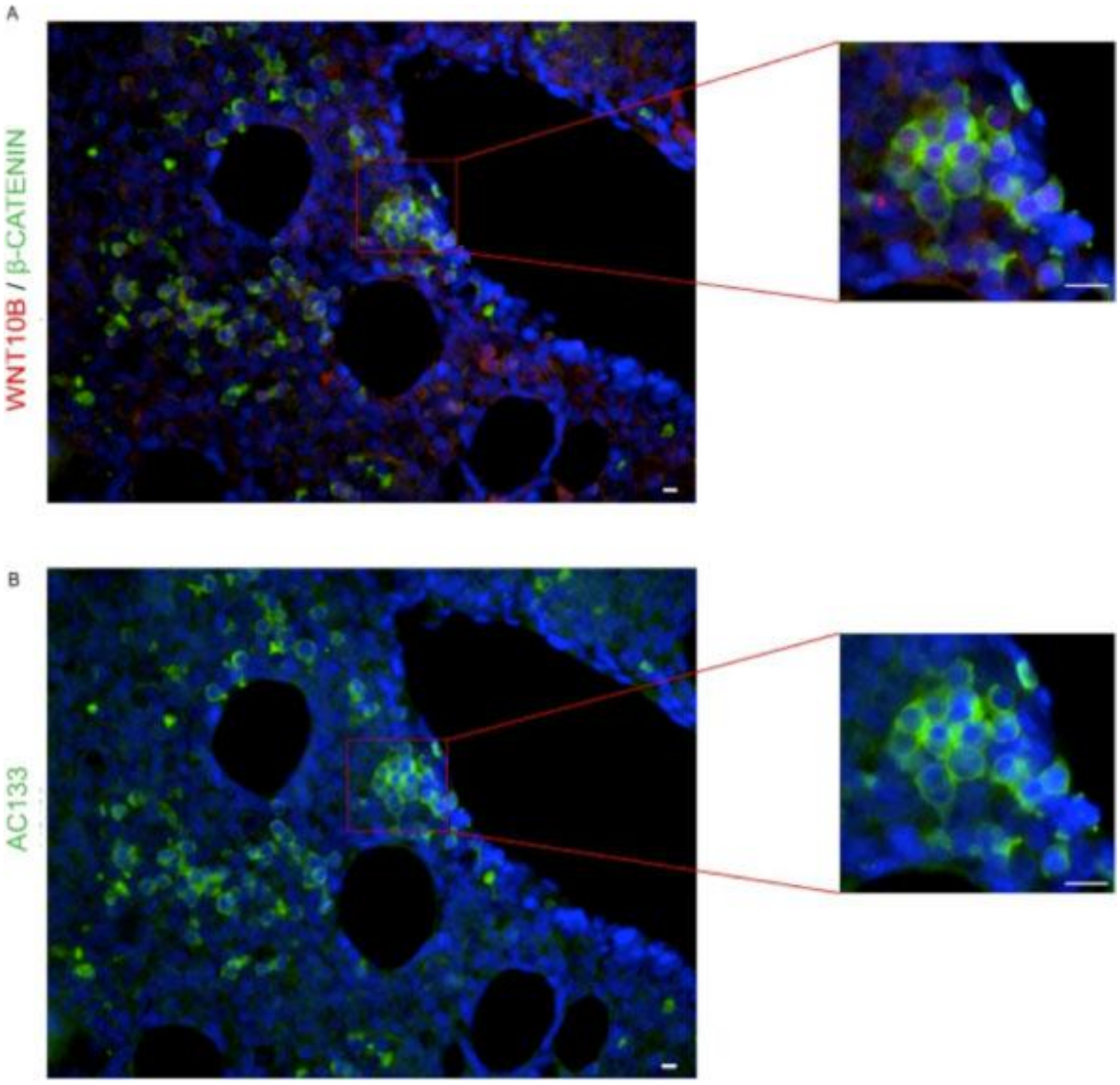


Figure 3.23 β-catenins activation in the subpopulation of AC133bright AML cells expressing WNT10B. A. The double immunostaining of bone marrow section from the AML9 for expression of active β-catenin (green), and WNT10B (red). Cell nuclei are shown in blue (DAPI). B. Direct immunolabeling in order to detect the AC133 positive cells (green). The immunostainings were performed on adjacent serial section Cell nuclei are shown in blue. Scale bar 10 μm.

We can observe that the β-catenin positive cells are also defined by the positivity for AC133bright signal. Besides, if we compare this two cell populations, we can note that nuclei dimater and DAPI intensity are the same. In order to define the spatial relationship between AC133 and ABC positive cells, in AML bone marrow cell population, we performed a double immunostaining AC133/ABC (Figure 3.24). In the Figure 3.24 we can observe the perfect correlation between AC133 and ABC signal, suggesting that the WNT signal responsiveness function is strictly associated to AC133bright cells. Using ImageJ, we noted that there are two types of signal positivity: the AC133 signals are localized around the membrane perimeter, while the ABC signals define the cytoplasmatic area.

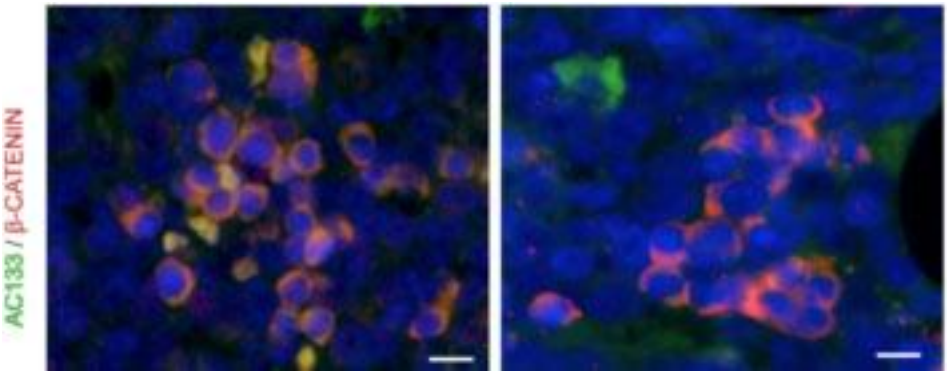


Figure 3.24 AC133^{bright} a marker of WNT signaling activation. Double immunostaining for AC133 and ABC detection, on bone marrow section derived from AML9 patient. Cell nuclei are shown in blue. Scale bar 10 μm.

In order to classify the cell populations that characterize the AML bone marrow

microenvironment, we performed a classification of AC133^{bright} and AC133/ABC cells, through CellProfiler Analyst. Image-based data is tremendously valuable in that multiple single-cell measurements are available. In many cases, a single measured feature (e.g., the total intensity of green stain or DAPI stain for the nucleus) can be used to score individual cells and the only challenge is to identify a suitable threshold for scoring positive cells. Cell image analysis allows accurate identification and measurement of cells' features, enabling automated analysis of certain phenotypes that were previously intractable. However, many interesting phenotypes require the assessment of several measured features of cells. Machine learning methods that select and combine multiple features for automated cell classification have been used to score many phenotypes. This can be accomplished in CellProfiler Analyst using plot of individual cell data *(Jones et al., 2008; Jones et al., 2009)*. For complex phenotypes, several features of each cell may be required for effective scoring. We considered, the double intensity signal of DAPI and the 6-8-µm nuclei diameter as common parameters. Then we performed and image analysis using CellProfiler Analyst, and defining two gates, called AC133 and AC133/ABC positive cells. The first step is the classification of ''interesting'' phenotype, followed by the scoring of the co-expression signals (Figure 3.25). The samples are scored with sequential gates following this approach: (1) score the entire population of cells from an experiment, defining the AC133 and AC133/ABC subpopulations (2) draw a gate around the data points representing potential cells of interest, (3) adjust the gate to include nearly all positive cells and exclude as many negative cells as possible, (4) plot the resulting gated subpopulation in a new density plot with two new measurement features as axes, (5) gate the subpopulation again based on these new features, and (6) calculate the percentage of each image's cells that fall within the final gate, with p-values.

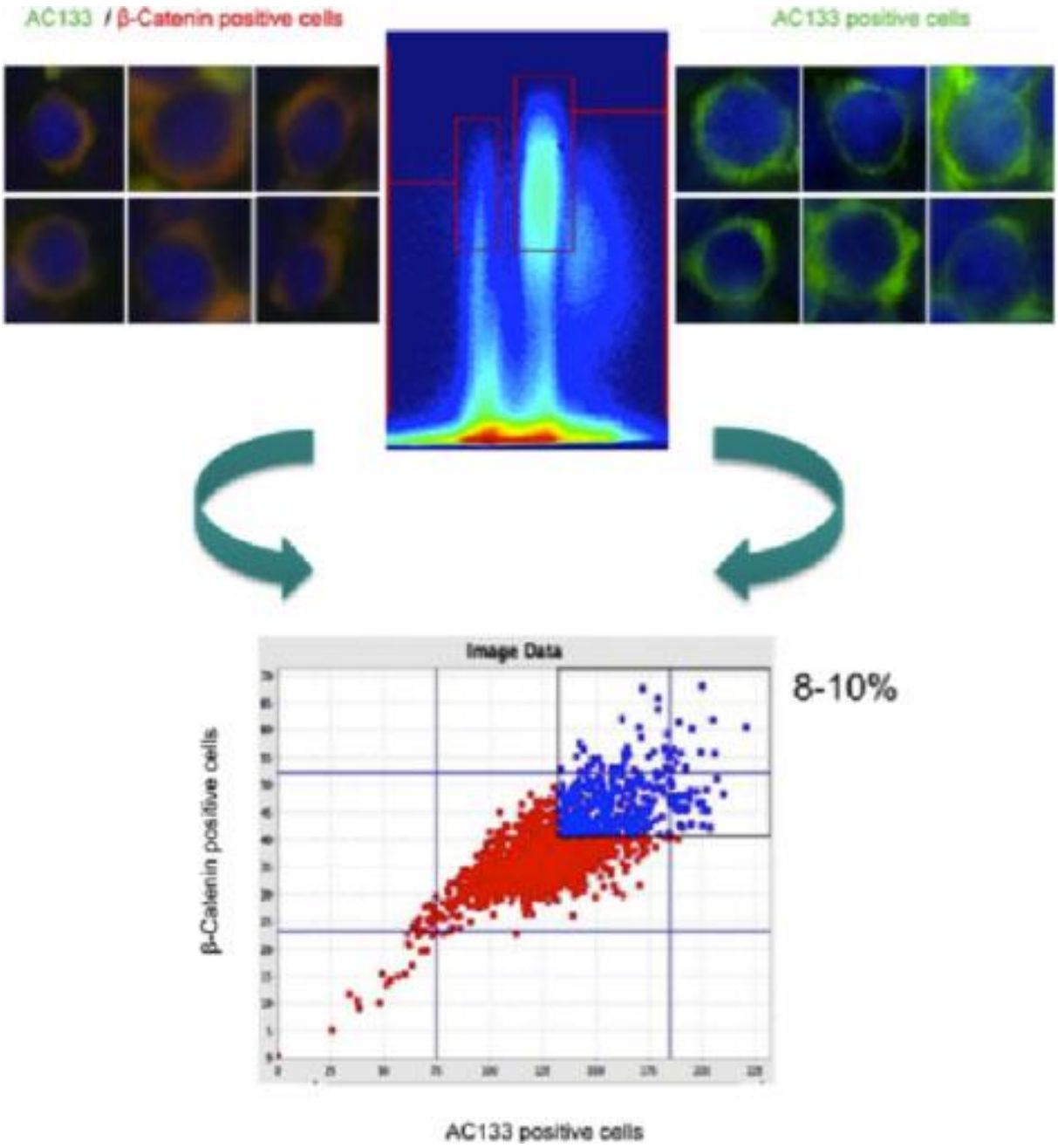


Figure 3.25 Definition of cell subpopulations. Samples can be scored for the number of gated cells and total cells in each sample, the enrichment of that percentage relative to the overall percentage of positive cells in the entire experiment and the left- and right-tail log10 p-values statistical significance of the enrichment, based on the number of

cells in the sample).

In these cases, a density plot showing individual cells can be useful for identifying interesting cell subpopulations, by delineating a section of the plot, called "gating". Whether the gate contains the cells of interest can be tested using two features: the "Show Object Montage" feature to see what individual cells within the gate look like, and the "Show Image" feature to see whether cells within a particular sample are appropriately marked as inside or outside the gate. Once the final, desired subpopulation of cells is gated, the number of cells that fall within that subpopulation is calculated for each image, for further statistical analysis. We considered two gates: one gate is based on the definition of $AC133^{bright}$ cells, and the other gate is established on the double positivity for AC133/ABC. The common gate is the double integrate signal for DAPI staining and the nuclei diamater. In the Figure 3.25 we can observe that $AC133^{bright}$/ABC positive cells are valued as 8-10% of cells in AML bone marrow environment and are characterized by 6-8 nm of nuclei diameter, and double DAPI integrated signal. Finally, through immunostaining followed by a cytometry image analysis, we underlined that $AC133^{bright}$ cells correlated with accumulation of active-β-catenin, demonstrating that activation of Wnt regenerative signaling marked by expression of the dephosphorylated β- catenin was restricted only to the smaller population of $AC133^{bright}$ leukemic cells.

3.7 IDENTIFICATION OF A WNT10B^{IVS1} VARIANT: A WNT-WNT SITUATION

In order to understand the cause of the highly WNT10B regenerative molecule expression in the leukemia environment, we performed a deeply characterization of the WNT10B mRNA. For the identification of the 5' end of a particular transcript, 5' Rapid Amplification of cDNA Ends (5'RACE) is a successful tool *(Frohman et al., 1988).* A gene-specific oligonucleotide that hybridises to a known sequence within a characterised coding region is used to prime reverse transcription. As the first step, we set up the 5' RACE that was carried out on RNA extracted from AML46 patient. Using a GSP2 primer, designed on WNT10B exon2, we obtained a product that is approximately 120bp long. To characterize the RACE product generated, the product was cloned into the pCR®IITOPO ® vector (Invitrogen), and inserts were analysed by EcoRI restriction digest. Inserts that correlated in size with the PCR products generated by 5'RACE were sequenced. We obtained three different results (Figure 3.26):

- Clones 8-11: the correct WNT10B sequence
- Cones 1-4 and clone 31: after the sequencing reaction, we noted the presence of 21bp at the beginning of the WNT10B transcript
- Clones 18-27: we sequenced clones 18 and 27, using universal primer M13 Fw and M13 Rw. Using BLAST and ASAPII, publicly available at http://blast.ncbi.nlm.nih.gov and http://www.bioinformatics.ucla.edu/ASAP2, we observed that the WNT10B transcript, has an Intron Retention IVS1 region of 77nt, and after the IVS1 sequence, there is a stop of cDNA, with the absence of

exon 1.

Using BLAST and ASAPII, publicly available at http://blast.ncbi.nlm.nih.gov and http://www.bioinformatics.ucla.edu/ASAP2, we observed that the WNT10B transcript, has an Intron Retention IVS1 region of 77nt, and after the IVS1 sequence, there is a stop of cDNA.

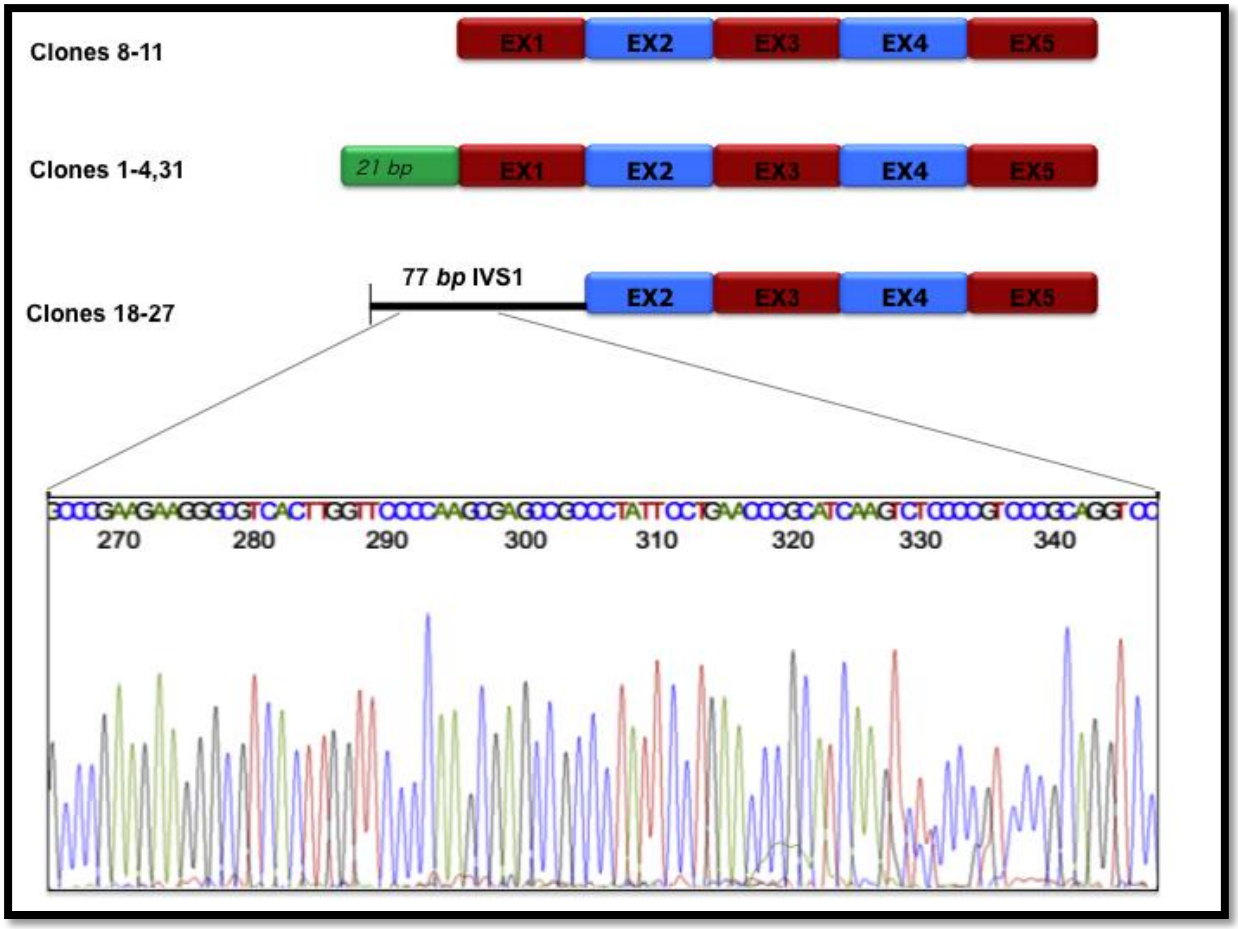


Figure 3.26 Scheme of 5' region of WNT10B.

In order to characterize the WNT10BI^{IVS1} region, we performed in silico analysis using free software as ASPIC (http://www.caspur.it/ASPIC/) and ESEFInder (http://exon.cshl.edu/ESE/). The in silico analysis, using our region as a "query", demonstrate that the Exon2-IVS1 splicing junction is correct, and that the end of the WNT10B^{IVS1} region corresponds with the end of the transcript. It's interesting to note that the ATG site, is localized in the exon 2, suggesting that this alteration doesn'involve

the protein expression. Then, the nature of the WNT10B^{IVS1} transcript remain unclear, but it will be the object for the future perspectives. In order to define the distribution and localization of WNT10B^{IVS1}, we performed the mRNA in situ detection on AML46 spotted and fixed cells, and on AML9 bone marrow biopsy. In our work Beghini et al., we demonstrated through establishment of a primary AC133+ AML cell culture (A46), that leukemia cells synthesize and secrete WNT ligands, increasing the levels of dephosphorylated β-catenin in vivo. Besides, the results of our experiments indicate that AC133 is expressed on AML-LSC in the A46 primary cells, suggesting that regeneration-associated Wnt expression signature is enriched in primary human AML LSC-containing fraction. In order to test the hypothesis that regeneration- associated Wnt signaling is involved in AML-LSC expansion, zebrafish embryonic model has used as a tool by examining the ability of A46 primary leukemia cells to modulate embryonic microenvironment. Therefore we used zebrafish embryos to show that A46 cells prompt secondary axis development inducing the formation of a dorsal-organizer-like structure, possibly via the secretion of different Wnt ligands. Considering this background data, we performed the mRNA in situ detection on A46 cells, derived from AML46 patient (Figure 3.27, β-actin mRNA *in situ* detection). In order to define tha ratio between WNT10B and WNT10B^{IVS1}, we set up the double detection in situ for both molecules. Using one common LNA primer for retrotranscription, we detected WNT10B and WNT10B^{IVS1}, through two specific padlock probe (Figure 3.28)

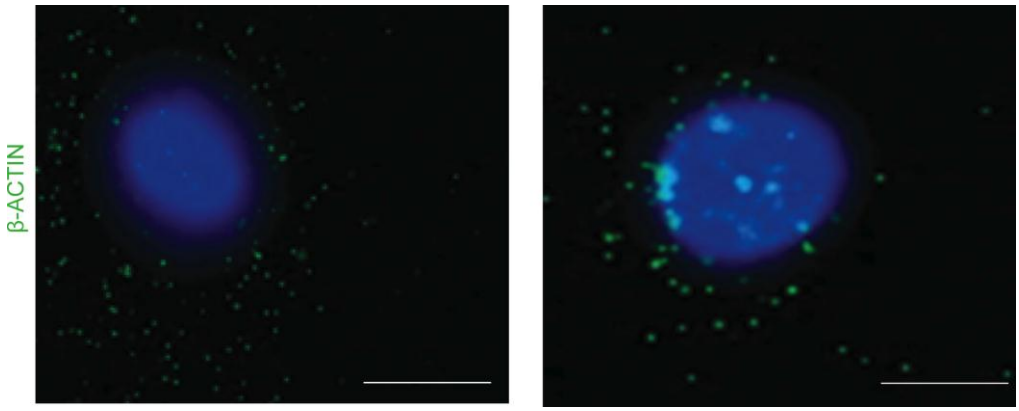


Figure 3.27 β-actin mRNA in situ detection on AML46 cells. β- actin RCPs are shown in green. Nulclei are shown in blue. Scal bar 10 µm.

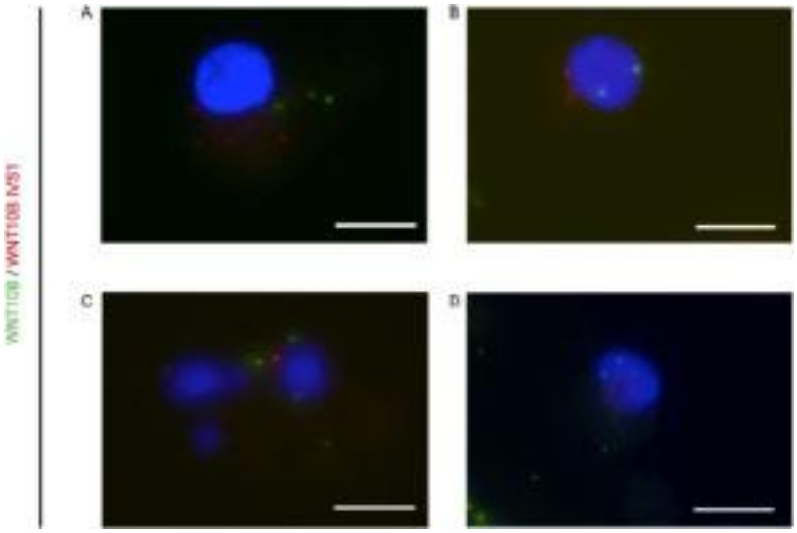


Figure 3.28 WNT10B and WNT10B^{IVS1} detection in situ. mRNA in situ detection of WNT10B and WNT10B^{IVS1} molecules on A46 cells. There are represented four fileds (A,B,C,D). Cell nuclei are shown in blue. WNT10B RCPs are shown in green and WNT10B^{IVS1} are shown in red. Scale bare 10 µm.

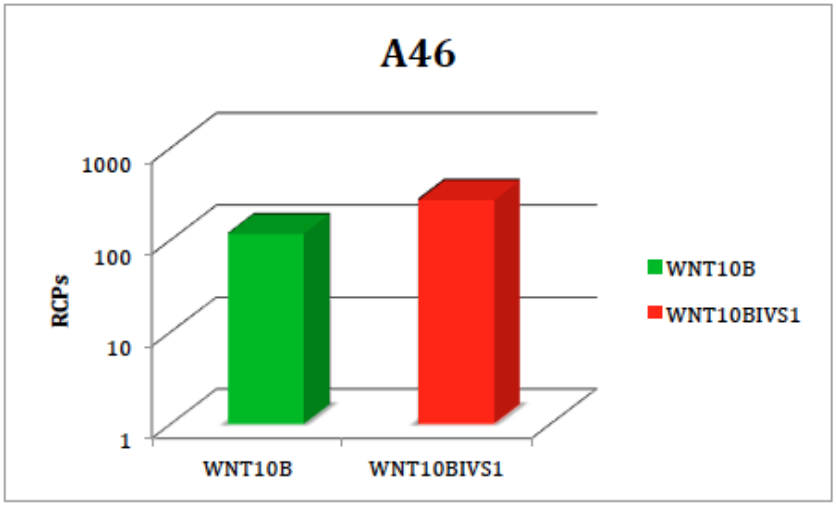


Figure 3.29 RCPs counting of WNT10B and WNT10B^{IVS1}. Data representing of CellProfiler output. The ratio between WNT10B RCPs (green) and WNT10B^{IVS1} (red) is close to 1.

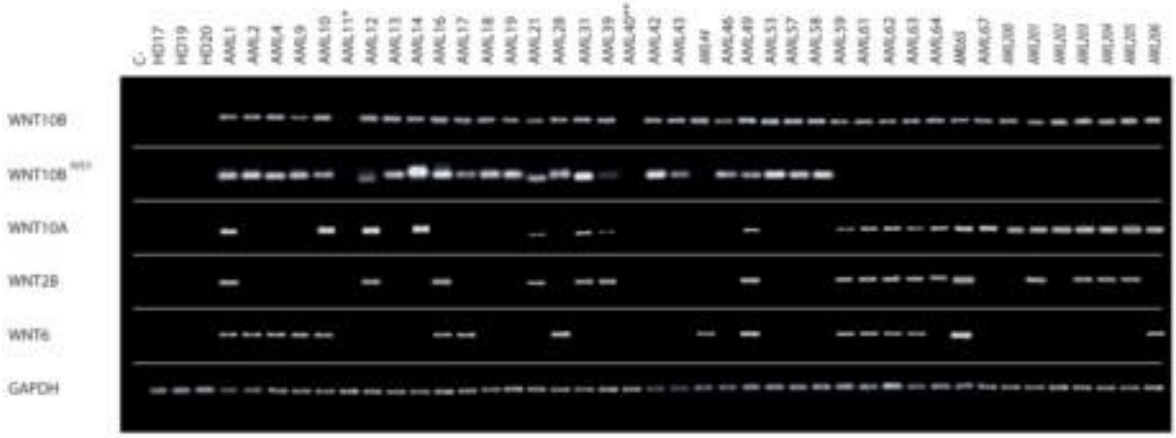
We can note that the ratio between WNT10B (green) and WNT10BIVS1 (red) RCPs is close to 1, suggesting the balance expression of this two expressed isoforms of transcript.

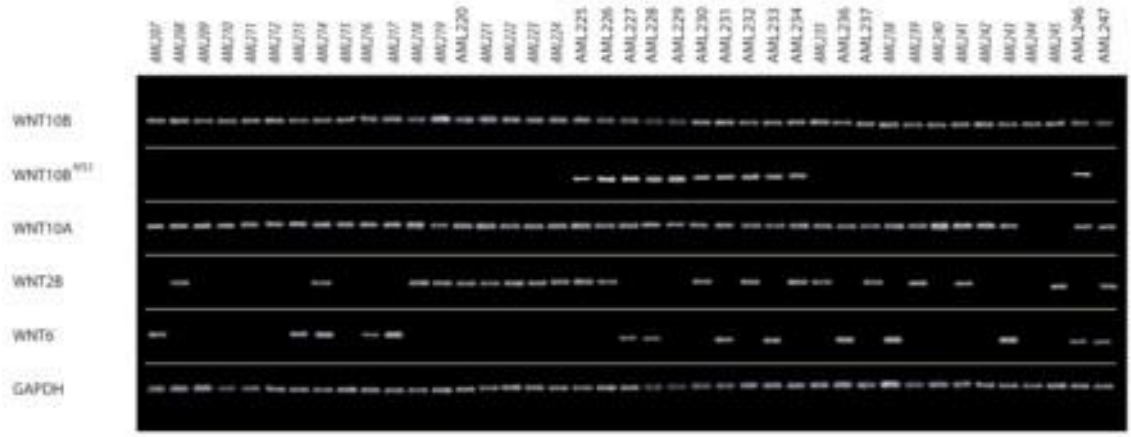
3.8 THE WNT10B EXPRESSION BY AML PATIENTS: CLINICAL RELEVANCE

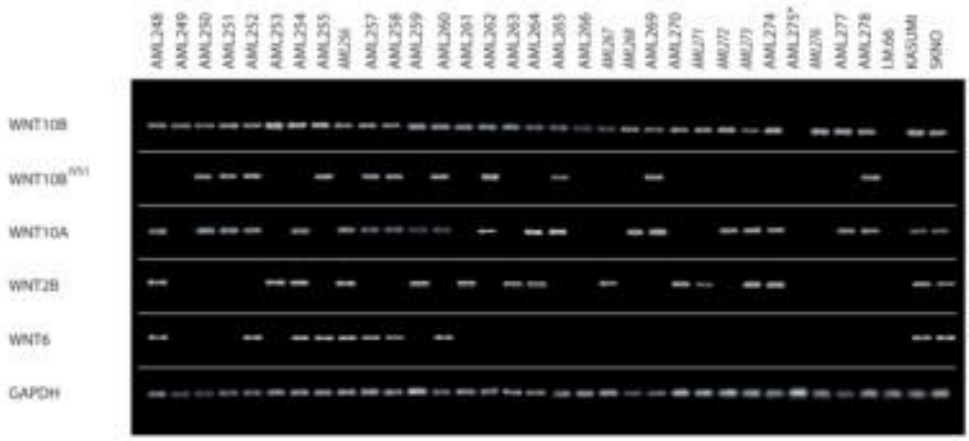
We previously revealed the ligand-dependent Wnt pathway activation in AML and showed a diffuse expression and release of WNT10B, a hematopoietic stem cells regenerative- associated molecule. According to the results presented in Beghini et al., the gene expression microarrays and pathway nalysis evidenced that the Wnt/β-catenin signaling is diffusely activated in the AC133+ AML population, with a specific transcriptional signature involving overexpression of the Wnt pathway agonists and down-modulation of the major antagonists. Besides, we noted that WNT2B, WNT6, WNT10A, and WNT10B, WNT ligands promoter of hematopoietic tissue regeneration *(Congdon et al., 2008; Katoh and Katoh, 2008)* are the WNT mediators specifically upregulated in the AC133+ AML cells. Considering this background data, we performed a gene expression analysis through RT- PCR on n= 112 AML samples and n=3 healthy donors, in order to define the profiling of WNT ligands for AML patient. We set up the detection of WNT10B, WNT10B^{IVS1}, WNT10A, WNT2B and WNT6 molecules, through RT-PCR using GAPDH, the housekeeping gene, as internal reference. In particular, we detect the WNT10B and WNT10B^{IVS1} transcripts, using one common RT primer, designed on Exon2-Exon3 junction. In the Figure 3.30 we can note that only WNT10B is expressed by all AML patients; conversly it's not expressed by AML11 and AML 275, diagnosed as AREB-T, by AML 40 a therapy related AML and by I.M.66 a myelofibrosis syndrome and the. These data suggest that the WNT10B ligand expression, is strictly

associated to the AML condition. Besides, we can note that WNT10B is espressed also by two AML cell lines, KASUMI1 and SKNO. It's important to observe that WNT10B^{IVS1} is not expressed by Core Binding Factor Leukemia patients (n=42), but it is a signature of only dyploid kariotype leukemia.

Figure 3.30 RT-PCR for WNT ligands WNT10B, WNT10B^{IVS1}, WNT10A, WNT2B and WNT6. n=120 AML samples and n=3 healthy donors were analyzed by RT-PCR. Product size: WNT10B 120 bp, WNT10BIVS1 80 bp, WNT10A 180 bp, WNT2B 260 bp, WNT6 300 bp. AML11* and AML275* are AREB-T, AML40** is a therapy related AML.


C-
HD17
HD19
HD20
AML1
AML2
AML4
AML9
AML10
AML11*
AML12
AML13
AML14
AML16
AML17
AML18
AML19
AML21
AML28
AML31
AML39
AML40**
AML42
AML43
IM14
AML46
AML49
AML51
AML57
AML58
AML59
AML61
AML62
AML63
AML64
AML65
AML67
WNT10B
WNT10B V51
WNT10A
WNT2B
WNT6
GAPDH


AMJ07
AMJ08
AMJ09
AMJ10
AMJ11
AMJ12
AMJ13
AMJ14
AMJ15
AMJ16
AMJ17
AMJ18
AMJ19
AML220
AMJ21
AMJ22
AMJ23
AMJ24
AML225
AML226
AML227
AML228
AML229
AML230
AML231
AML232
AML233
AML234
AMJ33
AML236
AML237
AMJ38
AMJ39
AMJ40
AMJ41
AMJ42
AMJ43
AMJ44
AML245
AML246
AML247
WNT10B
WNT10B V51
WNT10A
WNT2B
WNT6
GAPDH


AML248
AML249
AML250
AML251
AML252
AML253
AML254
AML255
AML256
AML257
AML258
AML259
AML260
AML261
AML262
AML263
AML264
AML265
AML266
AML267
AML268
AML269
AML270
AMJ71
AMJ72
AMJ73
AML274
AML275*
AMJ76
AML277
AML278
LM-66
KASUMI
SKNO
WNT10B
WNT10B V51
WNT10A
WNT2B
WNT6
GAPDH

In order to analyze the gene expression analysis, as the first step we used the Jaccard index for correlation; using GenMAPP (http://www.genmapp.org), a computer application designed to visualize gene expression and MappFinder module implemented in R envronment, we defined the correlation 1:1 between WNT ligands (see Figure 3.31). We defined a binary matrix of lecture, based on index ranges from 0 to 1 with 0=variables completely different, 1 =variables identical.

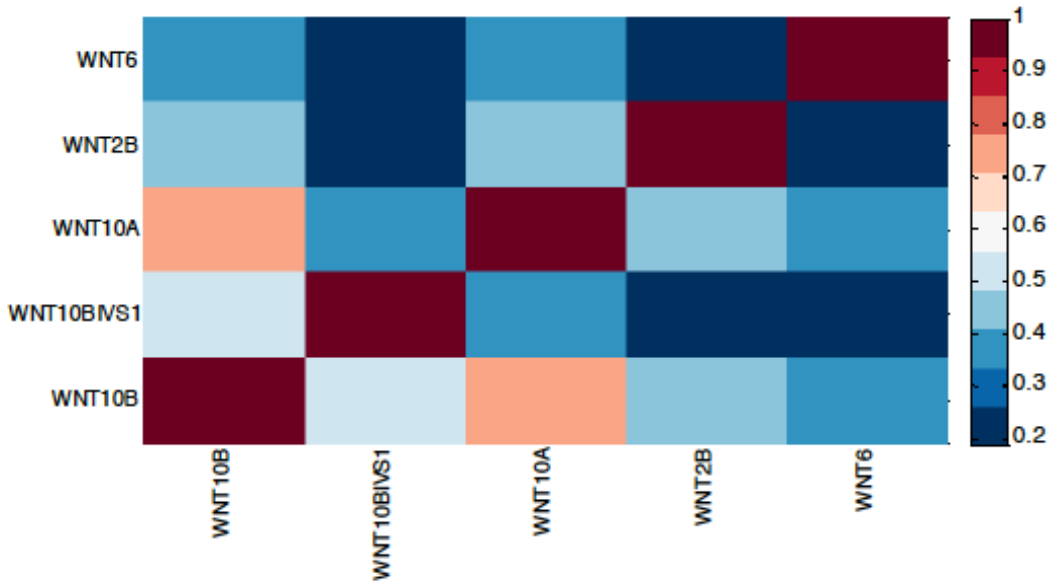


Figure 3.31 Heat color map for WNT ligands correlation. Using Jaccard index for correlation, we linked the WNT expression through the binary matrix. The results are represented through a gradient color map (high correlation: red, low correlation: blue).

We can observe that WNT10B and WNT10A ligands expression, are strictly correlated among AML samples. This parameter defines that, the AML signature profile is characterized by the associated expression of WNT10B and WNT10A ligands. Then, in summary, we can report that:

- AML patient are defined by the WNT10B molecule expression, strictly correlated with WNT10A ligand expression;
- Core Binding Factor Leukemia are defined by the assence of WNT10BIVS1.

This data suggest that, the WNT10B/WNT10B*IVS1* ligand expression are a potential marker of dyploid kariotype AML evidence, showing the role of $WNT10B^{IVS1}$ as a marker of Non Core Binding Factor Leukemia. In order to observe and evaluate the localiztion and distribution of $WNT10B^{IVS1}$ in the AML bone marrow context, we performed the mRNA in situ detection of WNT10B and $WNT10B^{IVS1}$ mRNA molecules. We planned the experiments using only one LNA primer, mapped to WNT10B exon3 and two Padlock Probes: the WNT10B mRNA was detected by a Padlock probe designed on the exon 1 while $WNT10B^{IVS1}$ was detected with a Padlock probe designed on IVS1-exon2 junction.

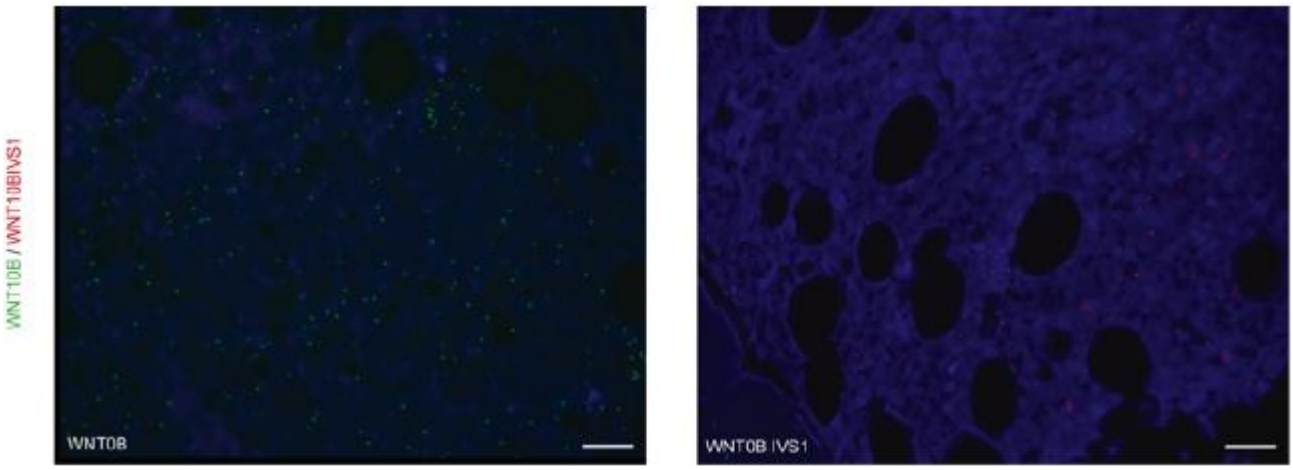


Figure 3.32 WNT10B and WNT10B^{IVS1} mRNA in situ detection. mRNA in situ detection. mRNA in situ detection on AML9 bone marrow biopsy. WNT10B RCPs are shown in green, WNT10B^{IVS1} RCPs are shown in red. Cell nuclei are shown in blue. Scale bare 20 µm.

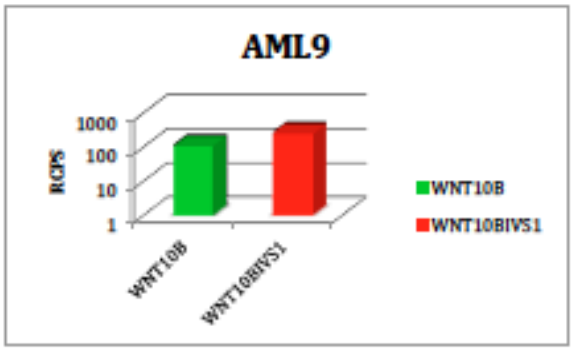


Figure 3.33 RCPs counting. The ratio between WNT10B RCPs (green) and WNT10B^{IVS1} (red) is close to 0.8.

Comparing the WNT10B^{IVS1} with the WNT10B mRNA distribution, we can note that the ratio results equal to 0.8, suggesting that there is a blance between WNT10B and WNT10B^{IVS1} expression. The quantitative comparative analysis between these two isoforms, is on going through the digital PCR, in order to define the ratio between WNT10B and WNT10B^{IVS1} in AML patients.

3.9 MOLT4: A WNT10B^{IVS1}-EXPRESSING CELL LINE

In order to define a model for the WNT10B and WNT10B ^{IVS1} molecules characterization , we studied their expression in hematopoietic cell lines of myeloid and lymphoid lineages. The expression analysis, conducted by RT-PCR following the experimental plan used In the WNT profiling expression analysis, reveals that the MOLT4 cell line expresses only WNT10B ^{IVS1} (Figure 3.34).

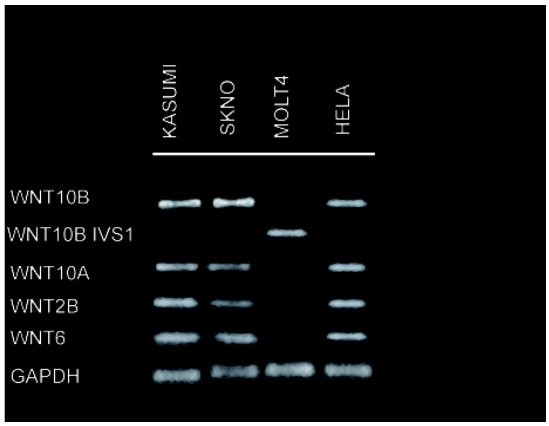
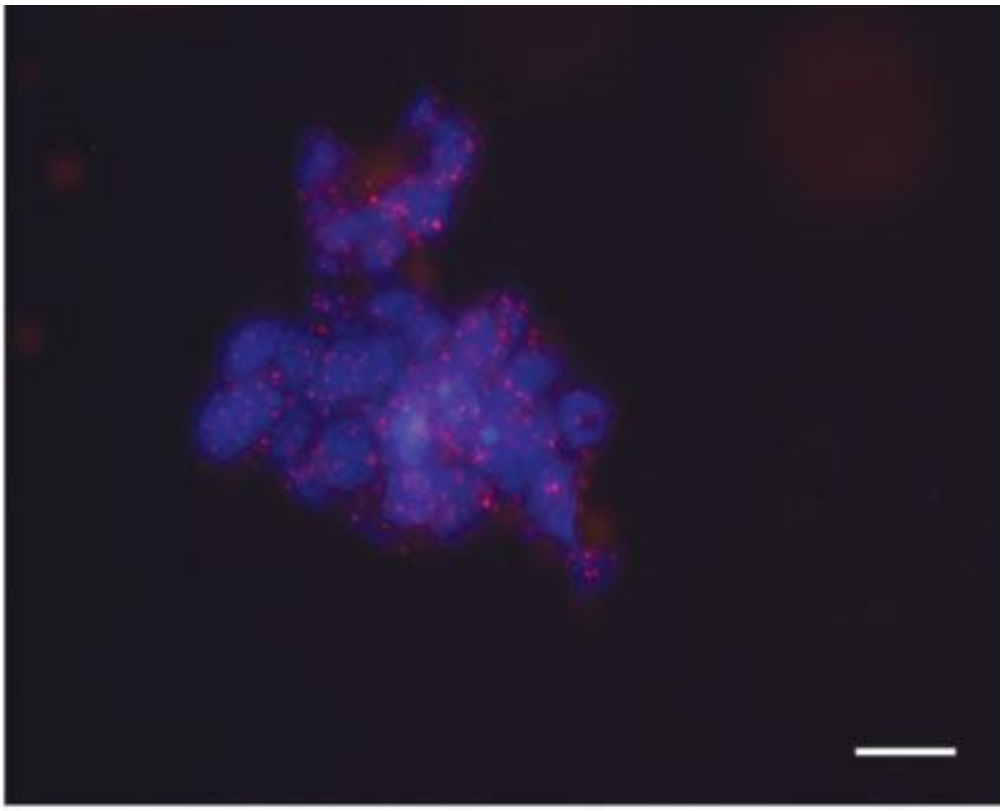


Figure 3.34 RT-PCR for WNT ligands WNT10B, WNT10B^{IVS1}, WNT10A, WNT2B and WNT6 performed on cell lines.

We can note in the figure 3.34 , that KASUMI and SKNO, two Core Binding Factor Leukemia cell lines, shared the same expression WNT ligands profile, characterized by the expression of WNT10B, WNT10A, WNT2B and WNT6 ligands, and by the $WNT10B^{IVS1}$ absence. Only the MOLT4 cell line, a human cell line of acute lymphoblastic leukemia, is characterized by the $WNT10B^{IVS1}$ expression, showing a "complementary" pattern profiling respect other cell lines. In order to define also the in situ distribution of the WNT10B ^{IVS1}.molecules, we performed the mRNA in situ detection on MOLT4 cells, spotted on slides (Figure 3.35).

A



B

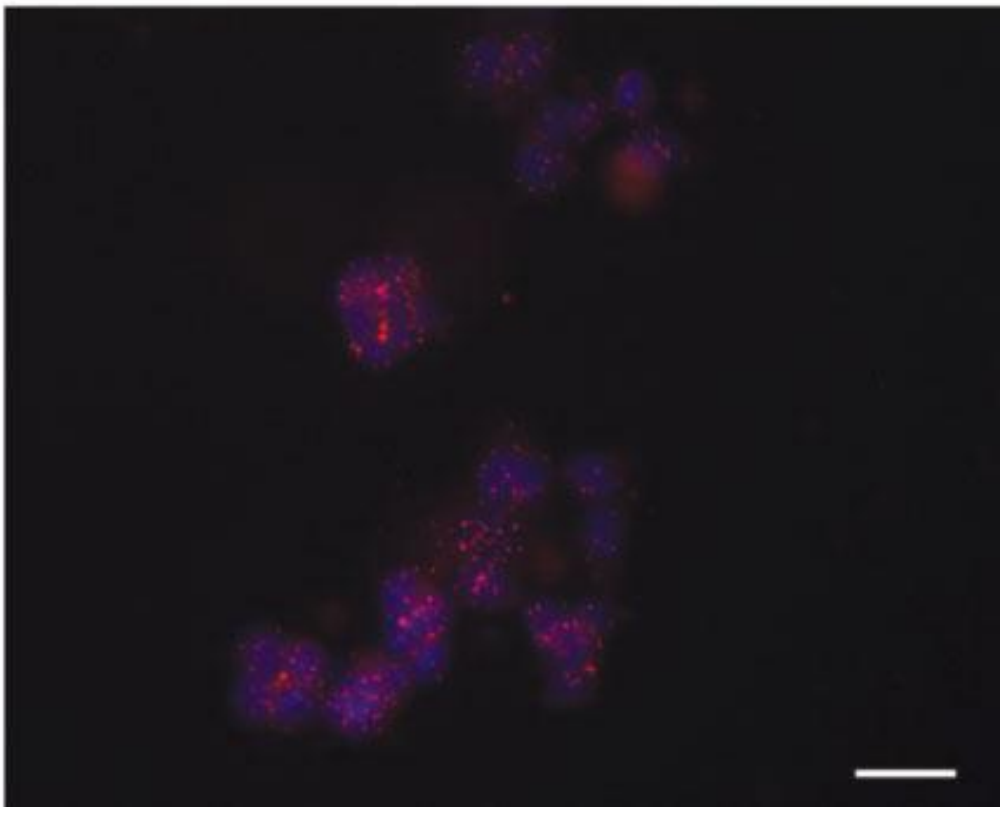


Figure 3.35 mRNA in situ detection on MOLT4 cells. β- actin and WNT10B^{IVS1}mRNA *in situ* detection . RCPs are shown in red. Nuclei are shown in blue. Scale bar 10 μm.

In the figure 3.35 we can note a diffuse expression of WNT10B^{IVS1}mRNA, similar to the housekeeping gene mRNA expression. The WNT10B mRNA *in situ* detection will be the next experiment, in order to support the RT-PCR result.

In order to evaluate the transcription and traduction processes, we also evaluate the WNT10B protein derived from WNT10B^{IVS1,} through a Western Blot analysis.

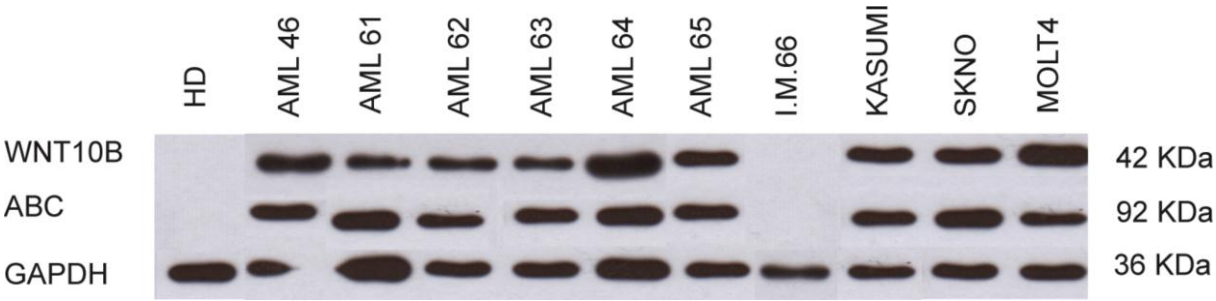


Figure 3.36 Western Blot analysis. Immunoblot analysis of ABC and WNT10B protein expression in cell fractions from 7 AML samples, 1 healthy donor introduced as control and three human cell lines.

We can observe that healthy donors doesn't present the WNT10B protein expression, as the signature of WNT signaling activation. We can underline that I.M.66 sample, a idiopathic myelofibrosis doesn't express WNT10B protein and ABC protein, confirming the RT-PCR results and evidencing that the activation of WNT signaling is restricted to *de novo* AML condition. Conversly, all AML patients and all cell lines analyzed shown the WNT10B protein expression and β-catenine expression as the signature of Wnt pathway activation. It's interisting to note, that also MOLT4 cell line, characterized by only the presence of $WNT10B^{IVS1}$ transcript in RT-PCR analysis, presented the same pattern of other AML samples. This preliminary data, suggests that this alteration is defined at the transcriptional level, with the finally correct protein production. In the future, we will try to identify the reason of this variant generation and also transcriptional and traductional regulation machanism.

4. DISCUSSION

According to recently evidences, the leukemia-initiating cell (LIC) properties occur in a self-renewing non-HSC progenitor cell population, preceded by the expansion of a preleukemic long-term hematopoietic stem cell (LT-HSC).

Recently, Wnt/β-catenin pathway requirement for LIC development in AML has emerged in mouse model, but the molecular function responsible for the preleukemic LT-HSC expansion and the acquisition of self-renewal ability in AML remain unclear. Although the long-term reconstituting human hematopoietic stem cell marker AC133 has been detected in a majority of CD34+ AML, no extensive data concerning the role of AC133 in AML had been available.

Taking into consideration that the AC133 antigen is the most specific marker available for the enrichment of rare hematopoietic stem component / progenitor population we provided in our project the initially purification through this marker of cell populations, obtained at diagnosis from 33 patients with AML and 10 healthy subjects.

The descriptive analysis of positive and negative fractions obtained by immunomagnetic separation of patients and healthy donors, followed by statistical analysis of comparative data, showed that CD133.1+ cell fraction was dramatically expanded among AML with respect to normal donors. In order to define the transcriptional programs induced in the AC133 + cell fraction, we performed a gene profiling analysis on the AC133 + fractions obtained from patients with AML and healthy subjects, took advantage of the Affymetrix platform (GeneChip ® system).

Unlike the study performed by Majeti et al., the gene expression analysis was performed on fractions AC133+ and not on the CD34+ cell population, using microarray Affymetrix platform ® technology, and followed by analysis of pre-processing for the validation data. In order to eliminate the bias intensity, the data obtained from the array were

normalized. The normalization of the signal intensity of arrays considered in our study was performed using an alternative algorithm than the one proposed by Affymetrix ®, the RMA (Robust Multy-Array Average) implemented in R environment. Unlike the non-parametric algorithm proposed by Affymetrix ®, RMA is instead parametric and based on a mathematical model capable of considering the intensity of each probe cell as the summary of a component of the useful signal, defined by an exponential distribution, with a noise component, molded, instead, by a Gaussian distribution. The analysis of Pre-Processing, allowed the comparison of different amount of samples used ensured the validity of the data obtained by eliminating the experimental errors.

The gene expression analysis of microarray data, was followed by the functional enrichment analysis through three computational tools: GOStats, DAVID and dysregulated pathway analysis according to Majeti et al *(Majeti et al., 2009)*.

The results presented in this thesis and in our work Beghini et al., using gene expression microarrays followed by pathway enrichment analysis provide direct evidence that the Wnt/β-catenin signaling is diffusely activated in the expanded AC133+ AML population, with a specific transcriptional signature involving overexpression of the Wnt pathway agonists and down-modulation of the major antagonists *(Beghini et al., 2012)*.

Taking into consideration that the transcription factors (TF), represent the primary executors of transcriptional programs with the function of establishing and maintaining certain cell phenotypes, we performed a bioinformatic analysis of gene profiling data using a set of genes defined for the function of transcriptional regulators. The entry point of our selection procedure was extraction of a set of 400 Pfam domains involved in transcription regulation. Next, we identified an initial set of 2,971 non redundant genes associated to a unique HGNC symbol, using this gene list to query the entire set of

ensemble human genes using the BioMART web interface. As the second step of a bioinformatics analysis, we then compiled an evolutionarily conserved set of TRs by arbitrarily requiring the presence of at least 5 one-to-one orthology relationships disregarding the involved species. We finally tested each of the 1,989 non redundant gene symbols obtained through the orthology filtering for the association with at least 1 Affymetrix HGU 133 Plus 2 microarray platform. This produced a final set of conserved putative transcriptional regulators composed by 1,919 genes, followed by the removal of all the gene symbols lacking an annotation in the Bioconductor annotation package associated with the Affymetrix HGU 133 Plus 2 platform. This filtering procedure produced the reference conserved TRs set composed by 1,611 evolutionarily conserved genes.

We found that at 0.05 significance level for all functional enrichment analysis, the term GO:0016055 associated with the function of "Wnt receptor signaling pathway", results the most specific dysregulated biological process. Thus, the three independent types of analysis, applied to all genes and to transcriptional regulators, selected the term "Wnt receptor signaling pathway" as the most specific self-renewal associated dysregulated pathway. We can note that WNT10B, WNT10A WNT2B and WNT6, known to promote the hematopoietic tissue regeneration, are the WNT mediators upregulated in the AC133+ AML cells. According to the literature's evidence, in the hematopoietic system WNT10B is significantly upregulated following an injury, increasing the growth of HSCs *(Angers & Moon, 2009)*. Taking this issues into consideration, we showed a dramatic increase of WNT10B expression and protein release within the microenvironment in the large majority of samples from AML patients recruited to this study. Analysis of freshly fractionated cells from AML patients showed that active Wnt signaling was predominant

in population highly enriched for the AC133 HSC marker. The hematopoietic regenerative-associated Wnt ligand WNT10B is expressed on both leukemic blasts and stromal-like cells, indicating a possible autocrine/paracrine involvement of Wnt in the bone marrow microenvironment. Conversely, we noted that the activation of Wnt signaling marked by expression of the dephosphorylated β-catenin has restricted to a smaller proportion of leukemic cells. The reasons for these results remain unclear, but it is possible that Wnt-induced β-catenin-activation, conferring a "responsive" phenotype, restricted to a rare population of cells.

In this thesis was presented an application of a new in situ technique using padlock probes and RCA for detection and genotyping of individual mRNA molecules in cells and tissues. The mRNA *in situ* detection allows selective and multiplex detection of individual transcripts and will serve as an important tool within gene expression studies *(Larsson et al., 2010)*.

The mRNA *in situ* detection, performed on AML bone marrow sections and on AML cells, allowed to detect and visualize the WNT10B mRNA molecules at their exact location in a tissue and in a single cell. The CellProfiler pipeline, used for a RCPs quantification referred to the housekeeping gene expression, showed a β-actin: WNT10B ratio close to 1.0 suggesting a diffuse constitutive activation of WNT10B transcription in the bone marrow.

It is worth to note, that we used image analysis software as Cellprofiler and ImageJ with a MATLAB bridge, in order to reduce the backgoround signals and to detect every RCPs in FPPE sections.

Since activation of WNT signaling can increase the HSC's ability to reconstitute the hematopoietic system of lethally irradiated mice, we transplanted the AC133+ A46 cells

into irradiated Rag2$^{-/-}\gamma c^{-/-}$ mice. The results of our experiments evidenced that AC133 is expressed on AML-LSC in the A46 primary cells, suggesting that regeneration-associated Wnt expression signature is enriched in primary human AML LSC-containing fraction.

According to Lu et al., the maternal Wnt ligands are fundamental for the establishment of the embryonic body plan, a milestone in the development of vertebrates. The Wnt factors induce the nuclear translocation of β-catenin, which triggers the formation of the dorsal organizer through the activation of zygotic dorsal-specific genes and they are recquired for proper tail development. In addition, Bowman et al., evidenced that developmental signal transduction pathways, such as Wnt, are often reactivated during regeneration. Taking this issues into consideration, we used zebrafish embryos to show that A46 cells, possibly via the secretion of different Wnt ligands, prompt secondary axis development inducing the formation of a dorsal-organizer-like structure. The mechanisms promoting organizer formation are known to involve cooperation between Nodal and Wnt signaling. In our paper Beghini et al., we demonstrated that the A46-dependent alteration of zebrafish development and activation of the organizer-specific gene *gsc* are not reliant on Nodal activity *(Beghini et al., 2012)*. The results of the first part of my phD program, provided a compelling evidence that regeneration-associated Wnt signaling exceeds the homeostatic range in human AML and affects responsive cells whose renewal is promoted by Wnt pathway activity.

Taking this results into consideration, during the second part of my phD project we focused our attention on a characterization of a regenerative function associated to WNT pathway induction. First of all, we detected *in situ* a AC133^{bright} subpopulation, that shared the WNT signaling activation signature.

In order to better characterize the LICs, we performed a series of direct immunolabeling

on AML bone marrow biopsies.

First of all, we evidenced that the activation of WNT signaling marked by the expression of active β-catenin was restricted to the smaller subpopulation of AC133^{bright} leukemic cells. We performed an interactive observation of the original cellular images using CellProfilerAnalyst, a new method for image-based screening and machine learning-based phenotype scoring, in order to define cell phenotypes. For complex phenotypes, several features of each cell may be required for effective scoring, measure nuclear features, staining's intensity and AC133/ β-catenin positivity. The multiple single-cell measurements analysis, define that only 8-10% of cells, are defined by a double AC133^{bright}/ β-catenin positivity, suggesting that only a small AC133^{bright} subpopulation shared the WNT signaling activation signature.

Focusing our attention on the major locus associated to the regenerative function, we performed a 5'RACE analysis on WNT10B mRNA, evidencing the presence of an alternative expressed WNT10B^{IVS1} variant. We evidenced *in situ* that the ratio between WNT10B and WNT10B^{IVS1} is close to 1, but the potential role and the mechanism of WNT10B^{IVS1} expression, will be the object of future studies.

In order to evaluate the expression of WNT ligands specifically up-regulated in the AC133+ AML cells and that promote hematopoietic tissue regeneration evidenced in our work *(Beghini et al., 2012)*, we performed the unsupervised expression analysis profiling on n=112 AML samples and n=3 healthy donors.

The first result, of this ongoing project, evidences that the WNT10B ligand is expressed by all AML patients, suggesting the potential role of WNT10B as a marker of AML condition. It is worth to note, that the regenerative WNT signaling associated function marks just AML condition, while myelodisplastic syndroms and AML therapy related are

completely negative for WNT regenerative ligands expression. Infact, performing the unsupervised expression analysis, in order to remove the biases, we had the possibility to re-evaluate the WNT regenrative ligands cases, as not AML conditions, suggesting the possible role of WNT profiling as a putative tool of diagnosis.

It's interesting to note that the expression of WNT10B^{IVS1} variant, marks only the Non Core Binding Factor Leukemias, suggesting its possible role as a biomarker of a diploid karyotype AML.

Taking this issues into consideration, as the future plan we will perform a WNT-regenerative-ligands profiling, in order to define a signature of AML patients, using WNT ligands expression as putative biomarkers. The first step in the WNT10B^{IVS1} molecule characterization, is the identification of the MOLT4 human cell line, that expressed only WNT10B^{IVS1} variant, , allowing the future characterization experiments. In summary, the findings we report in this thesis and in our work Beghini et al., define for the first time the regeneration-associated ligand-dependent WNT signaling induction in human AML cases. These studies suggest that the regenerative WNT signaling is an altered stem cell function, in AC133^{bright} AML leukemia stem cell fraction. The data presented, evidence that the WNT10B molecule is strictly associated to the de novo AML condition, suggesting its role as a putative disease biomarker and addressing the future experimental plans. As the future perspective, we are defining the quantitive ratio between WNT10B/ WNT10B^{IVS1}molecules in AML patients, through a Droplet Digital PCR technology that provides absolute quantification of nucleic acids,.

Future studies will be focused on a demonstration of a pathogenic association of AC133^{bright} LIC regeneration response with AML, as tumorigeicity, clinico-pathologic features and also patient outcomes.

5. REFERENCES

Aberle H, Butz S, Stappert J, et al. (1994). Assembly of the cadherin–catenin complex in vitro with recombinant proteins. *J Cell Sci* **107**, 3655–3663.

Abravaya K, Carrino JJ, Muldoon S, et al. (1995). Detection of point mutations with a modified ligase chain reaction (Gap-LCR). *Nucleic Acids Res* **23**, 675-82.

Agarwal KL, Buchi H, Caruthers MH, et al. (1970). Total synthesis of the gene for an alanine transfer ribonucleic acid from yeast. *Nature* **227,** 27-34.

Ailles LE, Gerhard B, Kawagoe H, et al. (1999). Growth characteristics of acute myelogenous leukemia progenitors that initiate malignant hematopoiesis in nonobese diabetic/severe combined immunodeficient mice. *Blood* **94** (5), 1761–1772.

Al-Hajj M, & Clarke MF (2004). Self-renewal and solid tumor stem cells. *Oncogene* **23**, 7274- 82.

Appelbaum FR, Kopecky KJ, Tallman MS, et al. (2006). The clinical spectrum of adult acute myeloid leukaemia associated with core binding factor translocations. *Br J Haematol* **135**, 165-173.

Arnoux P, Morosinotto T, Saga G, et al. (2009). A structural basis for the pH-dependent xanthophyll cycle in Arabidopsis thaliana. *Plant Cell* **21**, 2036- 2044.

Ashburner M, Ball CA, Blake JA, et al. (2000). Gene ontology: tool for the unification of

biology. The Gene Ontology Consortium. *Nat Genet* **25**, 25–29.

Bacher U, Kohlmann A, Haferlach C, et al. (2009). Gene expression profiling in acute myeloid leukaemia (AML). *Best Practice & Research Clinical Haematology* **22** (2), 169-180.

Bafico A, Gazit A, Pramila T, et al. (1999). Interaction of frizzled related protein (FRP) with Wnt ligands and the frizzled receptor suggests alternative mechanisms for FRP inhibition of Wnt signaling. *J Biol Chem*, **274** (23), 16180-16187.

Banerjee B, Miedema BE & Chandrasekhar HR (1999). Role of basement membrane collagen and elastin in the autofluorescence spectra of the colon. *J Invest Med* **47,** 326–331.

Banziger C, Soldini D, Schutt C, et al. (2006). Wntless, a conserved membrane protein dedicated to the secretion of Wnt proteins from signaling cells. *Cell* **125,** 509-522.

Bao S, Wu Q, McLendon RE, et al. (2006). Glioma stem cells promote radioresistance by preferential activation of the DNA damage response. *Nature* **444,** 756–760.

Barandon L, Couffinhal T, Ezan J, et al. (2003): Reduction of infarct size and prevention of cardiac rupture in transgenic mice overexpressing FrzA. *Circulation* **108** (18), 2282-2289.

Barrell D, Dimmer E, Huntley RP, et al. (2009). The GOA database in 2009: an integrated

Gene Ontology Annotation resource. *Nucleic Acids Research* **37** (Suppl 1), D396–D403.

Bartscherer K, Pelte N, Ingelfinger D, et al. (2006). Secretion of Wnt ligands requires Evi, a conserved transmembrane protein. *Cell* **125**, 523-533.

Bauer A, Chauvet S, Huber O, et al. (2000). Pontin52 and reptin52 function as antagonistic regulators of beta-catenin signalling activity. *EMBO J* **19,** 6121–6130.

Bauer N., Fonseca AV, Florek M, et al. (2008). New insights into the cell biology of hematopoietic progenitors by studying prominin-1 (CD133). *Cells Tiss Org* **188,** 127-138.

Becker AJ, Mc CE, & Till, JE. (1963). Cytological demonstration of the clonal nature of spleen colonies derived from transplanted mouse marrow cells. *Nature* **197**, 452-454.

Beghini A, Peterlongo P, Ripamonti CB, et al. (2000a). c-kit mutations in core binding factor leukemias. *Blood* **95**, 726–727.

Behrens J, von Kries JP, Ku¨hl M, et al. (1996). Functional interaction of beta-catenin with the transcription factor LEF-1. *Nature* **382**, 638–642.

Belenkaya TY, Wu Y, Tang X, et al. (2008). The retromer complex influences Wnt secretion by recycling wntless from endosomes to the trans-Golgi network. *Dev Cell* **14**, 120-131.

Belichenko PV, Fedorov AA, Dahlström, AB (1996). Quantitative analysis of immunofluorescence and lipofuscin distribution in human cortical areas by dual-channel confocal scanning microscopy. *J Neurosci Methods* **69** 155–161.

Benviste P, Frelin C, Janmohamed S, et al. (2010). Intermediate-term hematopoietic stem cells with extended but time-limited reconstitution potential. *Cell Stem Cell* **6,** 48–58.

Betz BL & Hess JL (2010). Acute myeloid leukemia diagnosis in the 21st century. *Arch Pathol Lab Med* **134,** 1427–1433.

Bhanot P, Brink M, Samos CH, et al. (1996). A new member of the frizzled family from Drosophila functions as a Wingless receptor. *Nature* **382**, 225-230.

Bhatia M, Wang JC, Kapp U, et al. (1997). Purification of primitive human hematopoietic cells capable of repopulating immune-deficient mice. *Proc Natl Acad Sci USA* **94** (10), 5320–5.

Bilic J, Huang YL, Davidson G, et al. (2007). Wnt induces LRP6 signalosomes and promotes dishevelled-dependent LRP6 phosphorylation. *Science* **316,** 1619–1622.

Blake J, Bult C, Kadin J, et al. (2011). The Mouse Genome Database (MGD): premier model organism resource for mammalian genomics and genetics. *Nucleic Acids Research* **39** (Suppl 1), D842–D848.

Blanco L & Salas M (1996). Relating structure to function in phi29 DNA polymerase. *J Biol Chem* **271,** 8509-8512.

Boissel N, Leroy H, Brethon B, et al. (2006a). Incidence and prognostic impact of c-Kit, FLT3, and Ras gene mutations in core binding factor acute myeloid leukemia (CBF-AML). *Leukemia* **20**, 965-970.

Bolstad BM, Irizarry RA, Astrand M, et al. (2003). A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics,* **19** (2), 185-193.

Bolte S & Cordelieres FP (2006). A guided tour into subcellular colocalization analysis in light microscopy. *Journal of Microscopy*, **224** (3), 213–232.

Bonnet D & Dick JE (1997). Human acute myeloid leukemia is organized as a hierarchy that originates from a primitive hematopoietic cell. *Nat Med* **3**, 730–737.

Borer PN, Dengler B, Tinoco I Jr et al. (1974). Stability of ribonucleic acid double-stranded helices. *J Mol Biol* **86**, 843–853.

Bowman TV, Trompouki E & Zon L (2012). Linking hematopoietic regeneration to

developmental signaling pathways: A story of BMP and Wnt. *Cell Cycle* **11**(3), 424-5.

Bradford Y, Conlin T, Dunn N, et al. (2011). ZFIN: enhancements and updates to the zebrafish model organism database. *Nucleic Acids Research* **39** (Suppl 1), D822–D829.

Bradley RS & Brown AM (1990). The proto-oncogene int-1 encodes a secreted protein associated with the extracellular matrix. *EMBO J* **9**, 1569-1575.

Brandts CH, Sargin B, Rode M, et al. (2005). Constitutive activation of Akt by Flt3 internal tandem duplications is necessary for increased survival, proliferation, and myeloid transformation. *Cancer Res* **65**, 9643-9650.

Breitenbuecher F, Markova B, Kasper S, et al. (2009). A novel molecular mechanism of primary resistance to FLT3-kinase inhibitors in AML. *Blood* **113**, 4063-4073.

Brioschi M, Fischer J, Cairoli R, et al. (2010). Down-regulation of micro-RNAs 222/221 in acute myelogenous leukemia with deranged core-binding factor subunits. *Neoplasia* **12**, 866-876.

Brown SD, Twells RC, Hey PJ, et al. (1998). Isolation and characterization of LRP6, a novel member of the low density lipoprotein receptor gene family. *Biochem Biophys Res Commun* **248**, 879-888.

Brunner E, Peter O, Schweizer L, et al. (1997). Pangolin encodes a Lef-1 homologue that acts downstream of Armadillo to transduce the Wingless signal in Drosophila. *Nature* **385**, 829–833.

Bullinger L, Krönke J, Schön C, et al. (2010). Identification of acquired copy number alterations and uniparental disomies in cytogenetically normal acute myeloid leukemia using high-resolution single-nucleotide polymorphism analysis. *Leukemia* **24**, 438–449.

Bustin SA (2002). Quantification of mRNA using real-time reverse transcription PCR (RT-PCR): trends and problems. *J Mol Endocrinol* **29**, 23-39.

Bustin SA, Benes V, Nolan T, et al. (2005). Quantitative real-time RT-PCR-a perspective. *J Mol Endocrinol* 3**4**, 597-601.

Cadigan KM & Nusse R (1997). Wnt signaling: a common theme in animal development. *Genes Dev* **11**, 3286-3305.

Cairoli R, Beghini A, Grillo G, et al. (2006). Prognostic impact of c-KIT mutations in core binding factor leukemias: an Italian retrospective study. *Blood* **107**, 3463-3468.

Cantrill S. Nobel prize 2008: Green fluorescent protein. *Nature Chemistry*.

Caramelo JJ & Parodi AJ (2007). How sugars convey information on protein

conformation in the endoplasmic reticulum. *Semin Cell Dev Biol* **18,** 732-742.

Care RS, Valk PJ, Goodeve AC, et al. (2003). Incidence and prognosis of c-KIT and FLT3 mutations in core binding factor (CBF) acute myeloid leukaemias. *Br J Haematol* **121**, 775-777.

Caricasole A, Ferraro T, Iacovelli L, et al. (2003): Functional characterization of WNT7A signaling in PC12 cells: interaction with A FZD5 x LRP6 receptor complex and modulation by Dickkopf proteins. *J Biol Chem*, **278** (39), 37024-37031.

Carpenter AE (2009). Extracting rich information from images. *Methods Mol Biol* **486**, 193-211.

Carpenter AE, Jones TR, Lamprecht MR, et al. (2006). CellProfiler: image analysis software for identifying and quantifying cell phenotypes. *Genome Biol* **7** (10): R100.

Carron C, Pascal A, Djiane A, et al. (2003). Frizzled receptor dimerization is sufficient to activate the Wnt/betacatenin pathway. *J Cell Sci* **116**, 2541-2550.

Causton H, Quackenbush J & Brazma, A (2003). Microarray gene expression data analysis: a beginner's guide. Massachusetts, USA: Blackwell.

Chase A, & Cross NC (2006). Signal transduction therapy in haematological malignancies: identification and targeting of tyrosine kinases. *Clin Sci (Lond)* **111**, 233-49.

Chelly J, et al (1990). Quantitative estimation of minor mRNAs by cDNA-polymerase chain reaction. Application to dystrophin mRNA in cultured myogenic and brain cells. *Eur J Biochem* **187**, 691-698.

Cherry JM, Adler C, Ball C, et al. (1998). SGD: Saccharomyces Genome Database. *Nucleic Acids Research* **26** (1), 73–79.

Chitalia VC, Foy RL, Bachschmid MM, et al. (2008). Jade-1 inhibits Wnt signalling by ubiquitylating betacatenin and mediates Wnt pathway inhibition by pVHL. *Nat Cell Biol* **10**, 1208–1216.

Chu SH & Small D (2009). Mechanisms of resistance to FLT3 inhibitors. *Drug Resist Updat* **12**, 8-16.

Civin CI, Almeida-Porada G, Lee MJ, et al. (1996). Sustained, retransplantable, multilineage engraftment of highly purified adult human bone marrow stem cells in vivo. *Blood* **88**, 4102–4109.

Clevers H, (2006). Wnt/beta-catenin signaling in development and disease. *Cell* **127**

(3)469-480.

Cong F, Schweizer L, & Varmus H (2004). Wnt signals across the plasma membrane to activate the beta-cateninpathway by forming oligomers containing its receptors, Frizzled and LRP. *Development* **131**, 5103-5115.

Coombs GS, Yu J, Canning CA, et al. (2010). WLS-dependent secretion of WNT3A requires Ser209 acylation and vacuolar acidification. *J Cell Sci* **123**, 3357-3367.

Corbeil D, Marzesco AM, Fargeas CA, et al. (2010). Prominin-1, a distinct cholesterol ? binding membrane protein, and the organisation of the apical plasma membrane of epithelial cells. *Subcell. Biochem*. **51**, 399-423.

Corbeil D, Röper K, Fargeas CA, et al. (2001b). Prominin: a story of cholesterol, plasma membrane protrusions and human pathology. *Traffic* **2**, 82-91.

Coudreuse DY, Roel G, Betist MC, et al. (2006). Wnt gradient formation requires retromer function in Wnt-producing cells. *Science* **312,** 921-924.

Couso JP & Martinez Arias A (1994). Notch is required for wingless signaling in the epidermis of Drosophila. *Cell* **79**, 259-272.

Croft D, O'Kelly G, Wu G, et al. (2011). Reactome: a database of reactions, pathways and biological processes. *Nucleic Acids Research* **39** (Suppl 1), D691–D697.

Curran S, McKay JA, McLeod HL et al. (2000). Laser capture microscopy. *Mol Pathol* **53**, 64-68.

Dajani R, Fraser E, Roe SM, et al. (2003). Structural basis for recruitment of glycogen synthase kinase 3beta to the axin-APC scaffold complex. *EMBO J* **22**, 494–501.

Daubendick SL, Ryan K, & Kool ET (1995). Rolling Circle RNA Synthesis: Circular Oligonucleotides as Efficient Substrates for T7 RNA Polymerase *J Am Chem Soc* **117**, 7818-7819.

David M, Petricoin E, Benjamin C, et al. (1995). Requirement for MAP kinase (ERK2) activity in interferon alpha- and interferon beta-stimulated gene expression through STAT proteins. *Science* **269**, 1721-1723.

Davidson G, Wu W, Shen J, et al. (2005). Casein kinase 1 gamma couples Wnt receptor activation to cytoplasmic signal transduction. *Nature* **438**, 867-872.

Davidson MW & Abramowitz M (2008). Optical microscopy. http://www.olympusmicro.com/ primer/microscopy.pdf.

Del Castillo P, Llorente AR & Stockert JC (1989). Influence of fixation, exciting light and section thickness on the primary fluorescence of samples for microfluorometric analysis. *Bas Appl Histochem* **33**, 251–257.

Delaunay J, Vey N, Leblanc T, et al. (2003). Prognosis of inv(16)/t(16;16) acute myeloid leukemia (AML): a survey of 110 cases from the French AML Intergroup. *Blood* **102**, 462-469.

Delisi C & Crothers DM (1971). Prediction of RNA secondary structure. *Proc. Natl Acad. Sci. USA* **68**, 2682–2685.

Denk W, Strickler J & Webb W (1990). Two-photon laser scanning fluorescence microscopy. *Science* **248** (4951), 73–76.

Dennis S, Aikawa M, Szeto W, et al. (1999). A secreted frizzled related protein, FrzA, selectively associates with Wnt-1 protein and regulates wnt-1 signaling. *J Cell Sci* **112**, 3815-3820.

Devereux J, Haeberli P & Smithies O (1984). A comprehensive set of sequence analysis programs for the VAX. *Nucleic Acids Res* **12**, 387–395.

Doherty AJ & Dafforn TR (2000). Nick recognition by DNA ligases. *J Mol Biol* **296** (1), 43-56.

Doherty AJ & Suh SW (2000). Structural and mechanistic conservation in DNA ligases. *Nucleic Acids Res* **28** (21), 4051-4058.

Döhner H, Estey EH, Amadori S, et al. (2010). Diagnosis and management of acute myeloid leukemia in adults: recommendations from an international expert panel, on behalf of the European LeukemiaNet. *Blood* **115,** 453-474.

Dohner K, Du J, Corbacioglu A, et al. (2006). JAK2V617F mutations as cooperative genetic lesions in t(8;21)-positive acute myeloid leukemia. *Haematologica* **91**, 1569-1570.

Doty P, Boedtker H, Fresco JR, et al. (1959). Secondary structure in ribonucleic acids. *Proc Natl Acad Sci USA* **45**, 482–499.

Draghici S, Khatri P, Eklund AC, et al. (2006). Reliability and reproducibility issues in DNA microarray measurements. *Trends Genet* **22**, 101-109.

Dubreuil D, Marzesco AM, Corbeil D, et al. (2007). Midbody and primary cilium of neural progenitors release extracellular membrane particles enriched in the stem cell marker prominin-1. *J Cell Biol* **176,** 483-49.

Dufourcq P, Couffinhal T, Ezan J, et al. (2002). FrzA, a secreted frizzled related protein, induced angiogenic response. *Circulation* **106** (24), 3097-3103.

Dupaˇc J & Hlav.ˇc V (2006). Stable wave detector of blobs in images. Proceedings of 28th Annual Symposium of the German Association for Pattern Recognition, Berlin, Germany, Springer Verlag, 760–769.

Edgren H, et al (2011). Identification of fusion genes in breast cancer by paired-end RNA-sequencing. *Genome Biol* **12**, R6.

Edwin EE & Jackman R (1981) Nature of autofluorescent material in cerebrocortical necrosis. *J Neurochem* **37**,1054–1056.

Emmert-Buck MR, et al (1996). Laser capture microdissection. *Science* **274**, 998-1001.

Engler MJ & Richardson CC (1982). DNA ligases. *The Enzymes* **15B**, 1-29. Academic Press, New York.

Esteban JA, Salas M & Blanco L (1993). Fidelity of phi 29 DNA polymerase. Comparison between protein-primed initiation and DNA polymerization. *J Biol Chem* **268,** 2719-2726.

Ettinger AW, Wilsch-Bräuninger M, Marzesco AM, et al. (2011). Proliferating versus differentiating stem and cancer cells exhibit distinct midbody release behavior. *Nature Commun* **2**, 503 (1-12).

Fagotto F, Glu¨ck U, Gumbiner BM (1998). Nuclear localization signal-independent and importin/karyopherin-independentnuclear import of beta-catenin. *Curr Biol* **8**, 181–190.

Falcon S & Gentleman R (2007). Using GOstats to test gene lists for GO term association. *Bioinformatics* **23**(2), 257–258.

Fareed GC, Wilt EM & Richardson CC (1971). Enzymatic breakage and joining of deoxyribonucleic acid. VIII. Hybrids of ribo- and deoxyribonucleotide homopolymers as substrates for polynucleotide ligase of bacteriophage T4. *J Biol Chem* **246**, 925-932.

Fargeas CA, Corbeil D & Huttner WB (2003b). AC133 antigen, CD133, prominin-1, prominin-2, etc.: prominin family gene products in need of a rational nomenclature. *Stem Cells* **21**, 506-508.

Fargeas CA, Fonseca AV, Huttner WB et al. (2006). Prominin-1 (CD133) from progenitor cells to human diseases. *Future Lipidology* **2**, 213-225.

Fargeas CA, Huttner WB, Corbeil D (2007). Nomenclature of prominin-1 (CD133) splice variants – an update. *Tissue Antigens* **69**, 602-606.

Fargeas CA, Joester A, Missol-Kolka E, et al. (2004). Identification of novel prominin-1/CD133 splice variants with alternative C-termini and their expression in epididymis and testis. *Journal of Cell Science* **117**, 4301-4311.

Fey P, Gaudet P, Curk T, et al. (2009). dictyBase - Dictyostelium bioinformatics resource update. *Nucleic Acids Research* **37** (Suppl 1), D515–D519.

Figueroa DJ, Hess JF, Ky B, et al. (2000). Expression of the type I diabetes-associated gene LRP5 in macrophages, vitamin A system cells, and the Islets of Langerhans suggests multiple potential roles in diabetes. *J Histochem Cytochem* **48**, 1357-1368.

Finch PW, He X, Kelley MJ, et al. (1997). Purification and molecular cloning of a secreted, Frizzled-related antagonist of Wnt action. *Proc Natl Acad Sci U S A*, **94** (13), 6770-6775.

Fire A. & Xu, SQ (1995). Rolling replication of short DNA circles**.** *Proc Natl Acad Sci USA* **92**, 4641- 4645.

Flower DR (1996). The lipocalin protein family: structure and function. *Biochem J* **318** (Pt 1), 1-14.

Fodor SP, Read JL, Pirrung MC, et al (1991). Light-directed, spatially addressable parallel chemical synthesis. *Science* **251** (4995), 767-773.

Forrester WC, Dell M, Perens E, et al (1999). A C. elegans Ror receptor tyrosine kinase regulates cell motility and asymmetric cell division. *Nature* **400** (6747):881-885.

Franch-Marro X, Wendler F, Griffith J, et al (2008a). In vivo role of lipid adducts on Wingless. *J Cell Sci* **121**, 1587-1592.

Franch-Marro X, Wendler F, Guidato S, et al (2008b). Wingless secretion requires endosome-to- Golgi retrieval of Wntless/Evi/Sprinter by the retromer complex. *Nat Cell Biol* **10**, 170-177.

Freier SM, Kierzek R, Jaeger JA, et al (1986). Improved free-energy parameters for predictions of RNA duplex stability. *Proc Natl Acad Sci USA* **83**, 9373–9377.

Fresco JR, Alberts BM & Doty P (1960). Some molecular details of the secondary structure of ribonucleic acid. *Nature* **188**, 98–101.

Frohman MA (1990). PCR Protocols: A Guide to Methods and Applications, *Academic Press, San Diego,* 28.

Frohman MA, Dush MK, Martin GR (1988). Rapid production of full-length cDNAs from rare transcripts: amplification using a single gene-specific oligonucleotide primer. *Proc Natl Acad Sci USA* **85**, 8998–9002.

Fu J, Jiang M, Mirando AJ, et al. (2009). Reciprocal regulation of Wnt and Gpr177/mouse Wntless is required for embryonic axis formation. *Proc Natl Acad Sci U S A* **106**, 18598-18603.

Fujino T, Asaba H, Kang MJ, et al (2003). Lowdensity lipoprotein receptor-related protein 5 (LRP5) is essential for normal cholesterol metabolism and glucose-induced insulin secretion. *Proc Natl Acad Sci U S A* **100**, 229-234.

Funayama N, Fagotto F, McCrea P, et al. (1995). Embryonic axis induction by the armadillo repeat domain of beta-catenin: evidence for intracellular signaling. *J Cell Biol* **128,** 959–968.

Galli LM, Barnes TL, Secrest SS, et al. (2007). Porcupine-mediated lipid-modification regulates the activity and distribution of Wnt proteins in the chick neural tube. *Development* **134**, 3339-3348.

Gellert M. (1967). Formation of covalent circles of lambda DNA by E. coli extracts. *Proc Natl Acad Sci U S A* **57**, 148-55.

Gerstein AV, Almeida TA, Zhao G, et al. (2002). APC/CTNNB1 (beta-catenin) pathway alterations in human prostate cancers. *Genes Chromosomes Cancer* **34**, 9-16.

Gibson UE, Heid CA, Williams PM (1996). A novel method for real time quantitative RT-

PCR. *Genome Res* **6** (10),995-1001.

Giese K, Kingsley C, Kirshner JR, et al. (1995). Assembly and function of a TCR alpha enhancer complex is dependent on LEF-1-induced DNA bending and multiple protein-protein interactions. *Genes Dev* **9**, 995-1008.

Glinka A, Wu W, Delius H, et al. (1998). Dickkopf-1 is a member of a new family of secreted proteins and functions in head induction. *Nature* **391**(6665), 357-362.

Goardon N, Marchi E, Atzberger A, et al. (2011). Coexistence of LMPP-like and GMP-like leukemia stem cells in acute myeloid leukemia. *Cancer cell* **19** (1), 138–52.

Goemans BF, Zwaan CM, Miller M, et al. (2005). Mutations in KIT and RAS are frequent events in pediatric core-binding factor acute myeloid leukemia. *Leukemia* **19**, 1536-1542.

Goidin D, Mamessier A, Staquet MJ, et al. (2001). Ribosomal 18S RNA prevails over glyceraldehyde-3-phosphate dehydrogenase and beta-actin genes as internal standard for quantitative comparison of mRNA levels in invasive and noninvasive human melanoma cell subpopulations. *Anal Biochem* **295**, 17-21.

Golling G, Li L, Pepling M, et al. (1996). Drosophila homologs of the proto-oncogene

product PEBP2/CBF beta regulate the DNA-binding properties of Runt. *Mol Cell Biol* **16**, 932-942.

Goodman RM, Thombre S, Firtina Z, et al. (2006). Sprinter: a novel transmembrane protein required for Wg secretion and signaling. *Development* **133**, 4901-4911.

Graham TA, Weaver C, Mao F, et al. (2000). Crystal structure of a beta-catenin/TCF complex. *Cell* **103**, 885–896.

Gumbiner BM. (1995). Signal transduction of beta-catenin. *Curr Opin Cell Biol* **7**, 634-640.

Haferlach C, Dicker F, Kohlmann A, et al. (2010). AML with CBFB-MYH11 rearrangement demonstrate RAS pathway alterations in 92% of all cases including a high frequency of NF1 deletions. *Leukemia* **24**, 1065-1069.

Haferlach T, Winkemann M, Loffler H, et al. (1996). The abnormal eosinophils are part of the leukemic cell population in acute myelomonocytic leukemia with abnormal eosinophils (AML M4Eo) and carry the pericentric inversion 16: a combination of May-Grunwald-Giemsa staining and fluorescence in situ hybridization. *Blood* **87**, 2459-2463.

Haralick RM & Shapiro L (1992). Computer and robot vision, volume 1 and 2. Addison-Wesley, Reading, MA, USA.

Harris TW, Antoshechkin I, Bieri T, et al. (2010). Wormbase: a comprehensive resource for nematode research. *Nucleic Acids Research* **38** (Suppl 1), D463–D467.

Hart M, Concordet JP, Lassot I, et al. (1999). The F-box protein beta-TrCP associates with phosphorylated betacatenin and regulates its activityin the cell. *Curr Biol* **9**, 207–210.

Hayakawa F, Towatari M, Kiyoi H, et al. (2000). Tandem-duplicated Flt3 constitutively activates STAT5 and MAP kinase and introduces autonomous cell growth in IL-3-dependent cell lines. *Oncogene 19*, 624-631.

He X, Saint-Jeannet JP, Wang Y, et al (1997). A member of the Frizzled protein family mediating axis induction by Wnt-5A. *Science* **275**(5306), 1652-1654.

Heid CA, Stevens J, Livak KJ, et al (1996). Real time quantitative PCR. *Genome Res* **6**(10), 986-994.

Heidel FH, Bullinger L, Feng Z, et al. (2012). Genetic and pharmacologic inhibition of beta-catenin targets imatinib-resistant leukemia stem cells in CML. *Cell Stem Cell* **10,** 412–424.

Herman MA, Vassilieva LL, Horvitz HR, et al. (1995). The C. elegans gene lin-44, which controls the polarity of certain asymmetric cell divisions, encodes a Wnt protein and acts cell nonautonomously. *Cell* **83**, 101–110.

Hernandez JM, Gonzalez MB, Granada I, et al. (2000). Detection of inv(16) and t(16;16) by fluorescence in situ hybridization in acute myeloid leukemia M4Eo. *Haematologica* **85**, 481-485.

Hertweck MK, Erdfelder F & Kreuzer KA (2011). CD44 in hematological neoplasias. *Ann Hematol* **90**, 493–508.

Hey PJ, Twells RC, Phillips MS, et al. (1998). Cloning of a novel member of the low-density lipoprotein receptor family. *Gene* **216**, 103-111.

Higgins NP, & Cozzarelli NR (1979). DNA-joining enzymes: a review. *Methods Enzymol* **68**, 50-71.

Hoffmans R, Sta¨deli R, Basler K (2005). Pygopus and legless provide essential transcriptional coactivator functions to armadillo/betacatenin. *Curr Biol* **15**, 1207–1211.

Hogan CJ, Shpall EJ & Keller G (2002). Differential long-term and multilineage engraftment potential from subfractions of human CD34+ cord blood cells transplanted into NOD/SCID mice. *Proc Natl Acad Sci USA* **99**, 413–418.

Holland PM, Abramson RD, Watson R et al. (1991). Detection of specific polymerase chain reaction product by utilizing the 5'→3' exonuclease activity of Thermus aquaticus

DNA polymerase. *Proc Natl Acad Sci USA* **88**, 7276–7280.

Holmen SL, Salic A, Zylstra CR, et al. (2002). A novel set of Wnt-Frizzled fusion proteins identifies receptor components that activate beta -catenin-dependent signaling. *J Biol Chem* 277, 34727-34735.

Hope KJ, Jin L & Dick JE (2004). Acute myeloid leukemia originates from a hierarchy of leukemic stem cell classes that differ in self-renewal capacity. *Nat Immunol* **5**(7), 738-43.

Houston DW, & Wylie C (2002). Cloning and expression of Xenopus Lrp5 and Lrp6 genes. *Mech Dev* **117**, 337-342.

Hsieh JC, Kodjabachian L, Rebbert ML, et al (1999). A new secreted protein that binds to Wnt proteins and inhibits their activities. *Nature* **398** (6726), 431-436.

Huang S, Bjornsti MA & Houghton PJ (2003). Rapamycins: mechanism of action and cellular resistance. *Cancer Biol Ther* **2**, 222-232.

Huang DW, Sherman BT & Lempicki RA (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* **4** (1), 55–57.

Huang G, Shigesada K, Ito K, et al (2001). Dimerization with PEBP2beta protects RUNX1/AML1 from ubiquitin-proteasome-mediated degradation. *Embo J* **20**, 723-733.

Huang S & Houghton PJ (2001). Mechanisms of resistance to rapamycins. *Drug Resist Updat* **4**(6), 378-91.

Huber AH & Weis WI (2001). The structure of the beta-catenin/Ecadherin complex and the molecular basis of diverse ligand recognition by beta-catenin. *Cell* **105**, 391–402.

Huber AH, Nelson WJ & Weis WI (1997). Three-dimensional structure of the armadillo repeat region of beta-catenin. *Cell* **90**, 871–882.

Huber O, Korn R, McLaughlin J, et al. (1996). Nuclear localization of beta-catenin by interaction with transcription factor LEF-1. *Mech Dev* **59**, 3–10.

Huisken J, Swoger J, Del Bene F, et al. (2004). Optical sectioning deep inside live embryos by selective plane illumination microscopy. *Science* **305**,1007 -1009.

Huisken J, Swoger J, Lindek S et al (2006). Handbook of Biological Confocal Microscopy , chapter Selective Plane Illumination Microscopy. Springer Verlag, Berlin, Germany, 2006.

Hunter S, Apweiler R, Attwood TK et al. (2009). InterPro: the integrative protein signature database. *Nucleic Acids Res.* **37**; D211-5.

Huntly BJ & Gilliland DG (2005). Leukaemia stem cells and the evolution of cancer-stem-cell research. *Nat Rev Cancer* **5**(4), 311-21.

Hur EM & Zhou FQ (2010). GSK3 signalling in neural development. *Nat Rev Neurosci* **11**, 539–551.

Hussain SZ, Sneddon T, Tan X, et al. (2004). Wnt impacts growth and differentiation in ex vivo liver development. *Exp Cell Res* **292**(1), 157-169.

Huttner HB, Janich P, Köhrmann M et al. (2008). The stem cell marker prominin-1/CD133 on membrane particles in human cerebrospinal fluid offers novel approaches for studying CNS disease. *Stem Cells* **26**, 698-705.

Ikeda H, Kanakura Y, Tamaki T, et al. (1991). Expression and functional role of the protooncogene c-kit in acute myeloblastic leukemia cells. *Blood* **78**, 2962-2968.

Ikuta K & Weissman IL. (1992). Evidence that hematopoietic stem cells express mouse c-kit but do not depend on steel factor for their generation. *Proc Natl Acad Sci USA* **89**, 1502–1506.

Illmer T, Schaich M, Ehninger G, et al. (2007). Tyrosine kinase mutations of JAK2 are rare events in AML but influence prognosis of patients with CBF-leukemias. *Haematologica* **92**, 137-138.

Irizarry RA, Hobbs B, Collin F, et al. (2003). Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatics* **4**(2), 249-64.

Jackson DA, Symons RH & Berg P (1972). Biochemical method for inserting new genetic information into DNA of Simian Virus 40: circular SV40 DNA molecules containing lambda phage genes and the galactose operon of Escherichia coli. *Proc Natl Acad Sci U S A* **69**, 2904-9.

Jaeger JA, Turner DH & Zuker M (1990). Predicting optimal and suboptimal secondary structure for RNA. *Methods Enzymol* **183**, 281–306.

Jaeger JA, Turner DH & Zuker M (1989). Improved predictions of secondary structures for RNA. *Proc Natl Acad Sci USA* **86**, 7706–7710.

Jaffe ES (2009). The 2008 WHO classification of lymphomas: implications for clinical practice and translational research. *Hematology Am Soc Hematol Educ Program*, 523-531.

Jaiswal P, Ni J, Yap I, et al. (2006). Gramene: a bird's eye view of cereal genomes. *Nucleic Acids Research* **34**, D717–723.

Jin L, Hope KJ, Zhai Q, et al. (2006). Targeting of CD44 eradicates human acute myeloid leukemic stem cells. *Nat Med* **12** (10), 1167–74.

Jing SQ & Trowbridge IS (1987). Identification of the intermolecular disulfide bonds of the human transferrin receptor and its lipid-attachment site. *EMBO J* **6**, 327-331.

Johnson SC (1967). Hierarchical clustering schemes. *Psychometrika* **32** (3), 241-254.

Jones SE & Jomary C (2002). Secreted Frizzled-related proteins: searching for relationships and patterns. *Bioessays* **24** (9), 811-820.

Jones TR, Carpenter AE & Golland P (2005). Voronoi-based segmentation of cells on image manifolds. Proceedings of the Workshop on Computer Vision for Biomedical Image Applications (CVBIA). *Lecture Notes in Computer Science 3765*. Published by Springer-Verlag, Berlin, p. 535-543.

Jones TR, Carpenter AE, Lamprecht MR et al. (2009). Scoring diverse cellular morphologies in image-based screens with iterative feedback and machine learning. *Proc Natl Acad Sci USA* **106**, 1826-1831.

Jones TR, Carpenter AE, Sabatini DM, et al. (2006). Methods for high-content, high-throughput image-based cell screening. *Proceedings of the Workshop on Microscopic Image Analysis with Applications in Biology (MIAAB).* Metaxas DN, Whitaker RT, Rittcher J, Sebastian T (Eds). Copenhagen, Denmark, October 5, pp 65-72.

Jones TR, Kang IH, Wheeler DB, et al. (2008). CellProfiler Analyst: data exploration and

analysis software for complex image-based screens. *BMC Bioinformatics* **9** (1), 482.

Jordan CT (2007). The leukemic stem cell. Best Pract Res Clin Haematol 20: 13–18.

Jordan CT, (2006). Searching for leukemia stem cells--not yet the end of the road? *Cancer Cell* **10**(4), 253-254.

Kallioniemi OP, Wagner U, Kononen J et al. (2001). Tissue microarray technology for high-throughput molecular profiling of cancer. *Hum Mol Genet* **10**, 657-662.

Kamachi Y, Ogawa E, Asano M, et al. (1990). Purification of a mouse nuclear factor that binds to both the A and B cores of the polyomavirus enhancer. *J Virol* **64**, 4808-4819.

Kania G, Corbeil D, Fuchs J, et al. (2005). Somatic stem cell marker prominin- 1/CD133 is expressed in embryonic stem cell-derived progenitors. *Stem Cells* **23**, 791-804.

Kapuscinski J (1995). DAPI: a DNA-specific fluorescent probe. *Biotechnic histochemistry official publication of the Biological Stain Commission* **70** (5), 220–233.

Karbanová J, Missol-Kolka E, Fonseca AV, et al. (2008). The stem cell marker CD133 (Prominin-1) is expressed in various human glandular epithelia. *J Histochem Cytochem* **56**, 977-993.

Kashyap RL, Koivo AJ, Mendel JM et al. (1986). In Memoriam King Sun Fu (1930–1985).

IEEE Transactions on Automatic Control **31** (4), 290.

Katanaev VL, Solis GP, Hausmann G, et al. (2008). Reggie-1/"otillin-2 promotes secretion of the long-range signalling forms of Wingless and Hedgehog in Drosophila. *EMBO J* **27**, 509-521.

Kaufmann BB & van Oudenaarden A (2007). Stochastic gene expression: from single molecules to the proteome. *Curr Opin Genet Dev* **17**, 107-112.

Kawano Y & Kypta R (2003). Secreted antagonists of the Wnt signalling pathway. *J Cell Sci*, **116** (Pt 13), 2627-2634.

Kemper K, Sprick MR, de Bree M, et al. (2010). The AC133 epitope, but not the CD133 protein, is lost upon cancer stem cell differentiation. *Cancer Res* **70**, 719–729.

Kim DH, Inagaki Y, Suzuki T, et al. (1998). A new low density lipoprotein receptor related protein, LRP5, is expressed in hepatocytes and adrenal cortex, and recognizes apolipoprotein. *E J Biochem (Tokyo)* **124**, 1072-1076.

Kim HG, Kojima K, Swindle CS, et al. (2008). FLT3-ITD cooperates with inv(16) to promote progression to acute myeloid leukemia. *Blood* **111**, 1567-1574.

Kim KT, Baird K, Ahn JY, et al. (2005). Pim-1 is up-regulated by constitutively activated FLT3 and plays a role in FLT3-mediated cell survival. *Blood* **105**, 1759-1767.

Kimelman D & Xu W (2006). Beta-catenin destruction complex: insights and questions from a structural perspective. *Oncogene* **25,** 7482–7491.

Kinzler KW & Vogelstein B (1997). Cancer-susceptibility genes. Gatekeepers and caretakers. *Nature* **386**, 761, 763.

Kirikoshi H, Katoh M (2002). Expression and regulation of WNT10B in human cancer: up-regulation of WNT10B in MCF-7 cells by beta-estradiol and down-regulation of WNT10B in NT2 cells by retinoic acid. *Int J Mol Med* **10**(4), 507-11.

Kitamura T, Matsuoka Y, Kimura T, et al. (2010). In vivo dynamic of human cord-blood derived CD34– SCID-repopulating cells using intra-bone marrow injection. Leukemia 24:162–168.

Kitamura T, Matsuoka Y, Kimura T, et al. (2010). In vivo dynamic of human cord-blood derived CD34-SCID- repopulating cells using intra-bone marrow injection. *Leukemia* **24**, 162-168.

Kittler J & Illingworth J (1986). Minimum error thresholding. *Pattern Recognition*, **19** (1), 41–47.

Klar TA, Engel E & Hell S (2001). W. Breaking abbe's diffraction resolution limit in fluorescence microscopy with stimulated emission depletion beams of various shapes. *Physical Review*, 64(066613).

Klaus A & Birchmeier W (2008). Wnt signalling and its impact on development and cancer. *Nat Rev Cancer* **8**, 387–398.

Kleppe K, Van de Sande JH & Khorana HG (1970). Polynucleotide ligase-catalyzed joining of deoxyribo-oligonucleotides on ribopolynucleotide templates and of ribooligonucleotides on deoxyribopolynucleotide templates. *Proc Natl Acad Sci U S A* **67** (1), 68-73.

Komekado H, Yamamoto H, Chiba T, et al. (2007). Glycosylation and palmitoylation of Wnt-3a are coupled to produce an active form of Wnt-3a. *Genes Cells* **12**, 521-534.

Kononen J, Bubendorf L, Kallioniemi A, et al. (1998). Tissue microarrays for high-throughput molecular profiling of tumor specimens. *Nat Med* **4**, 844-847.

Kosodo Y, Röper K, Haubensak W, et al. (2004). Asymmetric distribution of the apical plasma membrane during neurogenic divisions of mammalian neuroepithelial cells. *EMBO J* **23**, 2314-2324.

Kramps T, Peter O, Brunner E, et al. (2002). Wnt/wingless signaling requires BCL9/legless-mediated recruitment of pygopus to the nuclear b-catenin–TCF complex. *Cell* **109**: 47–60.

Krupnik VE, Sharp JD, Jiang C, et al. (1999). Functional and structural diversity of the human Dickkopf gene family. *Gene* **238** (2), 301-313.

Kurayoshi M, Yamamoto H, Izumi S, et al. (2007). Post-translational palmitoylation and glycosylation of Wnt-5a are necessary for its signalling. *Biochem J* **402**, 515-523.

Kurayoshi M, Yamamoto H, Izumi S, et al. (2007). Post-translational palmitoylation and glycosylation of Wnt-5a are necessary for its signalling. *Biochem J* **402**, 515-523.

Kurayoshi M, Yamamoto H, Izumi S, et al. (2007). Post-translational palmitoylation and glycosylation of Wnt-5a are necessary for its signalling. *Biochem J* **402**, 515-523.

Lamprecht MR, Sabatini DM & Carpenter AE (2007). CellProfiler: free, versatile software for automated biological image analysis. *Biotechniques* **42** (1), 71-75.

Landegren U, Kaiser R, Sanders J et al. (1998). A ligase-mediated gene detection technique. *Science* **241**, 1077-1080.

Lantuéjoul C & Beucher S (1981). On the use of geodesic metric in image analysis. *Journal of Microscopy*, **121**, 39–49.

Lapidot T, Sirard C, Vormoor J, et al. (1994). A cell initiating human acute myeloid leukaemia after transplantation into SCID mice. *Nature* **367**(6464), 645–648.

Larsson C, Grundberg I, Soderberg O. et al. (2010). In situ detection and genotyping of individual mRNA molecules. *Nat Methods* **7**, 395-397.

Larsson C, Koch J, Nygren A et al (2004). In situ genotyping individual DNA molecules by target-primed rolling-circle amplification of padlock probes. *Nat Methods* **1** (3), 227-232.

Lee LG, Connell CR & Bloch W (1993). Allelic discrimination by nick-translation PCR with fluorogenic probes. *Nucleci Acid Res* **21**(16), 3761-3766.

Lehman IR (1974). DNA ligase: structure, mechanism, and function. *Science* **186,** 790-797.

Levine M & Tjian R (2003). Transcription regulation and animal diversity. *Nature* **10**, 424(6945), 147-151. Review.

Levsky JM & Singer RH (2003). Gene expression and the myth of the average cell. *Trends Cell Biol* **13**, 4-6.

Ley TJ, Ding L, Walter MJ, et al. (2010). DNMT3A mutations in acute myeloid leukemia. *N Engl J Med* **363**(25), 2424-33.

Leyns L, Bouwmeester T, Kim SH, et al. (1997). Frzb-1 is a secreted antagonist of Wnt signaling expressed in the Spemann organizer. *Cell* **88** (6), 747-756.

Liang X, Nazarian A, Erdjument-Bromage H, et al. (2001). Heterogeneous fatty acylation of Src family kinases with polyunsaturated fatty acids regulates raft localization and signal transduction. *J Biol Chem* **276**, 30987-30994.

Lin K, Wang S, Julius MA, et al. (1997). The cysteine-rich frizzled domain of Frzb-1 is required and sufficient for modulation of Wnt signaling. *Proc Natl Acad Sci U S A* **94** (21), 11196-11200.

Lin X & Perrimon N. (1999). Dally cooperates with Drosophila frizzled 2 to transduce Wingless signalling. *Nature* **400,** 281–284.

Liu C, Li Y, Semenov M, et al. (2002). Control of beta-catenin phosphorylation/degradation by a dual-kinase mechanism. *Cell* **108**, 837–847.

Liu D, Daubendiek SL, Zillman MA, et al. (1996) Rolling circle DNA synthesis: Small circular oligonucleotides as efficient templates for DNA polymerases. *J Am Chem Soc* **118**, 1587–1594.

Liu G, Yuan X, Zeng Z, et al. (2006). Analysis of gene expression and chemoresistance of CD133+cancer stem cells in glioblastoma. *Mol Cancer*, **5**, 67.

Liu P, Tarle SA, Hajra A, et al. (1993). Fusion between transcription factor CBF beta/PEBP2 beta and a myosin heavy chain in acute myeloid leukemia. *Science* **261**, 1041-1044.

Liu, A (2010). Laser capture microdissection in the tissue biorepository. *J Biomol Tech* **21**, 120-125.

Livak KJ, Flood SJ, Marmaro J, et al. (1995). Oligonucleotides with fluorescent dyes at opposite ends provide a quenched probe system useful for detecting PCR product and nucleic acid hybridization. *PCR Methods Applic*. **4**, 357-362.

Lizardi PM, Huang X, Zhu Z, et al. (1998). Mutation detection and single-molecule counting using isothermal rolling-circle amplification. *Nat Genet* **19**, 225-232.

Logan CY, & Nusse R (2004). The Wnt signaling pathway in development and disease. *Annu Rev Cell Dev Biol* **20**, 781-810.

Loh E. (1991). Anchored PCR: amplification with single-sided specificity. *Methods Companion Methods Enzymol* **2**, 11–19.

Look AT (1997). Oncogenic transcription factors in the human acute leukemias. *Science*

278, 1059-1064.

Lu FI, Thisse C & Thisse B (2011). Identification and mechanism of regulation of the zebrafish dorsal determinant. *Proc Natl Acad Sci USA* **108**, 15876–15880.

Lu W, Yamamoto V, Ortega B, et al (2004). Mammalian Ryk is a Wnt coreceptor required for stimulation of neurite outgrowth. Cell **119** (1), 97-108.

Luo J, Bergstrom DE. & Barany F (1996). Improving the fidelity of Thermus thermophilus DNA ligase. *Nucleic Acids Res* **24** (15), 3071-3078.

MacDonald BT, Tamai K, He X (2009). Wnt/β-catenin signaling: components, mechanisms, and diseases. *Dev Cell* **17**(1), 9–26.

Magoori K, Kang MJ, Ito MR, et al. (2003). Severe hypercholesterolemia, impaired fat tolerance, and advanced atherosclerosis in mice lacking both low density lipoprotein receptor-related protein 5 and apolipoprotein E. *J Biol Chem* **278**, 11331-11336.

Majeti R, Becker MW, Tian Q, et al. (2009). Dysregulated gene expression networks in human acute myelogenous leukemia stem cells. *Proc Natl Acad Sci USA* **106**, 3396–3401.

Malpica N, De Solórzano CO, Vaquero JJ, et al (1997). Applying watershed algorithms to the segmentation of clustered nuclei. *Cytometry* **28** (4), 289–297.

Mao B, Wu W, Davidson G, et al. (2002). Kremen proteins are Dickkopf receptors that regulate Wnt/beta-catenin signalling. *Nature* **417**, 664-667.

Marcucci G, Mrozek K, Ruppert AS, et al. (2005). Prognostic factors and outcome of core binding factor acute myeloid leukemia patients with t(8;21) differ from those of patients with inv(16): a Cancer and Leukemia Group B study. *J Clin Oncol* **23,** 5705-5717.

Mardis ER, Ding L, Dooling DJ, et al. (2009). Recurring mutations found by sequencing an acute myeloid leukemia genome. *Engl J Med*, **361**, 1058-1066.

Marzesco AM, Janich P, Wilsch-Bräuninger M, et al. (2005). Release of extracellular membrane particles carrying the stem cell marker prominin-1 (CD133) from neural progenitors and other epithelial cells. *J Cell Sci* **118**, 2849-2858.

Marzesco AM, Wilsch-Bräuninger M, Dubreuil V, et al. (2009). Release of extracellular membrane vesicles from microvilli of epithelial cells is enhanced by depleting membrane cholesterol. *FEBS Lett*. **583**, 897-902.

Mathews DH, Andre TC, KimJ, et al. (1998). An updated recursive algorithm for RNA

secondary structure prediction with improved free energy parameters. chapter 15. In Leontis,N.B. and SantaLucia,J.,Jr (eds), American Chemical Society Symposium Series 682, American Chemical Society Washington, DC, pp. 246–257.

Mathews DH, Turner DH & Zuker M (2000). RNA secondary structure prediction. chapter 11.2. In Beaucage,S., Bergstrom,D.E., Glick,G.D. and Jones,R.A. (eds), Current Protocols in Nucleic Acid Chemistry, John Wiley & Sons New York, NY, 1–10.

Mathews DH, SabinaJ, Zuker M et al. (1999). Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. *J Mol Biol* **288**, 911–940.

MATLAB. version 7.5.0 (R2007b). The MathWorks Inc., Natick, Massachusetts, 2007.

Maurer CR, Qi R, & Raghavan V (2003). A Linear Time Algorithm for Computing Exact Euclidean Distance Transforms of Binary Images in Arbitrary Dimensions. *IEEE Transactions on Pattern Analysis and Machine Intelligence* , **25**(2), 265–270.

McCubrey JA, Steelman LS, Chappell WH, et al. (2007). Roles of the Raf/MEK/ERK pathway in cell growth, malignant transformation and drug resistance. *Biochim Biophys Acta* **1773**, 1263-1284.

McDermott SP, Eppert K, Lechman ER, et al. (2010). Comparison of human cord blood engraftment between immunocompromised mouse strains. *Blood* **116**, 193–200.

McKenzie JL, Gan OI, Doedens M, et al. (2006). Individual stem cells with highly variable proliferation and selfrenewal properties comprise the human hematopoietic stem cell compartment. *Nat Immunol* **7**, 1225–1233.

Meijering E, Smal I, Dzyubachyk O, et al. (2008). Microscope Image Processing-Time-Lapse Imaging, Q. Wu, F. A. Merchant, K. R. Castleman eds. *Elsevier Academic Press*.

Messina DN, Glasscock J, Gish W et al. (2004). An ORFeome-based analysis of human transcription factor genes and the construction of a microarray to interrogate their expression. *Genome Research* **14,** 2041–2047.

Meyer F & Beucher S (1990). Morphological segmentation. *J Visual Communication Image Representation* **1**, 21-46.

Miharada K & Karlsson S (2012). Common signaling networks characterize leukemia-initiating cells in acute myeloid leukemia. *Cell Stem Cell* **10** (2), 109-110

Mikels AJ &Nusse R (2006). Purified Wnt5a protein activates or inhibits beta-catenin-TCF signaling depending on receptor context. *PLoS Biol* **4** (4), e115.

Miller CL & Eaves CJ (1997). Expansion in vitro of adult murine hema- topoietic stem cells with transplantable lympho-myeloid reconstituting ability. *Proc Natl Acad Sci USA* **94**, 13648–13653.

Miller JR. (2002). The Wnts. *Genome Biol3*, REVIEWS3001.

Miraglia S, Godfrey W, Yin AH et al. (1997). A novel five-transmembrane hematopoietic stem cell antigen: Isolation, characterization, and molecular cloning. *Blood* **90**, 5013-5021.

Misselwitz B, Strittmatter G, Periaswamy B, et al. (2010). Enhanced CellClassifier: a multi-class classification tool for microscopy images. *BMC Bioinformatics* **11**, 30.

Mizrak D, Brittan M & Alison MR (2007). CD133: molecule of the moment. *J Pathol* **214**, 3–9.

Mizuki M, Fenski R, Halfter H, et al. (2000). Flt3 mutations from patients with acute myeloid leukemia induce transformation of 32D cells mediated by the Ras and STAT5 pathways. *Blood* **96**,3907-3914.

Moffett S, Brown DA & Linder ME (2000). Lipid-dependent targeting of G proteins into rafts. *J Biol Chem* **275**, 2191-2198.

Molenaar M, van de Wetering M, Oosterwegel M, et al. (1996). XTCF-3 transcription

factor mediates beta-catenin-induced axis formation in Xenopus embryos. *Cell* **86**, 391–399.

Moore MAS, Dorn DC, Schuringa JJ, et al. (2007). Constitutive activation of Flt3 and STAT5A enhances self-renewal and alters differentiation of hematopoietic stem cells. *Exp Hematol* **35**, 105-116.

Morozova O, Hirst M & Marra MA (2009). Applications of new sequencing technologies for transcriptome analysis. *Annu Rev Genomics Hum Genet* **10,** 135-151.

Morrison SJ & Weissman IL (1994). The long-term repopulating subset of hematopoietic stem cells is deterministic and isolatable by phenotype. *Immunity* **1**, 661-673.

Morrison SJ, Uchida N & Weissman IL (1995). The biology of hematopoietic stem cells. *Annu Rev Cell Dev Biol* **11**, 35-71.

Mosimann C, Hausmann G & Basler K (2009). Beta-catenin hits chromatin: Regulation of Wnt target gene activation. *Nat Rev Mol Cell Biol* **10**, 276–286.

Mrózek K, Marcucci G, Paschka P, et al. (2007). Clinical relevance of mutations and gene-expression changes in adult acute myeloid leukemia with normal cytogenetics: are we ready for a prognostically prioritized molecular classification? *Blood* **109**, 431- 448.

Nilsson M, Malmgren H, Samiotaki M, et al. (1994). Padlock probes: circularizing oligonucleotides for localized DNA detection. *Science* **265**, 2085-2088.

Ning ZQ, Li J & Arceci RJ (2001a). Signal transducer and activator of transcription 3 activation is required for Asp(816) mutant c-Kit-mediated cytokine-independent survival and proliferation in human leukemia cells. *Blood* **97**, 3559-3567.

Nishita M, Yoo SK, Nomachi A, et al. (2006). Filopodia formation mediated by receptor tyrosine kinase Ror2 is required for Wnt5a-induced cell migration. *J Cell Biol* **175** (4), 555-562.

Noonberg SB, Weiss TL, Garovoy MR, et al. (1992). Characterization and minimization of cellular autofluorescence in the study of oligonucleotide uptake using confocal microscopy. *Antisense Res Dev* **2**, 303–333.

Nusse R & Varmus HE (1982). Many tumors induced by the mouse mammary tumor virus contain a provirus integrated in the same region of the host genome. *Cell* **31**, 99-109.

Nusse R, Fuerer C, Ching W, et al. (2008). Wnt signaling and stem cell control. *Cold Spring Harb Symp Quant Biol* **73**, 59-66.

Nussinov R & Jacobson AB (1980). Fast algorithm for predicting the secondary structure of single-stranded RNA. *Proc Natl Acad Sci USA* **77**, 6309–6313.

Nussinov R, Pieczenik G, Griggs JR et al. (1978). Algorithm for loop matchings. SIAM *J. Appl. Math* **35**, 68–82.

Nusslein-Volhard C & Dahm R (2002). Zebrafish: A Practial Approach.

Nusslein-Volhard C & Wieschaus E (1980). Mutations affecting segment number and polarity in Drosophila. *Nature* **287**, 795-801.

Oishi I, Suzuki H, Onishi N, et al. (2003). The receptor tyrosine kinase Ror2 is involved in non-canonical Wnt5a/JNK signalling pathway. *Genes Cells* **8**(7), 645-654.

Oliphant TE (2007). Python for scientific computing. *Comput Sci Eng* **9**, 10-20.

Osawa M, Hanada K, Hamada H, et al .(1996). Long-term lymphohematopoietic reconstitution by a single CD34-low/negative hematopoietic stem cell. *Science* **273**, 242-245.

Otsu N (1979). A threshold selection method from gray-level histograms. *IEEE Transactions on System, Man and Cybernetics* **9** (1), 62–69.

Ozawa M, Baribault H & Kemler R (1989). The cytoplasmic domain of the cell adhesion molecule uvomorulin associates with three independent proteins structurally related in different species. *EMBO J* **8**, 1711–1717.

Ozsolak F & Milos PM (2011). RNA sequencing: advances, challenges and opportunities. *Nat Rev Genet* **12**, 87-98.

Pan CL, Baum PD, Gu M, Jorgensen, et al. (2008). C. elegans AP-2 and retromer control Wnt signaling by regulating mig-14/Wntless. *Dev Cell* **14**, 132-139.

Pardal R, Clarke MF & Morrison SJ (2003). Applying the principles of stem-cell biology to cancer. *Nat Rev Cancer* **3** (12), 895-902.

Parker DS, Jemison J & Cadigan KM (2002). Pygopus, a nuclear PHD-finger protein required for Wingless signaling in Drosophila. *Development* **129**, 2565–2576.

Paschka P (2008). Core Binding Factor Acute Myeloid Leukemia. *Semin Oncol* **35**, 410-

417.

Petersen M & Wengel J (2003). LNA: a versatile tool for therapeutics and genomics. *Trends Biotechnol* **21**, 74-81.

Pinson KI, Brennan J, Monkley S, et al. (2000). An LDL-receptor-related protein mediates Wnt signalling in mice. *Nature* **407**, 535-538.

Pirc-Danoewinata H, Dauwerse HG, Konig M, et al. (2000). CBFB/MYH11 fusion in a patient with AML-M4Eo and cytogenetically normal chromosomes 16. *Genes Chromosomes Cancer* **29**, 186-191.

Pollard JA, Alonzo TA, Gerbing RB, et al. (2010). Prevalence and prognostic significance of KIT mutations in pediatric patients with core binding factor AML enrolled on serial pediatric cooperative trials for de novo AML. *Blood* **115**, 2372-2379.

Port F & Basler K (2010). Wnt trafficking: new insights into Wnt maturation, secretion and spreading. *Traffic* **11**, 1265-1271.

Port F, Kuster M, Herr P, et al (2008). Wingless secretion promotes and requires retromer-dependent cycling of Wntless. *Nat Cell Biol* **10**, 178–185.

Poy F, Lepourcelet M, Shivdasani RA, et al. (2001). Structure of a human TCF4-beta-catenin complex. *Nat Struct Biol* **8,** 1053–1057.

Pozhitkov AE, Tautz D & Noble PA (1997). Oligonucleotide microarrays: widely applied poorly understood. *Brief Funct Genomic Proteomic* **6**, 141-148.

Pritchard CE & Southern EM (1997). Effects of base mismatches on joining of short oligodeoxynucleotides by DNA ligases. *Nucleic Acids Res* **25** (17), 3403-3407.

Quere' R, Andradottir S, Brun AC, et al. (2011). High levels of the adhesion molecule CD44 on leukemic cells generate acute myeloid leukemia relapse after withdrawal of the initial transforming event. *Leukemia* **25,** 515–526.

Raj A, Peskin CS, Tranchina D, et al. (2006). Stochastic mRNA synthesis in mammalian cells. *PLoS Biol* **4**, e309.

Ramo P. Sacher R, Snijder B, et al. (2009). CellClassifier: supervised learning of cellular phenotypes. *Bioinformatics* **25**, 3028-3030.

Rattner A, Hsieh JC, Smallwood PM, et al. (1997). A family of secreted proteins contains homology to the cysteine-rich ligand-binding domain of frizzled receptors. *Proc Natl Acad Sci U S A ,* **94** (7), 2859-2863.

Reuther GW (2008). JAK2 activation in myeloproliferative neoplasms: a potential role for heterodimeric receptors. *Cell Cycle* **7**, 714-719.

Reya T, Morrison SJ, Clarke MF et al. (2001). L. Stem cells, cancer, and cancer stem cells. *Nature* **414**, 105-111.

Riggleman B, Schedl P & Wieschaus E (1990). Spatial expression of the Drosophila segment polarity gene armadillo is posttranscriptionally regulated by wingless. *Cell* **63**, 549–560.

Rittscher J, Machiraju R, & Wong STC (2008). Microscopic image analysis for life science applications . *Artech House bioinformatics & biomedical imaging series*. Artech House.

Rizo A, Vellenga E & de Haan G (2006). Signaling pathways in self-renewing hematopoietic and leukemic stem cells: do all stem cells need a niche? *Hum Mol Genet* **15** Spec No 2: R210-9.

Roberts DM, Pronobis MI, Poulton JS, et al. (2011). Deconstructing the β-catenin destruction complex: mechanistic roles for the tumor suppressor APC in regulating Wnt signaling. *Mol Biol Cell* **22**, 1845–1863.

Rose JK, Adams GA & Gallione CJ (1984). The presence of cysteine in the cytoplasmic domain of the vesicular stomatitis virus glycoprotein is required for palmitate addition.

Proc Natl Acad Sci U SA **81**, 2050-2054.

Rosenfeld A (1969). Picture processing by computers. *Academic Press, New York*.

Rossi DJ, Jamieson CH, & Weissman IL (2008). Stems cells and the pathways to aging and cancer. *Cell* **132**, 681–696.

Rossi R, Montecucco A, Ciarrocchi G et al. (1997). Functional characterization of the T4 DNA ligase: a new insight into the mechanism of action. *Nucleic Acids Res* **25** (11), 2106-13.

Roth W, Wild-Bode C, Platten M et al. (2000). Secreted Frizzled-related proteins inhibit motility and promote growth of human malignant glioma cells. *Oncogene* 2000, **19** (37), 4210-4220.

Rowley JD (1973). Identificaton of a translocation with quinacrine fluorescence in a patient with acute leukemia. *Ann Genet* **16**, 109-112.

Rubinfeld B, Albert I, Porfiri E, et al. (1996). Binding of GSK3beta to the APC-beta-catenin complex and regulation of complex assembly. *Science* **272**, 1023-1026.

Ruiz-Herguido C, Guiu J, D'Altri TJ, et al. (2012). Hematopoietic stem cell development

requires transientWnt/beta-catenin activity, *J Exp Med* **209,** 1457–1468.

Saito T, Oda Y, Tanaka K, et al. (2001). Beta-catenin nuclear expression correlates with cyclin D1 overexpression in sporadic desmoid tumours. *J Pathol* **195**, 222-228.

Salser W (1977). Globin mRNA sequences: analysis of base pairing and evolutionary implications. *Cold Spring Harbor Symp Quant Biol* **42**, 985–1002.

Sankoff D, Kruskal JB, Mainville S et al. (1983). Fast algorithms to determine RNA secondary structures containing multiple loops. chapter 3. In Sankoff,D. and Kruskal,J.B. (eds), *Time Warps, String Edits, and Macromolecules: The Theory and Practice of Sequence Comparison* Addison-Wesley Reading, MA, pp. 93–120.

Sarry JE, Murphy K, Perry R, et al. (2011). Human acute myelogenous leukemia stem cells are rare and heterogeneous when assayed in NOD/SCID/IL2Rγc-deficient mice. *J Clin Invest* **121,** 384–395.

Satoh S, Daigo Y, Furukawa Y, et al. (2000). AXIN1 mutations in hepatocellular carcinomas, and growth suppression in cancer cells by virus-mediated transfer of AXIN1. *Nat Genet* **24**, 245-250.

Sawa H, Lobel L & Horvitz HR (1996). The Caenorhabditis elegans gene lin-17, which is required for certain asymmetric cell divisions, encodes a putative seven-transmembrane

protein similar to the Drosophila frizzled protein. *Genes Dev* **10**, 2189–2197.

Schessl C, Rawat VP, Cusan M, et al. (2005b). The AML1-ETO fusion gene and the FLT3 length mutation collaborate in inducing acute leukemia in mice. *J Clin Invest* **115**, 2159-2168.

Schnittger S, Bacher U, Haferlach C, et al. (2007a). Rare CBFB-MYH11 fusion transcripts in AML with inv(16)/t(16;16) are associated with therapy-related AML M4eo, atypical cytomorphology, atypical immunophenotype, atypical additional chromosomal rearrangements and low white blood cell count: a study on 162 patients. *Leukemia* **21**, 725-731.

Schnittger S, Dicker F, Kern W, et al. (2011). RUNX1 mutations are frequent in de novo AML with noncomplex karyotype and confer an unfavorable prognosis. *Blood* **117**, 2348-2357.

Schnittger S, Schoch C, Dugas M, et al. (2002). Analysis of FLT3 length mutations in 1003 patients with acute myeloid leukemia: correlation to cytogenetics, FAB subtype, and prognosis in the AMLCG study and usefulness as a marker for the detection of minimal residual disease. *Blood* **100**, 59-66.

Scholl C, Gilliland DG & Frohling S (2008a). Deregulation of signaling pathways in acute myeloid leukemia. *Semin Oncol* **35**, 336-345.

Schweizer L & Varmus H (2003). Wnt/Wingless signaling through beta-catenin requires the function of both LRP/Arrow and frizzled classes of receptors. *BMC Cell Biol* **4**, 4.

Segditsas S & Tomlinson I (2006). Colorectal cancer and genetic alterations in the Wnt pathway. *Oncogene* **25**(57), 7531-7.

Sharma M, Jamieson C, Johnson M, et al. (2012). Specific armadillo repeat sequences facilitate b-catenin nuclear transport in live cells via direct binding to nucleoporins Nup62, Nup153, and RanBP2/Nup358. *J Biol Chem* **287**, 819–831.

Shen Y, Zhu YM, Fan X, et al. (2011). Gene mutation patterns and their prognostic impact in a cohort of 1185 patients with acute myeloid leukemia. *Blood* **118**, 5593–5603.

Shih LY, Liang DC, Huang CF, et al. (2008). Cooperating mutations of receptor tyrosine kinases and Ras genes in childhood core-binding factor acute myeloid leukemia and a comparative analysis on paired diagnosis and relapse samples. *Leukemia* **22,** 303-307.

Shinkai Y, Rathbun G, Lam KP, et al. (1992). RAG-2-deficient mice lack mature lymphocytes owing to inability to initiate V(D)J rearrangement. *Cell* **68**, 855-867.

Shitashige M, Satow R, Honda K, et al. (2008). Regulation of Wnt signaling by the nuclear

pore complex. *Gastroenterology* **134**, 1961–1971.

Shmelkov SV, Butler JM, Hooper AT et al. (2008). CD133 expression is not restricted to stem cells, and both CD133+ and CD133− metastatic colon cancer cells initiate tumors. *J Clin Invest* 118, 2111–2120.

Shultz L, Ishikawa F & Greiner D. (2007). Humanizied mice in translational biomedical research. *Nat Rev Immunol* **7**, 108-130.

Siegfried E, Wilder EL & Perrimon N (1994). Components of wingless signalling in Drosophila. Nature 367: 76–80.

Silva FP, Swagemakers SM, Erpelinck-Verschueren C, et al. (2009). Gene expression profiling of minimally differentiated acute myeloid leukemia: M0 is a distinct entity subdivided by RUNX1 mutation status. *Blood* **114**, 3001-3007.

Siminovitch L, McCulloch EA & Till JE (1963). The Distribution of Colony-Forming Cells among Spleen Colonies. *J Cell Physiol* **62**, 327-336.

Simon RM, Korn EL, McShane LM et al. (2004). Design and analysis of DNA microarray investigations. New York: Springer.

Sonka M, Hlavac V, & Boyle R (2007). Image Processing, analysis, and machine Vision. *Thomson Engineering*, 3rd edition.

Southern, E.M (1975). Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J Mol Biol* **98**(3), 503-517.

Spangrude GJ, Heimfeld S & Weissman IL (1988). Purification and characterization of mouse hematopoietic stem cells. *Science* **241**, 58–62.

Spiekermann K, Bagrintseva K, Schwab R, et al. (2003). Overexpression and constitutive activation of FLT3 induces STAT5 activation in primary acute myeloid leukemia blast cells. *Clin Cancer Res* **9**, 2140-2150.

Sta¨deli R, Hoffmans R & Basler K (2006). Transcription under the control of nuclear Arm/beta-catenin. *Curr Biol* **16**, R378–R385.

Staal FJ & Clevers HC (2005). WNT signalling and haematopoiesis: a WNT-WNT situation. *Nat Rev Immunol* **5**(1), 21-30.

Stahl FW (1985). George Streisinger (December 27, 1927-August 11, 1984). *Genetics.*

Stapnes C, Gjertsen BT, Reikvam H, et al. (2009). Targeted therapy in acute myeloid leukaemia: current status and future directions. *Expert Opin Investig Drugs* **18**, 433-455.

Steelman LS, Abrams SL, Whelan J, et al. (2008). Contributions of the Raf/MEK/ERK, PI3K/PTEN/Akt/mTOR and Jak/STAT pathways to leukemia. *Leukemia* **22**, 686-707.

Stekel, D (2006). Microarray bioinformatics. *Cambridge: Cambridge University Press*.

Stephens DJ & Allan VJ (2003). Light microscopy techniques for live cell imaging. *Science* **300**, 82–86.

Stirewalt DL & Radich JP (2003). The role of FLT3 in haematopoietic malignancies. *Nat Rev Cancer* **3**, 650-65.

Strutt D (2003). Frizzled signalling and cell polarisation in Drosophila and vertebrates. *Development* **130**, 4501–4513.

Su Y, Fu C, Ishikawa S, et al. (2008). APC is essential for targeting phosphorylated beta-catenin to the SCFbeta-TrCP ubiquitin ligase. *Mol Cell* **32**, 652–661.

Suela J, Alvarez S & Cigudosa JC (2007). DNA profiling by arrayCGH in acute myeloid leukemia and myelodysplastic syndromes. *Cytogenet. Genome Res* **118**, 304–309.

Sugimura H, Mori H, Nagura K, et al. (2010). Fluorescence in situ hybridization analysis

with a tissue microarray: 'FISH and chips' analysis of pathology archives. *Pathol Int* **60**, 543-550.

Swarbreck D, Wilks C, Lamesch P, et al. (2008). The Arabidopsis Information Resource (TAIR): gene structure and function annotation. *Nucleic Acids Research* **36**, D1009–D1014.

Swiatek, W, Kang H, Garcia BA (2006). Negative regulation of LRP6 function by casein kinase I epsilon phosphorylation. *J Biol Chem* **281**, 12233-12241.

Takada R, Satomi Y, Kurata T, et al. (2006). Monounsaturated fatty acid modification of Wnt protein: its role in Wnt secretion. *Dev Cell* **11**, 791-801.

Tamai K, Zeng X, Liu C, et al. (2004). A mechanism for Wnt coreceptor activation. *Mol Cell* **13**, 149-156.

Tamai K, Semenov M, Kato Y, et al. (2000). LDLreceptor- related proteins in Wnt signal transduction. *Nature* **407**, 530-535.

Tamburini J, Elie C, Bardet V, et al. (2007). Constitutive phosphoinositide 3-kinase/Akt activation represents a favorable prognostic factor in de novo acute myelogenous leukemia patients. *Blood* **110**, 1025-1028.

Tanaka K, Kitagawa Y, & Kadowaki T. (2002). Drosophila segment polarity gene product porcupine stimulates the posttranslational N-glycosylation of wingless in the endoplasmic reticulum. *J Biol Chem* **277**, 12816-12823.

Tanaka K, Okabayashi K, Asashima M, et al. (2000). The evolutionarily conserved porcupine
gene family is involved in the processing of the Wnt family. *Eur J Biochem* **267**, 4300-4311.

Taussig DC, Miraki-Moud F, Anjos-Afonso F, et al. (2008). Anti-CD38 antibody–mediated clearance of human repopulating cells masks the heterogeneity of leukemia-initiating cells. *Blood* **112**, 568-575.

Taussig DC, Vargaftig J, Miraki-Moud F, et al. (2010). Leukemia-initiating cells from some acute myeloid leukemia patients with mutated nucleophosmin reside in the CD34(-) fraction. *Blood* **115**(10), 1976–1984.

ten Bosch JR, & Grody WW (2008). Keeping up with the next generation: massively parallel sequencing in clinical diagnostics. *J Mol Diagn* **10**, 484-492.

Tenen DG (2003). Disruption of differentiation in human cancer: AML shows the way. *Nat Rev Cancer* 3(2), 89-101.

Tenen DG, Hromas R, Licht JD, Zhang DE (1997). Transcription factors, normal myeloid development, and leukemia. *Blood* **90**, 489-519.

Thompson B, Townsley F, Rosin-Arbesfeld R, et al. (2002). A new nuclear component of the Wnt signalling pathway. *Nat Cell Biol* **4**, 367–373.

Till JE & McCulloch EA (1961). A direct measurement of the radiation sensitivity of normal mouse bone marrow cells. *Radiat Res* **14**, 213-222.

Tinoco I Jr & Uhlenbeck OC (1971). Estimation of secondary structure in ribonucleic acids. *Nature* **230**, 362–367.

Tinoco I Jr, Borer PN, Dengler B, et al. (1973). Improved estimation of secondary structure in ribonucleic acids. *Nature New Biol* **246**, 40–41.

Tirado CA, Valdez F, Klesse L, et al. (2010). Acute myeloid leukemia with inv(16) with CBFB-MYH11, 3'CBFB deletion, variant t(9;22) with BCR-ABL1, and del(7)(q22q32) in a pediatric patient: case report and literature review. *Cancer Genet Cytogenet* **200**, 54-59.

Tong J, Cao W & Barany F (1999). Biochemical properties of a high fidelity DNA ligase from Thermus species AK16D. *Nucleic Acids Res* **27** (3), 788-94.

Townsley FM, Thompson B, Bienz M (2004). Pygopus residues required for its binding to Legless are critical for transcription and development. *J Biol Chem* **279**, 5177–5183.

Trompouki E, Bowman TV, Lawton LN, et al. (2011). Lineage regulators direct BMP and Wnt pathways to cellspecific programs duringdifferentiation and regeneration. *Cell* **147,** 577–589.

Tsien RY (1998). The green fluorescent protein. *Annual Review of Biochemistry* , **67**, 509-544.

Tukey JW (1977). Exploratory Data Analysis. Addison-Wesley, Reading, MA.

Tweedie S, Ashburner M, Falls K, et al. (2009). The FlyBase Consortium. FlyBase: enhancing Drosophila Gene Ontology annotations. *Nucleic Acids Research* **37**, D555–D559.

Twells RC, Mein CA, Payne F, et al. (2003). Linkage and association mapping of the LRP5 locus on chromosome 11q13 in type 1 diabetes. *Hum Genet* **113**, 99-105.

Twigger SN, Shimoyama M, Bromberg S, et al. (2006). The Rat Genome Database, update 2007: Easing the path from disease to data and back again. *Nucleic Acids Research* **35** (Suppl 1), D658–D662.

Tyagi S and Kramer FR (1996). Molecular beacons: probes that fluoresce upon hybridization.

Nature Biotechnology **14**, 303-308.

Uchida N & Weissman IL (1992). Searching for hematopoietic stem cells: evidence that Thy-1.1lo Lin- Sca-1+ cells are the only stem cells in C57BL/Ka-Thy-1.1 bone marrow. *J Exp Med* **175** (1), 175-84.

Uchida N & Weissman IL (1992). Searching for hematopoietic stem. The Journal of experimental medicine. *J Exp Med* **175**, 175–184

Uchida N, He D, Friera AM, et al. (1997). The unexpected G0/G1 cell cycle status of mobilized hematopoietic stem cells from peripheral blood. *Blood* **89**:465-472.

Uhlenbeck OC, Borer PN, Dengler B et al. (1973). Stability of RNA hairpin loops: A6-Cm-U6. *J Mol Biol* **73**, 483–496.

Uren A, Reichsman F, Anest V, et al. (2000). Secreted frizzled-related protein-1 binds directly to Wingless and is a biphasic modulator of Wnt signaling. *J Biol Chem* **275** (6), 4374-4382.

Valenta T, Gay M, Steiner S, et al. (2011). Probing transcription-specific outputs of β-catenin in vivo. *Genes Dev* **25**,2631–2643.

Valenta T, Hausmann G & Basler K (2012). The many faces and functions of b-catenin. *The EMBO Journal* **31**, 2714–2736.

van de Wetering M, Cavallo R, Dooijes D, et al. (1997). Armadillo coactivates transcription driven by the product of the Drosophila segment polarity gene dTCF. *Cell* **88**, 789–799.

van den Heuvel M, Harryman-Samos C, Klingensmith J et al. (1993). Mutations in the segment polarity genes wingless and porcupine impair secretion of the wingless protein. *EMBO J* **12**, 5293-5302.

van Wijnen AJ, Stein GS, Gergen JP, et al. (2004). Nomenclature for Runt-related (RUNX) proteins. *Oncogene* **23**, 4209-4210.

Vardiman JW (2010). The World Health Organization (WHO) classification of tumors of the hematopoietic and lymphoid tissues: an overview with emphasis on the myeloid neoplasms. *Chem Biol Interact* **184**, 16-20.

Vardiman JW, Harris NL & Brunning RD (2002). The World Health Organization (WHO) classification of the myeloid neoplasms. *Blood* **100**, 2292-2302.

Vardiman JW, ThieleJ, Arber DA, et al. (2009). The 2008 revision of the World Health Organization (WHO) classification of myeloid neoplasms and acute leukemia: rationale and important changes. *Blood* **114**, 937-951.

Velculescu VE, Zhang L, Vogelstein B et al. (1995). Serial analysis of gene expression.

Science **270**, 484-487.

Verbunt RJAM, Fitzmaurice MA, Kramer JR, et al. (1992). Characterization of ultraviolet laser-induced autofluorescence of ceroid deposits and other structures in atherosclerotic plaques as a potential diagnostic for laser angiosurgery. *Am Heart* J **123**, 208–216.

Vincent L & Soille P (1991). Watersheds in digital spaces: An efficient algorithm based on immersion simulations. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **13**(6), 583–597.

Wählby C, Sintorn IM, Erlandsson F, et al. (2004). Combining intensity, edge and shape information for 2D and 3D segmentation of cell nuclei in tissue sections. *Journal of Microscopy* **215** (1), 67–76.

Wallace W, Schaefer LH & Swedlow JR (2001). A working person's guide to deconvolution in light microscopy. *Biotechniques* **31** (5), 1076–1078, 1080, 1082 passim.

Walter MJ, Payton JE, Ries RE, et al. (2009). Acquired copy number alterations in adult acute myeloid leukemia genomes. *Proc Natl Acad Sci USA* **106**, 12950–12955.

Walter, AE, Turner DH, KimJ, et al. (1994). Coaxial stacking of helixes enhances binding of oligoribonucleotides and improves predictions of RNA folding. *Proc Natl Acad Sci USA* **91**, 9218–9222.

Wang Z, Gerstein M, & Snyder M (2009). RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet* **10**, 57-63.

Wang C, Curtis JE, Geissler EN, et al. (1989). The expression of the proto-oncogene C-kit in the blast cells of acute myeloblastic leukemia. *Leukemia* **3**, 699-702.

Wang I, Kimura T, Asada R, et al. (2003). SCID-repopulating cell activity of human cord blood derived CD34– cells assured by intra-bone marrow injection. *Blood* **101**, 2924–2931.

Wang JC & Dick JE (2005). Cancer stem cells: lessons from leukemia. *Trends Cell Biol* **15**, 494-501.

Wang M, Hu Y, Amatangelo MD et al. (2011a). Role of ribosomal protein RPS2 in controlling let-7a expression in human prostate cancer. *Mol Cancer Res* **9**, 36-50.

Wang Q, Stacy T, Miller JD, et al. (1996b). The CBFbeta subunit is essential for CBFalpha2 (AML1) function in vivo. *Cell* **87**, 697-708.

Wang S, Krinks M, Lin K, et al. (1997). Frzb, a secreted protein expressed in the Spemann organizer, binds and inhibits Wnt-8. *Cell* **88** (6), 757-766.

Wang S, Wang Q, Crute BE, et al. (1993). Cloning and characterization of subunits of the T-cell receptor and murine leukemia virus enhancer corebinding factor. *Mol Cell Biol* **13**, 3324-3339.

Wang SW & Speck NA (1992). Purification of core-binding factor, a protein that binds the conserved core site in murine leukemia virus enhancers. *Mol Cell Biol* **12**, 89-102.

Wang Y, Krivstov AV, Sinha AU, et al. (2010). The Wnt/beta-catenin pathway is required for the development of leukemia stem cells in AML. *Science* **327** (5973), 1650–1653.

Wang Y, Lu J, Lee R, et al. (2002). Iterative normalization of cDNA microarray data. *IEEE Trans Inf Technol Biomed* **6** (1), 29-37.

Waterman MS (1978). Secondary structure of single-stranded nucleic acids. In Rota, G.-C. (ed.), Studies in Foundations and Combinatorics number 1 in Advances in Mathematics, Supplementary Studies. *Academic Press*, NY, pp. 167–212.

Waterman MS & Smith TM (1978). RNA secondary structure: a complete mathematical analysis. *Math Biosci* **42**, 257–266.

Wehrli M, Dougan ST, Caldwell K, et al. (2000). Arrow encodes an LDL-receptor-related protein essential for Wingless signalling. *Nature* **407**, 527-530.

Weigmann A, Corbeil D, Hellwig A et al. (1997). Prominin, a novel microvilli-specific

polytopic membrane protein of the apical surface of epithelial cells, is targeted to plasmalemmal protrusions of non-epithelial cells. *Proc Natl Acad Sci U S A* **94**, 12425-12430.

Weiss B, Jacquemin-Sablon A, Live TR, et al. (1968). Enzymatic breakage and joining of deoxyribonucleic acid. VI. Further purification and properties of polynucleotide ligase from Escherichia coli infected with bacteriophage T4. *J Biol Chem* **243**, 4543-55.

Weissman IL (2000). Stem cells: units of development, units of regeneration, and units in evolution. *Cell* **100**, 157-168.

Wend P, Holland JD, Ziebold U, et al. (2010). Wnt signaling in stem and cancer stem cells. *Semin Cell Dev Biol* **21**, 855–863.

Westerfield M (1995). The Zebrafish Book: the guide for the laboratory use of zebrafish.

Whitcombe D, Theaker J, Guy SP, et al. (1999). Detection of PCR products using self-probing amplicons and fluorescence. *Nat Biotechnol* **17**, 804–807.

Wieschaus E, Nu¨sslein-Volhard C & Ju¨rgens G (1984). Mutations affecting the pattern of the larval cuticle in Drosophila melanogaster III. Zygotic loci on the X-chromosome and fourth chromosome. *Rouxs Arch Dev Biol* **193**, 296–387.

Willert K, Brown JD, Danenberg E, et al. (2003). Wnt proteins are lipid-modified and can act as stem cell growth factors. *Nature* **423**, 448-452.

Willert K & Jones KA (2006). Wnt signaling: is the party in the nucleus? *Genes Dev*. **20**, 1394–1404.

Winston LA & Hunter T (1995). JAK2, Ras, and Raf Are Required for Activation of Extracellular Signal-regulated Kinase/Mitogen-activated Protein Kinase by Growth Hormone. *Journal of Biological Chemistry* **270**, 30837-30840.

Wray J, Kalkan T, & Smith AG (2011). Inhibition of glycogen synthase kinase 3 alleviates Tcf3 repression of the pluripotency. *Nat Cell Biol* **13**, 838- 845.

Wright MH, Calcagno AM, Salcido CD, et al. (2008). Brca1 breast tumors contain distinct CD44+/CD24− and CD133+ cells with cancer stem cell characteristics. *Breast Cancer Res* **10**, 1–16.

Wu D & Pan W (2010). GSK3: a multifaceted kinase in Wnt signaling. *Trends Biochem Sci* **35**, 161–168.

Wu DY & Wallace RB (1989). Specificity of the nick-closing activity of bacteriophage T4 DNA ligase. *Gene* **76** (2), 245-54.

Xin D, Hu L & Kong X (2008). Alternative promoters influence alternative splicing at the genomic level. *Plos One* **3**, e2377.

Xing Y, Clements WK, Kimelman D, et al. (2003). Crystal structure of a beta-catenin/axin complex suggests a mechanism for the beta-catenin destruction complex. *Genes Dev* **17**, 2753–2764.

Xing Y, Takemaru K, Liu J, et al. (2008) Crystal structure of a full-length beta-catenin. *Structure* **16**, 478–487.

Xu Q, D'Amore PA & Sokol SY (1998). Functional and biochemical interactions of Wnts with FrzA, a secreted Wnt antagonist. *Development* **125** (23), 4767-4776.

Yamashita Y, Yuan J, Suetake I, et al. (2010). Array-based genomic resequencing of human leukemia. *Oncogene* **29**, 3723–3731.

Yang PT, Lorenowicz MJ, Silhankova M, (2008). Wnt signaling requires retromer-dependent recycling of MIG-14/Wntless in Wnt-producing cells. *Dev Cell* **14**, 140-147.

Yang-Snyder J, Miller JR, Brown JD, et al. (1996). A frizzled homolog functions in a vertebrate Wnt signaling pathway. *Curr Biol* **6**, 1302–1306.

Yeung J, Esposito MT, Gandillet A, et al. (2010). Beta-Catenin mediates the establishment and drug resistance of MLL leukemic stem cells. *Cancer Cell* **18** (6), 606–618.

Yin AH, Miraglia S, Zanjani ED et al. (1997). AC133, a novel marker for human hematopoietic stem and progenitors cells. *Blood* **90**, 5002-5012.

Yokota S, Kiyoi H, Nakao M, et al. (1997). Internal tandem duplication of the FLT3 gene is preferentially seen in acute myeloid leukemia and myelodysplastic syndrome among various hematological malignancies. A study on a large series of patients and cell lines. *Leukemia* **11**, 1605-1609.

Yu Y, Flint A, Dvorin EL, et al. (2002). AC133-2, a novel isoform of human AC133 stem cell antigen. *J Biol Chem* **277**, 20711–20716.

Zhai L, Chaturvedi D & Cumberledge, S. (2004). Drosophila wnt-1 undergoes a hydrophobic modification and is targeted to lipid rafts, a process that requires porcupine. *J Biol Chem* **279**, 33220-33227.

Zhang B, Zerubia J & Olivo-Marin JC (2007). Gaussian approximations of fluorescence microscope point-spread function models. *Applied Optics* , **46** (10), 1819– 1829.

Zhang N, Wei P, Gong A, et al. (2011). FoxM1 promotes b-catenin nuclear localization and controls Wnt target-gene expression and glioma tumorigenesis. *Cancer Cell* **20**, 427–442.

Zhao H & Hardy RW (2004). Long chain saturated fatty acids induce annexin II translocation to detergent-resistant membranes. *Biochem J* **381**, 463-469.

Zimmerman SB, Little JW, Oshinsky CK, et al. (1967). Enzymatic joining of DNA strands: a novel reaction of diphosphopyridine nucleotide. *Proc Natl Acad Sci U S A* **57**: 1841-1848.

Zuker M (1989). Computer prediction of RNA structure. *Methods Enzymol* **180**, 262–288.

Zuker M (1989). On finding all suboptimal foldings of an RNA molecule. *Science* **244**, 48–52.

Zuker M (1994). Prediction of RNA secondary structure by energy minimization. chapter 23. In Griffin,A.M. and Griffin,H.G. (eds), *Computer Analysis of Sequence Data*, Vol. 25, Part II, Humana Press, Inc., Totowa, NJ, pp. 267–294.

Zuker M & Stiegler P (1981). Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. *Nucleic Acids Res* **9**, 133–148.

Zuker M, Mathews DH & Turner DH (1999). Algorithms and thermodynamics for RNA secondary structure prediction: A practical guide. In Barciszewski,J. and Clark,B.F.C. (ed.), RNA Biochemistry and Biotechnology, number 70 in NATO Science Partnership Sub-Series: 3: High Technology. chapter 2, Kluwer Academic Publishers Dordrecht, The Netherlands, pp. 11–43.

ACKNOWLEDGEMENTS

The work presented in this thesis has been carried out at the Department of Medical Biotechnology and Translational Medicine, Università degli Studi di Milano, Milano Italy.
A part of the work presented in this thesis has been carried out at the Department of Immunology, Genetics and Pathology at Rudbeck Laboratory Uppsala University, Sweden.
I would like to express my sincere gratitude to everyone who has supported me during these years and in different ways has contributed to the work presented herein. Especially, I would like to thank:

my PhD supervisor, ***Alessandro Beghini*** "My Prof": thank you for your leadership, support, for your kindness and constant enthusiasm about our work. Thank you for giving me this wonderful opportunity. Thank you a lot for the "Swedish experience", for sharing with me this adventure, sharing every moment, every experiment, every step. Thank you a lot, for our scientific and "philosophical" discussions, for our laughs and in particular for the complete freedom in the laboratory. Thank you a lot "My Prof"!

to Immunology, Genetics and Pathology at Rudbeck Laboratory A special thanks to ***Ulf Landegren, Mats Nilsson*** and ***Ola Soderberg*** for taking me on as a student and for

welcoming always the discussions. Thank you a lot for the hospitality and for your time. It has been a pleasure working with your group. I hope to work with you again!
Thank you a lot ***Martin Simonsson***, for your help and for your time with image analysis.
I would like to thank you ***Marco*****,** for your time, for taking care of me in the laboratory, for mail sharing and for our discussion. Thank you for being a friend. Finally, thanks to ***all Rudbeck Laboratory members***, thank you for turning the time spent there into a great experience. I miss you all!

to all my collaborators. First of all ***Luca DelGiacco*****:** many thanks for the collaboration, for your time and for your patience. Thank you a lot for our "skype talks", for our laughs and for the times. Thank you a lot for always teasing me! It has been a pleasure for me, to meet you! ***Roberto Brusamolino*** **:** for your friendship, for your time, for all the time you have spent working on the slide tissues. Thank you. ***Roberto Cairoli*** and ***Mauro Turrini***, for the collaboration, for your time, and for proposing the clinical view.
A thanks to ***Roberto DiLernia*** : for your friendship, for your time, for the funny dinners and for sharing your amazing anecdotes over dinner. Thanks to all the members of ***Cattaneo Laboratory***: thank you a lot for your time, for your help in the laboratory activities and for your friendship! A thanks also to ***Cristina:*** for your patience and for your work.

to ***my own family***: for all the love and support. Thank you for your never-ending caring, for the encouragement and for always believing in me; a special thanks to ***Valeria*****,** "my sister", the special one.

to ***my best friends*****:** many thanks to all my best friends, who have supported me during this period. A special thanks to ***Valeria*** the best friend ever. I've known you for a long time. Thank you for our coffees, for our talks, for our dinners. Thank you for being a fantastic friend! A particular thanks to ***Natalie*****:** I've known you for a short time, but you have become immediately a good friend. Thank you a lot for our coffees, our evenings, our discussions and laughs. Thank you a lot!

Many thanks to ***Davide, Giulia, Matteo, Josè, Matteo*** and to all my friends! Thank you ***Federico***, our "graphic designer"!

Special thanks to my ice-skating friends: ***Valentina T, Patrizia*** and ***Valentina DG***. Thank you a lot for our ice-skating evenings, for our memories and for our happy-hours! I'm very happy to see you again and share with you our skating passion.

Last but not least a special thanks to ***Andrea*****,** an "old-new" friend: Thank you a lot for our laughs, for our skype chats, for our evenings. Thank you a lot for being present in every moment. I'm very happy to see you again, after many years! Thank you a lot "engineer"!

Finally, a particular thanks to ***Francesco Sabatini***: thank you a lot for the cover of this thesis! Wonderful!

RINGRAZIAMENTI

Finalmente giunta alle ultime pagine della tesi, posso lasciar scorrere liberamente la penna sul foglio, per dire grazie a tutti coloro che mi hanno accompagnato durante il percorso di dottorato.

Non credo che ci siano parole per descrivere il mio stato d'animo in questo momento in cui mi trovo a scrivere le ultime pagine della tesi. Ultime pagine che rappresentano la fine di un percorso, ricchissimo di eventi ed emozioni, di incontri, di confronti, di esperienze, dove sicuramente il risultato netto in quest'altalena di emozioni, è la crescente passione per le attività di laboratorio, per la ricerca scientifica, per l'atmosfera che si respira seduti dietro al bancone di laboratorio.

Quasi con il desiderio di non chiudere questo importante capitolo e con la speranza di scrivere ancora tante pagine, mi trovo a dire grazie a tutte le persone che si sono inserite lungo questo percorso, accompagnandomi per mano fino a qui e condividendo ogni singolo istante.

Un grandissimo grazie:

ad ***Alessandro Beghini***, il "mio Prof", conosciuto ormai da tutti i laboratori come "your Prof". Sono tantissimi i "grazie" che devo rivolgerti e spero di non dimenticare nulla. Innanzitutto grazie per avermi dato la possibilità di intraprendere questa fantastica strada, per essere stato sempre presente in ogni momento, per le discussioni scientifiche

e non, per le risate e per aver sempre lasciato spazio e completa libertà in laboratorio. Grazie per aver avuto la pazienza di seguire ed indirizzare le mie passioni, osservando a distanza ed in silenzio ogni passaggio, ma lasciando completa libertà. Un grazie speciale per avermi dato la possibilità di confrontarmi con altri laboratori, intraprendendo un periodo all'estero e comprendendo la mia costante esigenza di confronto. Un grandissimo grazie per avermi accompagnato nell'"avventura svedese", condividendo ogni momento, ogni emozione, ogni singolo dettaglio, per essere stato sempre presente anche a distanza, attraverso skype.

Grazie anche per aver ascoltato pazientemente tutti i dubbi, le incertezze, per aver avuto sempre tempo per due parole e per sopportare i miei "sorry sorry". Infine un grandissimo grazie di cuore per essere un punto di riferimento e soprattutto per essere un "non prof".

ai Prof. ***Ulf Landegren, Ola Soderberg e Mats Nilsson***, del Rudbeck Laboratory, Uppsala University. Un grazie particolare per avermi accolta come un membro del vostro laboratorio, per avermi fatta sentire a casa, per esservi presi cura di me in ogni istante, per aver giocato e scherzato con le "blobs" e il "lock and roll". Grazie soprattutto per avermi dato la possibilità di confrontarmi con voi, permettendomi di crescere e di approfondire gli aspetti scientifici che più mi appassionano, attraverso approcci sempre più innovativi. Grazie anche a ***Martin Simonsson***, per la pazienza e il tempo dedicato all'analisi di immagine.

Un grazie speciale a ***Marco***, per esserti preso cura di me durante la mia permanenza ad Uppsala, per la pazienza avuta al microscopio, per il tempo che mi hai dedicato in laboratorio, per la pazienza e per il continuo contatto mail. Grazie soprattutto per la tua

amicizia!

Infine grazie a tutti i ***ragazzi del Rudbeck Laboratory***, per aver reso l'"esperienza svedese" veramente indimenticabile! Grazie a tutti! Spero di poter condividere con voi tante altre esperienze!

a tutti i ***collaboratori*** del laboratorio. Un grazie speciale a ***Luca DelGiacco***. Grazie per avermi accompagnato in questi anni, attraverso le infinite conversazioni in skype, e per aver sempre avuto il tempo di ascoltarmi, con tanta pazienza. Grazie soprattutto per avermi sempre preso in giro! Grazie a ***Roberto Brusamolino***, per esserti inserito con tanta pazienza nelle attività di laboratorio, supportando in ogni momento e con ogni mezzo il laboratorio.

Un grazie particolare a ***Roberto di Lernia***, per avermi dato la possibilità di partecipare a pranzi veramente divertenti e per avermi fatto conoscere con tanta ironia una parte di storia dell'Università e del Dipartimento. Devo ringraziare anche ***Mauro Turrini*** e ***Roberto Cairoli***, per la vostra collaborazione, la vostra disponibilità e per avermi dato la possibilità di conoscere il risvolto clinico delle attività di ricerca.

Grazie a tutti i ***ragazzi del gruppo Cattaneo***, per avermi sempre aiutato nelle attività di laboratorio con tanta pazienza, per aver rappresentato un punto di riferimento e soprattutto per la vostra amicizia.

Infine grazie anche a ***Cristina,*** per esserti inserita con tanta pazienza nell'ultima parte del percorso di dottorato.

a tutta la ***mia famiglia***, per avermi supportato lungo tutto il percorso, per esserci

sempre stati, in ogni momento, osservando in silenzio, ma rimanendo sempre vicini. Non ci sono parole per ringraziare ***Valeria***, la "mia sister", per esserci sempre in ogni momento e per essere decisamente più di una sorella.

a tutti gli ***amici***, che mi hanno accompagnato in questo percorso. Grazie per le tante risate! Un grazie di cuore a ***Valeria***, più di un'amica da tantissimi anni. Grazie mille per esserci sempre e per esserci sempre stata in ogni momento, per le quattro chiacchere serali davanti ad un caffè, per le conversazioni infinite e per la profonda amicizia che ci lega. Un grazie particolare a ***Natalie***; anche se ci conosciamo da poco tempo, sei diventata da subito una grande amica. Grazie per le nostre serate, per le tante risate, per i caffè, gli aperitivi e soprattutto per esserci sempre.

Grazie a ***Valentina T, Valentina DG*** e ***Patrizia.*** Sono felice di avervi ritrovato dopo tanti anni. Grazie per avermi dato nuovamente la possibilità di rimettere i pattini, di tornare sul ghiaccio e di riscoprire la passione e le emozioni per uno sport che ha rappresentato una pagina importante della mia e della nostra vita. Grazie soprattutto per la vostra "pazza" amicizia!

Un grazie particolare lo rivolgo ad ***Andrea***, "l'ingegnere", ritrovato dopo tanti anni con immenso piacere. Grazie per le nostre serate, per esserci sempre, per le nostre chiaccherate in skype, per farmi ridere sempre. Grazie per essere diventato pian piano nel tempo, un amico.

Grazie a ***Davide*** per far parte ormai della famiglia, a ***Giulia, Matteo, Josè, Rachele, Matteo*** e a tutti gli amici per esserci sempre stati, per aver condiviso con me ogni singolo istante del mio percorso di vita.

Infine devo ringraziare ***Federico***, per il supporto che mi ha fornito nell'elaborazione

grafica delle immagini e nella costruzione del famoso “pannello delle bandine”, che rappresenta il risultato di un esperimento. Un grandissimo grazie, per l’aiuto e soprattutto per la pazienza! Un grazie particolare a ***Francesco Sabatini***, per l’elaborazione grafica della copertina della tesi. Splendida! Grazie mille per il supporto e soprattutto per la pazienza!

Giunta proprio alle ultime righe della tesi, ringrazio tutti per avermi accompagnato in questo splendido e appassionante percorso!