TECHNISCHE UNIVERSITÄT MÜNCHEN

Institut für Experimentelle Genetik

Identification and characterization of a Y chromosome-encoded gene overexpressed in acute myeloid leukaemia cells

Antonio Gallo

Vollständiger Abdruck der von der Fakultät Wissenschaftszentrum Weihenstephan für Ernährung, Landnutzung und Umwelt der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften

genehmigten Dissertation.

Vorsitzender: Univ.-Prof. Dr. J. J. Hauner

Prüfer der Dissertation:
1. apl. Prof. Dr. J. Adamski
2. Univ.-Prof. Dr. A. Gierl
3. Univ.-Prof. Dr. H.-J. Kolb
(Ludwig-Maximilians-Universität München)

Die Dissertation wurde am 26.05.2009 bei der Technischen Universität München eingereicht und durch die Fakultät Wissenschaftszentrum Weihenstephan für Ernährung, Landnutzung und Umwelt am 11.09.2009 angenommen.

La scienza è sempre imperfetta. Ogni volta che risolve un problema, ne crea almeno dieci nuovi.
George Bernard Shaw

Tra tutte le cose sicure, la più sicura è il dubbio.
Bertolt Brecht

Alla mia famiglia, ai miei amici e ad Olga.

Table of contents

Abstract

Haematopoietic stem cell (HSC) transplantation is widely used in the treatment of acute myeloid leukaemia (AML) to establish a new donor-derived bone marrow in the patients and to remove the residual leukaemic cells left after chemotherapy. Haploidentical HSC donors, who have only a partial match with the patients at the human leukocyte antigen (HLA) locus, are the best choice when a full-matched donor is not available. Several studies indicate that male patients transplanted with female grafts have a better long-term survival probability than all other gender combinations. The better therapy outcome consists in a strong graft versus leukaemia (GvL) effect with reduced graft versus host disease (GvHD).

Basic hypothesis in the project is that, in the patients' leukaemic cells, Y chromosome genes encode for minor histocompatibility antigens which might be recognized as non-self by the T-cells of the female donor which might then generate a cell-mediated immune response against the tumoral cells. On the other hand, there is the possibility that certain Y chromosome genes encode for proteins which may interfere with the biology of tumoral cells.

In this context, the primary aim of this study was the comparison of Y chromosome-encoded genes between AML cells and their healthy counterpart and, subsequently, the characterisation of those genes, found to be up-regulated in AML cells, from an immunological and biological point of view. The study was focused on two similar subtypes of AML: M4 and M5, both characterized by high expression of CD14 on the leukaemic cell surface.

By using real time-PCR with low density arrays, I have identified 4 transcripts from uncharacterized genes up-regulated in AML cells: BCORL2, VCY, PCDH11, and TGIF2L, the latter two having a homolog on the X chromosome.

After prediction of candidate genes-derived peptides restricted to HLA-A0201 major histocompatibility complex molecules, I assessed their real HLA-A0201-binding capacity in an *in vitro* assay. Afterwards, I was able to identify a peptide, derived from the candidate gene VCY, which was able to weakly stimulate female cytotoxic T-cells.

Another line of research was focused on the characterization of the gene TGIF2LY, chosen because of its usual testis-restricted expression and of its transcription factor potential. The AML cell line THP-1 was used as model system, after assessment of TGIF2LY protein expression in these cells.

In order to get information about the TGIF2LY gene function, several experiments have been performed. Subcellular localisation studies showed that TGIF2LY is transported to the nuclei of the cells. Amino acid sequence and three-dimensional structure model analysis further suggested a transcription factor function for TGIF2LY. I implemented a protein-DNA binding assay from a previously established *in vitro* genomic selection and used it to identify genomic DNA regions to which TGIF2LY was able to bind. By

using an electrophoretic mobility shift assay (EMSA), I was able to identify the core nucleotide sequence recognized and bound by TGIF2LY. The analysis of the genomic DNA fragments bound by TGIF2LY suggests for this protein a role in differentiation and activation of antigen presenting cells.

This work provides new molecular biology and immunological data related to Y-chromosome controlled processes. Altogether, these data can be considered as starting point for new experimental work aimed to development of novel AML therapy strategies.

Zusammenfassung

Die hämatopoetische Stammzelltransplantation ist bei der Behandlung akuter myeloischer Leukämien weit verbreitet, um restliche, nach einer Chemotherapie verbliebene leukämische Zellen zu eliminieren und um in den Patienten ein neues Knochenmark, abgeleitet von den hämatopoietischen Stammzellen des Spenders, zu etablieren Haploidentische Stammzell-Spender, deren Human Leukocyte Antigen (HLA) Locus nur partiell mit dem des Patienten übereinstimmt, sind die beste Wahl, wenn kein komplett HLA-kompatibler Spender zur Verfügung steht.
Mehrere Studien weisen darauf hin, dass bei männlichen Empfängern eines weiblichen Transplantats langfristig eine bessere Überlebenswahrscheinlichkeit besteht als bei allen anderen Spender-Empfänger-Kombinationen. Bei solchen Patienten wurde ein starker Graft-versus-Leukaemia (GvL) Effekt mit reduziertem Graft-versus-Host Disease (GvHD) beobachtet.
Die Grundhypothese für die hier vorliegende Arbeit besteht darin, dass einige Y-chromosomale Gene für Minor Histokompatibilitätsantigene kodieren, welche von den T-Zellen des weiblichen Spenders als „nicht-eigen“ (non-self) erkannt werden. Dies führt zu einer spezifischen zellulären Immunantwort gegen die Empfänger-Tumorzellen. Andererseits könnte es sein, dass einige Y-chromosomal kodierte Proteine direkt mit den Tumorzellen wechselwirken.
In diesem Zusammenhang lag das Ziel dieser Arbeit im Vergleich der Y-chromosomal kodierten Gene zwischen männlichen AML- und gesunden Zellen, sowie weiterführend in der immunologischen und molekularbiologischen Charakterisierung derjenigen Gene, die in AML-Zellen überexprimiert waren. Die Arbeit konzentrierte sich auf Patienten mit den zwei sich ähnlichen Subtypen der AML M4 und M5, die beide durch eine hohe Expression von CD14 auf der Oberfläche der leukämischen Zellen gekennzeichnet sind.
Mittels Real-time-PCR wurden 4 Transkripte von Y-chromosomalen Genen identifiziert, die in AML Zellen überexprimiert waren: BCORL2, VCY, PCDH11 und TGIF2L.
Nach der Generierung einiger, von den Kandidaten-Genen abgeleiteten Peptiden, die für HLA-A0201 Haupthistokompatibilitätskomplex-Moleküle spezifsch sind, habe ich deren tatsächliche HLA Bindungskapazität in einem *in vitro* Experiment gemessen. Ein von Gen VCY abgeleitetes Peptid konnte dabei weibliche zytotoxische T Zellen schwach stimulieren.
Das Gen TGIF2LY wurde aufgrund seiner vermeintlich Testes-spezifischen Expression und wegen seiner potentiellen Funktion als Transkriptionsfaktor eingehend untersucht. Als Modellsystem wurde die AML Zelllinie THP-1 ausgewählt, da diese Zellen TGIF2LY exprimieren. Subzelluläre Lokalisierung des Proteins zeigte, dass TGIF2LY in den Zellkern transportiert wird. Sequenzhomologien und Strukturmodellierung deuten auf eine Funktion für TGIF2LY als Transkriptionsfaktor hin.

Mittels eines modifizierten Protein-DNA Bindungsassays wurden genomische DNA-Fragmente isoliert, die von TGIF2LY gebunden werden. Electromobility Shift Assay (EMSA) Experimente führten zur Identifizierung von Konsensus-Bindungssequenzen, spezifisch für TGIF2LY. Die eingehende Analyse der genomischen DNA Fragmente, mit denen TGIF2LY interagierte, deutet bei diesem Protein auf eine Rolle in Differenzierung und Aktivierung von antigenpräsentierenden Zellen hin.
Diese Arbeit liefert neue molekularbiologische und immunologische Erkenntnisse über Y-chromosomal kontrollierte Prozesse, die als Ausgangspunkte für die Entwicklung neuer Therapiestrategien gegen AML dienen können.

1 Introduction

1.1 Acute myeloid leukaemia

1.1.1 Definition, etiology, and epidemiology

Acute myeloid leukaemia (AML) is a highly heterogeneous clonal disorder of hematopoietic progenitor cells which are referred to as "blasts". These cells lose their ability to respond to normal proliferation regulators, like cytokines or hormones and, consequently, to correctly differentiate. Direct consequences of such disorder can be the total failure of the immune system which, in turn, can lead to fatal infection, bleeding, organ infiltration of the leukaemic blasts, typically, in absence of treatment, within one year from the diagnosis. The broad spectrum of consequences reflects the heterogeneity of the disease.

The most common cause of death in AML is bone marrow failure. The genetic reprogramming of AML blasts makes them unable to generate mature red cells, neutrophils, monocytes, and platelets, depending on the affected progenitor cell population. AML blasts also inhibit normal bone marrow cells from differentiating into mature blood cells. This inhibition is not only a result of the replacement of normal progenitor cells by the leukaemic blasts because there is no correlation between degree of cytopenia and marrow blast count. There are evidences that this inhibition may be mediated by various chemokines produced by AML blasts (Youn*, et al* 2000). Given the ability of AML to retard haematopoiesis, the rate of recovery after treatment of AML reflects not only the effect of the treatment on normal progenitor cells but also its effect on AML blasts (Estey*, et al* 2000).

The principal sign of marrow failure in AML is insurgence of infections. The mostly observed pathogens are common endogenous aerobic gram-positive and gram-negative bacteria like endogenous *Candida* species and, particularly in warm and humid climates, *Aspergillus* species, which are water-borne or enter via the respiratory tract (Anderlini*, et al* 1996). Studies on different antibiotics indicate that the main predictor for a successful treatment is an increasing neutrophil count. However, cytotoxic chemotherapy aimed to obtain such a rise, temporarily increases the risk of infection because it damages gastrointestinal mucosa. Potentially fatal organ infiltration, most ominously involving the lung and brain, becomes more likely as the white blood cell count gets higher, particularly in the monocytic subtype of AML (M4 and M5 subtypes;

see paragraph 1.1.2) or if blasts are positive for CD56 surface antigen (Seymour, *et al* 1994). In these situations, emergency cytotoxic chemotherapy is needed, although its use is often associated with life-threatening complications like acute respiratory distress syndrome and tumour lysis syndrome (Montesinos, *et al* 2008, Tsai, *et al* 2008).

Several risk factors for developing AML have been identified in the last years. Between them, it is important to mention pre-leukemic blood disorders, such as myelodysplastic (MDS) or myeloproliferative (MPS) syndromes which can evolve into AML. The risk for the patients depends on the type of MDS/MPS (Godley and Larson 2008, Sanz, *et al* 1989). Undergoing anti-cancer chemotherapy, in particular using alkylating agents, is also increasing the risk of developing AML in patients. The risk is highest about three to five years after chemotherapy (Godley and Larson 2008, Le Beau, *et al* 1986). Other chemotherapeutic agents, specifically epipodophyllotoxins and anthracyclines, have also been associated with onset of leukemia. These treatment-related leukemias are often associated with specific chromosomal abnormalities in the leukemic cells (Thirman, *et al* 1993).
Another important risk factor for developing AML is exposure to ionizing radiation: leukemia is common in workers of the nuclear industry (Cardis, *et al* 1995), but not in people living near nuclear power plants (von Muhlendahl 1998). People who have survived atomic bombs are also at high risk of AML development (Nakanishi, *et al* 1999).
Occupational chemical exposure to benzene and other aromatic organic solvents is controversial as a cause of AML. Benzene and many of its derivatives are known to be carcinogenic *in vitro*. While some studies have suggested a link between occupational exposure to benzene and increased risk of AML (Austin, *et al* 1988), others have suggested that the risk relative to this compound, if any, is minimal (Linet 1984).
Several congenital conditions may also increase the risk of leukemia. One of the most common is, probably, Down syndrome, which is associated with a 10- to 18-fold increase in the risk of AML (Evans and Steward 1972).
Although development of AML has been associated with the above-mentioned risk factors, generally, they account only for a small number of reported cases (Sandler and Collman 1987).
Leukaemogenesis is a multi-step process that requires the susceptibility of a hematopoietic progenitor cell to inductive agents at multiple stages. The different subtypes of AML may have distinct causal mechanisms, suggesting a functional link between a particular molecular abnormality or mutation and the causal agent (Crane, *et al* 1996). Several cases of AML arise de novo without relation to exposure to known leukaemogenic agents.
Acquired clonal chromosomal abnormalities are usually found in 50% to 80% of AML cases (Grimwade, *et al* 1998, Heim and Mitelman 1992, Moorman, *et al* 2001, Wahlin, *et al* 1991), with high incidence in patients with secondary leukaemia (Sanderson, *et al*

2006) or older age (Leith, *et al* 1997, Mauritzson, *et al* 1999, Rossi, *et al* 2000). Frequently reported abnormalities include loss or deletion of chromosome 5, 7, 9, and Y, translocations such as t(8;21)(q22;q22); t(15;17)(q22;q11), trisomy 8 and 21, and other abnormalities involving chromosomes 16, 9, and 11. Multiple studies have demonstrated the prognostic importance of cytogenetic abnormalities in AML, making this at present the most important predictor of short-term (Ferrant, *et al* 1997, Keating, *et al* 1988, Schiffer, *et al* 1989) and long-term (Estey, *et al* 1997) outcome. Patients with a good prognosis are those with functional inactivation of the core binding factors (CBFs) transcription factors AML1 and CBFb, which seem to be involved in development of normal haematopoiesis (Cho, *et al* 2004, Mao, *et al* 1999, Pelletier, *et al* 2002). These cases include patients with AML and t(8;21)(q22;q22) or inv(16)(p13;q22), two of the most frequent recurrent cytogenetic abnormalities in *de novo* AML in younger patients (Cheson, *et al* 2003). Poor-prognosis patients have a loss of all or parts of chromosome 5 or 7, translocations involving 11q23, or abnormalities of chromosome 3 (Grimwade, *et al* 2001).

Acute leukemias, in general, are rare diseases, but have a disproportionally large effect on cancer survival statistics (Jemal, *et al* 2008). AML is the most common type of myeloid leukemia in adults, yet continues to have the lowest survival rate of all leukemias. AML accounts for approximately 25% of all leukemias in adults in the Western world (Greenlee, *et al* 2001). Worldwide, the incidence of AML is highest in the United States, Australia, and Western Europe (Jemal, *et al* 2008). Adults aged 65 years or older represent the population in which the highest incidence of AML is observed. This group accounts for around 55% of AML cases (USA 2001-2005; data source: National Cancer Institute, http://seer.cancer.gov). Although survival rates have improved remarkably in the younger age patient group, the prognosis in older patients continues to be very poor (Redaelli, *et al* 2003, Sandler and Ross 1997).

1.1.2 Classification of acute myeloid leukaemia

Classification of AML has never been easy, due to the heterogeneity of the disease. In the 1970s, an international conference of leukemia experts was held to decide on the best system for classifying acute leukemias. The result was the so-called French-American-British (FAB) classification, which divides AML in nine subtypes, basing on morphological characteristics of the leukemic cells (Bennett, *et al* 1976). FAB AML subtypes are described in table 1.1.

Although the FAB classification is still widely used, it doesn't take in account many of the risk or prognostic factors which arose during the last two decades. For this reason, at the end of the 1990s the World Health Organization (WHO) re-defined the AML classification into broader categories, taking into account the presence or absence of

well-characterised cytogenetic markers and prognostic factors (Vardiman, *et al* 2002). The WHO classification is summarized in table 1.2.

FAB subtype	Definition	% of adult patients	Prognosis
M0	Undifferentiated acute myeloblastic leukaemia	5%	Worse
M1	Acute myeloblastic leukaemia with minimal maturation	15%	Average
M2	Acute myeloblastic leukaemia with maturation	20%	Better
M3	Acute promyelocytic leukaemia	10%	Best
M4	Acute myelomonocytic leukaemia	25%	Average
M4 eos	Acute myelomonocytic leukaemia with eosinophilia	5%	Better
M5	Acute monocytic leukaemia	10%	Average
M6	Acute erythroid leukaemia	5%	Worse
M7	Acute megakaryoblastic leukaemia	5%	Worse

Table 1.1. French-American-British classification of acute myeloid leukaemia. The percentage of adult patients is relative to AML cases in USA updated to 2007. Prognosis is relative to the average of AML cases and it is only indicative. Source: American Cancer Society (URL: www.cancer.org).

AML with recurrent genetic abnormalities	
	AML with t(8;21)(q22;q22), (AML1/ETO)
	AML with abnormal bone marrow eosinophils and inv(16)(p13q22) or t(16;16)(p13;q22), (CBF/MYH11)
	Acute promyelocytic leukemia with t(15;17)(q22;q12), (PML/RAR) and variants
	AML with 11q23 (MLL) abnormalities
AML with multilineage dysplasia	
	Following MDS or MDS/MPD
	Without prior MDS or MDS/MPD, but with dysplasia in at least 50% of cells in 2 or more myeloid lineages
AML and myelodysplastic syndromes, therapy related	
	Alkylating agent/radiation–related type
	Topoisomerase II inhibitor–related type (some may be lymphoid)
	Others
AML not otherwise categorized	
	AML minimally differentiated
	AML without maturation
	AML with maturation
	Acute myelomonocytic leukemia
	Acute monoblastic/acute monocytic leukemia
	Acute erythroid leukemia (erythroid/myeloid and pure erythroleukemia)
	Acute megakaryocytic leukemia
	Acute basophilic leukemia
	Acute panmyelosis with myelofibrosis
	Myeloid sarcoma

Table 1.2. World Health Organization classification of acute myeloid leukaemia. Genes affected by genetic abnormalities are indicated in brackets. Some cases of sarcoma and myelodysplastic syndromes are enclosed in this classification.

1.1.3 Diagnosis of acute myeloid leukaemia

The diagnosis of the different subtypes of AML is a major challenge. The correct diagnosis is not only essential for classification of this heterogeneous complex of disorders, but, in addition, it plays a central role for individual risk assessment and for therapeutic decisions. Diagnostics of acute leukemias underwent a complete change since the 1970s when cytomorphology and cytochemistry represented the only available diagnostic tools. Since then, the routine diagnostic protocols incorporated classical cytogenetics, molecular cytogenetics, including several fluorescence in situ hybridization (FISH) techniques and comparative genomic hybridization CGH), molecular genetics (mostly PCR-based techniques and sequencing), and immunophenotyping by multi-parameter flow cytometry (Haferlach*, et al* 2007). AML is associated with several chromosomal aberrations and molecular abnormalities (see paragraph 1.1.2) and the approach to this genetic complexity can only be achieved by the combination of the different methods mentioned above, in order to use these abnormalities as diagnostic and prognostic parameters (Schoch*, et al* 2004). All these techniques are complementary to each other: cytomorphology, cytochemistry, and multi-parameter flow cytometry allow the categorization of the acute leukemias to different lineages. Chromosome banding analysis provides information of all microscopically visible genetic abnormalities, whereas the FISH techniques together with cytogenetics can allow detection of sub-microscopic alterations. The molecular techniques are more sensitive than FISH with respect to minimal residual disease detection. They verify involved genes as detected by genetics and define several mutations not detectable by all the other techniques. These last techniques are a indispensable tool for both diagnosis and later follow-up of the patients (Kern*, et al* 2004, Schnittger*, et al* 2004, van der Velden*, et al* 2003).

In the last years, several algorithms have been implemented that collect information from each of the used diagnostic tools and elaborate them in order to provide a comprehensive diagnosis (Haferlach*, et al* 2005).

Nevertheless, there are some parameters, based on the percentage of leukaemic cells in the bone marrow of the patients, which can be used for a first and rapid diagnosis. Demonstration of accumulation of leukaemic blasts characteristic of AML (Frohling*, et al* 2005, Kelly and Gilliland 2002), which is a result of the block of differentiation of these cells, is one essential requirement for diagnosis. According to the FAB system (see paragraph 1.1.2), AML is confirmed when the bone marrow contains more than 30% leukaemic blasts. The WHO classification system has lowered the leukaemic blast minimum for diagnosis of AML to 20%. In this system, many cases classified as high-grade myelodysplastic syndromes in the FAB system are considered as AML (see table 1.2). Once the leukaemic blasts are demonstrated to reach 20% of bone marrow cells, they must be shown to be derived from the myeloid lineage. At least 20% of the

leukaemic blasts must have surface antigens associated with myeloid differentiation, most commonly CD33 or CD13. The following particular cases should be also taken into account: if more than 80% of the bone marrow is represented by monocytes, acute monocytic leukaemia is diagnosed. If more than 50% of the marrow consists of normoblasts or pronormoblasts (progenitors of erythrocytes) and if myeloid progenitor cells constitute more than 30% of the non-erythroid population, the diagnosis is erythroleukaemia. If it is not possible to aspire the bone marrow, a diagnosis of megakaryocytic leukaemia should be considered. These initial diagnoses should always be confirmed by biopsies using all the above-mentioned approaches.

1.1.4 Treatment of AML patients

Several protocols have been established to treat AML patients. Differences in such protocols depend on several factors. Of course, the first therapy decisions are taken depending on the subtype of AML that affects the patient and on its cytogenetic characterization. Other factors influencing the treatment of a patient concern age, clinical status (i.e. presence of infections), and availability of a proper donor in case of transplantation. Nevertheless, some therapeutic tools have been standardized and are now widely used for treatment of AML patients (Kohrt and Coutre 2008). Among them, the most used are chemotherapy, body irradiation, and bone marrow or hematopoietic stem cell transplantation.

Chemotherapy

The first step in the treatment of AML patients consists of an intensive and aggressive chemotherapeutic regimen, also called induction regimen. The aim of such chemotherapy is to induce remission in the patients. This is achieved when the blood counts are brought back to normal levels and there is no evidence of the presence of leukaemic cells in the blood and in the bone marrow of the patient.

Various strategies have been explored to eliminate the residual leukaemic cells which are not detectable with the usual diagnostic tools (minimal residual disease) from the bone marrow of patients in complete remission. These residual leukaemic cells, if not eradicated, might lead to relapse. Intensive consolidation therapy, high-dose or low-dose chemotherapy, or chemoradiotherapy with either allogeneic or autologous hematopoietic stem-cell transplantation (HSCT) are the most used treatments to control or eliminate minimal residual disease.

During the past three decades, a series of studies has established an induction regimen that has become a standard for the treatment of AML patients not participating on a clinical trial. A widely used combination for induction is the cell cycle–specific agent cytarabine by continuous infusion for seven days and the non–cell-cycle–specific

anthracycline antibiotic daunorubicin intravenously for three days (Dillman, *et al* 1991, Rowe and Tallman 1997, Yates, *et al* 1982). In order to improve the complete remission rate, several studies have tested alternatives and higher doses of anthracyclines (Berman, *et al* 1991, Buchner, *et al* 2001, Rowe, *et al* 2004, Vogler, *et al* 1992), higher doses of cytarabine (Plunkett, *et al* 1987, Weick, *et al* 1996), new agents combined with cytarabine and daunorubicin such as etoposide, the purine analog fludarabine or the cytotoxic quinoline alkaloid (camptothecin) topotecan (Bishop, *et al* 1990, Estey, *et al* 2001), or sequential standard therapy followed by high doses of cytarabine (Buchner, *et al* 1999, Mitus, *et al* 1995, Petersdorf, *et al* 2007). Despite theoretic advantages, none of these approaches showed to be definitively better than the standard regimen (Berman and McBride 1992).

The M3 subtype of AML, also known as acute promyelocytic leukemia (see paragraph 1.1.2), is almost universally treated with the drug all-*trans*-retinoic acid (ATRA) in addition to induction chemotherapy (Fenaux, *et al* 1999, Huang, *et al* 1988, Tallman, *et al* 1997). Care must be taken to prevent disseminated intravascular coagulation, complicating the treatment of AML M3 when the promyelocytes release the contents of their granules into the peripheral circulation. This AML subtype is effectively curable with well-documented treatment protocols.

Chemotherapy is also used to control the minimal residual disease in the so-called post-remission or consolidation therapy. Retrospective and prospective analyses of cooperative group studies show that increasing the intensity of post-remission therapy is beneficial in younger but not older adults (Bennett, *et al* 1997, Cassileth, *et al* 1992). Several studies have evaluated the impact of intensive post-remission consolidation therapy with high-dose cytarabine. A prospective study by the Cancer and Leukemia Group B (CALGB) demonstrated that four courses of such therapy are significantly better than four courses of intermediate- or standard-dose cytarabine consolidation treatments, confirming a dose-response effect in younger patients and a benefit in patients with good-risk cytogenetics (Mayer, *et al* 1994). However, in older patients undergoing such treatments, high toxicity has been observed.

Body irradiation and bone marrow or haematopoietic stem cell transplantation

Another method to get rid of or to control the minimal residual disease is to transplant the patient with bone marrow or haematopoietic stem cells from a compatible healthy donor (allogeneic transplantation). This procedure allows replacing the immune system of the patient with that of a healthy individual.

The idea to treat patients with leukaemia and other haematopoietic malignancies with total body irradiation followed by transplantation of bone marrow from healthy donors was developed already in the 1950s. However, evidences from murine leukaemia systems showed that if on one hand, it was possible to cure leukaemia with total body irradiation followed by allogeneic bone marrow transplantation (Barnes and Loutit 1957), on the other hand leukaemic cells couldn’t be totally eliminated by using

clinically tolerated doses of radiations (Burchenal, *et al* 1960). The success of treatment of acute leukaemias by using allogeneic bone marrow transplantation depended on several factors: selection of compatible sibling donors by human leukocyte antigen (HLA) typing, combination of tolerated total body irradiation with cyclophosphamide as pre-transplant conditioning regimen, and post-transplant immunosuppressive treatment with methotrexate (Thomas, *et al* 1975). In the so-treated patients it was observed that the occurrence of graft versus host disease (GvHD) had beneficial effects on therapeutic outcome and survival of the patients (Weiden, *et al* 1979, Weiden, *et al* 1981). GvHD is caused by T-cells derived from the transplanted bone marrow that react against the tissues of the patient (Reddy, *et al* 2008). The role of T-cells became evident when T-cell-depleted bone marrow (haematopoietic stem cells) transplantation was introduced in the clinical protocols for leukaemia treatment (Kolb, *et al* 1979, Rodt, *et al* 1979). It is now clear that the effect of the graft against leukaemia, the so-called graft versus leukaemia (GvL) effect, occurring when T-cells and natural killer (NK) cells of the graft recognize and kill the residual leukaemic cells in the patient, and the occurrence of GvHD are closely related (Porter and Levine 2006). Development and severity of GvHD is strongly related with post-transplant outcome. It may at the same time improve survival by decreasing the risk of disease relapse, due to GvL effect, and increase non-relapse mortality by causing organ failure and predisposing the recipient to life-threatening infections (Pasquini 2008).

The challenge of allogeneic haematopoietic stem cell transplantation for treatment of leukaemia and other haematological diseases is the prevention or the reduction of GvHD without eliminating the GvL effect (Kolb 2008).

One crucial step for the success of haematopoietic stem cell transplantation is the choice of the compatible healthy donor. There are several possibilities for the choice of the donor for allogeneic transplantation. One very important parameter is the HLA match degree. The HLA superlocus contains a large number of genes related to immune system functions and resides on chromosome 6. Among the proteins encoded by this superlocus, there are cell-surface antigen-presenting proteins. The proteins encoded by these genes are also known as “antigens”, and have been associated with rejection in organ transplantations. There are two classes of major HLA antigens which are essential elements in immune function. HLA class I molecules (A, B, and C) present endogenous peptides, including viral peptides if present, that are derived from protein degradation by the proteasome. These peptides are generally polymers of about nine amino acids. Foreign antigens are recognized by cytotoxic T-cells that kill the cells presenting them by inducing apoptosis. HLA class II molecules (DR, DP, and DQ) present antigens derived from the outer environment to T-helper cells, which are then stimulated to reproduce and to activate antibody-producing B-cells.

HLA molecules are among the most variable molecules in the organism because of the presence of a large number of alleles which code for their α- and β-chains. The HLA region of patients is usually genotyped in order to find the most compatible donor.

One of the best choices is to find an HLA-identical donor. In this category are enclosed siblings and relatives of the patients, with the same HLA genotypes. However, in about the 70% of cases, such donors are not available and alternative transplant options have to be found. The best option, in this case, is to find a related donor with only a partial match at the HLA superlocus, with one or more mismatched HLA alleles (haploidentical transplantation). In the last decades several attempt have been done in this direction (Mehta*, et al* 2004, Powles*, et al* 1983, Szydlo*, et al* 1997). The great advantages of haploidentical transplantation are that nearly all patients have an immediately available donor and that a stronger GvL effect can be realized (Aversa*, et al* 1998, Beatty*, et al* 1985). Nevertheless, the risk of developing severe GvHD is higher in patients who underwent haploidentical transplantation (Anasetti*, et al* 1990). It has already been mentioned that development of GvHD is closely related to the presence, in the graft, of cytotoxic T-cells which react against tissues of the patient. Steps towards reduction of GvHD occurrence and severity have been achieved by introduction of T-cell-depleted graft into transplantation protocols, as mentioned before. Haploidentical transplantation takes advantage of the anti-tumoral activity of NK cells. NK cells are inhibited by signals delivered through their surface killer immunoglobulin-like receptors (KIRs) following interactions with autologous HLA class I molecules expressed by target cells. This KIR–HLA class I interaction inhibits the cytotoxic activity of NK cells (Parham and McQueen 2003). In the absence of self-HLA class I molecules, autologous NK cells are not inhibited and they can mediate target cell lysis after proper activation, the so-called missing self hypothesis (Karre*, et al* 1986). In haploidentical transplantation, the lack of appropriate HLA class I ligands in the patient that were able to inhibit NK cells activity facilitates a strong GVL effect (Ruggeri*, et al* 1999).Co-stimulatory molecules, such as the lymphocyte function antigen 1 (LFA-1), highly expressed in AML cells, may direct NK cells to leukemic cells. NK cells also improve engraftment through elimination of hematopoietic cells of the patient. In addition, they decreased the incidence of GVHD, probably through elimination of dendritic cells in the patient that is essential for T-cell activation (Davies*, et al* 2002, Ruggeri*, et al* 2002).

1.2 Clinical observation

In the “Ludwig-Maximilian” university clinic in Grosshadern (Munich, Germany), the haematopoietic stem cell transplantation protocols with haploidentical donors are well established.

A retrospective analysis of the long-term survival of eighty-five AML patients who underwent haploidentical transplantation indicates that if a male patient is transplanted

with female graft, the patient has increased long-term survival probability. Figure 1.1 schematically represents the outcome of such study. The graph has been generated from unpublished data from the clinical cooperation group-haematopoietic cell transplantation (CCG-HCT), headed by Professor HJ. Kolb.

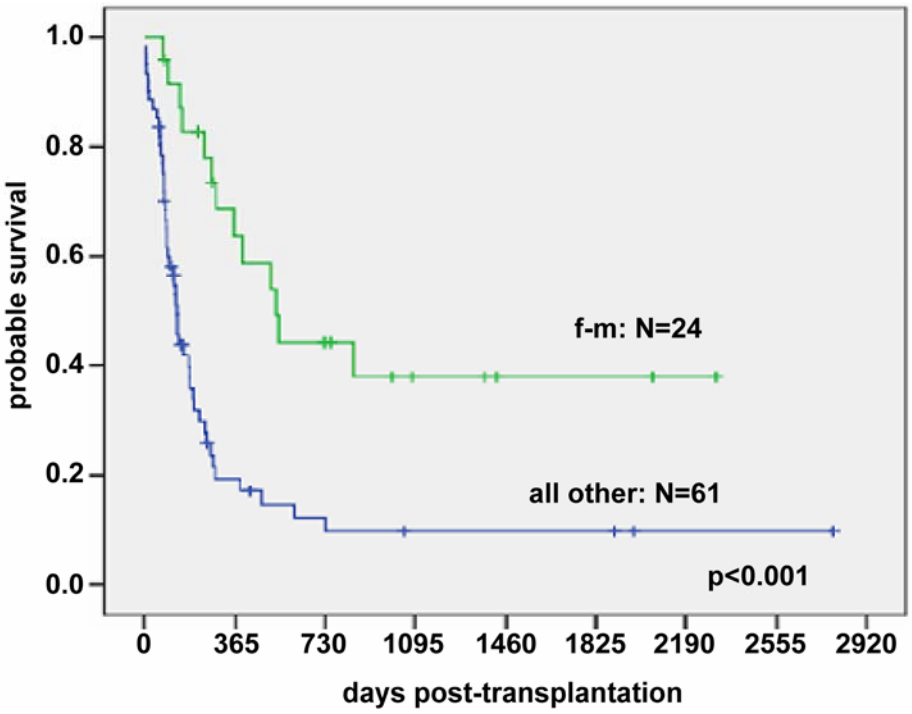

Figure 1.1. Probable survival of 85 AML patients who underwent haploidentical haematopoietic stem cell transplantation. The couples with male patient and female donor (green line) show more favourable outcome than all the other gender combinations (blue line). Unpublished data from Prof. HJ. Kolb, generated in the "Ludwig-maximilian,, university clinic; Grosshadern, Munich.

The survival rate is calculated by taking into account several parameters. Male patients who received haploidentical female graft mostly showed strong GvL effect and less complications derived from GvHD associated tissue damage. However, being GvHD and GvL effect different consequences of the same process, namely the reaction of T-cells of the graft against cells of the host, GvHD was also observed in this group of patients, although at lower degree.

This phenomenon seems to be not only a characteristic of haploidentical transplantation. It has been observed also in patients with various haematological malignancies transplanted with haematopoietic stem cells from an HLA-identical donor (Randolph*, et al* 2004). Several studies agreed that female-to-male transplantation results in occurrence of GvHD and strong GvL effect with increased long-term survival (Gahrton 2007).

Studies of allogeneic haematopoietic stem cell transplantation in murine models and in humans have highlighted the importance of minor histocompatibility antigens (minor H

antigens) of the recipient as the targets of donor T-cells that mediate GvL effect and GvHD (Faber, *et al* 1995, Fontaine, *et al* 2001, Marijt, *et al* 2003). Minor H antigens are peptides derived from cellular proteins encoded by polymorphic genes that differ between the transplant donor and recipient and that are presented to $CD8^+$ and $CD4^+$ T-cells by HLA class I and II molecules, respectively. In humans, only a small number of minor H antigens have been characterized, and the selective contribution of T-cells to GvL activity and GvHD remains speculative. It has been suggested that minor H antigens, which are limited in their expression to hematopoietic cells including leukaemic cells, may be targets for a selective GvL effect, whereas those that are broadly expressed are responsible for GvHD (Mutis and Goulmy 2002, Warren, *et al* 1998). However, the very large number of minor H antigens that are likely encoded by autosomal genes and are involved in allogeneic responses has made the analysis of the contribution of single factors of the female-to-male clinical setting more complicated.

Transplantation of stem cells from a female donor to a male recipient is a particular circumstance in which donor T-cells specific for minor H antigens, encoded by genes on the Y chromosome of the patient that are polymorphic with respect to their X chromosome homologues, may contribute to GvHD and GvL effect. Several genes on the Y chromosome have already been identified to encode minor H antigens, and the role of these antigens in GvHD and GvL responses has been investigated (Meadows, *et al* 1997, Miklos, *et al* 2005, Vogt, *et al* 2002, Wang, *et al* 1995). For example, haematopoietic and non-haematopoietic cells express antigens that are encoded by the SMCY gene and are presented by HLA-A2 and -B7 to $CD8^+$ T-cells. The T-cell response to the SMCY antigen has been put in relation with GvHD in small studies of male recipients of haematopoietic stem cell transplants from female donors (Dickinson, *et al* 2002, Mutis, *et al* 1999). In contrast, an HLA-B8-restricted peptide encoded by the UTY gene is not presented to $CD8^+$ T-cells by non-haematopoietic cells *in vitro* but is expressed on haematopoietic cells, including leukaemic progenitor cells (Bonnet, *et al* 1999, Warren, *et al* 2000). These observations suggest that it is possible that some minor H antigens encoded by the Y chromosome may be differentially associated with GvHD and GvL responses after transplanting stem cells from female donors into male recipients. The Y chromosome genes that encode minor H antigens are polymorphic with respect to their X homologues, suggesting that additional antigens may be presented by other HLA alleles and that they may also be targets for T-cells-mediated GvHD or GvL activity after female-to-male transplantation.

The human Y chromosome encodes for about fifty genes, the most of which are present in several copies. Several Y chromosome genes also generate RNA transcripts which undergo different splicing events, generating a high variety of gene products. There are Y chromosome genes that are involved in the development of male reproductive organs. Several known as well as putative transcription factors are also encoded on the Y chromosome. Some of them are supposed to be expressed only in male reproductive tissues, like testis or prostate. The effect of the expression of such genes in other tissues,

like blood or bone marrow and in particular in leukaemic cells, is poorly understood. Expression of Y chromosome-encoded genes in leukaemic cells may have a positive or negative influence on the treatment of the disease depending on the biological function of such genes.

1.3 Aim of the study

As reviewed in the introduction, haploidentical haematopoietic stem cell transplantation in AML patients has more chances of success when female graft is transplanted in male patients. The Y chromosome-encoded genes may play a primary role in this phenomenon either by generating minor histocompatibility antigens, having hence a direct anti-tumoral effect or by interfering with the biology of tumoral cells. The latter possibility may have several different implications depending on the biological function of the genes. In this context, the first aim of the present study is to investigate the expression level of all the known Y chromosome-encoded genes in male AML cells and in their healthy counterpart (aim 1).

AML is a very heterogeneous disease and the different affected cell populations can carry several genetic markers like chromosomal abnormalities, inversions, transpositions, etc. The present study is focused on male AML M4 and M5 patients. These subtypes are characterized by high expression of CD14 molecules on the leukaemic cell surface.

The choice of these particular subtypes has been taken for two main reasons. First, due to heterogeneity of this malignancy, I decided to investigate a leukaemic cell population which could be compared to the specific healthy counterpart (healthy $CD14^+$ cells). The second reason is due to an easier handling of the samples, as both AML and healthy cells were separated from peripheral blood or bone marrow using anti-CD14 beads, making the separation fast and easy and avoiding too much stress to the cells.

The Y chromosome gene expression analysis will lead to identification of those Y chromosome-encoded genes which are up-regulated in the leukaemic cells. After gene expression assessment, several aspects arose that had to be investigated. The second aim of this study is the characterisation of those genes, which were found to be differentially regulated in AML cells (aim 2). The understanding of the biological function of those genes is essential to get an idea about their possible role in the tumour biology.

Another important aspect concerns the possibility that those up-regulated genes generate antigenic peptides which might be involved in an anti-tumoral immune response. For this reason, the third aim of this study is the analysis of the antigenic potential of peptides derived from such genes (aim 3).

The general aim of this study was to contribute to the understanding of the basic mechanisms with which up-regulated genes can have an effect on the final therapy outcome of the patients. This is a prerequisite for a research aimed towards the development of new therapeutic strategies.

2 Results

The present section is divided in three parts. In the first part, the results of the Y chromosome genes expression analysis are presented. The second part is focused on the TGIF2LY gene and concerns the results of the experiments aimed to understand the function of this gene. The last part presents the data of the immunological analysis of the candidate genes, whose purpose is to understand the possible role of these genes in the immunology of the tumour.

2.1 Y chromosome gene expression profile in AML and healthy cells

2.1.1 Isolation of leukaemic and healthy $CD14^+$ cells

In order to obtain two comparable cell populations, one leukaemic and one healthy, $CD14^+$ cells have been separated from peripheral blood or bone marrow of male AML M4 and M5 patients as well as of male healthy individuals.
In the patients, such cell population represents the leukaemic monocytic cells, whereas in the healthy persons it corresponds to normal monocytes or macrophages.
Isolation of these cells was achieved by using anti-CD14 magnetic beads-conjugated antibodies and particular columns (see paragraph 4.8.2).
Before and after the magnetic separation, the samples were tested via flow cytometry to verify the enrichment of $CD14^+$ cells. Figure 2.1 shows the typical result of a flow cytometry analysis of healthy leukocytes, where it is possible to find a large amount of lymphocytes and a smaller amount of monocytes. AML M4 and M5 samples show a dramatically increased amount of monocytes.
Only samples that, after magnetic separation, showed not less than 90% of $CD14^+$ cells were used for the Y chromosome gene expression profiling analysis. Of course, a minimal contamination of $CD14^-$ cells is unavoidable. Some cell samples, obtained from leukaemic patients in the so-called blast crisis, containing more than 90% of leukaemic cells ($CD14^+$, in this case) have been directly used without prior cell sorting. More details about samples are provided in paragraph 4.8.1 and in table 4.1.

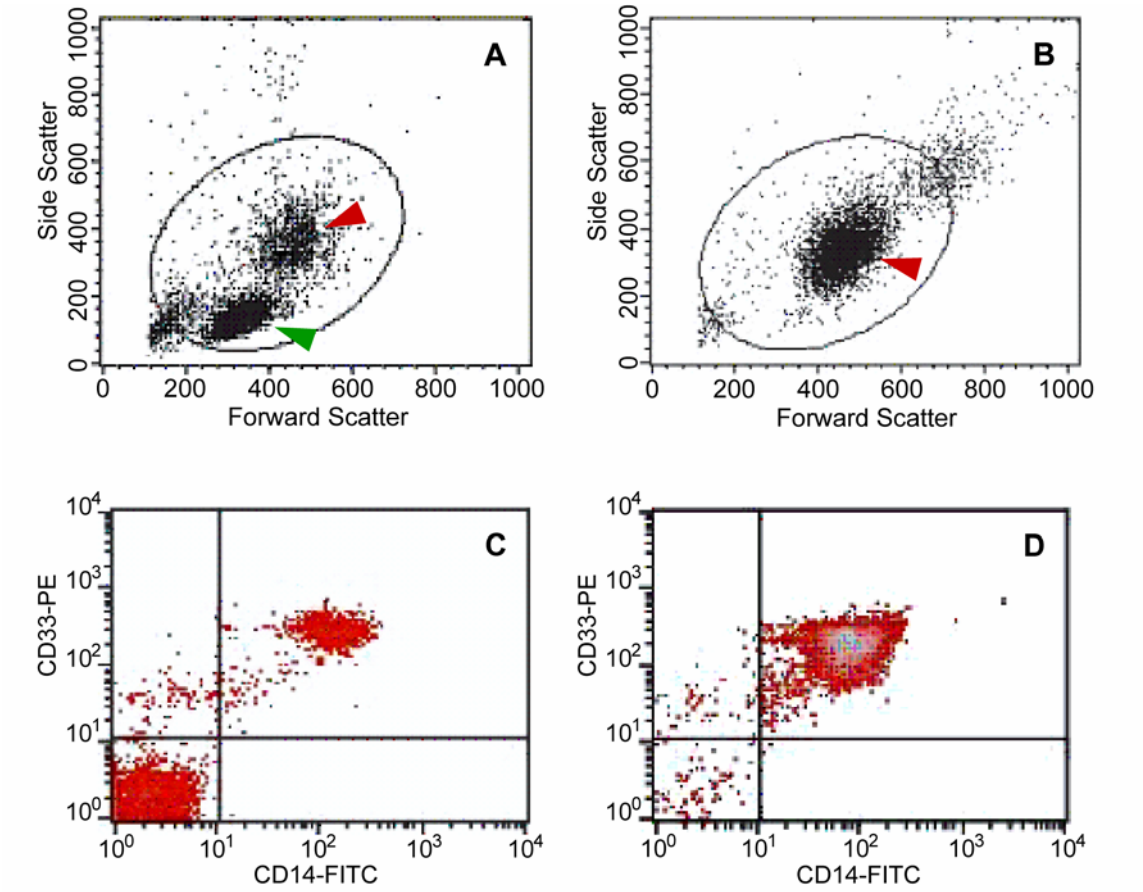

Figure 2.1. Flow cytometry profile of healthy peripheral blood mononuclear cells (PBMCs). A typical profile of non-separated healthy PBMCs (panel A) shows high abundance of lymphocytes (green arrow) and a lower amount of monocytes (red arrows). Monocytic cell population is dramatically increased after $CD14^+$ cell separation (panel B). Staining of CD33-CD14 double-positive cells allows to further verify the enrichment of $CD14^+$ cells before and after separation (panel C and D, respectively). After separation, almost no $CD14^-$ cells are detectable and the $CD14^+$ cell population reachs 97% of the total cells.

2.1.2 Low density arrays

In order to identify Y chromosome-encoded genes which might be regulated in the AML blasts, low density arrays (see paragraph 4.5.3) were used to analyse the Y chromosome gene expression in male AML M4 and M5 cells. Healthy male $CD14^+$ cells, isolated with the same procedure (see paragraph 4.8.2 for cell isolation procedure), were used as controls.

Figure 2.2 depicts the comparison of the expression of 46 Y chromosome-encoded genes, relative to the endogenous control hypoxanthine phosphoribosyltransferase 1 (HPRT1), in AML M4-M5 patients (n=11) and healthy volunteers (n=8; refer to paragraph 4.8.1 for information about patients and healthy volunteers). The histogram was calculated using the software Sequence Detection System 2.2.

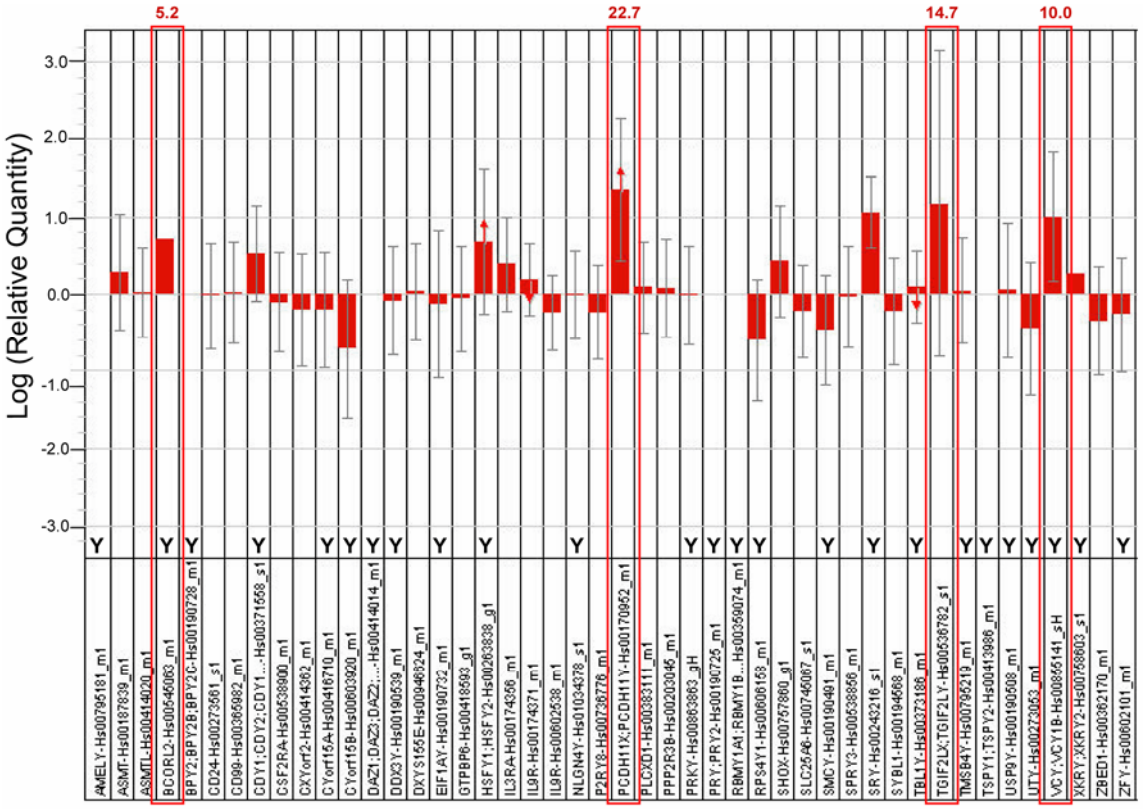

Figure 2.2. Y chromosome gene expression comparison between AML M4-M5 cells and healthy $CD14^+$ cells. The red bars refer to the gene expression level in the AML patients, relative to the expression level of genes in the healthy $CD14^+$ cells, considered as zero. Red frames indicate genes which were found to be up-regulated in the patients' probes. The numbers above them indicate their fold increase. Up-arrows atop the bars indicate the lack of expression of that gene in the healthy cells' group only. Down-arrows indicate the lack of gene expression in the patients' group. Letters "Y„ indicate that the transcript was amplified using probes specific for Y chromosome gene products.

In the histogram, all the known Y chromosome-encoded genes are set on the X-axis. The Y-axis represents the logarithm of relative quantity of template which indicates the fold difference of gene expression. The red bars represent the gene expression difference in the patients' group in comparison with the healthy volunteers' group, in which the expression levels are considered as zero. The red frames highlight genes which where found to be significantly up-regulated in the patients' group and the red numbers atop indicate the fold increase compared to the healthy samples' group. There are some genes, like HSFY, which were found to be expressed in only one patient sample and were not taken into account for further investigations (see also table 2.1). Similarly, the SRY gene was excluded as its expression was almost homogeneous among patients and healthy volunteers with the exception of one patient, in which SRY expression level was found to be very high. Therefore, in the histogram, the up-regulation of this gene is not derived from a general tendency within the patients' group, but from one patient only. The detailed analysis of expression of up-regulated genes, and their expression level in each sample is presented in table 2.1.

Gene	Patients											Healthy							
	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	D1	D2	D3	D4	D5	D6	D7	D8
BCORL2		34,0				31,3			35,1							35,2	35,9	34,3	
HSFY									36,0										
PCDH11				35,6	36,9			34,0											
SRY			37,6		35,7			36,7	35,9				35,9		35,1	35,9			
TGIF2L	32,5					30,3	35,9	27,0	34,7						36,3	31,7		34,3	
VCY		34,5				35,1	33,5		34,1						36,0			36,1	

Table 2.1. Single-sample analysis of transcript level of the six genes up-regulated in the patient group. Real time-PCR C_T values are displayed in green squares. Red squares indicate absence of expression.

It has to be mentioned that some probes included in the low density array are not able to discriminate between Y- and X-linked alleles of the gene. This is true for some genes located on the pseudoautosomal region of the two chromosomes and is due to the very high homology degree of this part of the chromosomes.

Four genes were found to be clearly up-regulated in AML M4 and M5 cells. These genes were BCORL2, PCDH11, TGIF2L, and VCY.

The first gene was BCL6 co-repressor-like 2 (BCORL2). It doesn't have an homolog on the X chromosome. BCORL2 was found to be expressed in three healthy donors and three patients. In this last group, detected fluorescence reached the threshold in average after 33.4 PCR cycles and after 35.1 cycles in the healthy group (these values are usually abbreviated to C_T). After normalization to the expression level of the endogenous control, HPRT1, the average relative expression (ΔC_T) of BCORL2, indicated as difference of the C_T values between target gene and endogenous control, was 7.5 for the healthy group and 5.2 for the patients' group. For each gene, the expression difference between the two groups of samples is calculated as ΔC_T value of the donor group subtracted from the ΔC_T value relative to the patient group ($\Delta\Delta C_T$; refer to paragraph 4.9.1). BCORL2 expression showed a 5.2-fold increase in the patients' group.

The second gene was protocadherin 11 (PCDH11). This gene possesses a homolog on the X chromosome and it was not possible to discriminate between the two homologs because the TaqMan probe used in this study was unspecific. PCDH11 expression was detected in three patients and it was not found to be expressed in any healthy sample. For this reason, it would be incorrect to talk about up-regulation of this gene although the software calculates that PCDH11 expression is 22.7-fold increased in the patients' group in comparison with the healthy samples' group.

The third gene strongly up-regulated in AML M4-M5 male cells was TGFB-induced factor homeobox 2-like (TGIF2L), a putative transcription factor. Its expression showed

a 14.7-fold increase in the patients' group. This gene is present both on chromosome Y and X. mRNAs transcripted form these two alleles are almost identical and it was not possible to use two different TaqMan probes to distinguish between the two transcripts.
The last gene found to be over-expressed in AML M4-M5 male blasts was variable charge Y-linked (VCY). The expression of this gene was 10-fold higher in the patients' group. VCY is Y chromosome-specific. It has a homolog on the X chromosome, VCX, but the sequence of its transcript is different enough from that of VCY mRNA. This made possible to use a TaqMan probe specific for the Y chromosome-derived transcript.

2.1.3 Candidate genes are not expressed in non-leukaemic cells of the patients

In order to verify that up-regulation of the previously described candidate genes was specific for the malignant cell fraction and not a general chracteristic of the patients' cells, healthy cell populations isolated from the same patients (the CD14-negative ($CD14^-$) blood cell fraction) were also investigated. For this purpose, the same real time-PCR procedure was used for the analysis as for $CD14^+$ cell samples. For most patients who underwent $CD14^+$ cell-selection, the CD14 negative cells (the healthy fraction) were available.
Figure 2.3 shows that expression of candidate genes was not observed in the healthy peripheral blood cell fraction of AML M4 and M5 patients. The graphs in figure 2.3 show the amplification plots of the four candidate genes.
The graphs describing the amplification plot of BCORL2, TGIF2LY, and VCY were obtained analyzing a sample from patient P6, who expresses all these genes. For PCDH11, a sample from patient P5 was analyzed. None of the patients' samples showed expression of all the four genes (see also table 2.1). The expression of the endogenous control gene, HPRT1, was comparable in both the $CD14^+$ and $CD14^-$ cell fractions. The C_T value of HPRT1 was 31.5 in the malignant cell population and 32.1 in the healthy cell population. In this experiment, no expression of any of the candidate genes was detected in the CD14-negative cell fraction of the patients, indicating that the expression of these genes is restricted to the leukaemic blasts.
It is possible to notice that the expression of the candidate genes in the $CD14^+$ cell fraction of the patients was very low. For each candidate gene, fluorescence levels reached the threshold always after at least 30 PCR cycle. At the present time, there are not too many available information concerning these genes (see also paragraph 3.2.4).

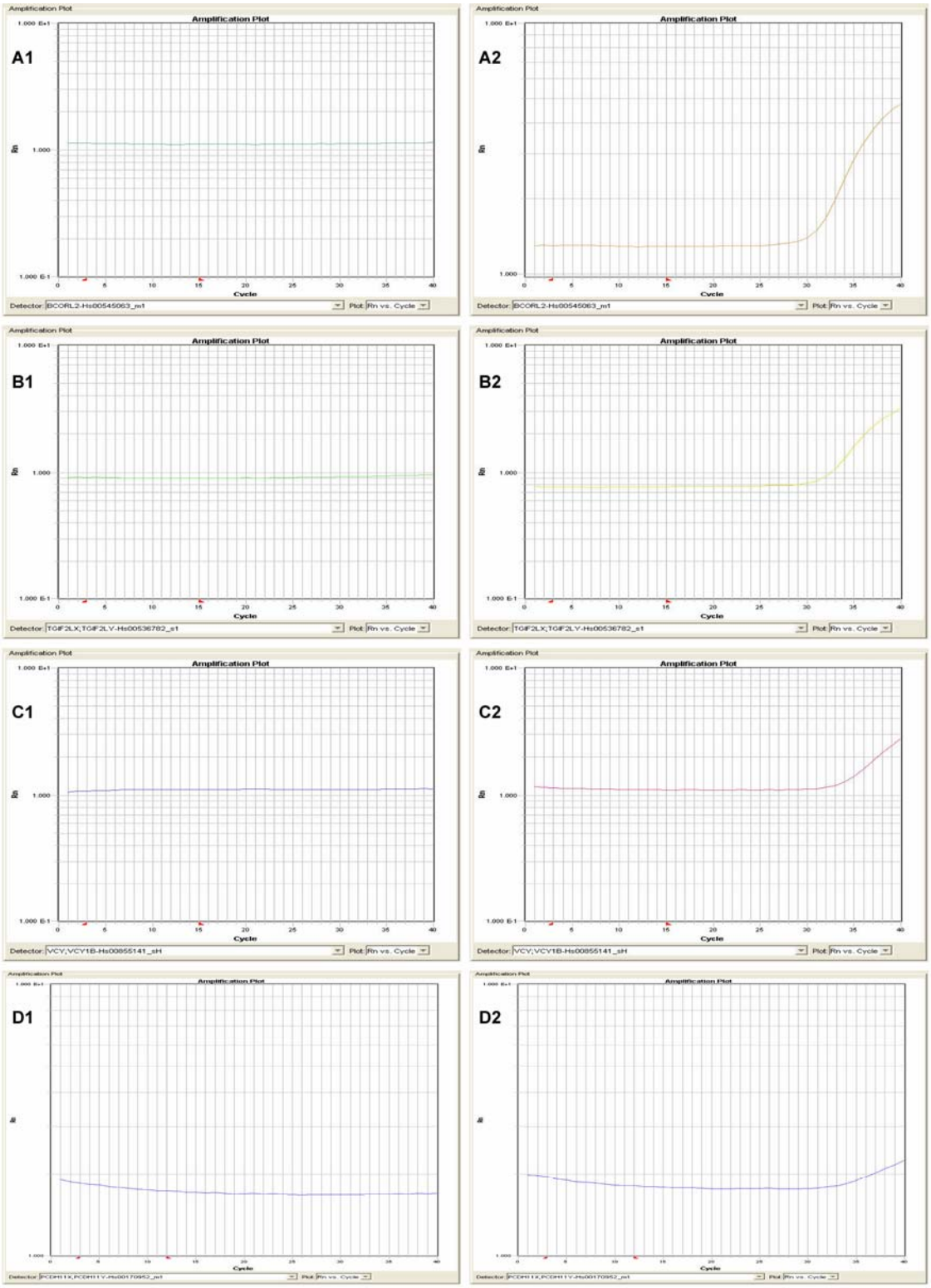

Figure 2.3. Expression of the candidate genes in the CD14-negative (panel A1-D1) and -positive (panel A2-D2) fractions of patient samples. The endogenous control, HPRT1 is expressed at comparable level in both cell populations. BCORL2 (A1-A2), TGIF2L (B1-B2), and VCY (C1-C2) are only amplifiable in the CD14-positive cell population, which corresponds to the leukaemic cells, of the patient P6. PCDH11 expression (D1-D2) is expressed in the leukaemic cell population of patient P5. On the X-axis are the PCR cycle numbers. On the Y-axis is the normalized reporter signal (R_n). This is the ratio of reporter dye fluorescence and the passive reference dye fluorescence. In this case, the inert dye ROX was used. its fluorescence emission does not change during the PCR reaction.

2.1.4 Candidate genes expression in THP-1 cells

In order to adopt a cell line model that could be used as a tool for further studies of the candidate genes in AML M4 and M5, I analysed the expression of the candidate genes in THP-1 cells. This cell line has been established from a young male patient affected by acute myelomonocytic leukaemia and it is characterized by expression of CD14 molecules on the cell surface (see also paragraph 4.2.1).
The used real time-PCR strategy was exactly the same as for the other samples. After RNA extraction and cDNA synthesis (see paragraphs 4.4.1 and 4.4.3), 100 ng of cDNA were loaded on a low density array card and the Y chromosome gene expression profile was evaluated.
Figure 2.4 and table 2.2 show the amplification plots of the four candidate genes in THP-1 cells and the comparison of Y chromosome gene expression between THP-1 cells and AML M4 and M5 patients' samples, respectively.

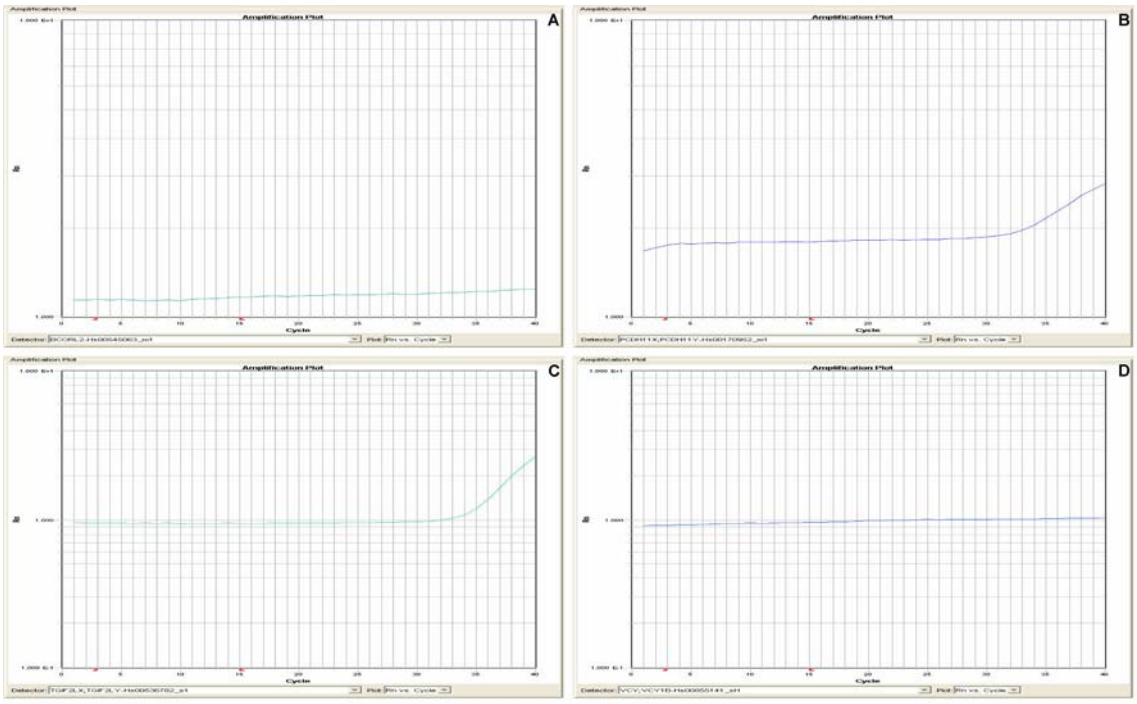

Figure 2.4. Analysis of the expression of BCORL2 (panel A), PCDH11 (B), TGIF2L (C), and VCY (D) in THP-1 cells. The same real time-PCR procedure as for the patent and healthy samples was used; endogenous control, HPRT1, was used as positive control. THP-1 cells do express PCDH11 and TGIF2L and can be used as a model system for further analysis on these two genes.

The Y chromosome gene expression profile of THP-1 cells fairly reflects that of the patients used in this study. Anyhow, there are some differences.

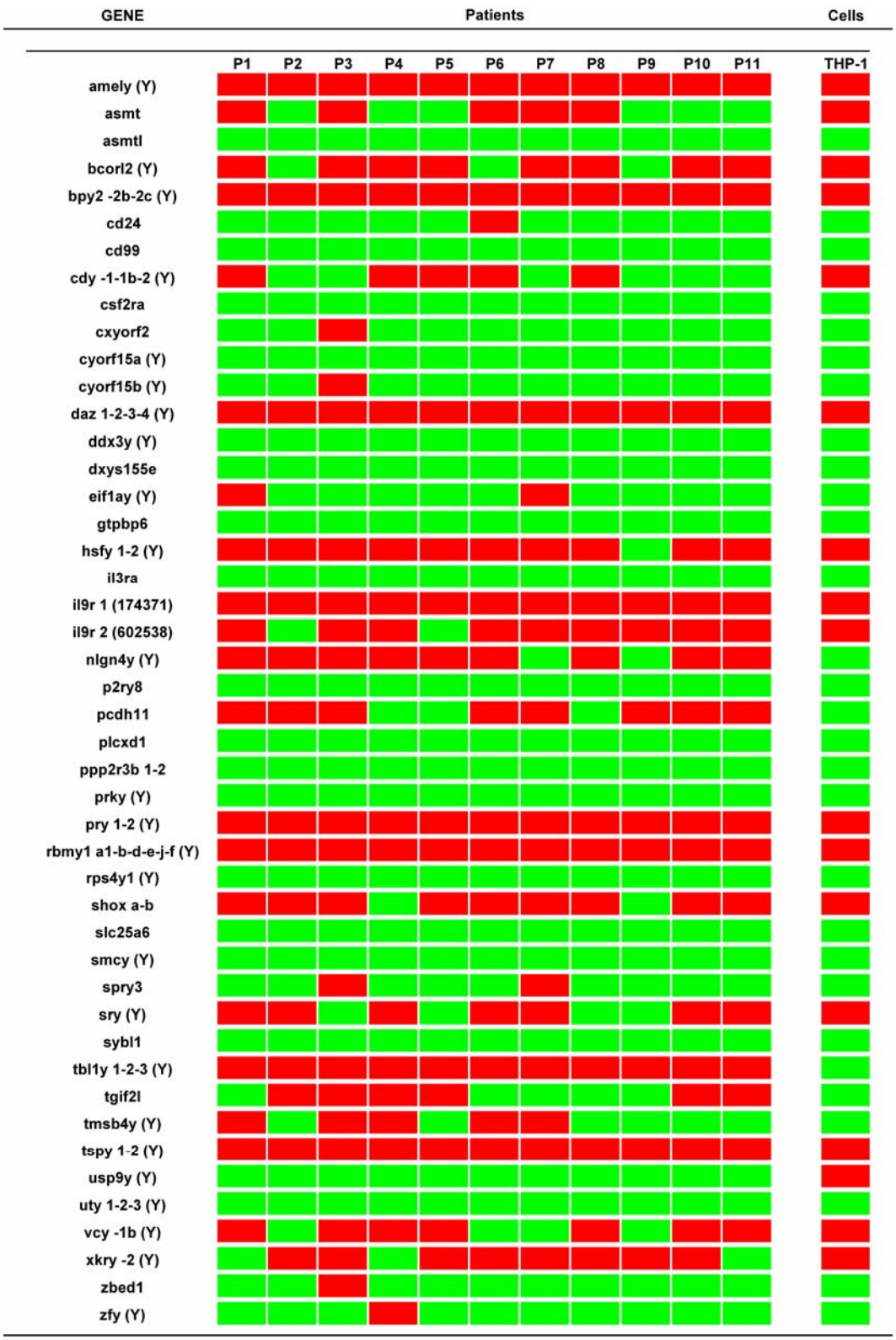

Table 2.2. Y chromosome gene expression in AML M4 and M5 patients and THP-1 cells. Green squares indicate that the corrispondent gene is expressed, while red squares indicate that it is not. the (Y) symbols indicate that the Taqman probes were specific for the Y chromosome-encoded transcripts.

All the genes found to be expressed in all the patients, excluding the USP9Y gene, were expressed also in THP-1 cells and the genes which were not expressed in all the patients were also not detectable in the cell line.Among the above-mentioned candidate genes, THP-1 cells were found to express PCDH11 and TGIF2L, athough at very low level (figure 2.4, panel B and C, respectively). Differently from AML M4 and M5 samples, the BCORL2 and VCY genes (figure 2.4, panel A and D, respectively) were found to be not expressed in THP-1 cells. The endogenous control HPRT1 was used to verify the correct functioning of the real time-PCR procedure.The results of this experiment indicate that THP-1 cells can be used as a model for AML M4 and M5 and the fact that they also express the PCDH11 and the TGIF2L genes makes them a good model to investigate the function of these genes. The low expression levels of the two candidate genes are fairly comparable with the level of expression of PCDH11 and TGIF2L in the patients' samples.

2.2 Investigating TGIF2LY function

2.2.1 Introduction

First I started to analyse more deeply all the four genes which were found to be up-regulated in AML cells, but finally, for time reasons, I decided to focus only on one of them. I chose TGIF2LY. One reason for which TGIF2LY has been chosen was because of its tissue specificity. The restriction of its expression to adult testis is supported by experimental evidences (Blanco-Arias*, et al* 2002, Skaletsky*, et al* 2003) and it would be very interesting to understand the role of this protein when it is expressed in another tissue. Another reason is that bioinformatics analysis of the protein suggested the possibility for TGIF2LY to have transcription factor activity. The fact that a transcription factor is expressed out of the only tissue where it should be present provides one more reason to deeply investigate TGIF2LY. A third reason is that there are plenty of articles in which it is shown that several genes whose expression is limited to testis are expressed also in cancer cells, often with immunogenic potential.
Therefore, the following part of this thesis refers to the characterisation of TGIF2LY.
To understand the function of a gene, it is essential to gain information about several aspects of it. For example it is important to know whether the RNA is translated into a protein and, if it happens, where the protein does localize within the cells. This will give essential indications about the gene function. In fact, it would be possible to speculate on the function of a protein basing on data concerning its expression and localisation. If

a given protein is located in a particular cellular compartment, it will probably exert a function which is related to that district. Another important issue is to understand whether the protein under investigation has one or more interactors. Their identification can provide information about the function of the gene of interest, which can be related to the function of the protein complex to which it belongs to. If the protein of interest is found to be DNA-binding, it is possible that it is involved in gene expression regulation. Therefore, in order to get information about the function of such a protein, it would be necessary to identify the gene or the genes that it regulates. Figure 2.5 presents schematically the approach used to better understand the function of TGIF2LY.

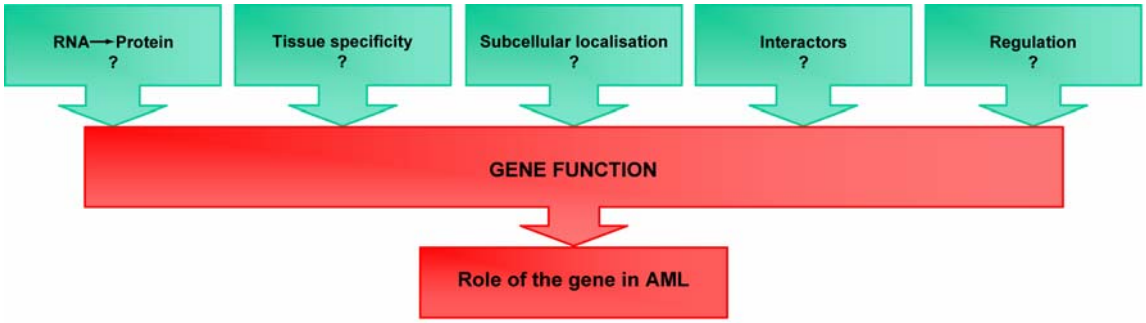

Figure 2.5. Molecular biology approach to define the potential interaction partners and the possible biological function of TGIF2LY. In order to get hints about its possible role in the disease, it is necessary to answer to several key questions.

The understanding of TGIF2LY function will be essential to understand its possible role in AML. In order to do that, the experiments were focused on TGIF2LY protein expression, subcellular localisation, and its DNA binding capacity.

2.2.2 TGIF2LY protein expression in THP-1 cells

The data concerning TGIF2LY obtained from the real time-PCR experiment (see paragraph 2.1) referred to the presence and the relative amount of mRNA transcribed from the gene. In order to verify that the mRNA is subsequently translated into protein, I used THP-1 cell lysate in a western blot experiment. The male cell line THP-1 showed expression of TGIF2L mRNA. However, the TaqMan probe used for real time-PCR experiments was not able to distinguish between Y and X chromosome homologs, because their mRNA are nearly identical.

To get an idea about the expression of TGIF2LX in AML M4 and M5, low density arrays have been used to analyze a sample of $CD14^+$ cells derived from an AML M4 female patient (patient code: AMP3) and the same cell population from healthy female peripheral blood (table 2.3).

GENE	AMP3	FHD	GENE	AMP3	FHD
amely (Y)			pcdh11		
asmt			plcxd1		
asmtl			ppp2r3b 1-2		
bcorl2 (Y)			prky (Y)		
bpy2 -2b-2c (Y)			pry 1-2 (Y)		
cd24			rbmy1 a1-b-d-e-j-f (Y)		
cd99			rps4y1 (Y)		
cdy -1-1b-2 (Y)			shox a-b		
csf2ra			slc25a6		
cxyorf2			smcy (Y)		
cyorf15a (Y)			spry3		
cyorf15b (Y)			sry (Y)		
daz 1-2-3-4 (Y)			sybl1		
ddx3y (Y)			tbl1y 1-2-3 (Y)		
dxys155e			tgif2l		
eif1ay (Y)			tmsb4y (Y)		
gtpbp6			tspy 1-2 (Y)		
hsfy 1-2 (Y)			usp9y (Y)		
il3ra			uty 1-2-3 (Y)		
il9r 1 (174371)			vcy -1b (Y)		
il9r 2 (602538)			xkry -2 (Y)		
nlgn4y (Y)			zbed1		
p2ry8			zfy (Y)		

Table 2.3. Expression of Y and X chromosome genes in CD14+ cells from a AML M4 female patient (AMP3) and a female healthy donor (FHD). The symbol (Y) indicate genes which are present exclusively on the Y chromosome; they are obviously not expressed. TGIF2LX (TGIF2L probe) is expressed in patient cells

As expected, no expression of Y chromosome-specific genes was detected. Interestingly, TGIF2LX was expressed in the patient. If activation of transcription of TGIF2LX and TGIF2LY is due to the same mechanisms, this can provide an explanation to these findings. Moreover, this result suggests the possibility that also in male patients, transcription of TGIF2LX occurs together with transcription of TGIF2LY. After this analysis, protein expression of TGIF2LY in THP-1 cells was assessed.

Using a commercially-available anti-TGIF2LY antibody, the western blot would reveal whether the Y chromosome homolog of the TGIF2L gene is expressed in this leukaemic cell line.

Proteins were extracted from three cellular compartments: cytoplasm, nucleus, and cell membranes (see paragraphs 4.6.3 to 4.6.5 for all the procedures). Pure TGIF2LY-GST

fusion protein (see paragraphs 4.6.1 and 4.6.2) was used as positive control. Figure 2.6 shows the result of the western blot.

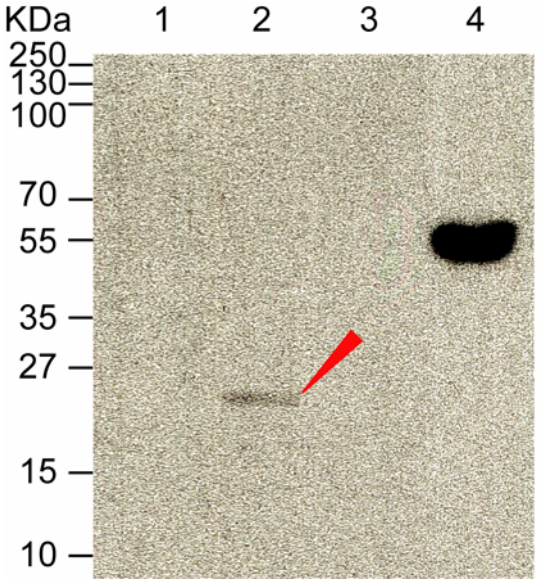

Figure 2.6. Western blot to assess TGIF2LY expression at protein level. The first three lanes represent the cytoplasmic (1), the nuclear (2) and the pellet (3) protein fraction of THP-1 cells (3×10^6). The red arrow indicates the TGIF2LY band. TGIF2LY fused with GST was used as positive control (lane 4). Primary antibody was rabbit-anti-TGIF2LY (diluted 1:4000); the secondary antibody was goat-anti-rabbit peroxidase-conjugated (1:4000). TGIF2LY expected weight: ~21 KDa. TGIF2LY-GST expected weight: ~47 KDa.

The western blot demonstrated that TGIF2LY is expressed, although weakly, in leukaemic cells. Moreover, the fact that the protein was detected only in the nuclear fraction of the cells fits perfectly to its potential transcription factor activity. However, if the protein would be present also in other cell compartments at lower level, it would probably be not possible to detect it with this method. Due to low expression levels, the TGIF2LY band would be not visible. The low protein level was somehow expected. Besides the number of cells used for the protein extraction, it has to be remembered that the RNA level of the TGIF2L transcript was also found to be rather low. In addition, several transcription factors are usually expressed at low levels in cells and their half-life could be rather short, according to the fact that their expression and function is usually limited to the moment in which the cells have to regulate particular genes. Another point that has to be kept in mind is that there is the possibility that a portion of the fluorescence signal in the real time-PCR experiment was generated by amplification of TGIF2LX RNA. In this case, the TGIF2LY protein would be derived only from one fraction of the RNA detected before. The protein expression of TGIF2LX has not been assessed in this study.

2.2.3 Subcellular localization of TGIF2LY

One very helpful method to get information about the function of a protein, as mentioned before, is to express that protein in cells and to identify the cell compartment where it is located. In order to verify the nuclear localisation of TGIF2LY, I first cloned its coding sequence, retrieved from the NCBI server, into a C-terminally MYC-tagged vector. Next, I transiently transfected HeLa cells with this construct and localized the so-expressed protein by using immunostaining and fluorescence microscopy (see paragraph 4.2.4 for the procedure).
As figure 2.7 shows, no clear co-localization of TGIF2LY and one of the used cell compartment staining could be observed. The protein seemed to locate in several cell compartments to different extent. In some cells it was located in the nuclei (figure 2.7, yellow arrows), in others it seemed to be located in the cytoplasm (white arrow), and other cells showed the protein localized in several cell compartments at the same time (red arrow). No clear information about the localization of TGIF2LY came from this experiment. It may be that the presence of the MYC-tag at the C-terminus of the protein was somehow interfering with the correct transporting of TGIF2LY in the proper cell compartment, maybe masking the signal peptide making it impossible to be recognized from the protein transporting machinery of the cells.
In order to understand and to solve this issue, I repeated the experiment by cloning the TGIF2LY coding sequence into an N-terminally MYC-tagged vector, and transfecting HeLa cells with this construct (figure 2.8).
With this setup, the experiment showed up to successful. The protein showed the same distribution pattern in all the transfected cells. TGIF2LY co-localize perfectly with the nuclear staining, as shown in figure 2.8.
These experiments provided two important evidences: the first is that the TGIF2LY protein is localized in the nuclei of the cells. This finding is perfectly in agreement with the transcription factor activity that TGIF2LY might have: after synthesis, transcription factors are usually transported to the nuclei of the cells, where they can exert their regulatory function.
The second evidence is that its signal peptide might be located at the C-terminus of the TGIF2LY protein. This information can be gathered from the ambiguous distribution pattern observed when the protein is expressed with a C-terminal MYC-tag. This tag is very small, ten amino acids, but it is still possible that its presence interferes with the recognition of the signal peptide when the tag is too close to it. In fact, a short peptide, KKKRK, typical for nuclear localization was found in position 50-54 of TGIF2LY sequence (see also paragraph 3.3.2).

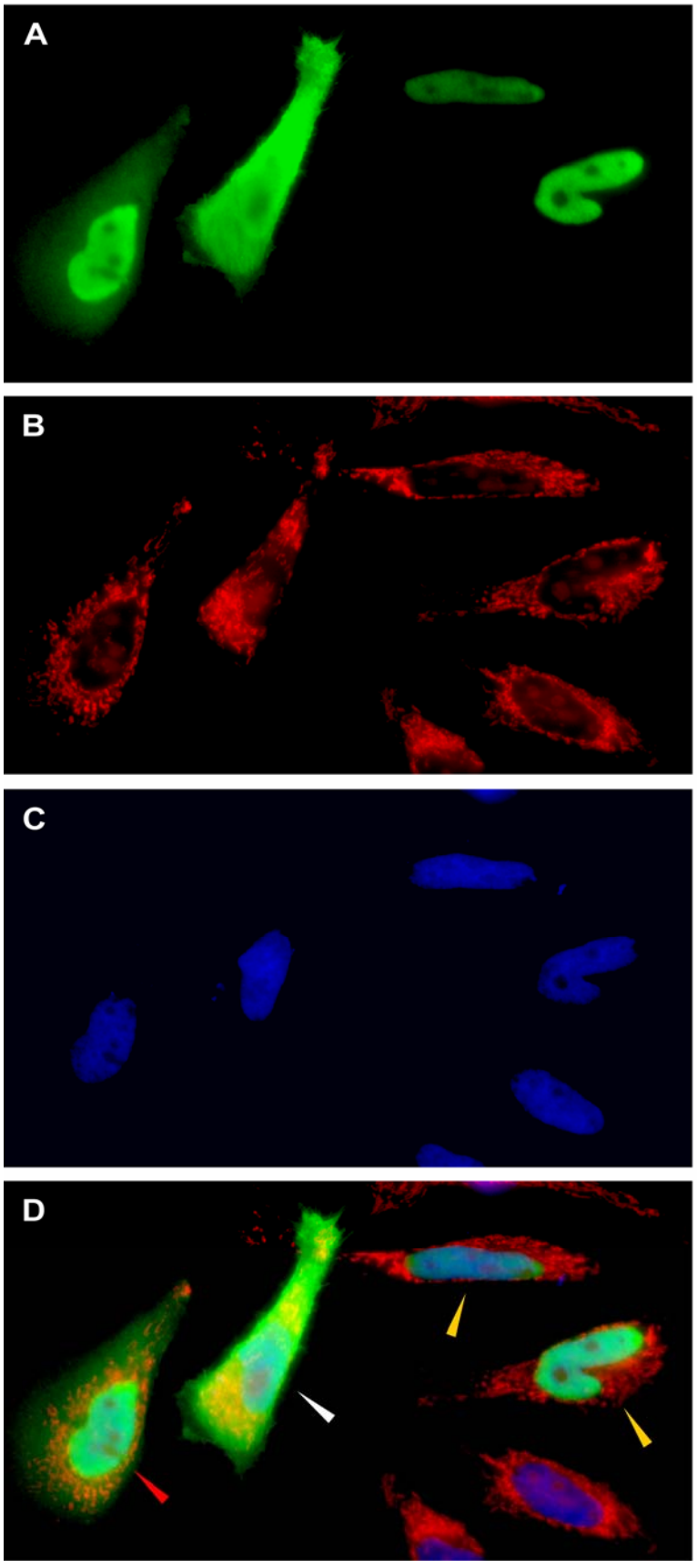

Figure 2.7. Localisation of TGIF2LY in HeLa cells after expression from C-terminal-Myc-tagged pcDNA 4 vector (panel A). The protein localizes in several cell compartments like cytoplasm and nucleus giving no clear information about protein distribution. Panel B: mitochondrial staining; C: Hoechst nuclear staining; D: overlap of all the three staining. Protein distribution, yellw arrows: nuclear; white arrow: cytoplasmic; red arrow: undefined.

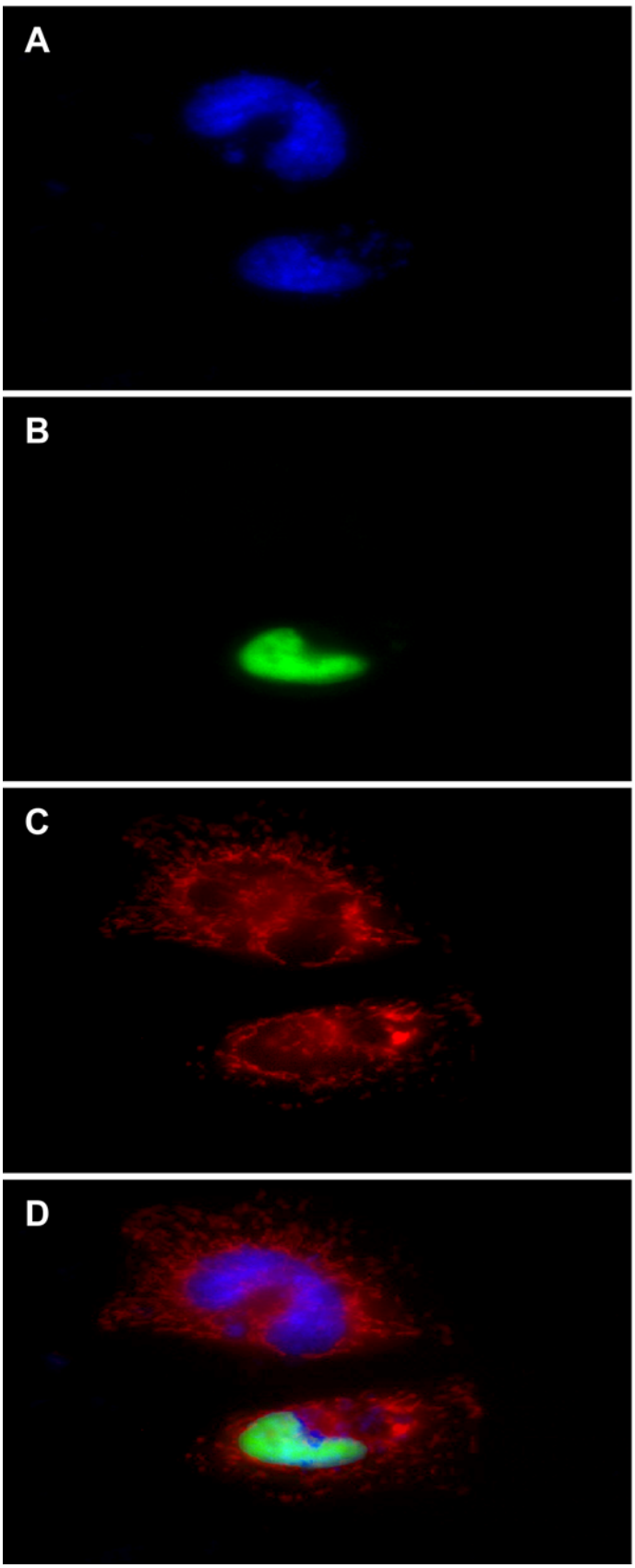

Figure 2.8. Localisation of TGIF2LY in HeLa cells after expression from an N-terminal-Myc-tagged pcDNA 3 vector. Panel A: Hoechst nuclear staining; B: localisation of TGIF2LY; C: mitochondrial staining. TGIF2LY and the nuclear staining completely overlap (panel D) in the transfected cell. TGIF2LY is distributed to the nuclei.

2.2.4 Protein function prediction using bioinformatics

Another good method to get hints about the function of a protein is to compare it with related proteins which are better characterized. If important characteristics like amino acid sequence and structure of the protein of interest show high homology with those of related proteins, it would be possible to hypothesize that they exert also a similar function.

TGIF1 and TGIF2 are two other members of the three amino acid loop extension (TALE) superfamily of transcription factors, to which also TGIF2LY belongs to. TGIF1 is a repressor of retinoic acid-induced genes (Bertolino, *et al* 1995). The second protein, TGIF2, is also a transcription repressor that exerts its function by binding directly to its target DNA sequence and recruiting histone deacetylase (Melhuish, *et al* 2001). Both proteins can also directly interact with SMAD proteins and inhibit TGF-β-induced gene expression (Chen, *et al* 2003, Wotton, *et al* 1999).

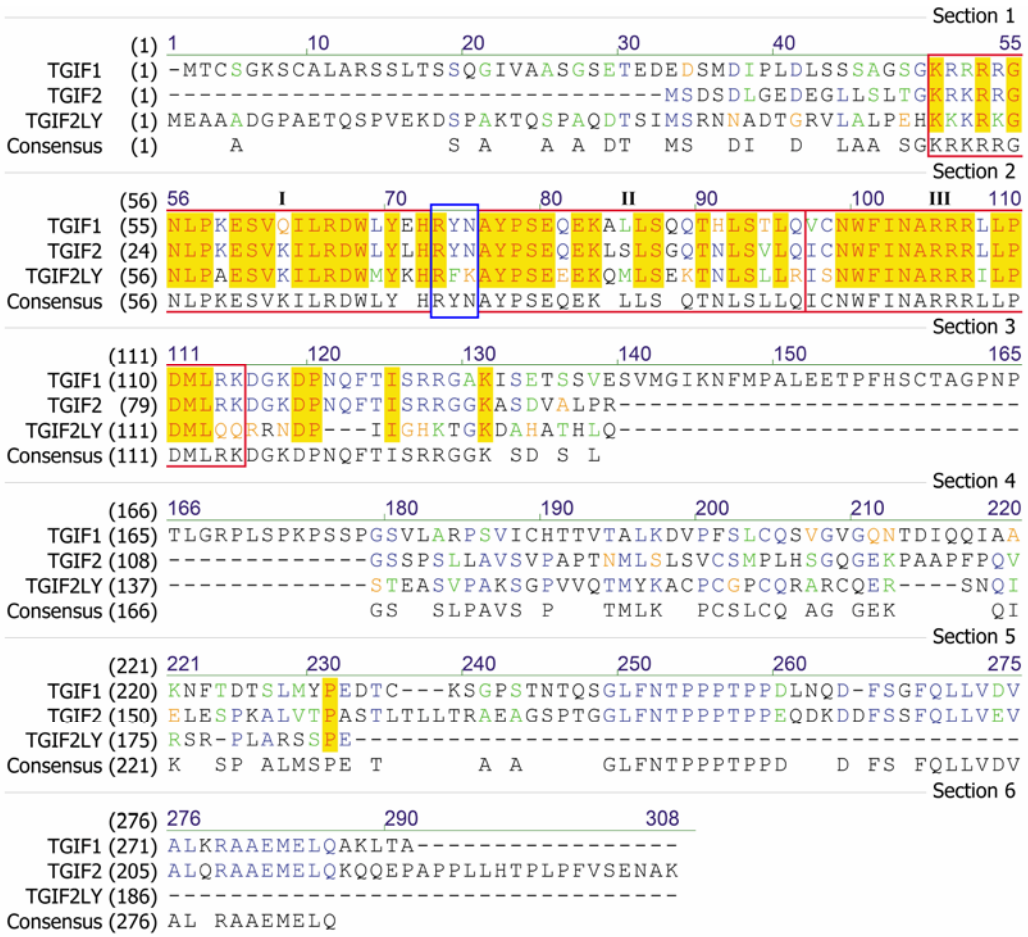

Figure 2.9. Alignment of TGIF2LY with TGIF1 and TGIF2, other two members of the three amino acid loop extension (TALE; indicated from the blue frame) homeodomain-containing transcription factors. Red frames indicate the homeodomain, with the three helices, numbered from I to III. Color scheme (color = amino acid substitution type): black = non-similar; blue = conservative; green = block of similar; red on yellow background = identical; orange = weakly similar. Dashes indicate missing amino acids.

In order to evaluate the homology of TGIF2LY and these two other TALE proteins, the first step was the comparison of their amino acid sequences. This analysis was carried out using the AlignX software (see paragraph 4.9.2).
The comparison, presented schematically in figure 2.9, shows not only that, as expected, the three proteins are closely related, but also that their homeobox regions share high homology.

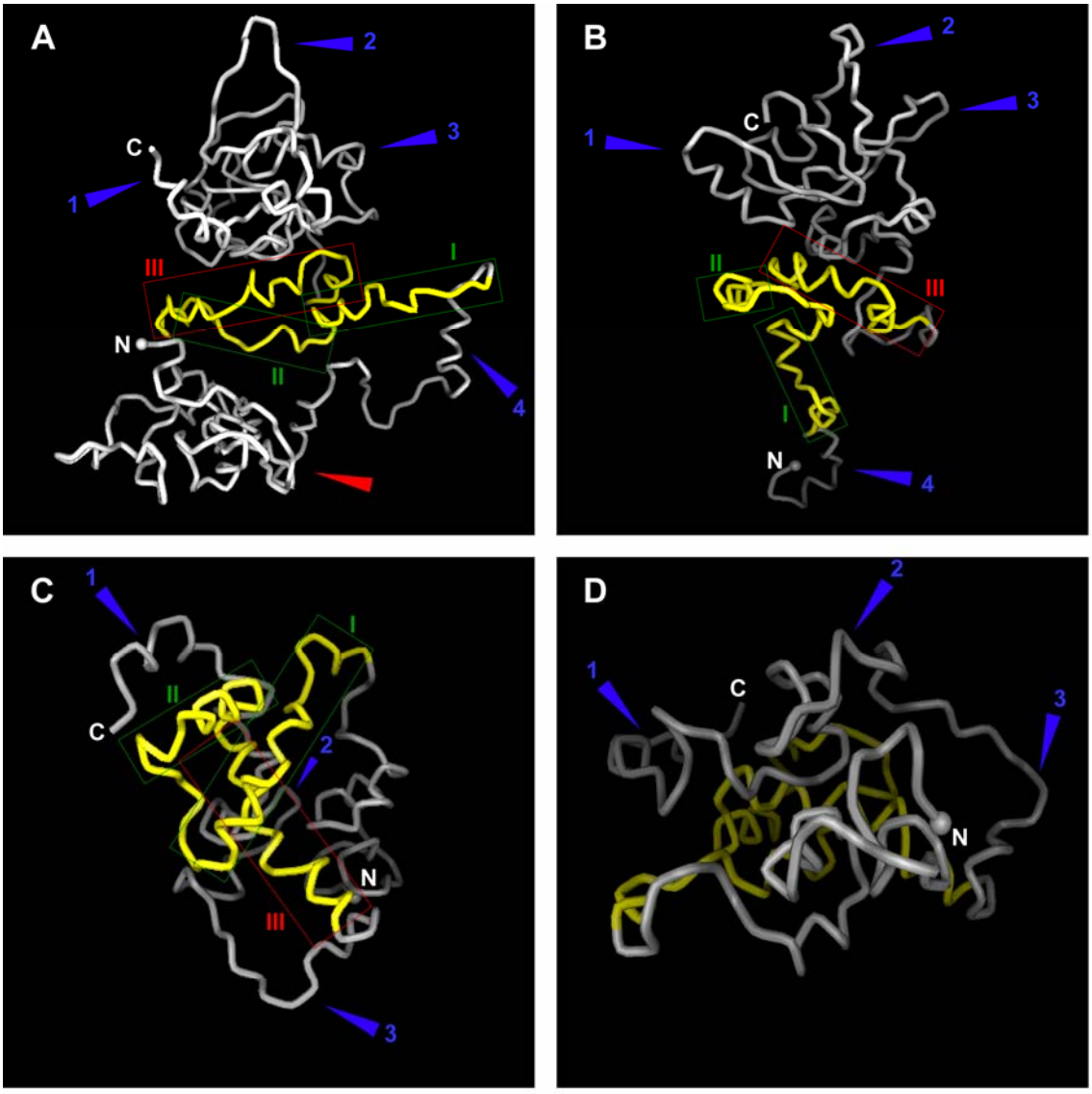

Figure 2.10. 3D models of TGIF1 (panel A), TGIF2 (B) and TGIF2LY (C and D) predicted using a publicly available prediction software. The green frames indicate helix one (I) and two (II) of the homeobox domain (in yellow). The third helix (III) is highlighted by a red frame. This last helix is the one through which many homeodomain-containing proteins bind to DNA. N- and C- termini are indicated by the letters N and C, respectively. Blue arrows and numbers indicate structures that show similar spatial organization in the three proteins. The red arrow indicates a protein region which is present in TGIF1 only. It corrsponds to 31 amino acids that are missing at the N-terminus of TGIF2 and that are dissimilar in TGIF2LY (see figure 2.8 for sequence homology in the three proteins). Predicted three-dimensional structure of TGIF2LY diverges considerably.

The three helices of the homeobox are framed in red in figure 2.9. The C-terminus of TGIF2LY diverges profoundly from those of TGIF1 and TGIF2, which appear to be relatively homologous to each other. TGIF1 has a relatively large central region, amino acid 139 to 177, that is not present in the other two proteins. The N-terminus of TGIF2 is 31 amino acids shorter that that of TGIF1.
In order to calculate and analyse the three-dimensional models of the three proteins, I used the software SCRATCH Protein Predictor (See paragraph 4.9.3). These models, presented in figure 2.10, have been positioned so that it is possible to distinguish the three helices of their homeodomains.
It is possible to see from the three-dimensional models that the organization of the homeodomains, which appear in yellow in figure 2.10, is very similar in all the three proteins. The structural organization of the C-termini of TGIF1 and TGIF2 seems to be also similar to a certain degree, especially for the structures indicated by the blue arrows and numbers in figure 2.10. The same region seems to have a slightly different organization in TGIF2LY, although the structures numbered from 1 to 3 in figure 2.10 are still recognizable in this protein. This reflects the amino acid sequence dissimilarity of the C-terminus of TGIF2LY with respect to the other two proteins.
These findings suggest that TGIF1 and TGIF2 may have the same regulatory function within the cell, because both their DNA-binding and regulatory domains appear highly conserved. TGIF2LY might actually recognize the same, or a similar, DNA sequence as the other two proteins, because of the conservation of the DNA-binding domain, but it is probable that its regulatory function is different, because of the different N- and C-terminal regions.

To direct next experiments into realistic design, I verified the homology degree of the homeobox region of TGIF2LY with other related proteins, TGIF1, TGIF2 and the protein vismay (vis), from *Drosophila melanogaster*. The function of the latter protein is not completely understood. It seems to play a role in spermatogenesis in the fruit fly (Wang and Mann 2003). The DNA sequences that are recognized and bound by these three proteins have been already identified.
The alignment of the homeobox regions of the four proteins is shown in figure 2.11.

Figure 2.11. Alignment of the homeobox regions of human TGIF2LY, TGIF1, TGIF2, and the protein vis (Vismay) from *Drosophila melanogaster*, for which the DNA consensus sequences are described. The homology between these Homeoboxes is very high, expecially for the helix three, which is responsible for DNA sequence recognition. The red and the blue frames indicate the homeodomain, with the three helices, and the TALE region respectively. The amino acid positions which are important for DNA recognition are highlighted by green frames. Color scheme (color = amino acid substitution type): black = non-similar; blue = conservative; green = block of similar; red on yellow background = identical; orange = weakly similar. Dashes indicate missing amino acids.

According to figure 2.11, it is possible to observe that the portions of the homeobox, which are crucial for DNA sequence-specificity, are rather identical in these proteins. In TGIF2LY at the positions 2 and 5 of the homeobox, the arginine is conservatively substituted by a lysine; all the five key-positions within the third helix, marked in red in figure 2.12, are conserved.

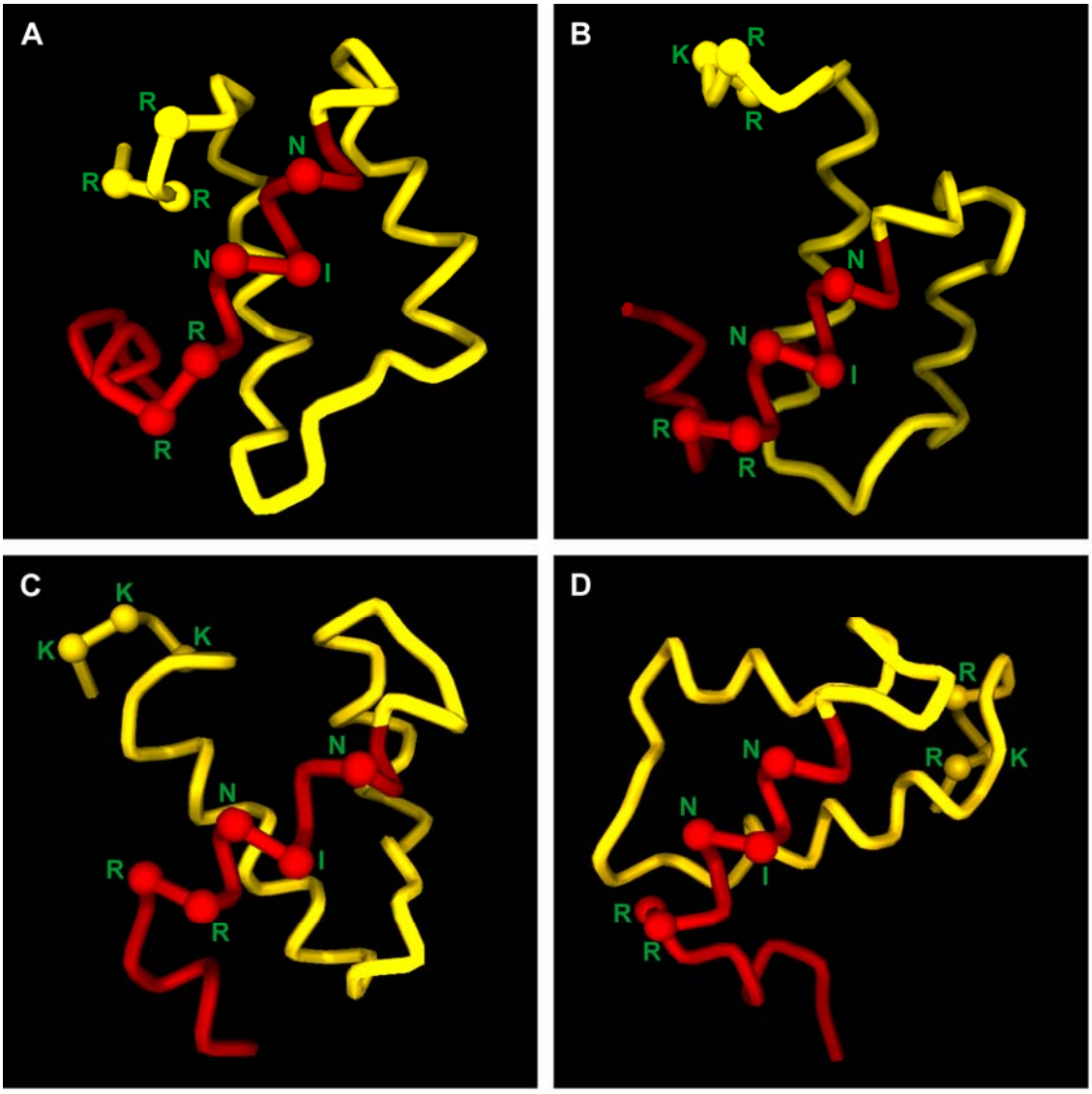

Figure 2.12. Three-dimensional structure of the homeobox region of TGIF1 (panel A), TGIF2 (B), TGIF2LY (C), and vismay (D). Third helices, mainly responsible for DNA sequence recognition, are highlighted in red. Residues involved in DNA recognition are indicated by green letters. The three-dimensional structure of the four proteins is highly similar and amino acids involved in DNA recognition are conserved.

The DNA consensus sequence of the protein vis (5'-TGACA-3') has been experimentally verified and the DNA sequence recognized by TGIF1, as well as by TGIF2, (5'-TGTCA-3') has also been published (Bertolino, *et al* 1995, Melhuish, *et al*

2001). The two DNA sequences look very similar and, due to the similarity of the homeobox region of TGIF2LY with those of TGIF1, TGIF2, and vis, it might be possible that TGIF2LY binds to a similar DNA sequence or even to the same one.
Another hint in this direction is provided by the models of the three-dimensional structure of the four proteins, presented in figure 2.12. In the models, the third helix, mainly responsible for the DNA consensus sequence recognition, is highlighted in red and the amino acids which are important for such a process are indicated by green letters (Noyes*, et al* 2008).
It is possible to notice that the three helices of the homeoboxes of the four proteins have a similar spatial organization. The five key-amino acids on helix three are identical in all the four proteins. A difference is represented by the three amino acids on helix one, in the positions 2, 3, and 5. These residues are also involved in DNA recognition. The motif arginine-lysine-arginine (R-K-R), present in TGIF2 and vis, is different, although conserved, in the other two proteins. TGIF1 contains three arginines (R-R-R) in these positions whereas TGIF2LY has three lysines (K-K-K). It is hence possible that the four proteins recognize a similar DNA consensus sequence.

2.2.5 TGIF2LY-DNA interaction

There are several techniques which allow the investigation of the interactions between proteins and DNA. One of the most used is chromatin immunoprecipitation and all its variants. In this method, cells are usually transfected with an expression vector containing the protein of interest. Afterwards, the protein is covalently bound to DNA through cross-linking and the DNA is mechanically fragmented. Antibodies specific for the protein under investigation are used to precipitate DNA-protein complexes, from which the protein is then removed. The so-purified DNA fragments are then ready to be amplified and sequenced.
Limitations for the use of this technique are the need of high amount of specific antibody and the high time need. In fact, besides the time needed for preparation of the expression vector, the time necessary for cell culture has to be taken into account.
During this study I had no time to generate and test antibodies for chromatin immunoprecipitation. For this reason I decided to implement another method for the identification of the DNA consensus sequence of TGIF2LY.
In Shostak *et al* (2004), a method called "*in vitro* genomic selection" has been described. Briefly, in this protocol the protein of interest was expressed and purified as GST-fusion protein. Genomic DNA from cells was digested and the fragments were incubated with the fusion protein. Glutathione (GT)-sepharose beads were then used to isolate protein-DNA complexes (see also paragraph 4.6.6).

I decided to partially modify this method and, in this study the term "*in vitro* genomic selection" will always refer to the modified protocol. The use of magnetic beads-conjugated with anti-GST antibodies and a magnetic separator, allowed a faster procedure. Figure 2.13 presents schematically the method used in this experiment.

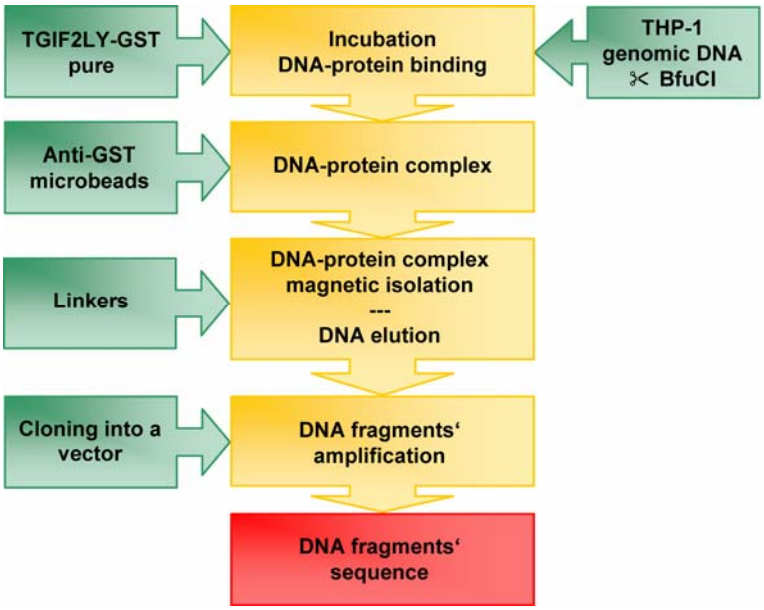

Figure 2.13. Schematic presentation of the procedure of the modified *in vitro* genomic selection. Using Anti-GST magnetic beads and a magnetic separator, the procedure results faster. DNA linkers are necessary for the amplification of the isolated DNA fragments. Two vectors have been used to sequence the isolated DNA fragments: pCR II, for shorter fragments and pCR-XL-TOPO, for longer fragments. For both vectors the TOPO-cloning procedure was used.

I cloned the TGIF2LY coding sequence into a pGEX vector, expressed the TGIF2LY-GST fusion protein in *E. coli* BL21 DE3 codon plus RP, and purified it (figure 2.14; see also paragraph 4.6.2 for the description of the method).

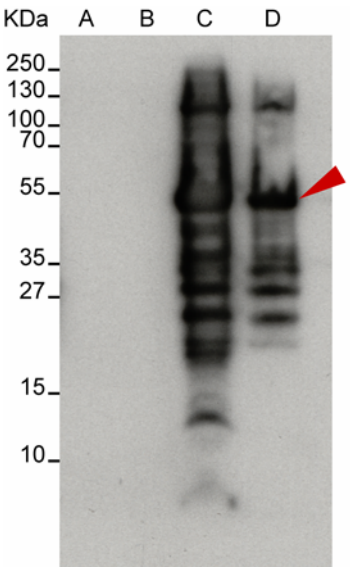

Figure 2.14. Western blot to verify the expression of TGIF2LY-GST fusion protein. The coding sequence of TGIF2LY was cloned into pGEX delta-BamHI and expressed in E. coli BL21 DE3 codon plus RP fused with an N-terminal GST. Lane A and B represent the non-induced bacteria lysate pellet and supernatant respectively, whereas lane C and D represent pellet and supernatant, respectively, of induced bacteria lysate. The red arrow indicate the band corresponding to GST-TGIF2LY protein.
1st antibody: mouse anti-GST (1:4000); 2nd antibody: goat anti mouse peroxidase conjugated (1:4000). Chemiluminescent staining (ECL). TGIF2LY-GST expected weight: ~47 KDa.

For identification of TGIF2LY DNA consensus sequence and, subsequently, of genes regulated by TGIF2LY, genomic DNA from THP-1 cells was restricted with BfuCI and incubated with the fusion protein. DNA-protein complexes were "fished" out using the magnetic beads-conjugated with anti-GST antibodies and a magnetic separator (see paragraph 4.6.6). Isolated DNA was amplified, cloned into two different vectors, and sequenced. In order to obtain information about annotated genes which could be bound by TGIF2LY, the sequences of the so-isolated DNA fragments were blasted against a human genomic database. BLAST results are shown in table 2.4 for fragments cloned into pCR-XL-TOPO vector, which allows cloning of longer fragments and in table 2.5 for those cloned into pCRII-TOPO vector.

Clone ID	BLAST result (Human G+T)	Involvement in Cancer
XL 4	Homo sapiens chromosome 10 genomic contig, alternate assembly - Features in this part of subject sequence: inositol polyphosphate-5-phosphatase A (INPP5A)	Possible
XL 8	Homo sapiens chromosome 18 genomic contig, alternate assembly - Features in this part of subject sequence: sialic acid binding Ig-like lectin 15 (SIGLEC15)	No info
XL 14	Homo sapiens chromosome 15 genomic contig, alternate assembly - Features flanking this part of subject sequence: 10098 bp at 5' side: GRINL1A combined protein isoform 8 (GCOM1); 34095 bp at 3' side: cingulin-like 1 (CGNL1)	No info
XL 15	Homo sapiens chromosome 10 genomic contig, reference assembly - Features flanking this part of subject sequence: 21619 bp at 5' side: dickkopf homolog 1 precursor (DKK1); 429229 bp at 3' side: soluble mannose-binding lectin precursor (MBL2)	DKK1 - Yes MBL2 - Possible
XL 17	Homo sapiens chromosome 6 genomic contig, alternate assembly - Features in this part of subject sequence: kelch-like 32 (KLHL32)	No info

Table 2.4. BLAST identity of DNA fragments recognized from TGIF2LY. DNA fragments were cloned into pCR-XL-TOPO vector via TA-TOPO-cloning. Their sequences were blasted against a human genomic database (NCBI). Annotated genes located within the fragments or in their proximity are highlighted in blue. Information about involvement in cancer were retrieved from the Atlas of Genetics and Cytogenetics in Oncology and Haematology (http://atlasgeneticsoncology.org).

Clone ID	BLAST result (Human G+T)	Involvement in Cancer
2	Homo sapiens chromosome 19 genomic contig, reference assembly - Features flanking this part of subject sequence: 3864 bp at 5' side: eukaryotic translation elongation factor 2 (EEF2); 18159 bp at 3' side: protein inhibitor of activated STAT, 4 (PIAS4)	EEF2 - Possible PIAS4 - Possible
3	Homo sapiens chromosome 17 genomic contig, alternate assembly - Features flanking this part of subject sequence: 25449 bp at 5' side: ubiquitin B precursor (UBB); 2066 bp at 3' side: proline rich 6 (PRR6)	UBB - Possible
6	Homo sapiens chromosome 14 genomic contig, reference assembly - Features in this part of subject sequence: telomerase-associated protein 1 (TEP1)	Possible
8	Homo sapiens chromosome 17 genomic contig, alternate assembly - Features flanking this part of subject sequence: 340210 bp at 5' side: NLR family, pyrin domain containing 1 isoform 2 (NLRP1); 157629 bp at 3' side: WSC domain containing 1 (WSCD1)	NLRP1 - Possible
10	Homo sapiens chromosome 1 genomic contig, alternate assembly - Features in this part of subject sequence: coiled-coil domain containing 19 (CCDC19)	No info
19	Homo sapiens chromosome 4 genomic contig, reference assembly - Features in this part of subject sequence: shroom family member 3 protein (SHROOM3)	No info
23	Homo sapiens chromosome 2 genomic contig, alternate assembly - Features flanking this part of subject sequence: 21131 bp at 5' side: protein C (inactivator of coagulation factors Va and VIIIa) (PROC); 55476 bp at 3' side: mitogen-activated protein kinase kinase kinase 2 (MAP3K2)	PROC - Possible MAP3K2 - Possible
25	Homo sapiens chromosome 8 genomic contig, reference assembly - Features in this part of subject sequence: cadherin 17 precursor (CDH17)	Possible
19 B	Homo sapiens chromosome 9 genomic contig, alternate assembly - Features in this part of subject sequence: CDK5 regulatory subunit associated protein 2 isoform b; CDK5 regulatory subunit associated protein 2 isoform a (CDK5RAP)	Possible
37 B	Homo sapiens chromosome 17 genomic contig, alternate assembly - Features in this part of subject sequence: forkhead box K2 (FOXK2)	Possible
45 B	Homo sapiens chromosome 2 genomic contig, reference assembly - Features flanking this part of subject sequence: 98004 bp at 5' side: hypothetical protein; 61467 bp at 3' side: similar to hCG17037	No info
5 B	Homo sapiens chromosome 3 genomic contig, reference assembly - Features flanking this part of subject sequence: 26248 bp at 5' side: hypothetical protein LOC131831; 11484 bp at 3' side: seven in absentia homolog 2 (SIAH2)	Possible

Table 2.5. BLAST identity of DNA fragments recognized from TGIF2LY. DNA fragments were cloned into pCRII-TOPO vector via TA-TOPO-cloning. Their sequences were blasted against a human genomic database (NCBI). Annotated genes located within the fragments or in their proximity are highlighted in blue. Information about involvement in cancer were retrieved from the Atlas of Genetics and Cytogenetics in Oncology and Haematology (http://atlasgeneticsoncology.org).

The “Atlas of genetics and cytogenetics in oncology and haematology” (see paragraph 4.9.5) was used to find out described involvement in cancer of annotated genes located

within the isolated DNA fragments or in their proximity. Such analysis was performed to get an idea about a possible direct involvement of TGIF2LY in carcinogenesis.
This result provides information about the DNA sequence recognized *in vitro* from TGIF2LY.

2.2.6 Identification of TGIF2LY binding site

The next step was to analyse all the sequences obtained from the *in vitro* genomic selection experiment in order to find short conserved DNA sequences, shared by all the isolated fragments, where it is possible for TGIF2LY to bind.
Plain alignment of the fragments' sequences did not show any relevant hit that could be the core binding site for TGIF2LY. Therefore, these sequences were analysed with the software AlignX, with which it is possible to put two sequences on the X- and Y- axis of a plot and identify similarities which are, otherwise, not noticeable (see paragraph 4.9.2 and figure 4.7 for the analysis description).
As mentioned before (see paragraph 2.2.4), TGIF1, TGIF2, and vis do recognize a similar nucleotide sequence, 5'-TGTCA-3' for the first two proteins and 5'-TGACA-3' for the last one. For this reason the first step was to look for a similar sequence in all the fragments obtained with the *in vitro* genomic selection.
By this, the only short DNA sequence, present in all the isolated fragments and similar to the binding site of TGIF1, TGIF2, and vis was found to be 5'-TGA(C/T)(A/C/T)-3'.
The presence of this sequence in all the isolated fragments (see appendix A.2) and the strong similarity of this sequence with the consensus sequence of the protein vis made it a good candidate as consensus sequence of TGIF2LY.
In order to verify the hypothesis that TGIF2LY is able to bind the sequence 5'-TGA(C/T)(A/C/T)-3', an electrophoretic mobility shift assay was performed. The results are shown in figure 2.15.

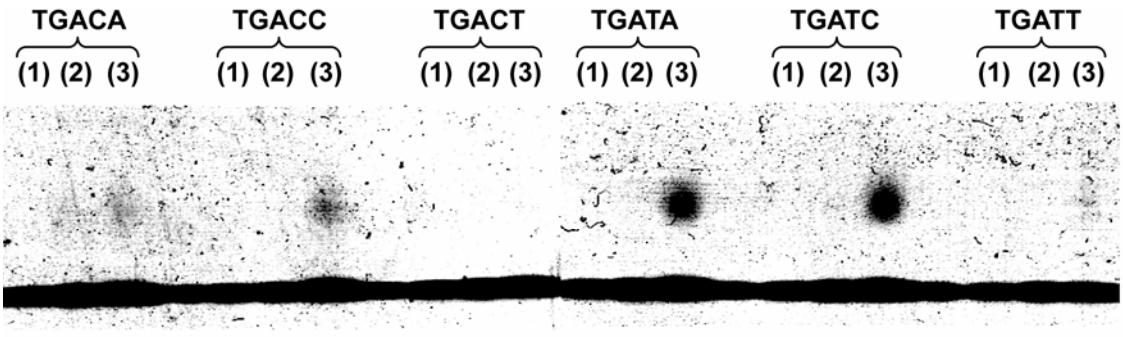

Figure 2.15. Identification of TGIF2LY binding site through EMSA. TGIF2LY shows high affinity for the sequences TGATA and TGATC and low affinity for the sequences TGACC and TGACA. (1): no protein; (2): purified GST; (3): purified TGIF2LY-GST fusion protein.

Every oligonucleotide was incubated with the purified fusion protein TGIF2LY-GST, with a GST control, and without any protein in separate reactions. TGIF2LY showed high affinity for the sequences 5'-TGATA-3' and 5'-TGATC-3'. The protein was also able to recognize the sequences 5'-TGACC-3' and 5'-TGACA-3', although with much lower affinity. No shift was observed when the oligonucleotides were incubated with the GST protein, indicating absence of unspecific binding.

2.3 Immunological potential of the candidate genes

2.3.1 Introduction

The other line of research presented in this study was focused on understanding whether the candidate genes might have an immunological impact.
Therefore, an assessment of antigenicity of candidate genes-derived peptides was necessary. The experiments described in the following paragraphs were based on the idea that an antigenic peptide derived from a protein which is over-expressed in leukaemic cells might be a good target for specific adoptive immune therapy. If such antigenic peptide would be presented on the cell surface in association with MHC molecules, it could be possible to isolate peptide-specific T-cells from an HLA-matched donor and to infuse them into the patient, where they would be able to recognize cells expressing that peptide and to initiate an immune response against them. Putative MHC-binding peptides were identified through bioinformatics prediction. The real binding capacity of these candidate peptides as well as their antigenicity was then analyzed *in vitro*.

2.3.2 Prediction of HLA-A0201-restricted peptides

In order to identify immunogenic candidate genes-derived peptides, the amino acid sequences of BCORL2, PCDH11Y, TGIF2LX, TGIF2LY, and VCY were submitted to three publicly available programs ("Bimas", "Syfpeithi", and "Paproc" see paragraphs 4.7.1 and 4.9.4). These literature-based algorithms provide information about HLA molecule-restricted peptides, which can be generated from a given amino acid chain. TGIF2LX was enclosed in this analysis because it was not possible to exclude its

expression in AML M4 and M5 cells, due to non-specificity of the TaqMan probe used for the real time-PCR analysis (see paragraphs 2.1.2 and 3.2.2).
Table 2.6 shows the high-scoring HLA-A0201-restricted peptides, derived from the candidate genes, which have been further investigated (see paragraph 4.9.4 for scoring system).

Peptide code	Sequence	Gene
F217	FLLLVDAAV	TGIF2LX
G185	GIAQPKKKV	
LQ94	LLQISNWFI	
H134	HLQSTEASV	TGIF2LY
LR94	LLRISNWFI	
N56	NLPAESVKI	
Q84	QMLSEKTNL	
G54	GAATKMAAV	VCY
K37	KVAEKGEAV	
P113	PLSKGRPST	
Q106	QESELEEPL	

Table 2.6. HLA-A0201-restricted peptides derived from genes overexpressed in AML patients' samples. The peptide code refers to the first amino acid of each peptide and to its position within the protein sequence.

High-scoring HLA-A0201-binding peptides were identified only for the genes TGIF2LX, TGIF2LY, and VCY. No high-scoring peptides, HLA-A0201-restricted, were identified for the genes BCORL2 and PCDH11Y.

2.3.3 Peptide-HLA-A0201 binding

In order to confirm the predicted HLA-A0201 binding potential of the selected peptides, a HLA-A0201-peptide binding assay was performed by using the HLA-A0201^{+} TAP-deficient T2 cell line (see paragraph 4.7.2 for the assay's description and paragraph 4.2.1 for information about T2 cells).
All the eleven peptides were screened as well as a known HLA-A0201-binding peptide derived form the influenza virus matrix protein (IMP, positive control) and two non-HLA-A0201-restricted peptides (A98ID and I540S, negative controls). Peptides were

titrated (2.5, 5, and 10 µg/mL, see paragraph 4.7.2) and the best binding efficacy was observed with a peptide concentration of 5 µg/mL. The results are shown in figure 2.16.

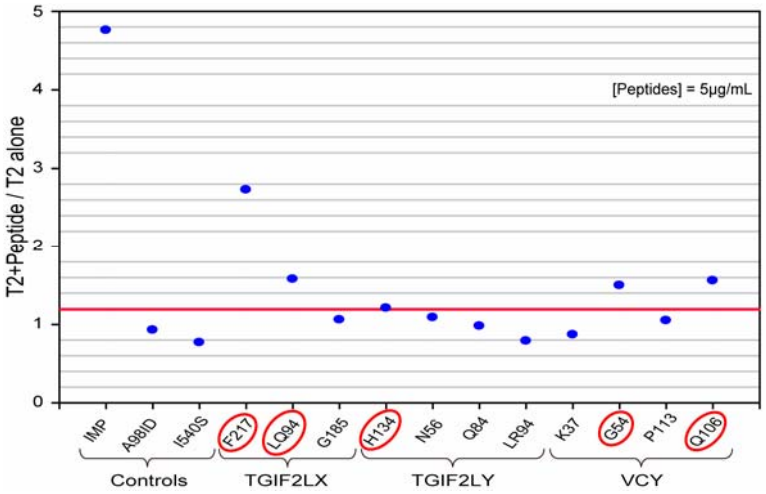

Figure 2.16. HLA-A0201-peptide binding assay. 11 candidate peptides were tested for their HLA-A0201 binding capacity. Two peptides derived from TGIF2LX (F217 and LQ94), one derived from TGIF2LY (H134), and two derived from VCY (G54 and Q106), circled in red, as well as the positive control (IMP) were able to bind HLA-A0201 molecules, whereas the negative controls (A98ID and I540S) were not. The red line represents the treshold value (1.2) under which the peptides are considered as non-HLA-A0201-binding.
Peptide concentration: 5µg/mL.

The binding efficacy was calculated by the ratio of T2 cells incubated with the peptides and T2 cells alone. Presence of the peptides on the cell surface was determined by using an anti-HLA-A0201 PE-conjugated antibody and flow cytometry.
Although none of the screened peptides reached the binding efficacy of the positive control (IMP value: 4.8), two peptides derived from TGIF2LX (F217: 2.8 and LQ94: 1.6) and another two derived from VCY (G54: 1.5 and Q106: 1.6) were clearly above the threshold value of 1.2, thus they were able to bind the HLA-A0201 molecules. The peptide H134, derived from TGIF2LY, showed very low HLA-A0201 binding capacity (value: 1.25), but it was anyhow selected for further analysis.
In addition, the peptides G185 (derived from TGIF2LX), N56, Q84, LR94 (TGIF2LY), K37, and P113 (VCY) showed binding efficacy values (range 0.8-1.1) comparable with those of the negative controls A98ID (0.9) and I540S (0.8).

2.3.4 Antigenicity of the peptides

In further experiments, the immunogenicity of the HLA-A0201-binding peptides F217, LQ94 (derived from TGIF2LX), H134 (TGIF2LY), G54, and Q106 (VCY) was analysed. Therefore, these peptides were tested for their ability to generate a $CD8^+$ T-cell response. $CD3^+$ T-cells were separated from the PBMCs of female volunteers (who gave birth to at least one male child) and cultivated in presence of T2 cells loaded with the different peptides. After four weeks of *in vitro* stimulation, the cytotoxic capacity of

the so-expanded $CD8^+$ T-cells was assessed by a classical chromium release assay (see paragraph 4.7.3), representing the antigenicity of the peptides.

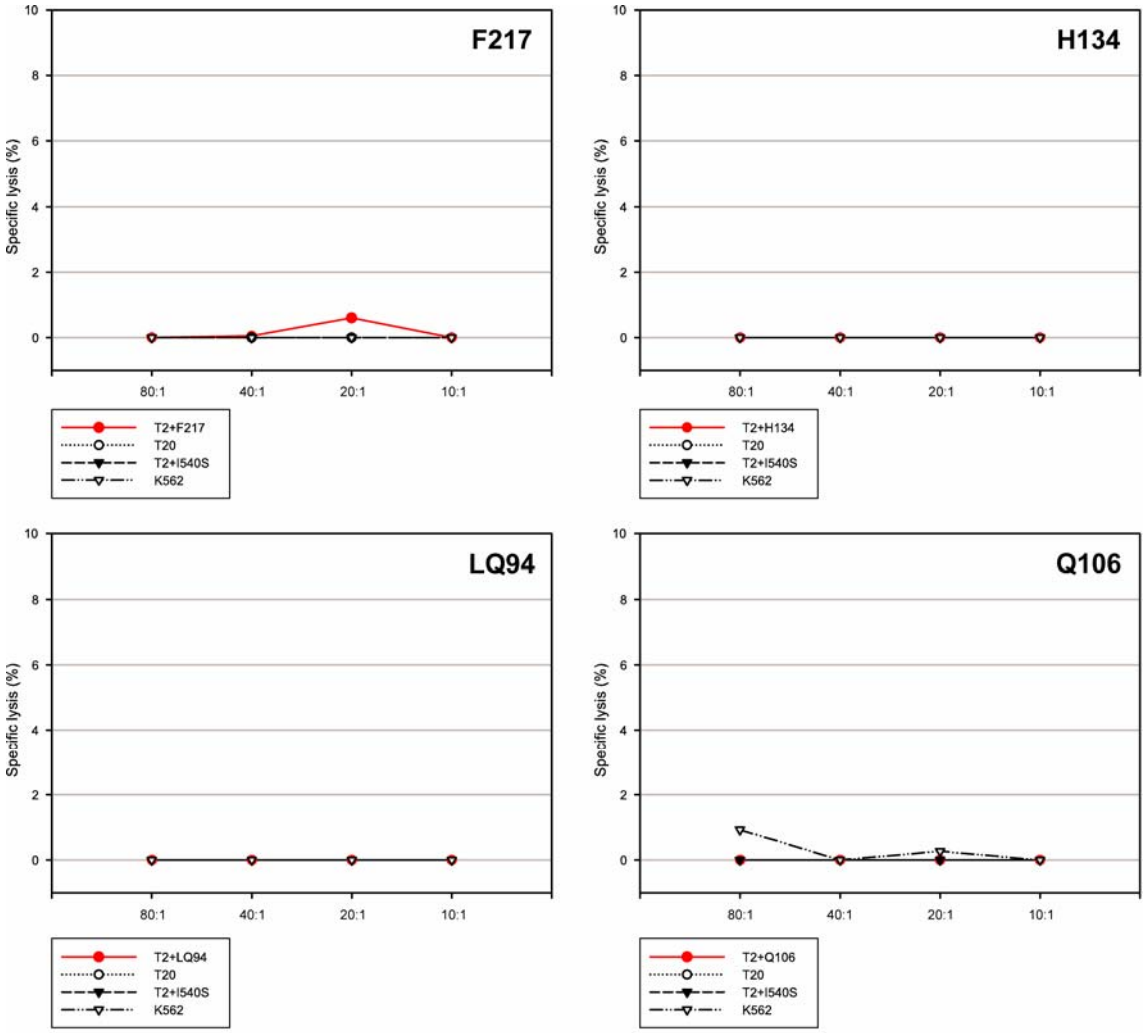

Figure 2.17. Antigenic potential of the candidate peptides. F217 shows some antigenicity, but the percentage of lysed cells is less than 1. X-axis: ratio of effector cells and target cells. Y-axis: percentage of specifically lysed target cells.

The peptides derived from the genes TGIF2LX and TGIF2LY (F217, LQ94, and H134) did not show any ability to stimulate the growth of $CD8^+$ cytotoxic lymphocytes, as shown in figure 2.17.

The peptide F217, derived from the TGIF2LX gene, was considered as non-immunogenic because the percentage of lysed target cells, presenting this peptide, was less than 1.

The peptide G54 (figure 2.18), derived from the VCY gene, was the only one which showed antigenic potential. The chromium release assay result shows that T-cells

stimulated with this peptide are able to specifically lyse a detectable percentage of target cells.

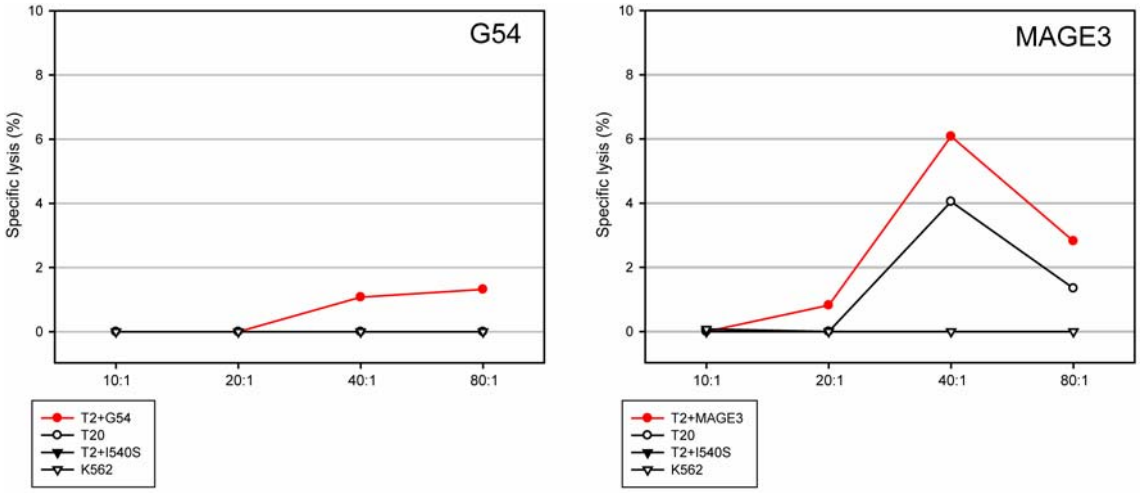

Figure 2.18. Antigenicity of the peptide G54. T2 cells loaded with the peptide G54 (GAATKMAAV) and incubated with potentially peptide-specific $CD8^+$ T cells (panel A), are lysed. The positive control is represented by the peptide MAGE3 (melanoma antigen family A-3; FLWGPRALV), loaded on T2 cells incubated with MAGE3-specific $CD8^+$ T cells derived from the same individual (panel B). Negative controls: T20 are T2 cells incubated without any peptide; I540S is a peptide which does not bind HLA-A0201 molecules (HFLLWKLIA); K562 cells are thipically a target for NK cells. The peptide G54 shows only weak antigenicity. X-axis: ratio of effector cells and target cells. Y-axis: percentage of specific lysis.

In each case, T2 cells loaded with the control peptide I540S were not recognized by cytotoxic T-cells. In the case of the positive control peptide, MAGE3, unloaded T2 cells showed susceptibility to be lysed by cytotoxic T-cells. Interestingly, the percentage of MAGE3-loaded T2 cells specifically lysed by cytotoxic cells was found to be rather low and this can have impact on the outcome of the experiment, suggesting that the lysing capacity of these cells is not that high. No NK-cells specific lysis was observed in this experiment, as K562 cells loaded with the different peptides were not specifically lysed. The only exception occurred in the case of K562 cells loaded with the Q106 peptide (figure 2.17, panel D), where a little specific lysis (around 1%) was observed.

3 Discussion

3.1 Study design

3.1.1 General remarks

In the last years, an increasing number of studies focused on gene expression profiling of cells from AML patients, or on comparison of gene expression between AML patients' samples and healthy blood cells have been published. All these studies provide a huge amount of data, which are, sometimes, contradictory (Court*, et al* 2004, Levy-Nissenbaum*, et al* 2003, Park*, et al* 2007, Raspadori*, et al* 2002).
AML is a very heterogeneous disease, affecting several cell populations of the myeloid lineage (see paragraphs 1.1.1 and 1.1.2). The affected cells show different morphologic as well as genetic characteristics. The classical French-American British classification, based mostly on morphological features of the leukaemic blasts, has given the way, in the last years, to a more precise classification which takes into account morphological characteristics as well as the karyotype of the blasts.
Since gene expression profiling experiments became widely used to identify differentially expressed genes in several fields, a huge number of studies focusing on almost all the biological processes have been published. For example, there are papers in which the effects of different diets on gene expression profile have been investigated using microarrays for mouse models (Newmark*, et al* 2009) as well as for human cells (Nguyen*, et al* 2006), correlating the different nutrition parameters of the investigated populations with the expression of cancer-related genes.
Also the effect of alcohol on gene expression has been widely studied in animal models (Sharma*, et al* 2008) and in human (Wallerath*, et al* 2003), either performing a genome-wide screen or examining expression levels of single genes.
Similar studies have been published for an incalculable number of different aspects. Differential gene expression has been evaluated between different human ethnic groups (Miao*, et al* 2008) and similar experiments have been performed with cells treated with all the possible available drugs or infected with various viruses and bacteria, for clinical purposes. Examples of gene expression profiling experiments can fill several books and this is not the aim of this thesis. The main information derived from this short excursus is that gene expression can be altered from plenty of different factors, especially in complex systems. Therefore, the achievement of the perfect gene expression

comparison is not possible, but, in order to get valuable and, especially, useful information the experiment should be accurately designed. For this purpose, to perform an exhaustive gene expression comparison between AML and healthy cells, the choice of the cell population to be investigated is a critical step.
In a lot of studies concerning expression profiling of AML cells, all these small but crucial differences within samples' group (heterogeneity of the disease itself, but also for the previously mentioned characteristics-diet, ethnic group, presence of inflammations, etcetera) were not really taken into account. For this reason a relevant amount of data generated by gene expression profile has to be re-examined in the light of the actual knowledge about AML and considering all the possible parameters to cluster the utilized samples.
The present study is based on unpublished clinical data from patients who underwent haploidentical haematopoietic stem cell transplantation (see paragraph 4.8 for information about patients). It has been reported that prognosis is more favourable when female hematopoietic stem cells are transplanted into male patients (See paragraph 1.2). According to this finding and to the fact that the presence of the Y chromosome is probably the major genetic difference between male and female cells, the expression of Y chromosome-encoded genes in AML blasts as well as in a comparable healthy cell population was analysed. The aim was to identify genes up-regulated in AML cells which can have a role in the biology of the tumour.

3.1.2 Samples' choice

The most critical step in designing this study was the choice of the samples which should represent AML cells and healthy cells. For a correct evaluation of Y chromosome gene expression levels, two comparable cell populations, one tumoral and the other healthy, had to be found, possibly taking into account all the previously discussed factors which can influence the transcriptome of cells.
The idea was to focus on a particular AML subtype for which it was possible to isolate a comparable healthy cell population from healthy blood. There were several reasons for which AML M4 and M5 cells have been chosen for this project. First of all, these two AML subtypes are well characterised. Affected cells show monocytic morphology, and they highly express the CD14 molecule on the cell surface. For these reasons, the term acute myelomonocytic leukaemia is often used to refer to subtypes M4 and M5 of AML. Another important reason for which these particular subtypes have been chosen is that it is possible to positively select and isolate $CD14^+$ cells (see paragraphs 2.1.1 and 4.8.2) from AML patients' blood, as well as from healthy blood samples using magnetic beads-conjugated antibody. Increased expression of the CD14 molecule in AML M4 and M5 blasts makes it a good target for a positive selection strategy. Using

one antibody and one separation step it is possible to isolate AML blasts from patient blood samples and normal monocytes (expressing CD14 molecules) from healthy blood. A similar procedure could be applied to other AML subtypes' samples, but the absence of a single marker, not expressed in other cell populations, would make necessary to perform several separation rounds and this would be too stressful for the cells.
Cells isolated through CD14-selection remain viable after the separation, avoiding sample loss and stress-induced gene expression alteration, as specified from the manufacturer of the microbeads-conjugated antibodies (Miltenyi Biotec).
If, on one hand the strict uniformity of the characteristics of the probes to be enclosed in the study is essential to get better and clearer results, on the other hand it can be seriously problematic. The problem is that it becomes difficult to collect a large number of suitable samples.
In the present study several filters have been imposed for the patients' eligibility.
Due to the restriction of expression analysis to Y chromosome genes, all the AML probes to be analyzed had to be obviously from male patients. The focus on analysis of CD14-positive AML cells only made necessary to select patients' blood probes according to the clinical data about their immunophenotype. If it was not possible to get such information, the samples had to be first screened for the abundance of CD14 molecules. Another filter was imposed by the decision to exclude from the study all the patients who showed other diseases in concomitance with AML. Leukaemia-induced immunosuppression leads often to viral and bacterial infections and this had also an impact on eligibility of the patients for the study. Some of the samples enclosed in the study were kindly provided from Professor Anne Dickinson, from the University of Newcastle upon Tyne (see also table 4.1 for more information about the utilized samples).

3.2 Y chromosome gene expression evaluation

3.2.1 Why looking at the Y chromosome

Several genes encoded on the Y chromosome have been associated with various disease and various types of cancers, but their role has not yet been understood (Chen*, et al* 2002, Ousati Ashtiani*, et al* 2009). Several studies and case reports highlighted in the last year the fact that, in some types of cancer, male tumoral cells show sometimes the tendency to loose their Y chromosome (Center*, et al* 1993, Wolff*, et al* 2005, Wong*, et al* 2008). The reasons and the processes are still unknown; it is not clear yet whether

this may be a mechanism with which tumoral cells protect them, or another consequence - or, maybe, a cause - of the cancer.
Expression levels of some of the Y chromosome-encoded genes has been evaluated in prostate cancer (Dasari, *et al* 2001, Lau and Zhang 2000), but no studies including all the known Y chromosome genes and involving other cancer types have been published.
From the immunological point of view, the Y chromosome is also very interesting. In the last decade, Studies aimed to identification and analysis of HLA molecules-restricted antigens provided a large amount of data about minor histocompatibility antigens derived from genes encoded from the Y chromosome (Rosinski, *et al* 2008, Simmons, *et al* 1997, Spierings, *et al* 2003, Warren, *et al* 2000). The importance of expression of particular minor histocompatibility antigens in tumoral cells has been highlighted in several studies (Falkenburg, *et al* 2003, Meklat, *et al* 2007, Melloni, *et al* 2004).
Altogether, these findings have raised the interest about several Y chromosome-encoded genes.

3.2.2 Specificity of the TaqMan probes

The first step for the customization of low density arrays was the selection of the TaqMan probes to be spotted in the microfluidic cards (See paragraph 4.5.3).
Among the probes inventoried in the Applied Biosystems database, forty-six assays have been chosen, covering all the known Y chromosome encoded transcripts.
Where it was possible to choose between several probes for one transcript, the decision was taken for that probe which was not able to bind to the X chromosome allele of that gene. Normally, probes binding to an exon-exon junction of the mRNA were selected in order to reduce the possibility for other RNA forms, like unprocessed transcripts, to be amplified during the real time-PCR. Unfortunately this was not possible for all the genes of interest.
Some genes on the pseudoautosomal region of the Y chromosome share high homology with their counterparts on the X chromosome. RNA of some homologous genes, for example in the case of TGIF2LY and TGIF2LX, is transcribed from a single exon. The differences between the exons of the homologs are limited to single-nucleotide substitutions; this makes it very difficult to discriminate between the X and the Y allele and a real time-PCR probe for such genes should be designed very carefully. For some genes, there were no probes available which would be able to bind to the Y chromosome-encoded mRNA only. Having no other choice, such probes have been enclosed in the plate and, during data analysis, it was kept in mind that both the X and the Y chromosome-encoded proteins were potentially present in the analysed cells.

3.2.3 Y chromosome gene expression in AML and healthy cells

The global analysis of Y chromosome gene expression in the patients' group, normalized to the expression in the healthy volunteers' group showed relatively strong up-regulation of six genes: BCORL2, PCDH11, TGIF2L, VCY, SRY, and HSFY. After a detailed analysis, the latter two genes were found to be not up-regulated. Table 2.1 reports the expression levels of the above mentioned genes in all the single samples used in this study. Such single-sample analysis was necessary for the assessment of the real level of expression of the candidate genes in the two groups of samples.
The transcripts of PCDH11 and TGIF2L were amplified with probes unable to discriminate between transcripts encoded by the X and Y chromosome.
The gene expression levels and the calculation of the differences in gene expression between the two groups of patients were generated by using the Sequence Detection System software, which showed several limitations. It seems that this software evaluates only the transcript level abundance, represented by the C_T value, in the two groups of samples, without considering the expression in single samples. For the gene HSFY, for example, it was noticed that its mRNA was amplified only in one sample out of eleven. No expression at all was found in the group of healthy samples. For these reasons HSFY has been considered as an outlier and excluded from further investigations.
A similar speculation can be applied to the SRY gene. Its transcript was present in four out of eleven patients' samples and in three out of eight healthy samples. The software calculates the percentage of samples in which the transcript was found within each group, without taking into account the expression level in each sample. In the patients' group, fluorescence became detectable after 36.5 PCR cycles in average, this means one cycle after fluorescence detection in the healthy group, where the average was 35.6 cycles. After a detailed analysis, SRY should be considered down-regulated in the patients' group. For this reason, it has not been taken into account for further studies.
Only four genes resulted to be really up-regulated in the patient-derived AML cells: BCORL2, PCDH11, TGIF2L, and VCY.
Among the two endogenous controls enclosed in the experiment, GAPDH and HPRT1 (see also paragraph 4.5.3), only the latter showed to be suitable for this study. It was expressed at comparable levels among all the samples whereas GAPDH expression levels showed more variability. Furthermore, HPRT1, in contrast with GAPDH, is expressed at very low level in these cells making it a better control for genes which are low expressed as those encoded on the Y chromosome.

3.2.4 Candidate genes

At the present time, there are not too many information available about the four genes regulated in AML. When the human genome sequence was published, the nucleotide sequences of the candidate genes were annotated and can be found in several publicly available databases. Almost all the available information about the candidates were generated from *in silico* analysis. What is known about them is summarised in the following paragraphs.

BCL6 co-repressor-like 2 (BCORL2)
BCORL2 (GeneID: 286554) is localized at the position q11.222 of the Y chromosome and doesn't have any homolog on the X chromosome. When the present study was started, BCORL2 was thought to be a protein-coding gene, but, afterwards, it was reclassified as pseudogene. The function of this gene is still unknown and there are no recent publications about it. Some *in silico* analysis indicate that its RNA might be weakly expressed in lung and prostate; higher RNA levels should be found in spleen and thymus (source: NCBI-UniGene).

Protocadherin 11 (PCDH11)
The PCDH11 gene (GeneID: 27328 for PCDH11X; 83259 for PCDH11Y) is located in a region of the X and Y chromosomes with very high degree of homology. The Y-linked homolog is localized at position p11.2 and it can generate four transcript variants, which differ slightly in the C-terminal region. PCDH11Y is probably the best characterised among the four candidate genes, although its function is not completely clear. It has a trans-membrane domain and potentially spans the cell membrane, with the N-terminus lying on the extra-cellular side. PCDH11Y seems to be potentially involved in calcium-dependent cell-cell adhesion, although there are no experimental data concerning it. This protein is expressed predominantly in adult and fetal brain and, at a very low level in testis (Blanco*, et al* 2000).
PCDH11Y has been described to be related to several diseases. It has been demonstrated that a chromosomal aberration involving this gene, the translocation t(Y;3)(p11;p12) to ROBO2 leads to multiple congenital abnormalities (Lu*, et al* 2007). PCDH11Y expression has also been associated with the acquisition of apoptosis-resistance by prostate cancer cells (Chen*, et al* 2002) and it has been suggested as a possible cause of androgen-independent growth of prostate cancer cells inducing WNT signalling, which, in turn, promotes the malignant behaviour and hormone-resistance of prostate cancer cells (Terry*, et al* 2006). The expression of the gene in AML cells, reported in this study, might have the same effect on leukaemic cells.

TGFB-induced factor homeobox 2-like (TGIF2L)

TGIF2LY (GeneID: 90655) can be found in the same Y chromosome region as pcdh11y, p11.2. It has as well a homolog on the X chromosome (TGIF2LX; GeneID: 90316). The RNAs transcribed from the two homologs are nearly identical, but a single-base insertion almost in the middle of the coding sequence of the X-linked homolog causes a frame-shift, leading to a 56 amino acids-longer protein. As consequence, the C-terminus of the two proteins is totally different.

Both TGIF2LX and TGIF2LY contain, in the N-terminal region, a homeodomain, which makes them belong to the three amino acids loop extension (TALE) superfamily of transcription factors. Other members of this superfamily are TGIF1 and TGIF2, to which TGIF2LY and TGIF2LX are closely related (see paragraph 2.2.4 and figure 2.9). TGIF1 and 2 have been already characterised and their function has become clearer in the last years.

TGIF1 (GeneID: 7050) has been found to bind to a retinoid X receptor-responsive element, functioning as a competitor of the retinoic acid receptor in the binding to retinoic acid response elements (Bertolino*, et al* 1995). TGIF1 also interacts with SMAD proteins and is an inhibitor of TGF-β signalling pathway (Chen*, et al* 2003, Wotton*, et al* 1999). Several mutations of TGIF1 have been associated with diseases like holoprosencephaly (Aguilella*, et al* 2003, El-Jaick*, et al* 2007). TGIF1 expression has been also associated with several cancers, like bladder transitional cell carcinoma (Voorter*, et al* 1995), ovarian carcinoma (Luo*, et al* 2002), and esophageal carcinoma (Nakakuki*, et al* 2002).

TGIF2 (GeneID: 60436) is expressed in several tissues. Its expression is relatively high in heart, kidney and testis, and low in brain and prostate (Imoto*, et al* 2000).

This gene is a transcription repressor, which functions either by binding directly to its target DNA sequence or by interacting with TGF-β activated SMAD proteins, like TGIF1. TGIF2 is probably responsible of the recruitment of histone deacetylase proteins (Melhuish*, et al* 2001).

TGIF2 has also been associated with disease. Its expression has been found to be up-regulated in several ovarian cancer cell lines (Imoto*, et al* 2000).

TGIF2LY is certainly the less characterised member of the TALE transcription factors superfamily; almost all the information about this gene have been generated *in silico*, by comparison with related or similar proteins. For this reason the biological function of this gene is not clear at all. It has been proposed that TGIF2LY may play a role in transcription (Blanco-Arias*, et al* 2002). Its expression, as well as TGIF2LX expression, has been reported to be restricted to adult testis only (Blanco-Arias*, et al* 2002, Skaletsky*, et al* 2003), but here appreciable transcript level has been observed in AML M4 and M5 cells as well as in healthy $CD14^+$ cells, although in the latter the average transcript level was very low.

Variable charge Y-linked (VCY)

The VCY gene (GeneID: 9084) is located on the Y chromosome at the position q11.221, and has been found to be expressed in testis only, and, specifically, in germ cells (Lahn and Page 2000, Zou*, et al* 2003). Moreover, its expression seems to be limited to nucleoli and interaction of VCY with the acidic ribosomal protein PO has been demonstrated using yeast two-hybrid system analysis; this finding suggests a possible role of VCY in regulation of ribosome assembly during spermatogenesis (Zou*, et al* 2003). It is interesting to notice that, like in the case of TGIF2LY, the RNA encoded from a gene which is supposed to be expressed in testis only has been detected at relatively high level in leukaemic cells.

3.2.5 Candidate genes are up-regulated in leukaemic cells

One of the key points to be addressed after the analysis of the Y chromosome gene expression was to demonstrate that higher expression of the identified candidate genes was a peculiarity of the tumoral cells. Beside PCDH11, expression of all the other candidate genes has been detected also in some healthy samples.

An experiment to compare the expression of those genes between the leukaemic fraction of the patients' blood cells, $CD14^+$, and the non-leukaemic cells of the same patient, $CD14^-$, was necessary to get an answer to this question.

As described in paragraph 2.1.3 and in figure 2.3, expression of all the candidate genes within patients' samples is restricted to leukaemic cells. However, these genes showed some expression, although at very low level, in some healthy cell samples as well, with the exception of the PCDH11 gene. This might mean that BCORL2, TGIF2L and VCY genes can be normally expressed at basal levels in $CD14^+$ cells and that they are up-regulated when leukaemia affects this cell population. The few information available for these three candidate genes make it impossible to further speculate the function that BCORL2, TGIF2L, and VCY may have in the tumour biology and to understand whether their up-regulation is to be included in the possible tumour causes or in its consequences. BCORL2 is a pseudogene: its mRNA is usually not translated into a protein in normal cells. It has been demonstrated that pseudogenes can be activated, and a pseudogene-derived protein can consequently be expressed, in several types of cancer (Sun*, et al* 2008, Zou*, et al* 2009). Analysis of transcription, showing presence of the BCORL2 transcript in both healthy and leukaemic $CD14^+$ cells, doesn't exclude the possibility of the occurrence of BCORL2 pseudogene activation in tumoral cells. This could lead, in turn, to the presence of one or more unusual protein in AML cells, due to the presence, within the BCORL2 RNA, of several open reading frames (ORFs). The commercial unavailability of anti BCORL2 antibodies made not possible, for time

reasons, to investigate the presence of the protein in healthy and leukaemic cells in this study.
A similar comment can be referred to TGIF2L and VCY. The function of the VCY gene is not known. There are some experimental data which put this protein into relation with the regulation of ribosomal assembly in spermatogenesis (Zou*, et al* 2003). These data are based on the assumption that the gene is expressed only in adult testis (Lahn and Page 2000, Zou*, et al* 2003). The fact that VCY showed to be expressed also in healthy $CD14^+$ blood cells (see table 2.1) is puzzling. Again, it remains to be determined whether the VCY protein is also present and, if yes, in which cell population. Also for the VCY protein there are no commercial antibodies available and, for lack of time, it was not possible to generate antibodies, test them, and use them for such interesting verifications. Deeper investigations have been carried out to understand the function and the possible role in leukaemic cells of the TGIF2L gene (see paragraph 3.3).
A different comment has to be given with regards to the PCDH11 gene. A transcript derived from this gene was not detected in cells from healthy individuals and its expression was restricted to leukaemic cells of the patients. It has been proposed that expression of PCDH11Y is involved with the acquisition of apoptosis-resistance by prostate cancer cells. Although the mechanism through which this gene induces apoptosis-resistance remains still unclear, PCDH11Y might have the same effect in AML cells. Moreover, if the gene is involved in the immune recognition of tumoral cells - either directly, generating immunogenic peptides, or indirectly, inducing expression of other genes which generate immunogenic peptides - the fact that PCDH11Y is expressed usually only in brain, an organ which is not under immunosurveillance, would make it the perfect tumour-associated gene. In fact, it is known that immunogenic peptides derived from genes expressed in tissues like brain or testis, normally not controlled by the immune system, are sometimes present also in cells of several tumours. Such antigens can be used as targets for immune therapy (Gavert*, et al* 2008, Meklat*, et al* 2007).

3.2.6 THP-1 as cell line model

Because of to the scarce availability of patient material and to the above-mentioned heterogeneity of those probes, a cell line model was needed. THP-1 cells are derived from a young male patient with acute myelomonocytic leukaemia (Tsuchiya*, et al* 1980) and they do express CD14 molecules on the cell surface. They would be a very useful model to study $CD14^+$ AML. In order to understand whether THP-1 cells are really comparable with the patients' cells enclosed in this study, I decided to evaluate the Y chromosome gene expression of the cell line and to compare it with the patients' probes (table 2.2).

Besides some differences, which are actually not that rare when a cell line is compared with primary cells, it is possible to notice that the Y chromosome gene expression profile of THP-1 cells fairly reflects that one of the patients. Among the candidate genes, THP-1 cells do express PCDH11 and TGIF2L, which are the genes that showed the highest up-regulation in the patients' group. To understand the role of these two genes in the pathogenesis, or more in general in the tumour biology, of this subtype of leukaemia, more information about the gene function are needed. According to the expression pattern of the Y chromosome genes, THP-1 cells can be a good cell line model for AML M4 and M5, and the fact that they also express the PCDH11 and the TGIF2L genes, makes them the perfect model to investigate the function of these genes.

3.3 TGIF2LY – molecular analysis

3.3.1 Introductory remarks

At the beginning a deeper analysis of all the four candidate genes was started. Afterwards, for time reasons, I decided to focus only on one of those genes: TGIF2LY.
Before starting any functional analysis, the main problem was to understand whether the Y-linked homolog of this gene was expressed in the leukaemic cells. It is very difficult to distinguish between TGIF2LY and TGIF2LX referring to mRNA only: their transcripts are nearly identical. A single nucleotide deletion in TGIF2LY mRNA leads to a premature stop codon and the resultant protein is shorter than TGIF2LX.
The only possibility to be sure about the presence of the Y-encoded protein was using an anti-TGIF2LY antibody in a western blot experiment, where proteins extracted from THP-1 cells have been used. Unfortunately, unavailability of patient material didn't allow assessment of protein expression in patients.
Figure 2.6 shows the result of the western blot. TGIF2LY is expressed in leukaemic cells although its level is quite low.
As mentioned before, the low protein level was expected. It has to be kept in mind that the mRNA level was also very low: in real time-PCR, fluorescence was detected after around thirty-five cycles with 100 ng of cDNA from total RNA of THP-1 cells. This RNA level is comparable with that of the patients' samples with the same amount of cDNA tested. Moreover, it shouldn't be forgotten that the TaqMan probe used for the real time-PCR is not able to discriminate between TGIF2LY and TGIF2LX mRNA. If these genes are both expressed, the level of the total transcript will be higher than the

level of only one of the two proteins. Because of focusing on TGIF2LY only, the protein expression of TGIF2LX was not assessed in this study.
To get an idea about the eventual expression of TGIF2LX in AML M4 and M5, low density arrays have been used to analyze via real time-PCR a sample of $CD14^{+}$ cells derived from an AML M4 female patient (patient code: AMP3) and the same female healthy cell population as well (refer to table 2.3). As expected, expression of genes that do not have an X chromosome homolog was not observed. The interesting result, in this experiment, was the detection of TGIF2LX expression in the patient. One possible explanation for this is that transcription of TGIF2LY and TGIF2LX is activated through the same mechanisms and that it is not Y chromosome-dependent. This result suggests as well the possibility that, also in male patients, transcription of TGIF2LX occurs together with transcription of TGIF2LY.
The homeodomain region of TGIF2LX is highly homolog to that of TGIF2LY. Only one amino acid is substituted in the second helix of the DNA binding region (figure 3.1).

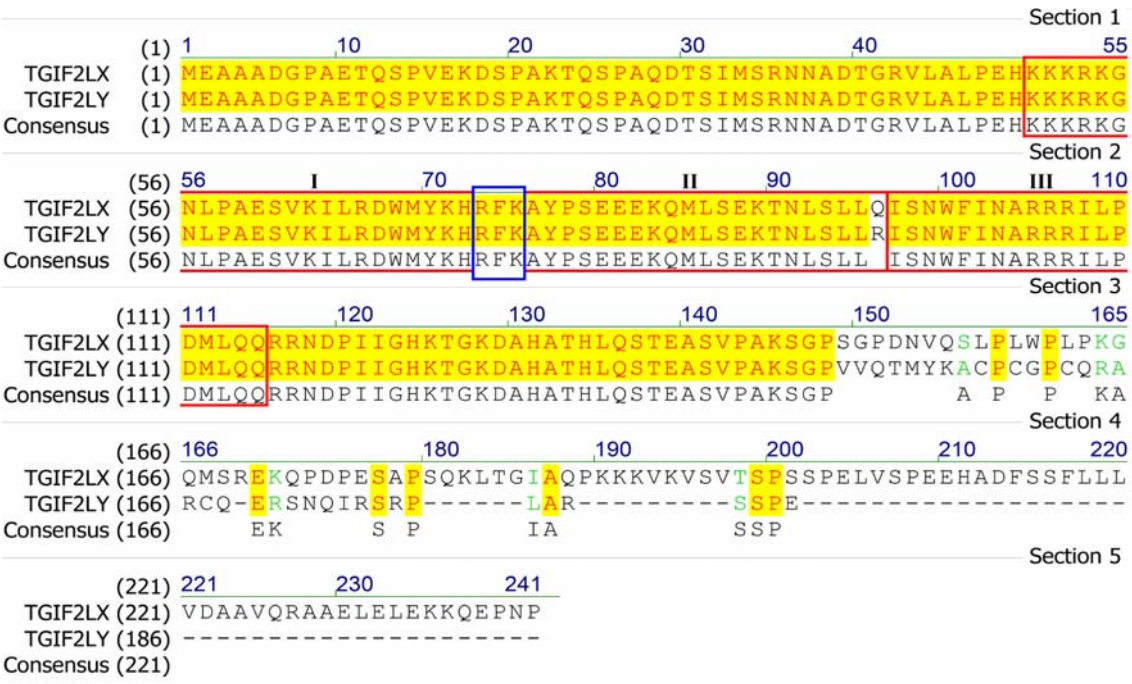

Figure 3.1. Alignment of TGIF2LX and TGIF2LY proteins. Red frames indicate the three helices of the homeodomain, numbered from I to III. This region shows nearly 100% homology. Blue frame indicates the three amino acid loop extension (TALE). C-terminal portions of the proteins are completely divergent.
Color scheme (color = amino acid substitution type): black = non-similar; green = block of similar; red on yellow background = identical. Dashes indicate missing amino acids.

As shown in figure 3.1, the N-terminal part in the two proteins, including the homeobox region, is nearly identical. This suggests that the two proteins recognize and bind the same DNA consensus sequence. The C-terminal portion of the proteins, on the other

hand, is highly divergent. These evidences suggest that the two proteins either have a different regulatory activity, if interaction with other proteins occurs through their C-terminal portion, or the same one, if the protein-protein interaction region is localized at the N-terminus.
TGIF2LY was chosen for in-depth studies for several reasons. First of all because, as mentioned before, its expression is supposed to be restricted to adult testis only and it would be very interesting to understand the role of this gene when it is expressed in a different tissue. The fact that TGIF2LY is a potential transcription factor makes it even more interesting: what is the function of a transcription factor when it is expressed in an unusual tissue?

3.3.2 TGIF2LY can have transcription factor activity

Information about the function of TGIF2LY gene are scarce and the most of them have been generated *in silico*, basing only on DNA and protein sequence analysis. Here, we were interested in understanding the role that this gene can play in AML. Of course, in order to speculate on the role of a gene in tumoral cells, it is essential to get hints about its normal biological function.

Three-dimensional structure of TGIF2LY
First step was to look for homologies between TGIF2LY and other homologous and already-characterised proteins. Blasting the amino acid sequence of TGIF2LY against human and *Drosophila melanogaster* databases, it was possible to identify closely related proteins. As shown in figure 2.9 and 3.1, TGIF2LX showed the highest homology degree, followed by TGIF2 and TGIF1, in human and the vismay gene (vis), of *Drosophila melanogaster*. They all belong to the three amino acid loop extension (TALE) superfamily, a group of homeodomain-containing transcription factors.
Looking carefully at the modelled three-dimensional structure of the homeobox of these proteins (figure 2.12), it can be noticed that the organization of the three helices is similar in all of them. Very interestingly, the amino acidic residues, essential for the recognition of the DNA consensus sequence (Noyes*, et al* 2008), are highly conserved and located more or less in the same positions on the helices. The five residues on the third helix of the homeobox, which plays a critical role in DNA sequence recognition, are identical in all the four proteins (see figures 2.11 and 2.12). These amino acids are positioned in such a way that they can easily access double-stranded DNA to detect the right sequence. The three amino acids at the beginning of the first helix, at the positions 2, 3, and 5 of the homeodomain are also involved in DNA recognition. The original motif arginine-lysine-arginine (R-K-R), present in TGIF2 and vis, is modified in the

other two proteins. It is R-R-R in TGIF1 and K-K-K in TGIF2LY. However arginine and lysine have similar properties.
The spatial organization of the three helices of the proteins has also to be taken into account. It is possible to notice, from the models in figure 2.12, that helices one and two are parallel to each other. Helix three is positioned on another layer and forms an angle with the other two helices which is different in all the proteins. The initial portion of the first helix, with its three amino acids important for DNA recognition, has a different shape in all the proteins. Especially in the vis protein, this region seems to be located behind the remaining portion of helix three, potentially reducing its DNA accessibility.
It has to be mentioned that all these three-dimensional structures have been generated using modelling software: it is possible that the real three-dimensional organization of the proteins is different from the one presented here, especially for the terminal portions of the sequence submitted to the software. In fact, providing only the homeobox region to the software, it doesn't take into account the surrounding, which might be crucial for a correct calculation of the spatial organization of a protein. However, it is possible to see that the organization of the third helix, probably the most important for a correct DNA recognition, is similar in all the four proteins.

Subcellular localisation of TGIF2LY

Subcellular localisation was another essential step to further elucidate the function of TGIF2LY.
There are more than twenty publicly available servers which are used for predicting the localization of proteins at subcellular level, but, for several reasons, it is not possible to completely rely upon them to assess this crucial aspect of protein biology. First of all, the prediction method is different in different algorithms. Some software, like TargetP, analyze only the N-terminal sequence of the submitted protein identifying there the possible signal peptide (Emanuelsson*, et al* 2000) and predicting the localisation of the protein. Some other software, like SubLoc, analyze the overall amino acid sequence to calculate the possible localisation of the protein of interest (Hua and Sun 2001).
The low reliability of these algorithms has several causes: many targeting signals have not yet been understood and, in addition, unconventional targeting signals have been identified. Furthermore, co-translational pathways for transportation of proteins into organelles have been described (Crowley and Payne 1998), but so far they cannot be recognized by the prediction algorithms (see Schneider and Fechner 2004 for a review of the predicting algorithms).
Another aspect which should be taken into account is that subcellular localisation of proteins doesn't have to be static but can change according to the status of the cells and to the protein function. For example, there are proteins, like receptors, located in the cell membrane or in the mitochondria or in other organelles. Upon activation (by phosphorylation, for example), these proteins or parts of them, are translocated to the nucleus where they act as or interact with transcription factor to modulate cell responses

to compounds or environmental changes. It is therefore not impossible that predictions considered as incorrect are, actually, alternative localizations under different circumstances. However, the method applied for this investigation is standardly used and should therefore provide greater reliability than predictions using bioinformatics.

HeLa cells have been chosen to be transfected and stained. These cells possess good adhesive properties, growing normally as monolayer. This feature is essential in this kind of experiment because of the numerous washes which have to be performed during the staining procedure. Another advantage in using HeLa cells for subcellular localisation studies derives from their relatively large volume. Under the microscope, this allows to easily distinguish between the organelles, providing a clearer result. One aspect to be taken into account, using these cells, is that they are derived from a female being and TGIF2LY is supposed, according to databases, to be expressed in adult testis only. It would be interesting to repeat this experiment using a testis-derived cell line.
However, for these first studies I was interested only in mere subcellular localisation of TGIF2LY in cells and, for this purpose, HeLa cells revealed to be a useful tool.
One of the most used methods for intracellular localisation of proteins involves the use of the green fluorescent protein (GFP) as tag. In the present study this method was not used because of the size of the protein: TGIF2LY is a quite small protein, with a weight of around 21 KDa. The GFP protein is larger than TGIF2LY, weighing around 27 KDa. The presence of such a large protein can totally mask the targeting peptides present within TGIF2LY, influencing dramatically the translocation of the protein. To avoid this problem, a vector encoding a MYC-tag was chosen. This tag is only 10 amino acids long and its interference with the translocation of TGIF2LY was supposed to be lower.
However, once the construct TGIF2LY-C-terminal MYC-tag was expressed in HeLa cells, the localisation of the protein was rather hard to determine. Figure 2.7 shows that, in HeLa cells transfected with this construct, TGIF2LY was translocated in several cell compartments (figure 2.7, panels A and D).
The localisation of the protein became clearer when HeLa cells were transfected with a TGIF2LY-N-terminal MYC-tag construct (figure 2.8). In these cells, TGIF2LY was only detectable in the nuclei.
This result was a bit puzzling. In the first helix of the homeodomain, close to the N-terminus of the protein, there is a short peptide typical for nuclear localization (KKKRK in position 50-54 of the protein). For this reason, it would be expected that the presence of the MYC-tag at the C-terminus of TGIF2LY doesn't interfere with the correct translocation of the protein. However, this result can have two different explanations. On one hand, it is possible that another peptide is necessary for the correct distribution of the protein. If this second signal is contained in the C-terminal portion of TGIF2LY, the presence of the MYC-tag could interfere with it and lead to the ambiguous distribution seen in figure 2.7. On the other hand, taking a look to the three-dimensional structure of TGIF2LY (figure 3.2), it is possible to note that, due to the folding of the

protein, the C-terminus lays in proximity of the nuclear signal peptide KKKRK, more than the N-terminus.

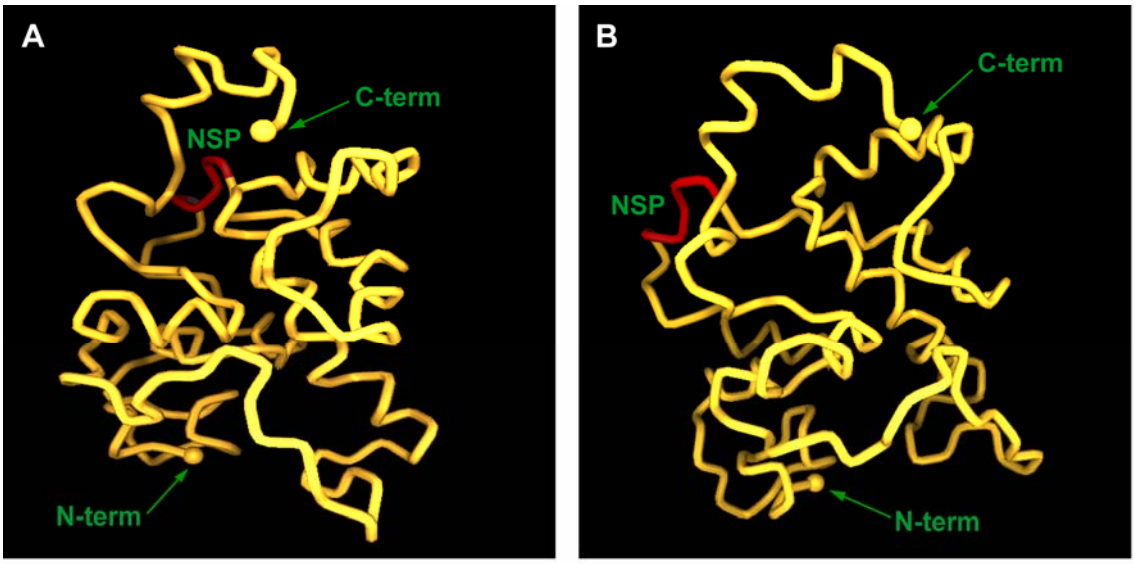

Figure 3.2. Position of a nuclear translocation signal peptide (NSP), highlighted in red, within TGIF2LY. Panel A and B show two different orientation of the protein. The C-terminus is closer to the signal peptide than the N-terminus: the presence of a tag at the C-terminus may interfere with the correct translocation of the protein within the cell.

Due to this fact, it can be that the presence of the MYC-tag at the C-terminus of TGIF2LY interferes with the normal distribution of the protein within the cells, maybe modifying its three-dimensional conformation, so that the signal peptide is masked.
Altogether, these data suggests that the nuclear localization is the most probable for TGIF2LY.
Bioinformatic analysis of TGIF2LY provided several important information. The protein possesses a homeodomain, whose sequence and three-dimensional conformation is similar to the homeodomains of other proteins which have been characterized and which are functioning as transcription factors, like TGIF1, TGIF2LY, and vismay. Moreover, it has experimentally demonstrated, for the first time in this study, that TGIF2LY localizes in the nuclei of cells. All this findings lead to the conclusion that this protein can function as transcription factor.

3.3.3 Interaction of TGIF2LY with DNA

Once the transcription factor potential of TGIF2LY was assessed, the next step was to analyze the DNA-binding capacity of the protein and, eventually to find out the nucleotide sequence which it recognizes.

One of the most used methods to investigate interactions between proteins and DNA is chromatin immunoprecipitation. In this method, cells are usually transfected with an expression vector in which the protein of interest is cloned. Once the protein is expressed, it is covalently bound to DNA through cross-linking, usually using formaldehyde, and the DNA is mechanically fragmented, through sonication. Antibodies, specific for the protein under investigation, are used to precipitate DNA-protein complexes, from which, after reversing of the cross-link, the protein is removed. The so-purified DNA fragments are then ready to be amplified and sequenced.
There are several limitations for this technique which made it not possible to be used in this study. One is the need of high amount of antibody specific, in this case, for TGIF2LY. The antibody which has been used in this study, to assess the protein expression with a western blot experiment, was not suitable for immunoprecipitation. It binds to a region in the C-terminal portion of TGIF2LY. This region is close to the homeobox in the folded protein (see figure 3.2) and the presence of DNA would mask the binding site of the antibody.
The second reason concerns the high time need. In fact, generation of specific antibodies and their test require quite a long time. Beside this, the time needed for optimization of the expression vector and for all the cell culture procedures has also to be taken into account.
For this study, I had no time to generate and test antibodies for chromatin immunoprecipitation. For this reason this method was discarded and another technique for the identification of the binding site of TGIF2LY was implemented.
A method called *in vitro* genomic selection has been described (Shostak*, et al* 2004). In this protocol the coding sequence of the protein of interest is cloned into an expression vector, expressed and purified as GST-fusion protein. Genomic DNA from cells is digested, using Sau3AI, a restriction enzyme whose DNA consensus sequence occurs rather frequently, and the fragments are incubated with the fusion protein. GT-sepharose beads are then used to isolate protein-DNA complexes, from which, afterwards, DNA is purified and sequenced.
I decided for a modification of this method. In order to proceed faster, I used magnetic beads-conjugated anti-GST antibodies and a magnetic separator instead of purifying the GST-fusion protein using GT-sepharose beads (see also paragraph 4.6.6 and figure 2.13 for the method description).
The modified protocol is working nicely: several genomic DNA fragments where bound from TGIF2LY, as shown in tables 2.4 and 2.5.
It has to be kept in mind that these data are generated *in vitro*; there are no hints about the protein behaviour *in vivo*. Both fusion protein and genomic DNA were purified and co-incubated in artificial binding conditions. These parameters are probably quite different from physiological conditions. For example, it is possible that the protein needs to form complexes with other proteins to exert its function as transcription factor or that some DNA regions might not be accessible, for example due to chromatin

condensation, to proteins. Different experiments, like chromatine immunoprecipitation, are needed to investigate this aspect. The aim of this experiment, as first step, was to individuate the nucleotide sequence that is recognized by TGIF2LY and this characteristic should depend solely on the homeodomain. Once the binding site is identified, it is of course possible to design experiments aimed to investigate other aspects of TGIF2LY-DNA interaction.

Several studies on homeodomain-containing transcription factors demonstrated that these proteins bind the specific DNA sequence through their third helix. Moreover, there are fixed positions for amino acids which are crucial for the correct and specific DNA recognition (Ades and Sauer 1995, Damante*, et al* 1996, Passner*, et al* 1999, Piper*, et al* 1999); these are, for members of TALE superfamily, the amino acids in position 2-3-5-50-53-54-57-58 of the homeobox. Recently, an extensive prediction of DNA sequences recognized from homeodomain-containing protein has been proposed (Noyes*, et al* 2008); for some transcription factors, the prediction has been verified with experimental work (supplemental material of the cited article). Among these verified proteins there was also vismay (vis), from *Drosophila melanogaster*. This protein shows high similarity with TGIF-family members, as revealed from BLAST analysis of TGIF1, TGIF2 and TGIF2LY sequences against *Drosophila melanogaster* protein database (figure 2.10). At homeobox level, the amino acid sequence of the four proteins is almost identical. The third helix of the homeodomain, in particular, shows the highest homology degree among these proteins. Here all the amino acid substations, if present, are conservative and none of the previously mentioned amino acids in position 50-53-54-57-58 is substituted (refer to figure 2.12). The amino acids in position 2-3-5 are different. TGIF1 has R-R-R in these positions, TGIF2 and vis have R-K-R, and TGIF2LY has K-K-K. Due to the difference between these regions, the DNA sequence recognized from the proteins can be different. It has been experimentally demonstrated that TGIF1 and TGIF2 are able to bind to the same nucleotide sequence 5'-TGTCA-3' (Bertolino*, et al* 1995). The protein vis binds to the sequence 5'-TGACA-3' (Melhuish*, et al* 2001). Surprisingly, TGIF2 and vis, the proteins in which the eight amino acids responsible for DNA recognition are identical, bind to different nucleotide sequences and TGIF1 and TGIF2, which have different amino acids in position 3, bind to the same DNA site. These observations suggest that the amino acid sequence is not the only factor influencing the recognition of DNA sequences from the homeobox. Probably, the three-dimensional conformation of these domains plays also a role in this process. In fact, it could be important also the orientation that these amino acids have when they get in contact with DNA.

Speculating on these observations, it might be possible that TGIF2LY does recognize a nucleotide sequence very similar to those recognized by TGIF1, TGIF2, and vis. For this reason a global check among all the DNA fragments previously obtained was performed, in order to identify a nucleotide sequence highly similar to 5'-TGTCA-3' and 5'-TGACA-3' and present, with low variability, in all the DNA fragments.

The sequences of all the seventeen DNA fragments were analyzed as described in paragraph 4.9.2. Interestingly the only sequence fulfilling the previously mentioned conditions was 5'-TGA(C/T)(A/C/T)-3'.
Double-strand DNA oligos for an electrophoretic mobility shift assay (EMSA) were designed containing four repetitions of each putative binding site (table 3.1). Using such oligonucleotides, it would be theoretically possible for the protein to bind to a sequence lying on the junction between two repetitions of the putative binding site (table 3.1). In order to overcome this problem, all the previously isolated DNA fragments were checked for the presence of all the possible five-bases sequence combinations contained in each oligonucleotide. Only the putative binding sites were found to be present in all the fragments.

Oligo code	Possible binding sequences
TGA-C-A/C/T_1	Cy5-TGACA TGACA TGACA TGACA
	GACA T
	ACA TG
	CA TGA
	A TGAC
TGA-C-A/C/T_2	Cy5-TGACC TGACC TGACC TGACC
	GACC T
	ACC TG
	CC TGA
	C TGAC
TGA-C-A/C/T_3	Cy5-TGACT TGACT TGACT TGACT
	GACT T
	ACT TG
	CT TGA
	T TGAC
TGA-T-A/C/T_1	Cy5-TGATA TGATA TGATA TGATA
	GATA T
	ATA TG
	TA TGA
	A TGAT
TGA-T-A/C/T_2	Cy5-TGATC TGATC TGATC TGATC
	GATC T
	ATC TG
	TC TGA
	C TGAT
TGA-T-A/C/T_3	Cy5-TGATT TGATT TGATT TGATT
	GATT T
	ATT TG
	TT TGA
	T TGAT

Table 3.1. Possible five-bases sequences contained in the oligonucleotides used for EMSA experiments.

The electrophoretic mobility shift assay showed that TGIF2LY binds with high affinity to the sequences 5'-TGATA-3' and 5'-TGATC-3' and with low affinity to the sequences 5'-TGACC-3' and 5'-TGACA-3' (see paragraph 2.2.6 and figure 2.15). All the DNA fragments isolated through *in vitro* genomic selection were then analyzed for the presence of the high- and low-affinity TGIF2LY-binding sequences (table 3.2).

Sequence	Fragment code	Description	Notes
5'-TGATA-3'	XL14	GCOM1/CGNL1	In brain only - excluded
	XL15	DKK1/MBL2	
	XL17	Within a gene	Excluded
	3	UBB/PRR6	
	25	Within a gene	Excluded
	19B	Within a gene	Excluded
	45B	Uncharacterised	Excluded
5'-TGATC-3'	8	NLRP1/WSCD1	
	19	Within a gene	Excluded
	19B	Within a gene	Excluded
	45B	Uncharacterised	Excluded
5'-TGACA-3'	XL17	Within a gene	Excluded
	8	NLRP1/WSCD1	
	10	Within a gene	Excluded
	19	Within a gene	Excluded
	23	PROC/MAP3K2	
	25	Within a gene	Excluded
	5B	SIAH2/Hyp. Protein	
	37B	Within a gene	Excluded
	45B	Uncharacterised	Excluded
5'-TGACC-3'	XL8	Within a gene	Excluded
	XL14	GCOM1/CGNL1	In brain only - excluded
	XL17	Within a gene	Excluded
	2	EEF2/PIAS4	
	3	UBB/PRR6	
	6	Within a gene	Excluded
	8	NLRP1/WSCD1	
	10	Within a gene	Excluded
	23	PROC/MAP3K2	
	5B	SIAH2/Hyp. Protein	
	37B	Within a gene	Excluded
	45B	Uncharacterised	Excluded

Table 3.2. High-affinity (5'-TGATA-3' and 5'-TGATC-3') and low-affinity (5'-TGACA-3' and 5'-TGACC-3') TGIF2LY-binding sequences in the DNA fragments isolated through *in vitro* genomic selection. Fragments corresponding to intragenic DNA regions and to uncharacterised genomic regions were not taken into account.

Fragments corresponding to uncharacterised genomic regions or to regions laying within a gene were not taken into account, according to the fact that transcription repressors or enhancers usually bind to DNA sequences laying outside the coding region of a gene and, often, rather far away from the promoter of the regulated gene. One DNA fragment (code: XL14), represents a genomic region flanked by two genes which are reported to be expressed in brain only. For this reason, this fragment was excluded from the analysis.
As mentioned before, because of the experimental setting of the *in vitro* genomic selection, the information gained concern only the capacity of TGIF2LY to bind DNA *in vitro*. It is not possible to speculate which gene, among those identified with this experiment, is effectively regulated *in vivo* from TGIF2LY.

3.3.4 Genes potentially regulated by TGIF2LY

Among the genomic DNA fragments isolated through *in vitro* genomic selection, one was found to contain only a high-affinity TGIF2LY consensus sequence. Two fragments possessed both the high- and the low-affinity sequences, and three were found to contain only a low-affinity consensus sequence.
In order to understand the regulatory pathway in which TGIF2LY may exert its function, information about the genes flanking these six genomic DNA regions have been collected. A summary is presented schematically in table 3.3.

Fragment "XL15"
This fragment contains a genomic region flanked 5' by the gene dickkopf homolog 1 (DKK1), and 3' by the gene soluble mannose binding lectin 2 (MBL2). It might be possible for TGIF2LY to regulate the expression of both genes. TGIF2LY showed high affinity for the DNA sequence present in this genomic fragment, 5'-TGATA-3'. Plenty of articles concerning the DKK1 gene have been published in the last years. It has been demonstrated that DKK1 takes part in the Wnt signalling, inhibiting it (Fedi*, et al* 1999). There is evidence that activation of Wnt signaling can contribute to the neoplastic process. Inappropriate expression of these proteins causes mammary tumour formation in mice (Lin*, et al* 1992). Moreover, in cell culture, several Wnt family members have been shown to induce altered morphology and increased saturation density of epithelial (Wong*, et al* 1994) and fibroblast (Bafico*, et al* 1998) cell lines. Furthermore, genetic abnormalities correlating with altered Wnt signalling have been observed in human colon cancer (Morin*, et al* 1997), melanomas (Rubinfeld*, et al* 1997), and hepatocellular carcinomas (de La Coste*, et al* 1998), indicating that aberrations of Wnt signaling pathways are critical to the development of these human cancers. Silencing of DKK1 stops the inhibition of Wnt signalling which, in turn, can

contribute to tumoral transformation of cells. DKK1 plays a role in both cell proliferation and programmed cell death (Mukhopadhyay, *et al* 2001). Interestingly, DKK1 silencing, due to hypermethylation, has been suggested as risk factor for poor overall survival in AML (Suzuki, *et al* 2007); suppression of DKK1 expression has been proposed to be involved in AML progression.
The other gene flanking this genomic region is MBL2. This protein is an essential component for the innate immune system, being one of the first lines of defence against pathogens. It binds to carbohydrate patterns found on the surface of a wide variety of microorganisms, including bacteria, protozoa and fungi. Binding of the protein to its ligands leads to opsonization as well as activation of the complement system (Ohlenschlaeger, *et al* 2004). Mutations or down-regulation of this gene has been put in relation with increase of susceptibility to infections (Garred, *et al* 1995, Summerfield, *et al* 1997, Walport 2001), evidencing the importance of this protein in innate immunity.

Fragment "3"

This fragment contains a genomic region flanking the ubiquitin B (UBB) gene on the 5' side and the proline rich 6 (PRR6) gene on the 3' side. Both the high-affinity (5'-TGATA-3') and the low-affinity (5'-TGACC-3') TGIF2LY consensus sequences are present in this fragment.
Ubiquitin B is involved in ATP-dependent protein degradation in the cells (Tanaka 2009) and is important in a lot of different biological processes, like proteasomal protein degradation, occurring when a protein is labelled with a polyubiquitin chain (Chau, *et al* 1989, Thrower, *et al* 2000) and activation of transcription factors (Kaiser, *et al* 2000, Kuras, *et al* 2002). The presence of this protein is important in antigen presentation by antigen presenting cells. In case of infections, degradation of intracellular protein leads to generation of pathogen-specific short peptides which can be presented on the cell surface of dendritic cells and other antigen presenting cells; such peptides, in turn, can be recognized by B and T lymphocytes leading to activation of adoptive immune responses.
The function of the PRR6 gene (also known as centromere protein V; CENPV) is not completely understood. It seems to be associated with the kinetochore and might play a role in spatial organization of the chromosomes during cell division, as demonstrated through studies of mutations (Tadeu, *et al* 2008).

Fragment "8"

This fragment contains the high-affinity TGIF2LY-binding sequence 5'-TGATC-3'. The genomic DNA region represented by this fragment is flanked by the gene NLR family, pyrin domain containing 1 (NLRP1) on the 5' side and the gene WSC domain containing 1 (WSCD1) on the other side. There are almost no information about the latter gene. It is supposed to code for a membrane protein which contains a WSC domain, potentially binding to carbohydrates (Ponting, *et al* 1999).

The NLRP1 gene is better characterised. This protein appears to be expressed rather ubiquitously (Hlaing, *et al* 2001) and it has been demonstrated to be expressed in CD14 positive cells and up-regulated in acute myeloid leukaemia blasts (Sanz, *et al* 2004). Its expression was found to be high in THP-1 cells (Kummer, *et al* 2007). NLRP1 is important for the innate immune system, being able to initiate apoptosis. NLRP1 protein contains a nucleotide-binding domain, a leucine-rich repeat region, and a caspase-recruiting domain (Chu, *et al* 2001). Such structure makes this protein similar to nucleotide-binding oligomerization domain (NOD) containing proteins, which are key components of innate immunity. These proteins are able to recognize bacterial peptido-glycans, by binding to muramyl dipeptide, a component of bacterial wall, and to direct the cells either to apoptosis or to activation as antigen presenting cells (Inohara and Nunez 2001). NLRP1 has been demonstrated to activate caspase 1 (Chu, *et al* 2001, Faustin, *et al* 2007). This protein has also been associated with autoimmune and autoinflammatory diseases (Jin, *et al* 2007).

Fragment "2"

The low-affinity TGIF2LY-binding sequences 5'-TGACA-3' and 5'-TGACC-3' were found to be present in this fragment. It represents a genomic region which is flanked by the eukaryotic translation elongation factor 2 (EEF2) gene on the 5' side and the gene protein inhibitor of activated STAT 4 (PIAS4) on the 3' side.

The first of these genes codes for a cytoplasmatic protein which is essential, as suggested by its name, for the elongation phase of RNA translation. EEF2 protein promotes the GTP-dependent translocation of the nascent protein chain from the A-site to the P-site of the ribosome (Browne and Proud 2002).

The second gene, PIAS4, codes for a protein which is involved in a large number of pathways of modulation of the immune system. PIAS4 plays its role downstream of cytokine-mediated signalling by down-regulating several transcription factors which, in turn, regulate subsets of cytokine-induced genes. More in details, PIAS4 is involved in response to interferon. It functions here as a corepressor of the STAT1 protein (Liu, *et al* 2001). STAT1 is involved in activation of several interferon-induced genes. Response to interferon produces several important effects like increase of antigen presentation by up-regulating MHC class I molecules expression and inducing the expression of specialized β-subunits of the proteasome, promoting in this way the assembling of the so-called immunoproteasome (Wang and Maldonado 2006). Activation of lysosomes in macrophages is another interferon-induced response. PIAS4 plays also a role in the NF-kB pathway. NF-kB is a cytoplasmic transcription factor which migrates to the nucleus upon activation. This process can be promoted by several stimuli, like Toll-like receptor 4 activation, and leads to activation of several genes promoting very different responses, depending on the cell in which it is activated, like inflammation, cellular proliferation, apoptosis and response to infection (Bonizzi and Karin 2004, Viatour, *et al* 2005). In this context, PIAS4 functions as repressor of TRIF-

mediated transcription (Zhang, *et al* 2004). TRIF is a key protein in activated-Toll-like receptor signal transduction. Another pathway in which PIAS4 plays also a role is the transforming growth factor β signalling. This pathway regulates, through the SMADs transcriptional activators, several biological processes including cellular differentiation and proliferation, tumorigenesis and immune response (Attisano and Wrana 2000, Yingling, *et al* 2004). PIAS4 negatively regulates this pathway by repressing SMAD3-mediated transcription either by recruiting histone deacetylase 1 (Imoto, *et al* 2003), or by sumoylating SMAD3 and SMAD4 (Lee, *et al* 2003). PIAS4 seems also to be involved in several ways in tumorigenesis by inhibiting, through sumoylation, the activity of some key proteins for cell differentiation and cell cycle regulation like MYB (Dahle, *et al* 2003) - which plays an important role in differentiation of hematopoietic cells (Oh and Reddy 1999) - PLAG1 (Van Dyck, *et al* 2004), and the Wnt-responsive transcription factor LEF1 (Sachdev, *et al* 2001).

Fragment "23"

This fragment contained both the low-affinity TGIF2LY-binding sequences. This genomic region is located on chromosome 2 and is flanked by the protein C (inactivator of coagulation factors Va and VIIIa) gene (PROC) on the 5' side and by the mitogen-activated protein kinase kinase kinase 2 (MAP3K2) gene on the other side.

The PROC protein is an anticoagulant serine protease, which inactivates the coagulation factors Va and VIIIa by proteolytic cleavage (Dahlback and Stenflo 1980, Marlar, *et al* 1982, Vehar and Davie 1980). Mutations of this gene have been associated with several pathologies like thrombosis and thrombophilia (Grundy, *et al* 1992a, Grundy, *et al* 1992b, Romeo, *et al* 1987).

The MAP3K2 protein plays a crucial role in immune response by activation of cytokine-induced NF-kB. This protein has been demonstrated to directly phosphorylate and activate NF-kB inhibitor kinases (IKKs), which consent the activation of NF-kappa B (Blank, *et al* 1996, Zhao and Lee 1999). Its expression has been found to be regulated in autoimmune diseases like rheumatoid arthritis (Hammaker, *et al* 2004).

Fragment "5B"

This fragment contains both the low-affinity TGIF2LY-binding sequences. It represents a genomic region located on chromosome 3 and is flanked by a locus which code for an hypothetical protein (LOC131831) on the 5' side and by the seven in absentia *Drosophila* homolog 2 (SIAH2) gene on the 3' side. SIAH2 is a small protein that functions as E3 ubiquitin-protein ligase, mediating in this way ubiquitination and subsequent proteasomal degradation of target proteins (Hu, *et al* 1997a, Hu, *et al* 1997b). It has been demonstrated that SIAH2 is involved in down-regulation of B-cell-specific transcription factors, which are usually induced upon B-cell activation (Boehm, *et al* 2001). SIAH2 also targets the degradation of DCC, BAG-1 and nuclear receptor co-repressors (Matsuzawa, *et al* 1998, Sourisseau, *et al* 2001, Zhang, *et al* 1998).

Different evidences indicate that SIAH2 is rather important in immune response: on one hand, it has been demonstrated that SIAH2 pro-apoptotic function is dramatically reduced upon interaction with proteins that inhibit or sequestrate SIAH2 (Okabe*, et al* 2003). On the other hand, it has been shown that SIAH2 directly interacts with proteins like TRAF2, a central regulator of cellular response to cytokines and stress, and promotes their degradation. TRAF2 is usually induced by the cytokine TNF-α. Expression of SIAH2 accelerates degradation of TRAF2 upon stimulation by TNF-α and promotes TNF-α-induced apoptosis, a crucial pathway in immune response (Habelhah*, et al* 2002).
The functions of all the previously described genes are summarized in table 3.3.

Fragment code	TGIF2LY-binding sequence	Genes	Function
XL15	5'-TGATA-3' high affinity	DKK1	↓ Wnt signaling: ↓ cell proliferation
		MBL2	↑ complement activation: ↑ innate immune response ↑ activation of adoptive immune response
3	5'-TGATA-3' high affinity	PRR6	kinetochore: cell division
	5'-TGACC-3' low affinity	UBB	ubiquitination: ↑ antigen presentation
8	5'-TGATC-3' high affinity	WSCD1	carbohydrates binding (not clear)
	5'-TGACA-3' and 5'-TGACC-3' low affinity	NLRP1	↑ innate immune response either apoptosis or antigen presentation
23	5'-TGACA-3' and 5'-TGACC-3' low affinity	PROC	↓ blood coagulation
		MAP3K2	↑ NF-kB: ↑ inflammatory and immune response
5B	5'-TGACA-3' and 5'-TGACC-3' low affinity	SIAH2	↑ubiquitination ↑ antigen presentation ↑ TNF-α-induced apoptosis
2	5'-TGACC-3' low affinity	EEF2	protein chain elongation during protein syntesis
		PIAS4	↓Inf response; ↓ NF-kB response ↓TGF- response; ↓ Wnt signaling

Table 3.3. Function of the genes potentially regulated by TGIF2LY. Almost all the genes play a role in pathways involved in immune response. TGIF2LY might also take part in modulation of immune response.

In this study, it was not possible to assess the biological function of TGIF2LY. For this reason two possibilities have to be taken into account: one is that TGIF2LY functions as repressor of gene transcription and the second is that this protein enhances gene transcription.

In the first case, assuming that TGIF2LY binds to the previously described genomic regions also *in vivo*, it would mean that all the mentioned genes will be down-regulated. This condition would bring several advantages to the leukaemic cells. They would be stimulated to proliferate, due to lack of inhibition of Wnt signalling and would be partially refractory to cytokine stimulation. It has to be kept in mind that THP-1 and AML M4-M5 blasts are derived from normal monocytes, a cell population which has a crucial importance in the immune system being responsible, through differentiation into macrophages, of initiating the innate immune response upon infection and activating the adoptive immune response through differentiation into antigen presenting cells. A consequence of expression of TGIF2LY within these cells, assuming it as a repressor, would dramatically reduce the ability of these cells to initiate immune response, inducing the cells to stay undifferentiated. In this case, expression of TGIF2LY in patients with AML M4-M5 would be a bad prognosis marker.

The second possibility is that TGIF2LY functions as an enhancer of the genes mentioned above. Up-regulation of all these genes would bring a disadvantage to the leukaemic cells. First of all, being the Wnt signalling inhibited, their proliferation capacity would be strongly reduced. Moreover, several genes among those described above are involved in stimulation of antigen presentation. If these genes get up-regulated, the AML M4-M5 cells would be forced to become antigen presenting cells, with the risk for them to present tumour-associated antigens which are usually rather antigenic. This would actually help the immune system to localize and kill selectively leukaemic cells. From this point of view, TGIF2LY should be considered as a good prognosis marker when it is expressed in AML M4-M5 patients.

According to the fact that the gene NLRP1 has been found to be highly expressed in THP-1 cells (Kummer*, et al* 2007), which express also TGIF2LY at protein level, would actually exclude the hypothesis that TGIF2LY repress the expression of NLRP1. It could also mean that the gene NLRP1 is not regulated by TGIF2LY in physiological conditions.

The hypothesis of TGIF2LY as an expression enhancer would also fit with the findings of Wang and Mann, who observed a loss of expression of specific target genes in vismay mutants suggesting that vis might directly or indirectly activate these genes (Wang and Mann 2003). However, as yet, none of the TGIF family members have been shown to activate transcription. Thus, it remains an open question whether genes like TGIF2LY and vismay play a direct role in the activation of target genes or, alternatively, if they act indirectly by repressing the expression of a repressor of these genes. To further understand these issues, other experiments will be necessary to completely elucidate the regulation mechanisms of TGIF2LY. It would be interesting, for example, to knock down the gene in THP-1 cells, usually expressing it, and analyze the expression of those genes potentially regulated by TGIF2LY.

3.4 Immunological studies on the candidate genes

3.4.1 Introductory remarks

Investigation of the function of genes, which showed up-regulation in AML patients' samples, is essential to assign them a role, as cause or effect, in tumour biology. However, there is another important aspect which has to be taken into account.
It is known that tumour cells express on their surface antigens which are normally not presented by healthy cells. AML cells are not an exception. In the last years, several tumour-associated antigens have been identified and characterized (Greiner*, et al* 2006). Moreover, several protocols for leukaemia treatment have been developed, which exploit the expression of tumour-associated antigens on the leukaemic cells (Duncan and Roddie 2008, Porter and Antin 2006, van de Loosdrecht*, et al* 2009).
Recently, the interest has been captured by particular tumour-associated antigens: the so-called cancer-testis antigens. These antigens are derived from proteins which are normally expressed in testis only. The fact that testis aren't normally under the surveillance of the immune system makes such antigens the perfect targets for immunotherapy (Guinn*, et al* 2007, Meklat*, et al* 2007). Among the candidate genes identified in this study, the genes TGIF2L (both X- and Y-linked) and VCY are usually expressed only in testis, although their transcripts have been found at very low levels in some healthy monocytes samples.
For these reasons, in parallel to investigation of the function of the protein TGIF2LY, all the candidate genes were analyzed for their immunological potential.

3.4.2 Immunological potential of the candidate genes

Peptide prediction
The amino acid sequences of BCORL2, PCDH11Y, TGIF2LX and TGIF2LY, and VCY were submitted to three different algorithms: HLA Peptide Binding Predictions (BIMAS), Syfpeithi, and Paproc. The first two programs provide information about peptides, present within the given amino acid sequence, which are predicted to bind HLA molecules. Both algorithms analyze the sequence evaluating, for each possible peptide, how stable the bond to the HLA molecules would be, taking into account each amino acid in each position of the peptide. For this study, the programs were asked to predict HLA-A0201-restricted peptides. In this way, it was possible to obtain a list of possible peptides derived from the sequences of the candidate proteins ranked by predicted HLA-A0201 binding capacity.

The hit-list was then re-ordered using the data derived from the proteasomal cleavage prediction software, Paproc. For example, if a peptide showed high affinity to HLA molecules, but there were several possible proteasomal cleavage sites within it, it was discarded.
Several remarks have to be done about the prediction method.
A reasonable question would concern the choice of HLA-A0201-restricted peptides only. This HLA type was chosen for several reasons. First of all, the HLA-A0201 allele is relatively abundant in European populations; its frequency can vary according to the populations taken into account, but in Western and Eastern Europe this allele is present in about 30% of individuals, according to the New Allele Frequency Database, http://www.allelefrequencies.net (Middleton*, et al* 2003). Working with samples collected only in Europe, in Germany and in England precisely, made the choice of this HLA allele reasonable.
Another reason led to the choice of HLA-A0201 allele for this study. The main problem, using other HLA alleles, is to find a reliable read-out system. In this case T2 cells have been used to investigate the real binding capacity of the selected peptides first, and to analyze their antigenicity then. T2 cells, due to TAP deficiency, are not able to load endogenous antigens on HLA molecules on their cell surface. T2 cells express, among the MHC class I molecules, HLA-A0201 alleles. If no antigens are loaded onto MHC molecules, they are unstable and are rapidly internalized and degraded. It is then possible, by adding HLA-A0201-restricted peptides to the culture medium, to stabilize the MHC class I molecules and to measure, through antibodies, the affinity of those peptides for the MHC complexes (see also paragraph 4.7.2).
It is of course possible to analyze the immunological potential of peptides restricted to other HLA alleles, but it would be necessary, for this purpose, to plan and optimize a different read-out system.
Concerning the use of the algorithm Paproc to predict the possible proteasomal cleavage of the candidate proteins, it has to be mentioned that its results have to be handled carefully. This software bases its cleavage prediction on experimental cleavage data of yeast and human 20 S proteasomes (Kuttler*, et al* 2000). The role of proteasome in antigen-presenting cells is crucial. Protein processed within these cells, are cleaved by the so-called immunoproteasome, a variant of the proteasome present in all the other cells, with different β-subunits (Wang and Maldonado 2006). There is a slight difference of cleavage site specificity between the two proteasomes. For this reason, results obtained from the algorithm Paproc have to be taken just as an indication. In this case, nonamers with three or more possible cleavage sites were rejected, but those with only one cleavage site were kept for *in vitro* analysis.

HLA-A0201-binding potential of the peptides

In order to assess antigenicity of the so-predicted peptides, the first step was to analyze their real binding capacity to HLA molecules. As mentioned before, only peptides with

high probability to bind those molecules have been taken into account. It has to be pointed out that a high predicted binding probability is not necessarily reflecting the real binding capacity of a given peptide: several factors can influence, *in vivo*, the binding ability of a peptide. The software used for predicting HLA-peptide complex stability take into account what amino acid is present in each position within the nonamer, in case of HLA-A0201. One of the other factors which can influence binding capacity is the three-dimensional conformation of the peptide of interest. If this structure is not spatially compatible with the HLA molecules' groove in which the peptide should be placed, the overall affinity of that peptide for HLA molecules will be strongly reduced, if not nullified. These are the reasons which make a peptide-HLA binding assay essential.

Referring to figure 2.16, it is possible to see that, among the eleven tested peptides, only five were really able to stably bind to HLA-A0201 molecules. A deeper analysis of all the peptides shows that there are motifs which are probably important for a stable bond of the peptide to HLA-A0201 molecules (figure 3.3).

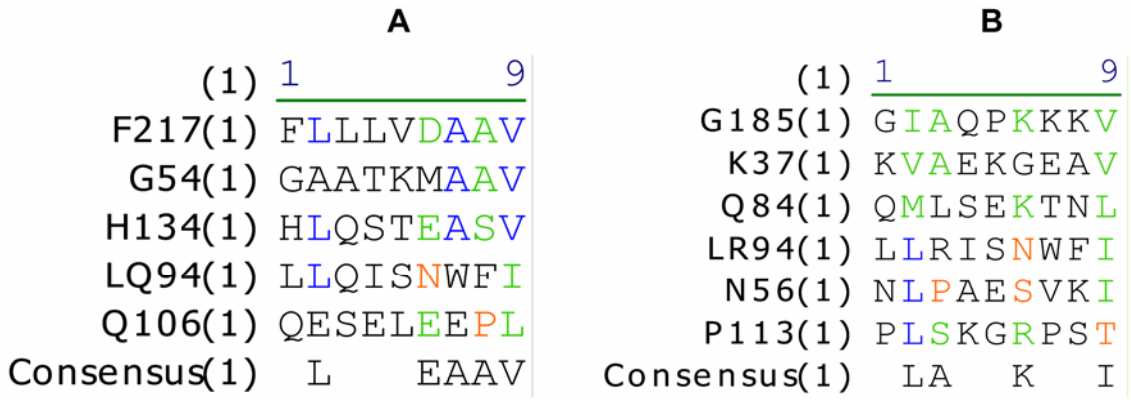

Figure 3.3. Search for peptide consensus sequence basing on their capacity to bind to HLA-A0201 molecules. Panel A: peptides which are able to bind to HLA-A0201 molecules. Panel B: peptides which are not able to bind to these molecules. It is possible to see that the C-terminal end of the binding peptides is conserved. Color scheme (color = amino acid substitution type): black = non-similar; blue = conservative; green = block of similar; orange = weakly similar.

It is possible to notice that the C-terminal portion of the peptides which are able to bind to HLA molecules shows some conservation. It has been shown that the most important amino acid for HLA-A0201 affinity are those in position 2 and 9 (Parker, *et al* 1995). However, this is probably not the only condition influencing binding capacity. Looking at the peptides LQ94 and LR94, it is possible to see that they differ only in one amino acid in position 3, a glutamine for LQ94 and an arginine for LR94. No arginine is found at this position among the binding peptides. It is possible that this amino acid in this position of the nonamer is not compatible with the binding groove of the HLA-A0201 molecules. It is interesting to notice that, although the five binding peptides reached a

binding value above the 1.2 threshold, none of them is comparable, for binding potential, to the positive control derived form influenza virus matrix protein (IMP: GILGFVFTL). This peptide is known to stably bind HLA-A0201 molecules and to be strongly immunogenic. Only the peptide F217, derived from the TGIF2LX gene, shows high binding potential, with a value of about 2.8 (against a value of about 4.8 for IMP; figure 2.16). The worst-binding peptide, H134, had a coefficient of about 1.25, slightly above the threshold. Comparing the IMP control with all the other binding peptides (figure 3.4), it is possible to evaluate all the similarities between them and to speculate about the amino acids important for HLA-binding.

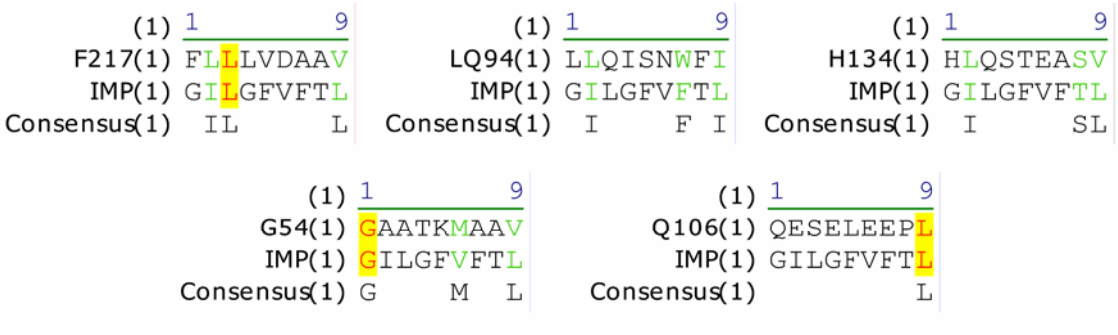

Figure 3.4. Search for key amino acids for peptide-HLA-A0201 binding. Alignment of the positive control peptide from influenza virus matrix protein (IMP) with the HLA-A0201-binding peptides. F217 showed the highest binding capacity among these peptides; H134 the lowest. Color scheme (color = amino acid substitution type): black = non-similar; green = block of similar; red on yellow background = identical.

Being F217 the best-binding among the candidate peptides, it is possible to get other evidences that position 3 is crucial for HLA-A0201 binding. In fact, leucine or similar amino acids in positions 2 and 9 are present also in non-binding peptides; the only significant difference concerns amino acids in position 3. Of course, it is possible that, for highly stable binding several conditions have to be fulfilled simultaneously.

Antigenicity of the peptides

The five HLA-A0201-binding candidate peptides were then tested for their antigenicity. To do so, T2 cell were loaded with these peptides and used to stimulate the growth of peptide-specific T-lymphocytes isolated form a HLA-matched healthy female blood (see paragraph 4.7.3).

The results of this experiment (figures 2.17 and 2.18) were not that encouraging. Even for T2 cells loaded with a known immunogenic peptide, melanoma antigen family A-3 (MAGE3: FLWGPRALV), only a small percentage of specific lysis could be seen. This result can be explained by taking into account the conditions with which the chromium release assay has been performed. First of all, it is possible that radioactive labelling of the target cells is somehow affecting the cells' viability increasing the background of

released ^{51}Cr and reducing, for this reason, the signal derived from the specific ^{51}Cr release. On the other hand, it has to be taken into account that T2 cells are not the perfect antigen-presenting cells. This cell line is derived by a fusion of a B- and a T-lymphoblastoid cell lines (Salter and Cresswell 1986). B-cells are normally functioning as antigen-presenting cells, expressing on their cell surface the co-stimulatory molecule CD80, expressed also by monocytes, which interacts with CD28 molecules on the surface of T-cells, promoting their activation and survival. T-cells, both the CD4 and CD8 positive, do not have this function. Their task is to recognize antigens presented by cells and, if necessary, to start a cell-mediated immune response. T2 cells are able to stably load their HLA-A0201 molecules with peptides, as shown by the binding assay (figure 2.16), but, probably, the stimulation of T-cells is not as effective as that of natural antigen-presenting cells. Nevertheless, it is possible to see that the peptide G54 is recognized by T-cells and a small percentage of the cells presenting it are killed by specific T-cells.

The small percentage of specifically lysed cells is not necessarily a bad result. First of all, it has to be kept in mind that, also in the case of the positive control peptide, the percentage of specifically lysed cells is quite small. Probably, by optimizing the procedure, it is possible to make the assay more sensitive and to see whether to an increase of lysis of control peptide-loaded cells corresponds also an increase of lysis of cells loaded with the G54 peptide.

Furthermore, according to the result of the binding assay, G54 is one of the peptides which showed the lowest affinity to the HLA-A0201 molecules, with a coefficient of about 1.5. This means that the bond of this peptide is rather weak. Having said that, it is clear that, among the target cells loaded with the peptide G54 and used to expand a specific T-cell population, only a relatively small percentage presents a large number of HLA-A0201-peptide complexes. This can have influence both on T-cell stimulation, and thus expansion, and on recognition of the antigen by the so-expanded T-cells.

In order to assess the real antigenic potential of the G54 peptide it would be necessary either to optimize the chromium release assay presented here, or to modify it by using a different target for expansion and specificity assessment of the T-cells. One possibility, for this purpose, can be the generation of dendritic cells from an AML M4-M5 male patient's sample. These cells can be either transfected with an expression vector encoding the full-length protein of interest - G54 is derived from the VCY protein - or incubated together with the purified peptides. Protocols for dendritic cells generation from AML cells have already been established (Schuster*, et al* 2008). The challenge here would be to find a full-matched female donor to obtain T-cells. Peptide-specific cytotoxic T-cells can be expanded from such a donor and used to test their capacity to lyse peptide-presenting target cells. Another possibility can be provided by THP-1 cells. It is possible to induce differentiation of THP-1 cells into dendritic cells by cultivating them with a cytokine cocktail containing granulocyte-monocyte colony stimulation factor (GM-CSF) (Berges*, et al* 2005). Differentiated THP-1 cells could be a useful tool

for peptide antigenicity assessment. It would be necessary to optimize the method by adding the peptides of interest, or the full-length protein, to the medium, in order to make the THP-1 cells presenting that antigen. Here as well, it would be necessary to find a full HLA-matched female donor to get T-cells.

3.5 Outlook

3.5.1 The "time factor"

As mentioned several times before, the so-called "time factor" limited dramatically the possibilities of the experimental work for this study. The present paragraph is not meant to be a complaint; it is just exposing the problems arising during the ongoing project, which influenced, sometimes heavily, the time schedule.
There were several reasons for that. First of all, the initial part of the study consisted in establishing and optimizing all the methods. The choice of the AML subtype to investigate was a crucial step and several trials have been made to establish the optimal cell separation strategy, according to the idea of obtaining the same cell population from both leukaemic patients and healthy volunteers' blood. The next step, the collection of AML samples, took also quite a while and several problems arose during this phase. Although AML M4-M5 cases are not so rare, the fact that eligible samples should show high expression of CD14 molecules and be from male patients restricted considerably the possibilities. Some samples had not been immunophenotyped at the time of diagnosis and, in order to assess their suitability, they had to be checked through flow cytometry for presence and abundance of CD14 molecules on the leukaemic cells. The limited availability of some samples, intended as low number of cells frozen at the diagnosis or as loss of cells after thawing, made the whole sampling procedure longer than expected.
The protocol for low density arrays needed to be optimized as well. Several pilot experiments had to be performed in order to understand how reliable this method was. The analysis of the samples with this tool provided an impressing amount of data, which had to be evaluated very carefully sample by sample. Analysis of the so-obtained data took time also because of the scarce performance of the software used for the evaluation.
Another phase of the project that required more time than planned was the evaluation of the data obtained with the TGIF2LY-DNA binding assay. The search of a possible DNA binding sequence for the protein took quite some time. The main reason is that

comparing all the nucleotide sequences of the previously-obtained TGIF2LY-binding fragments with each other, in order to find all the possible short common sequences, was a long and difficult procedure.
Time problems influenced also the immunological part of the project. One example concerned the delivery of peptides to be analyzed for HLA binding ability and for antigenicity. This alone required longer than one month, becoming one of the "speed-limiting step" for the schedule of the planned experiments.
All the mentioned factors influenced, in different ways, the time schedule of the project. Nevertheless, they were all necessary steps to get data which could, and can, be used for further experimental work.

3.5.2 Further steps

The present study has opened several research topics and provided the basis for new experimental works.
From a biological point of view, it would be very interesting to fully understand the role of TGIF2LY in healthy and leukaemic cells. Some indications about the function of this protein have been provided from the present study; other important points, concerning the biology of TGIF2LY, have to be elucidated. It will be essential to perform other *in vitro* and *in vivo* experiments to finally understand whether TGIF2LY plays a role in regulation of genes involved in innate and adoptive immune response and how the regulation mechanisms do function.
The function of the other candidate genes, in both healthy and leukaemic cells, remains also to be elucidated. Once these points are assessed, it will be possible to understand whether their over-expression in leukaemic cells has to be seen as cause or as consequence of the tumour. Further studies are necessary to assess the biological role of these proteins within the leukaemic cells and to implement their use, if possible as prognostic tools or as therapeutic targets.
The immunological aspect of the candidate genes is also very important and deserves further studies. The direct involvement of candidate gene-derived peptides in the recognition of leukaemic cells by the immune system is a possibility which might bring enormous advantages for the treatment of the disease. New therapy strategies can be designed based on the use of peptide specific T-cells for the complete removal of residual leukaemic cells after transplantation. It will be necessary to perform further analysis, in the case of the G54 peptide, focused on the assessment of the real antigenicity of this peptide, maybe using patient-derived dendritic cells.
Further research should also be focused on the analysis of candidate gene-derived peptides restricted to other HLA types.

4 Material and Methods

4.1 Working with *Escherichia coli*

4.1.1 Culture media

After preparation, all the media have been autoclaved for 20 minutes at 121°C. For solid media, agar-agar was added at a concentration of 1.5% (w/v) before sterilisation. All the temperature-sensitive ingredients have been sterile filtrated and added after autoclaving.

LB medium (Bertani 1951):	10 g casein hydrolysate 5 g yeast extract 4 g NaCl H_2O_{bidest} ad 1 L, pH 7.4
SOC medium (Sambrook and Russell 2001):	2 g tryptone 0.5 g yeast extract 0.2 g $MgCl_2 \times 6\ H_2O$ 0.25 g $MgSO_4 \times 7\ H_2O$ 0.36 g glucose H_2O_{bidest} ad 0.1 L

4.1.2 Inhibitory and selective media supplements

All the used supplements have been sterile filtrated. For liquid culture media, they were added directly before inoculation. For solid culture media, they were added once the media temperature was lowered to ~45°C. Ampicillin (final concentration: 50-100 µg/mL) and kanamycin (final concentration: 30 µg/mL) have been used as selection antibiotics, both dissolved in H_2O_{bidest}.

4.1.3 Growing of bacteria

Inoculation of *E. coli* on solid media was conducted with inoculating loops, previously sterilised by heating. For inoculation in liquid media, both from colonies on agar plates and glycerol stocks, sterile pipet tips were used. Liquid cultures of more than 30 mL volume were inoculated 1:100 from overnight grown liquid cultures (5 mL). All the cultures were growth at 37°C and, in case of liquid cultures, under vigorous shaking (≥ 200 rpm).

4.1.4 Short- and long-term storage of bacteria cultures

After overnight incubation at 37°C, *E. coli* cultures can be stored on agar plates for several months at 4°C when wrapped into Parafilm (American National Can, USA). Glycerol stocks were prepared for bacteria long-term storage by mixing sterile 80% glycerol with the overnight grown liquid cultures (1:1), and stored at -80°C in Cryo Tube™ vials (Nunc, Denmark).

4.1.5 Production of chemocompetent *E. coli*

2.5 mL overnight culture were added to 250 mL LB medium with 20 mM $MgSO_4$ and grown under vigorous shaking at 37°C up to OD_{600} 0.6-0.8. After centrifugation (4000 × g; 4°C; 15 minutes), the bacterial pellet was resuspended in 100 mL ice-cold TFB1 buffer, placed on ice for 5 minutes and centrifuged again. After resuspension in 10 mL ice-cold TFB2 buffer and an incubation of 30 minutes on ice, bacteria were split into 100 µL aliquots, shock-frozen in liquid nitrogen and stored at -80°C.

Buffers:

TFB1 (pH 5.8): 30 mM KAc
100 mM RbCl
10 mM $CaCl_2$
50 mM $MnCl_2$
15% glycerol

TFB2 (pH 6.5): 10 mM MOPS
75 mM $CaCl_2$
10 mM RbCl
15% glycerol

4.1.6 Transformation of chemocompetent *E. coli* by heat shock

The chemocompetent bacteria were thawed on ice and gently mixed with 1-5 ng of plasmid DNA or a maximum of 5 µL DNA from a ligation reaction. After an incubation of 20 minutes on ice, the cells were heat-shocked at 42°C in a water bath for 30-60 seconds (depending on the strain) and immediately placed back on ice for at least 1 minute. After addition of 400 µL SOC medium, cells were incubated under vigorous shaking at 37°C for 60 minutes. Subsequently, ~100 µL cell suspension were plated onto LB agar plates containing the appropriate selective antibiotics. The plates were incubated overnight at 37°C.

4.2 Working with eukaryotic cell lines

4.2.1 Cultivation of cell lines

293T
This is a subline of human embryonic kidney (HEK) 293 cells, which has been transformed with adenovirus and contains the SV-40 T large antigen. These cells were cultured in High Glucose Dulbecco's Modified Eagle Medium (D-MEM; PAA Laboratories, Austria) supplemented with 10% Fetal Bovine Serum (FBS; Biochrom, Germany), 2 mM L-glutamine, 100 U/mL penicillin and 100 µg/µL streptomycin (all from Gibco, USA). 293T cells were grown as a monolayer, at 37°C and 5% CO_2 in a humidified atmosphere.

HeLa
This is a well-established human cervix carcinoma cell line. Good adhesive and transfection properties and a relatively large cell volume make these cells optimal for subcellular localisation studies. HeLa cells were cultured in Modified Eagle Medium (MEM; PAA Laboratories, Austria) supplemented with 10% FBS, 2 mM L-glutamine,

100 U/mL penicillin and 100 µg/µL streptomycin. They were grown as a monolayer, at 37°C and 5% CO_2 in a humidified atmosphere.

174CEM.T2 (T2)
This T-lymphocytic cell line is unable to load endogenous peptides on the surface MHC complex, due to lack of expression of the TAP protein. The cells were cultured in RPMI-1640 medium (PAA Laboratories, Austria), supplemented with 10% FBS, 2 mM L-glutamine, 100 U/mL penicillin and 100 µg/µL streptomycin. T2 cells were grown in suspension, at 37°C and 5% CO_2 in a humidified atmosphere.

THP-1
This monocytic cell line has been established from the peripheral blood of a one year old male patient with acute myelomonocytic leukaemia (Tsuchiya*, et al* 1980). The cells were cultured in RPMI-1640 medium, supplemented with 10% FBS, 2 mM L-glutamine, 100 U/mL penicillin and 100 µg/µL streptomycin. THP-1 cells were grown in suspension, at 37°C and 5% CO_2 in a humidified atmosphere.

4.2.2 Maintenance of cell culture: Splitting, Thawing, Freezing and long-term storage of eukaryotic cells

Splitting
293T and HeLa cells, which grow as a monolayer, adhering to the bottom of the culture flasks, need to be transferred to a new culture flask, before they reach confluence in the old one. For this purpose, the culture medium is removed; the cells are washed with warm (~30°C) phosphate buffered saline solution (PBS) and then covered with a trypsin/EDTA solution (3 mL for T75 culture flasks; 0.8 mL for T25). After a short incubation at 37°C, the cells are detached from the bottom of the flasks and resuspended in warm fresh culture medium. HeLa cells were usually split 1:4 – 1:8 in medium; 293T cells 1:8 – 1:16, because of growing more rapidly. The cells were split every 3-4 days.
Cells which grow in suspension (T2 and THP-1) were splitted 1:4 – 1:8 once or twice a week. To do so, the cells are resuspended by gently pipetting and an aliquot is diluted in fresh culture medium.

PBS: 10 mM sodium-phosphate buffer (pH 7.4)
150 mM NaCl

Thawing

To thaw eukaryotic cells, one vial was taken from liquid nitrogen and quickly put in a 37°C water bath. The cell suspension was subsequently diluted in 9 mL of the appropriate culture medium, centrifuged for 10 minutes at 1000 × g. Cell pellet was resuspended and transferred to a T75 culture flask. The thawing procedure should be conducted as fast as possible, as the freezing medium contains dimethylsulfoxid (DMSO), which is toxic for the cells, when the temperature is raised.

Freezing and long-term storage

Cells at about 80-90% confluence in a T75 flask were trypsinized, resuspended in the appropriate culture medium and subsequently collected by centrifugation (10 minutes at 500×g). Culture medium with the usual amount of antibiotics, 90% FBS and 10% DMSO was used as freezing medium. The cell pellet was resuspended in 2 mL of freezing medium, which was then pipeted into two Cryo Tube™ vials. In order to avoid shock-freezing, potentially lethal for eukaryotic cells, a Nunc freezing box was used, filled with isopropanol and put at -80°C. This allows a temperature decrease of 1°C/minute and avoids crystallisation of cell content.

For long-term storage, the vials were transferred to liquid nitrogen; for storage up to six months, they were put at -80°C.

4.2.3 Transfection of eukaryotic cells

To express proteins in eukaryotic system, cells were transfected with FuGene6 reagent (Roche). FuGene6 is based on the principle of lipofection and acts by including plasmid DNA into lipidic droplets. The delivery of the transfectant plasmid is achieved by fusion of those lipidic droplets with the plasmatic membrane of the cells.

Based on the manufacturer's protocol, the transfection procedure was optimized. For all cell lines used in this work, a 3:1 ratio of µL Fugene6 per µg DNA was found to yield the best results. As described in the original protocol, FuGene6 was first mixed with the culture medium without FBS and incubated at room temperature for 5 minutes. DNA was subsequently added and the mix was gently mixed by inversion. Drops were collected by brief centrifugation. Before transfection, the mixture was incubated for another 20-25 min at room temperature to allow the formation of the FuGene6-DNA complexes.

4.2.4 Growing of HeLa cells for subcellular localization studies

For subcellular localization studies, 5×10^4 to 10^5 cells were seeded into each well of a 6 well plate containing a microscope glass cover slip. Transfection was performed with 1 µg of DNA per well. HeLa cells have been chosen for their rather large cellular volume, which allows a better visualization.

Staining of mitochondria using compartment-specific tracker
For staining of mitochondria, fluorescence-coupled tracker (Invitrogen, USA) was used. This tracker is transported actively to the mitochondria within the living cells. Therefore, the tracker was diluted to 300 nM in culture medium and then added to the cells. After incubation at 37°C for 30-35 minutes, culture medium was removed and cells were washed twice with PBS.

Fixation
Fixation of the cells was conducted applying 1 mL of 3.7% formaldehyde in PBS and incubating at 37°C for 10-15 minutes. After fixation, cells were washed twice with PBS.

Permeabilization
If necessary for antibody staining, permeabilization of the cells was carried out by adding 1 mL 0.5% Triton-X (Merck, USA) in PBS to each well and incubating at room temperature for 5-10 minutes. After that, cells were washed three times in PBS.

Antibody staining
For antibody staining of distinct cellular compartments or MYC-tagged proteins under investigation, permeabilized cells were first blocked, incubating them with 3% bovine seroalbumine (BSA; New England Biolabs, USA) in PBS, at room temperature for 30 minutes. Primary and secondary (where necessary) antibody solutions were prepared in 1:2000 dilutions (or as indicated) in blocking solution; cells were incubated with them at room temperature for 1 hour for each antibody solution. Before and after application of the antibody solutions, cells were washed three times with PBS. In this study, in order to localise TGIF2LY, the primary antibody was a mouse monoclonal anti-MYC 9B11 (New England Biolabs, USA) and the secondary antibody was an AlexaFluor 488 Goat anti-Mouse IgG (Molecular Probes, USA).

Nuclei staining
Nuclei staining was performed with either DAPI (4*'*,6-diamidino-2-phenylindole,dihydrochloride) or Hoechst33342 (both from Invitrogen, USA). DAPI was diluted in H_2O_{bidest} and applied on the cells with a final concentration of 300 nM.

The cells were incubated for 1-2 minutes. Hoechst33342 staining was diluted 1:5000 or 1:10000 in PBS and the cells were incubated for 1-5 minutes.

Microscope slides preparation
After all necessary staining steps, the cells were washed three times with PBS. The glass cover slips were removed from the plate's wells and pasted upside down on Super-Frost Plus cover slides (Menzel Glaeser, Germany) using Vectashield (Vectalabs, USA).

Detection of fluorescence and documentation
Fluorescence was detected using a Zeiss Axiophot fluorescence microscope with respective filters. Camera was controlled for taking photographs by Axiovision 4.6 software (Zeiss, Germany). Adobe Photoshop CS was used to process the photographs, namely, to define lower and upper thresholds, to separate the red, green, and blue channel.

4.3 DNA-based molecular biological methods

4.3.1 Isolation and purification procedures

Isolation of plasmid DNA from bacteria
Isolation of plasmid DNA from *E. coli* strains was carried out in two different scales: mini-preparation, using 1 to 5 mL of an overnight bacteria culture and the NucleoSpin Plasmid kit (Macherey & Nagel, Germany), and midi-preparation, using 50 mL of an overnight culture and the NucleoBond PC100 Kit (Macherey & Nagel, Germany). Bacteria were harvested by centrifugation for 15 minutes at 9000×g at 4°C and the pellets resuspended in the appropriate lysis buffer of the kits. DNA was isolated according to the manufacturer's protocols. In the mini-preparation, DNA was eluted from the columns with 50 µL of H_2O_{bidest}. The elution volume for the midi-preparation was 100-200 µL H_2O_{bidest}.

Isolation of genomic DNA from cultured cells
Genomic DNA from cultured cell lines was obtained using the Wizard SV genomic DNA Purification System (Promega, USA). According to the manufacturer's specifications, the protocol was optimized. After cell lysis, binding of the genomic DNA to the provided silica columns and the washing steps, the elution volume was

reduced to 50 µL of H_2O_{bidest} for 5×10^6 lysed cells in order to concentrate the eluted DNA.

Purification of linear dsDNA from aqueous solutions and agarose gels
DNA purification from solutions containing enzymes, dNTPs, salts etc. was carried out by use of the Wizard SV Gel and PCR Clean-Up System (Promega, USA) according to the manufacturer's protocols. For elution, 35-40 µL H_2O_{bidest} were applied.
For DNA purification from agarose gels, TAE buffer based 1-2% agarose gels (w/v) were used.

50× TAE: 2 M Tris
1 M acetic acid
0.1 M EDTA
H_2O_{bidest} up to 1 L

Precipitation of DNA
Highly diluted or impure DNA can be precipitated in order to increase concentration or purity. Therefore, 0.1 volume of 3 M sodium acetate (pH 5.2) and 2.5 volumes of ethanol p.a. were added to the DNA-containing solution. This was incubated for at least 2 hours at -20°C. After centrifugation (15 minutes at 75000×g and 4°C), the supernatant was removed and the pellet washed in -20°C cold 70% ethanol (centrifugation: 10 minutes at 75000×g and 4°C). After removal of the supernatant, the DNA pellet could be dried and resuspended in the optimal volume of H_2O_{bidest}.

4.3.2 Measurement and quality assessment of DNA solutions

Separation and monitoring by agarose gel electrophoresis
For qualitative measurement of DNA or subsequent isolation, DNA-containing solutions were applied to an agarose gel for electrophoresis. Different w/v percentages of agarose in 1× TBE buffer were used to optimize the analysis according to the expected size of the DNA fragments: 0.5% for dsDNA ≥ 5 Kb; 1% for dsDNA ~2-5 Kb; 2% for dsDNA ~0.5-2 Kb and 3% for dsDNA ≤ 0.5 Kb. For analysis, 6× Loading Dye was added to the DNA solution, which was mixed and applied to the gel containing ethidium bromide (0.5 µg/mL). The DNA was separated on the basis of the fragments' weight in a constant electric field of 80-160 V, depending on gel size and percentage of agarose. DNA bands on the gel were monitored and photographed with UV trans-illumination (λ=254 nm) on a Bio Vision gel documentation system (PeqLab, Germany).

By adding 6× Loading Dye to the sample, the gel run could be optically monitored and stopped according to the separation. The included dyes, bromophenol blue and xylene cyanol FF behave on a 0.8% agarose gel like dsDNA fragments of 0.3 and 4 Kb size, respectively.

10× TBE: 108 g Tris
55 g boric acid
9.3 g EDTA
H_2O_{bidest} up to 1 L

6× Loading Dye: 15% Ficoll 400 (Pharmacia)
0.25% bromophenol blue
0.25% xylene cyanol FF
or purchased as ready-to-use solution from Fermentas

DNA fragment size was determined taking into account that, within an agarose gel matrix, running velocity of dsDNA fragments is inversely proportional to the decadic logarithm of their molecular weight. It is therefore possible to estimate the length of a fragment of unknown size by comparison with the R_f values of length of standardized fragments. Usually, this is achieved by inclusion of DNA standards to gel electrophoresis. In this work, λ-Markers 3 and 8 were used (both from Fermentas, Canada).

Quantification by optical density (OD)

To monitor quantity and quality of nucleic acid preparation, DNA (or RNA) containing solutions were diluted in H_2O_{dest} and the optical density (OD) of the resulting solution was measured at λ=260 nm for quantification and, additionally, at λ=280 nm for qualification. The concentration of the nucleic acid solution was calculated according to the following formula (adapted from (Cryer*, et al* 1975)):

OD_{260} × dilution factor × 50 ng/µL = concentration [ng/µL]

The quality was assessed by the ratio value R (OD_{260} nm/OD_{280} nm) where R = 1.8 for dsDNA represents the optimal value.
All the OD measurements were conducted using the Beckman DU 530 Life Sciences UV/Vis Spectrophotometer (Beckman, Germany).

4.3.3 Cloning strategies

Cloning via restriction sites

For most applications, DNA fragments were inserted into vectors via restriction sites. Digestion of DNA fragments and vector with restriction endonucleases yields complementary (so-called sticky) DNA ends which can then be used for ligation. In contrast to TA-cloning, utilization of two different restriction sites additionally allows a site-directed insertion.

TOPO-TA cloning

For direct cloning of DNA fragments from PCR reactions with no prior cleanup, the TOPO-TA cloning Kit (Invitrogen, USA) was used. For this method, A-overhangs produced in the elongation step of the PCR by the Taq-polymerase are ligated to T-overhangs in the respective vector (TA-cloning). The yield of ligation products is enhanced by topoisomerase, attached to the vector's cloning sites (TOPO-cloning). After amplification of the DNA by PCR 4 µL of the reaction mixture were subjected to the ligation procedure into vector pCRII or pCR2.1 and transformed into chemocompetent TOP 10 cells as recommended by the manufacturer.
For cloning of DNA fragments longer than 3 Kb, the TOPO-XL cloning kit (Invitrogen, USA) was used, according to the manufacturer's protocol.

Cleavage of DNA by restriction endonucleases

For cloning purposes or analysis of plasmids, DNA fragments and vectors were digested with the desired restriction enzymes. 1-10 µg DNA were digested in 20-50 µL reaction volume containing appropriate concentrations of buffer, BSA (if required) and 5-20 U enzyme for 2-16 h at 37°C. The adequate amount of enzyme was calculated according to the assumption that 1 U of enzyme digests 1 µg DNA in 1 hour under optimum conditions. The reaction was stopped by heat inactivation or removal of the enzyme using the Wizard SV Gel and PCR Clean-Up System (Promega, USA; see also chapter 4.3.1). All the restriction enzymes used for this study were purchased from New England Biolabs, USA.

Dephosphorylation of linear DNA fragments by alkaline phosphatase

Alkaline phosphatase can be used to cleave terminal 5'-phosphate groups. This is especially suitable to inhibit religation of restricted vectors, in order to increase the final amount of linearized vectors. Dephosphorylation can be conducted in common restriction enzyme buffers or in a separate buffer provided with the enzyme. Usually, calf intestine alkaline phosphatase (CIAP; Fermentas, Canada) was added directly to vector digestions for the last 30 minutes of incubation and heat-inactivated or cleaned-up with the restriction reaction.

Ligation of DNA fragments

For ligation of inserts and respective vectors, DNA fragments were mixed in a 4:1 molar ratio (insert:vector) and incubated with 1× ligase buffer and T4-DNA-Ligase (New England Biolabs or Promega). Incubation was conducted at room temperature for 2-4 hours or at 14°C overnight.

DNA amounts were adjusted according to the manufacturer's recommendations. For transformation protocol, see paragraph 4.1.6.

4.4 RNA-based molecular biological methods

4.4.1 Isolation and purification procedures RNA methods

Isolation of total RNA from eukaryotic cells

In order to isolate total RNA from cultured cells and from both fresh and frozen leukocytes samples, the RNeasy Mini Kit (Qiagen, USA) was used. According to the manufacturer's specifications, the protocol was optimized. Before starting, β-mercaptoethanol was added to the lysis buffer of the kit (RLT buffer) up to a final concentration of 10 µL per mL of buffer. 600 µL of this lysis buffer were used to lyse up to 10^7 cells. The lysate homogenization was performed by passing the cell lysate 5 to 10 times through a 20-gauge needle (0.9 mm diameter). After binding the RNA to the silica columns provided from the kit, RNA was washed and eluted in 40-45 µL of H_2O_{bidest} in order to concentrate it.

Only RNase-free equipment was used for handling of cells from which the RNA should be isolated.

Precipitation of RNA

Highly diluted or impure RNA can be precipitated in order to increase concentration or purity with the same protocol used for the precipitation of DNA (see paragraph 4.3.1)

Determination of RNA concentration by measurement of OD

RNA concentrations were determined in the same way as described for DNA (see paragraph 4.3.2) but calculated with the following formula:

OD_{260} × dilution factor × 40 ng/µL = concentration [ng/µL]

4.4.2 Determination of RNA integrity

In order to assess the integrity of isolated RNA, the RNA 6000 Nano Kit (Agilent Technologies, USA) was used. According to the manufacturer's protocol, 1 µL of each RNA sample was loaded into the sample wells of a RNA Nano Chip, provided with the kit. The chips were analysed with an Agilent 2100 Bioanalyzer (Agilent Technologies, USA). Using microcapillary electrophoretic separation on these chips, RNA samples are separated and subsequently analysed via laser induced fluorescence detection. The bioanalyzer software (2100 Expert software) generates an electropherogram and a gel-like image and displays results such as sample concentration and the ratio between 18S to 20S ribosomal RNA, which, in turn, gives a first hint about the RNA degradation within the sample. The electropherogram provides a detailed visual assessment of the RNA sample quality as well as the RNA integrity number (RIN) relative to each sample. The RIN is determined by the ratio of the ribosomal bands and by the entire electrophoretic trace of the RNA sample, for example the presence or absence of degradation products. The RIN scale goes from a value of 10, for intact RNA, to a value of 1, for totally degraded RNA (Schroeder et al., 2006). In this study only RNA samples with a RIN value ≥ 8 were used for gene expression analysis.

4.4.3 Reverse transcription of RNA into cDNA

Transcription to cDNA was performed by use of RevertAid First Strand cDNA Synthesis Kit (Fermentas, Canada) according to manufacturer's protocol. cDNA used in gene expression analysis was generated from 1 µg total RNA with random hexamer primers in a total volume of 20 µL.

4.5 PCR-based methods

4.5.1 Polymerase chain reaction (PCR)

In order to amplify specific DNA fragments, the polymerase chain reaction (PCR) was used. PCRs were conducted on DNA templates in 20 or 50 µL volume formats containing 200 nM dNTP-mix, 0.5 μM each forward and reverse primer and 1 or 2 U DNA polymerase in 1× PCR Buffer. Usually, in-house made Taq-polymerase was used.

For cloning purposes, Taq was substituted by proof-reading polymerases as Pfu Turbo DNA-polymerase (Stratagene, USA) or Pfx Polymerase (Invitrogen, USA) for better performance and accuracy. DNA templates included genomic DNA, plasmid DNA, linear dsDNA, cDNA, primary PCR (without prior cleanup) or bacterial culture. The standard program on a Robo-Cycler PCR machine (Stratagene, USA) implied an initial denaturing for 5 minutes at 95°C followed by 35 cycles with 30 seconds at 95°C, 35 seconds at annealing temperature (T_a), 1 minute for each amplified kb at 72°C, where T_a is the melting temperature (T_m) specific for the primer pair used lowered by 5°C. This cyclic reaction was usually followed by a 10 minutes final elongation step at 72°C. For bacterial culture as PCR template, initial denaturing was prolonged to 10 minutes.

10× PCR Buffer:	100 mM Tris-HCl, pH 9.0
	500 mM KCl
	15 mM $MgCl_2$

PCR optimization

In case of unspecific or too little yielded PCR product, three strategies were followed. Mainly, this was the case for PCR reactions where primers contained large overhangs due to the addition of restriction sites useful for cloning purposes. First, DMSO was added to the PCR reaction in minimum amounts recommended in the Stratagene Pfu Turbo protocol. Second, cycling during PCR program was split into two parts: 10 cycles at 50°C as T_a, followed by 25 cycles with the optimum temperature calculated. If possible, a primary PCR was conducted with primers containing no overhangs yielding a product of slightly bigger size than initially needed. The secondary PCR-reaction was then performed on 1 µL of the initial PCR reaction with the cloning primers now binding closer to the core of the primary PCR product (nested PCR).

4.5.2 Real time-PCR

The combination of the classical polymerase chain reaction technique with a system that allows obtaining data from each step of the PCR (real time) is a very powerful tool to evaluate the expression - relative or absolute - of genes of interest in biological samples. This technique is based on the usage of a fluorochrome which emits a fluorescent signal once bound to double strand DNA in order to quantify the amount of amplicons at each PCR cycle.

In the past years, several fluorochromes have been developed. The most used real time-PCR fluorochromes are the following two:

Sybr Green

It is a cyanine dye which, once bound to double strand-DNA, absorbs blue light, at λ=488 nm and emits green light at λ=522 nm. Besides Taq polymerase, its buffer and dNTPs, the reaction mix consists of the Sybr Green dye and a couple of primers specific for the gene of interest. When the PCR goes on and amplicons are formed, the dye is incorporated into the double strand-DNA and emits green fluorescence (figure 4.1, panel A).

TaqMan probe

It is a short oligonucleotide that contains a fluorescent dye (typically at the 5' end) and a quenching dye (typically located at the 3' end). When irradiated, the excited fluorescent dye transfers energy to the nearby quenching dye molecule rather than fluorescing (FRET), resulting in a non-fluorescent substrate. TaqMan probes are designed to hybridize to an internal region of a PCR product. During PCR, when the polymerase replicates a template on which a TaqMan probe is bound, the 5' exonuclease activity of the polymerase cleaves the probe. This separates the fluorescent and quenching dyes and FRET no longer occurs. Fluorescence increases in each cycle, proportional to the rate of probe cleavage (figure 4.1, panel B).

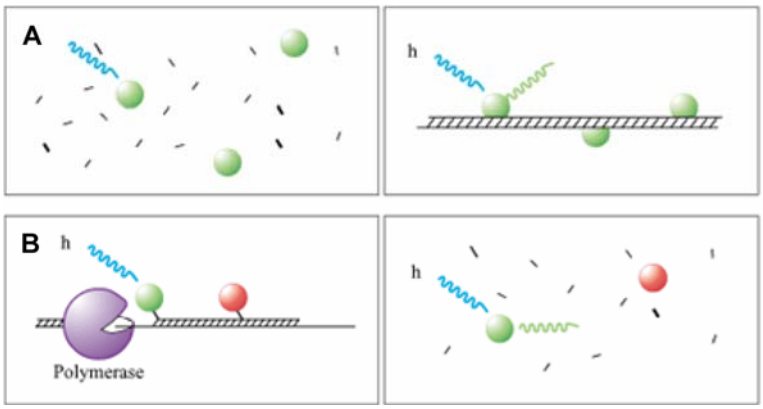

Figure 4.1. Real time-PCR strategies. A. Sybr Green: emits a green fluorescence when irradiated (h) only if bound to double strand-DNA. B. TaqMan probe: the green fluorescence is quenched as long as the probe is intact. During the elongation step of the PCR, the probe is degraded from the polymerase and the fluorescence is released. Modified from: http://www.gene-quantification.de/chemistry.html

With both these fluorochrome systems, the intensity of fluorescence will increase during the elongation step of each PCR cycle. In case of TaqMan probes, a computer algorithm compares the amount of reporter dye emission (R) with the quenching dye emission (Q) every 8.5 seconds during the PCR amplification, generating a ΔRn value (R/Q) (also called ΔRQ). The ΔRn value reflects the amount of hybridization probes that have been degraded. The algorithm fits an exponential function to the mean ΔRn values of the last three data points of every PCR extension cycle, generating an amplification plot. A relative fluorescent emission threshold is set based on the baseline of the ΔRn during the first 10-15 cycles (Heid, *et al* 1996). The algorithm calculates the

cycle at which each PCR amplification reaches a significant threshold, C_T (usually 10 times the standard deviation of the baseline). The fluorescence is measured during each cycle and its intensity is related to the amount of product synthesized until that moment. The more transcript of interest is present in the sample, the less PCR cycles will be necessary to reach a point where the emitted signal will be statistically significant in comparison with the background noise (Gibson*, et al* 1996).

4.5.3 Real time-PCR with Low Density Arrays

In order to evaluate the Y chromosome genes' expression in the samples, TaqMan low density arrays (Applied Biosystems, USA) have been used.
The low density array microfluidic card is an array of reaction vessels for the real time-PCR step. Typically, each reaction well of an array contains a pair of PCR primers and a TaqMan probe, specific for one gene of interest, that detect the amplification rate of that target. Relative levels of gene expression are determined from the fluorescence data generated during PCR using the ABI PRISM 7900HT Sequence Detection System software (Applied Biosystems, USA). The arrays used in this study have been customized with the real time-PCR primers and TaqMan probes specific for all the known Y chromosome encoded genes, pre-loaded in the arrays' spots.

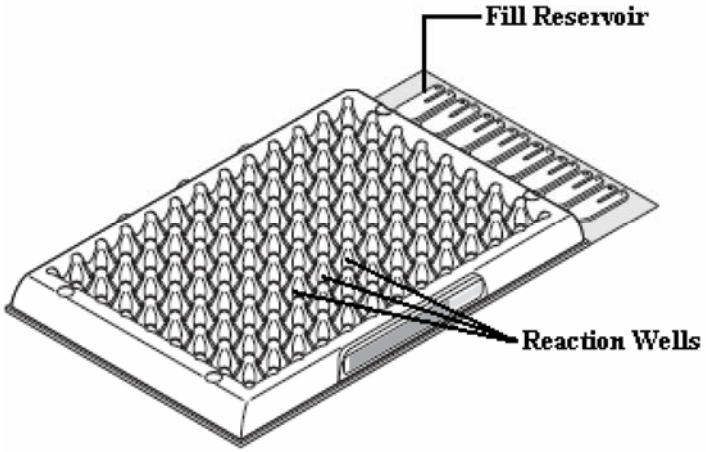

Figure 4.2. TaqMan Low Density Arrays. The sample is loaded into the fill reservoirs. Into each reaction well a lyophilized pair of primers and a TaqMan probe, specific for the gene of interest, is pre-loaded.
This tool allows screening a sample for a maximum of 384 genes. In this study 47 genes of interest (all encoded on the Y chromosome) and one endogenous control were chosen.

Low density array design
The annotated genes encoded on the Y chromosome are forty-six. The low density array cards used in this study were customized as mentioned before with TaqMan probes specific for those genes (figure 4.3). Two different endogenous controls: glyceraldehyde-3-phosphate dehydrogenase (GAPDH; NCBI accession number: NM_002046) and hypoxanthine phosphoribosyltransferase 1 (HPRT1; NM_000194)

(Liu, *et al* 2005) were added to the cards. The choice of the endogenous control is a critical step in a gene expression comparison experiment (Vandesompele, *et al* 2002). GAPDH expression was too variable in the examined samples and its relatively high expression level in comparison to HPRT1 made it unsuitable as reference gene. HPRT1 has been chosen here for evaluation of expression levels of genes of interest. Its homogeny in expression between patient and healthy samples and the fact that it is normally low expressed made it the best choice for this study.
The other forty-six TaqMan probes for each lane of the arrays where chosen among those available in the Applied Biosystem's database. Unfortunately some Y chromosome encoded genes share high homology -in some cases more than 90%- with the respective homologs on the X chromosome. For this reason, some of the available probes were not able to discriminate between X and Y alleles.

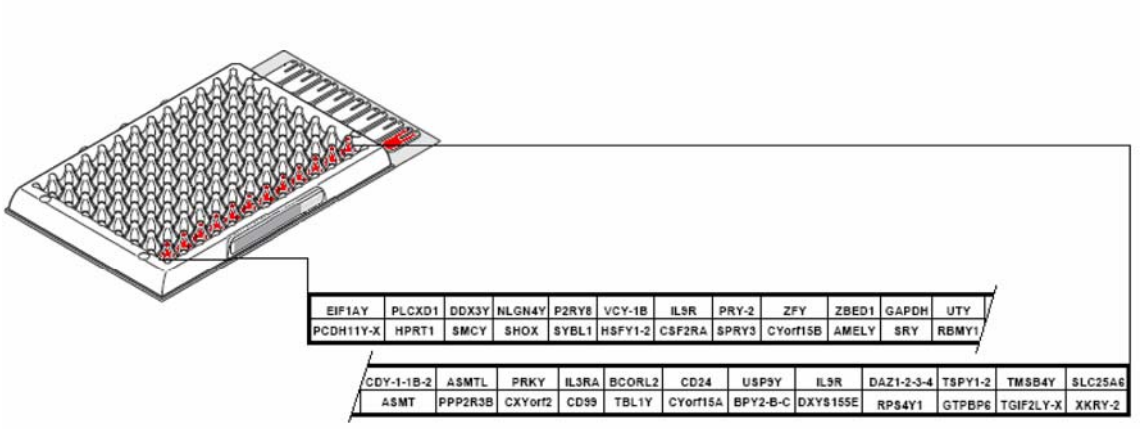

Figure 4.3. Customized TaqMan Low Density Array. In each lane (highlighted in red), corresponding to one fill reservoir, the TaqMan probes specific for all the known Y chromosome encoded genes were pre-loaded. In the box are reported the symbols relative to those genes.

Experimental procedure

100 ng of cDNA from each sample were adjusted to a volume of 50 µL with H_2O_{bidest} and mixed with 50 µL of TaqMan Universal PCR Master Mix (Applied Biosystems, USA). The mix was loaded into one (of eight) fill reservoir of an array. The arrays were then centrifuged twice, for one minute at 750×g each time. After centrifugation, the arrays were sealed in order to isolate each reaction well. The fill reservoirs were then removed from the arrays using scissors and the arrays were finally loaded into the ABI PRISM 7900HT (Applied Biosystems, USA) a thermocycler equipped with a fluorescence scanner. Thermal cycling conditions were the following: 2 minutes at 50°C, 10 minutes at 94.5°C followed from 40 cycles of denaturation at 97°C for 30 seconds and annealing and extension at 59.7°C for 1 minute.

Data evaluation
The real time-PCR data collected with the ABI PRISM 7900HT Sequence Detection System software (Applied Biosystems, USA), as described before, were processed using the $2^{-\Delta\Delta C_T}$ method (see paragraph 4.9.1).

4.5.4 DNA sequencing

Identification or verification of DNA fragments or plasmids was achieved by using the ABI PRISM 3730 DNA Analyzer (Applied Biosystems). The sequencing-PCR reaction was prepared by using the BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems), according to the manufacturer's protocol. The probes were purified by using the Dye Ex 2.0 Spin Kit (Qiagen, USA), in case of a small number of probes, or the Montage^{TM} Seq_{96} Sequencing Reaction Cleanup Kit (Millipore), in case of a larger number of probes. Both kits were used according to the manufacturers' protocol.

4.6 Protein chemistry

4.6.1 Production of recombinant protein in *E. coli*

The full-length coding sequence of the genes of interest was cloned into the expression vector pGEX 2T PL2 (Leenders*, et al* 1996); this vector encodes for the protein glutathione-S-transferase (GST), which is fused to the N-terminal of the protein of interest. Constructs were transformed into *E. coli* BL21 DE3 codon plus RP (Stratagene; see paragraph 4.1.6 for transformation procedure) and stored as glycerol-stocks (see paragraph 4.1.4) after verification of correct cloning by sequencing.
For production of recombinant proteins, a 5 mL pre-culture was grown overnight at 37°C in LB-amp culture medium (see paragraph 4.1.1 for LB medium recipe). 500 µL of this culture were transferred to 50 mL fresh LB-amp medium and bacteria were grown at 37°C for another 3-4 hours until OD_{600} reached 0.6-0.8. The culture was then split evenly and IPTG was added to one aliquot for induction of protein expression at a final concentration of 1 mM. Cultures were grown for another 3-4 hours. Bacteria were harvested by centrifugation at 75000×g and 4°C for 15 minutes and resuspended in lysis buffer. Cell lysis was performed by five cycles of 10 seconds of sonication followed by 10 seconds of chilling on ice. The genomic DNA was digested by addition of 1 U

endonuclease (Benzonase from Sigma) and $MgCl_2$ (5 mM final concentration). Samples were centrifuged to separate soluble and insoluble proteins. The supernatant was transferred to a fresh tube and the pellet fraction resuspended in an equal amount of lysis buffer.

Lysis buffer:	PBS 0.1 mg/mL lysozyme protease inhibitor mix
PBS:	10 mM sodium-phosphate buffer (pH 7.4) 150 mM NaCl
Protease inhibitor mix (1000×):	20 mg/mL Antipain 0.2 mg/mL Aprotinin 0.2 mg/mL Leupeptin Or purchased as ready-to-use tablets from Roche

4.6.2 GST-fusion protein purification

In order to purify TGIF2LY-GST fusion protein from the soluble protein fraction of expression-induced bacteria, the cell lysate was added to glutathione (GT)-sepharose 4B beads (Amersham Biosciences, USA) in a ratio of 1:0.002 (200 µL beads for 100 mL of bacteria culture). The GT-sepharose beads were previously washed with 10 volumes of ice-cold PBS and centrifuged for five minutes (100×g, 4°C) three times in total.
The mixture of protein extract and GT-sepharose beads was incubated overnight at 4°C on a nutator. This mixture was then centrifuged for five minutes at 100×g and 4°C and washed three times with ice-cold PBS as described before.
The elution of the GST-fusion protein from the GT-sepharose beads was achieved by incubating the beads with glutathione-elution buffer in a buffer-beads ratio of 1:1. Then, beads were incubated overnight at 4°C on a nutator. After centrifugation (10 minutes at 100×g and 4°C) the supernatant containing the eluted GST-fusion protein was collected. The elution procedure was then repeated twice after 2 hours-incubations, each time collecting the supernatant.
The so-eluted protein was injected into a Slide-A-Lyzer Dialysis Cassette 0.5-3 mL (Thermo Scientific, USA), in order to remove glutathione contamination. Dialysis was performed in 500 mL of protein-DNA binding buffer (see paragraph 4.6.6) overnight at 4°C, according to manufacturer's protocol.

Glutathione-elution buffer: 50 mM Tris (pH 8,0)
10 mM glutathione

4.6.3 Protein isolation from eukaryotic cells

In order to isolate proteins from cultured cells or from other eukaryotic cell samples, the NE-PER® Nuclear and Cytoplasmic Extraction Reagents Kit (Pierce, USA) was used according to the manufacturer's protocol.

Protein quantification
Pure protein concentration as well as concentration of cellular protein extracts was measured using the Protein Assay kit (Bio-Rad, USA) according to the manufacturer's protocol.

4.6.4 SDS-Polyacrylamide Gel Electrophoresis (SDS-PAGE)

Proteins of the fractions from the isolation procedure were electrophoretically separated under denaturing conditions by SDS-PAGE (gel size: 8.6 × 6.8 cm circa). Separation was visualized by Coomassie staining. 10 µl of each fraction was added to 10 µL 5× Laemmli loading buffer, heated for 5 minutes at 95°C and loaded on a Tricin-Gel (10 % resolving gel, 4 % stacking gel). Proteins were separated in a constant electric field. For detection of the proteins the resolving gels were incubated in Coomassie staining solution for 30 minutes. Destaining was obtained by washing with 10 % acetic acid at room temperature until background staining was removed. For documentation, gels were equilibrated in H_2O and dried between two sheets of cellophane.

Acrylamide solution (AA): acrylamide/bisacrylamide, 30 % (30:0.8)

Anode buffer: 0.2 M Tris-HCl, pH 8.9

Cathode buffer: 0.1 M Tris
0.1 M Tricin
0.1 % SDS

Coomassie staining solution: 200 mL methanol
5 mL acetic acid
295 mL H_2O
500 mg Coomassie blue G250 (filtrate)

Gel buffer:	3 M Tris-HCl, pH 8.45 0.3 % SDS
5× Laemmli buffer:	50 % glycerol 4 % SDS 0.1 % Coomassie blue G250 (filtrate) 0.2 M Tris-HCl, pH 6.8
Resolving gel (10 %):	3.3 mL AA 3 mL Gel buffer 1 mL H_2O 2.5 mL glycerol (50 %) 20 µL TEMED 50 µL APS
Stacking gel (4 %):	0.67 mL AA 0.67 mL Gel buffer 3.67 mL H2O 10 µL TEMED 40 µL APS

4.6.5 Western blot

For visualization of the fusion proteins, monoclonal antibodies against their specific tags were used. Therefore, the separated protein fractions from the SDS gels were transferred onto a PVDF-membrane Hybond-P (Amersham Biosciences, USA) by semi-dry blotting. The gel was equilibrated in blotting buffer for 10 minutes. The membrane was pre-wet with methanol, then immersed in blotting buffer and filter paper was soaked in blotting buffer. The blot was set up in the sequence "anode – filter paper – membrane – gel – filter paper– cathode" (figure 4.4) and the proteins were transferred onto PVDF-membrane in 30 minutes at 20 V in the Transblot (Bio-Rad, USA). After blotting, the membrane was blocked in PBS with 5% milk powder at RT for 30 minutes, washed three times for 5 minutes with PBS and incubated overnight at 4°C with the primary antibody diluted 1:4000 in PBS with 0.5 % milk powder. The membrane was washed three times for 5 minutes in PBS and incubated for 2 hours at room temperature with the peroxidase-conjugated secondary antibody diluted 1:4000 in PBS with 0.5 % milk powder. After washing the membrane three times for 5 minutes in PBS, the Western Lightning® Chemiluminescence Reagent Plus kit (PerkinElmer, USA) was used according to the manufacturer's protocol, in order to visualize peroxidase-

conjugated antibody-marked proteins (ECL staining) on a developed film (Kodak). Alternatively, or subsequently to ECL staining, 25 mL freshly prepared developing solution were added (DAB staining). Colour development was stopped by removal of developing solution and rinsing the membrane with H_2O.

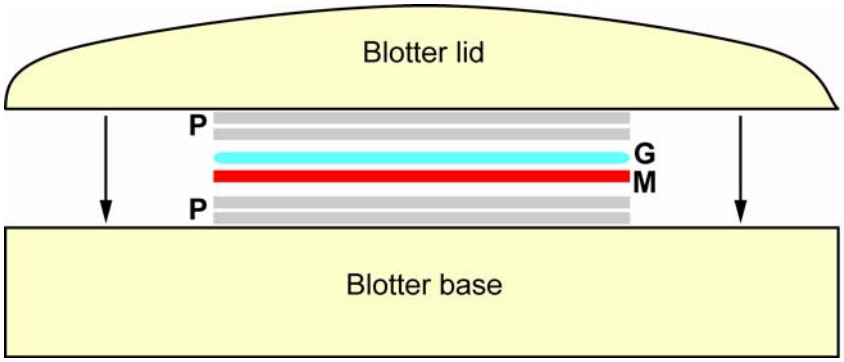

Figure 4.4. Assembled blotter. Two double layers of filter paper (P) soaked in blot buffer are used to separate the acrylamide gel (G) and the membrane (M) from the blotter elements. The two arrows indicate the direction in which proteins are migrating once voltage is applied.

5× Developing buffer (pH 7.5):	0.225 g NaH_2PO_4
	1.59 g Na_2HPO_4
	11 g NaCl
	8.5 g Imidazol
	625 µL Tween-20
	ad 250 mL with H_2O
Blotting buffer:	48 mM Tris
	33 mM Tricin
	1.3 mM SDS (10 %)
	20 % Methanol
Developing solution:	1× Developing Buffer
	100 µL CoCl2 (10 mg/ml)
	1 mg Diaminobenzidine (DAB)
	10 µL H_2O_2
	ad 25 mL with H_2O
PBS:	10 mM Sodium-phosphate buffer (pH 7.4)
	150 mM NaCl

4.6.6 Protein-DNA interaction assay

In order to investigate the capacity of the putative transcription factor TGIF2LY to bind specific DNA sequences, a protein-DNA binding assay needed to be established.
The developed assay is based on the *in vitro* genomic selection method (described in Shostak*, et al* 2004). The procedure was modified and it is schematically described in figure 2.13.
Two micrograms of purified recombinant GST-tagged TGIF2LY protein were incubated with 5 µg of THP-1 cells genomic DNA, previously digested overnight at 37°C with 8 units of BfuCI restriction endonuclease (New England Biolabs, USA), in 150 µL of protein-DNA binding buffer for 45 minutes at room temperature on a nutator. Protein–DNA complexes were then incubated with 110 µL of Anti-GST MicroBeads (Miltenyi Biotec, Germany) for 30 minutes on ice. This mix was then loaded onto a MACS µColumn, previously rinsed with 200 µL of protein-DNA binding buffer. The flow through was discarded. The column was washed with washing buffer 1 and afterwards with washing buffer 2. Washes were performed at 4°C incubating the column for 5 minutes with each washing buffer. DNA was then eluted from the column by adding 25 µL of DNA elution buffer and by incubating for 10 minutes at room temperature; after application of 100 µL of the same buffer to the columns, the flow through was collected. A second elution step was performed with an incubation time of 15 minutes. The so-collected DNA was precipitated (see paragraph 4.3.1) and resuspended in 15 µL of H_2O_{bidest}. Then, double-strand DNA BfuCI-linkers (see appendix A.1) were added to the isolated genomic DNA fragments. 14 µL of DNA were incubated for 2 hours at room temperature with 3 µL of hybridised linkers, 1 µL of T4 DNA ligase and 2 µL of T4 DNA ligase 10× buffer (final volume 20 µL). 5 µL of the ligation mix were used to perform a PCR using primers designed to bind to the linkers.

PCR mix:	10× Taq-buffer	5 µL
	10× dNTPs mix	5 µL
	10 µM primer mix	2 µL
	Template DNA	5 µL
	Taq polymerase	1.5 µL
	H_2O_{bidest}	31.5 µL
	Final volume	50 µL

Thermal profile:	95°C	5 minutes	
	95°C	1 minute	30 cycles
	57°C	45 seconds	
	72°C	3 minutes	
	72°C	7 minutes	

After PCR, the DNA was purified (see paragraph 4.3.1) and 2 µg of it were used for a second binding round with the purified protein. Basically, it was performed as the first one, but for both washing steps washing buffer 1 was used.

After the second binding round, eluted DNA was directly cloned into the vectors pCR-XL-TOPO or pCRII-TOPO (both from Invitrogen, USA). DNA fragments were sequenced in order to have a first hint about genes regulated by TGIF2LY.

Protein-DNA binding buffer:	20 mM Hepes (pH: 7.9) 8% Glycerol 10 mM $MgCl_2$ 100 mM KCl 10 µM $Zn(CH_3COO)_2$
Washing buffer 1:	20 mM Hepes (pH: 7.9) 8% Glycerol 10 mM $MgCl_2$ 250 mM KCl 10 µM $Zn(CH_3COO)_2$
Washing buffer 2:	20 mM Hepes (pH: 7.9) 8% Glycerol 10 mM $MgCl_2$ 500 mM KCl 10 µM $Zn(CH_3COO)_2$
DNA elution buffer:	20 mM Hepes (pH: 7.9) 8% Glycerol 10 mM $MgCl_2$ 1 M KCl 10 µM $Zn(CH_3COO)_2$

Sequence analysis

The sequences of the so-isolated DNA fragments were blasted against the human genomic database (NCBI; http://blast.ncbi.nlm.nih.gov/Blast.cgi) and analysed using the AlignX software (see paragraph 4.9.2).

4.6.7 Electrophoretic mobility shift assay (EMSA)

Electrophoretic mobility shift assays (EMSA) have been used to investigate the binding potential of TGIF2LY to motifs of DNA. 5'-Cy5-labelled oligonucleotides were used as TGIF2LY interaction partners.
Complement oligonucleotides (concentration: 100 µM) were mixed together in a 1:1 ratio in 1× TE buffer and incubated for 10 minutes at 95°C and, subsequently, for 15 minutes at their melting temperature decreased of three centigrades (T_m-3°C). Double-strand oligonucleotides were left at RT for at least two hours to chill and then diluted to a final concentration of 20 ng/µL in 1× TE buffer.
The subsequent mixtures were prepared for each of the used double-strand oligonucleotides:

	Oligos alone	Oligos plus protein
H_2O_{bidest}	42 µL	37 µL
10× binding buffer	5 µL	5 µL
500 mM $MgCl_2$	1 µL	1 µL
dI-dC (1 mg/mL)	1 µL	1 µL
Purified GST-TGIF2LY (0.5 µg/µL)	-	5 µL
Oligos (20 ng/µL)	1 µL	1 µL

The so-prepared mixtures were incubated at RT for 30 minutes in the dark; binding reaction was stopped by adding 5 µL of 6× loading buffer. A 5% polyacrylamide gel was prepared according to the following protocol:

H_2O_{bidest}	3.49 mL
30% Acrylamide/bisacrylamide 30:1	1.16 mL
3× Gel buffer	2.3 mL
TEMED	12.5 µL
APS	30 µL

Samples were loaded onto the gel, which was run for around 35 minutes at 170 V.
Fluorescence on the gel was then read with a Typhoon scanner (Amersham Biosciences, USA).
Cy5-labelled (5'-3' strand) and non-labelled (3'-5' strand) single-strand oligonucleotides were purchased lyophilized from Metabion (Germany) and

resuspended in H_2O_{bidest} to a final concentration of 100 µM. See also appendix A.1 for oligonucleotide sequences.

10× TE buffer:	100 mM Tris/HCl pH 7.5 10 mM EDTA
10× Running buffer:	250 mM Tris 1.92 M Glycine
6× Loading buffer:	40% Sucrose (w/v) Orange G dye
3× Gel buffer:	1.5 M Tris/HCl pH 8.85
10× Running buffer:	250 mM Tris 1.92 M Glycine

4.7 Immunological methods

4.7.1 HLA-A0201-restricted peptides

In order to investigate a possible direct involvement of those genes overexpressed in the AML samples in the improved therapy response, the possible antigenicity of candidate genes-derived peptides was analysed.
Three different publicly available software were used in order to predict HLA-A0201-restricted peptides, able to generate an immune response and in turn, to make leukaemic cells, which presents these peptides on their surface, recognizable by the immune system.
Two of these software, the “HLA Peptide Binding Predictions” (Bimas) and the “Syfpeithi”, rank nonameric peptides which potentially bind to HLA molecules.
The potential natural processing by the proteasomal cleavage of the proteins encoded from the candidate genes was checked using the “Prediction Algorithm for Proteasomal Cleavages” software (Paproc). See paragraph 4.9.4 for more detailed information about these software.

All predicted peptides were synthesized and purified by HPLC by Peptide Speciality Laboratories (Germany), dissolved in DMSO up to a final concentration of 10 mg/mL and stored at -20°C until use.

4.7.2 Peptides–HLA-A0201 binding assay using T2 cells

For this assay (Nijman*, et al* 1993), a T2 cell line was used (Elvin*, et al* 1993). Because of a deletion in the MHC region on the chromosome 6, these cells are not able to express the TAP protein and, in turn, can not form the ternary complex of MHC-α-chain, β_2-microglobulin and endogenous peptides. For this reason, only dimers of MHC-α-chain and β_2-microglobulin are transported to the cellular membrane. These dimers are not stable and, once the β_2-microglobulin dissociates, the heavy α-chain is internalised and degraded. Therefore, T2 cells express a strongly reduced number of MHC class I molecules on the cell surface.
The principle of this assay is based on the fact that a HLA-A0201-restricted peptide added to the culture medium can bind to a MHC-α-chain and β_2-microglobulin dimer and recompose a stable ternary complex. Therefore, the stronger the bond between a peptide and the MHC molecule is, the more stable ternary complexes will be formed and transported to the cell surface where, using a specific fluorescent antibody, they can be detected.
Using this method, the previously predicted peptides were analysed for their real ability to stably bind to the HLA-A 0201 molecules.
For each peptide of interest, five different conditions were prepared:

T2 cells alone
T2 cells alone
T2 + β_2-microglobulin + 10µg peptide
T2 + β_2-microglobulin + 5µg peptide
T2 + β_2-microglobulin + 2,5µg peptide

For each condition, $1{,}5\times10^5$ cells were seeded into each cavity of a 24 well-plate, in 1 mL RPMI with glutamine and antibiotics, but without FCS, because the presence of the proteins of the serum might interfere with the peptide bond or might unspecificly block the MHC molecules.
Several controls were used for each experiment: a peptide which is known to bind strongly on the HLA-A0201 molecules, the influenza virus matrix protein (IMP; peptide sequence: GILGFVFTL) as positive control and two negative controls: a peptide which is derived from scrambled sequence of human telomerase reverse transcriptase (I540S;

sequence: HFLLWKLIA) and a peptide which is restricted for the HLA-B37 (A98ID; sequence AHTKDGFNF). The method is presented schematically in figure 4.5.

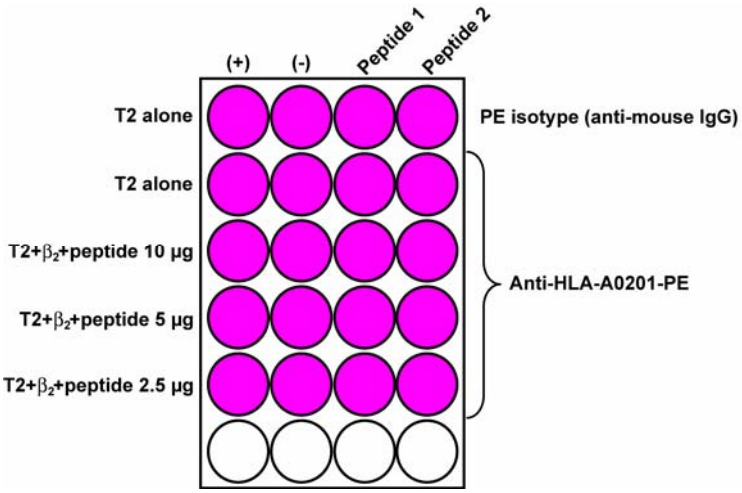

Figure 4.5. Schematic representation of peptide-HLA-A0201 binding assay preparation. The peptides, the positive (+) and negative (-) controls are plated together with T2 cells and β_2 microglobulin, at the above indicated concentrations. The plates are incubated overnight (see text for incubation condition). Fluorescent antibodies are then added and the cells analysed via flow cytometry.

10, 5, and 2.5 µg of both β_2-microglobulin and peptide were added to one aliquot of T2 cells, keeping the ratio between β_2-microglobulin and peptide always of 1:1.
The plate was incubated at 37°C with 5% CO_2 for 16-20 hours. Afterwards, the cells were transferred to flow cytometry tubes and centrifuged for 5 minutes at 2500×g. After removing the supernatant, 3 µL of anti HLA-A0201-phycoerythrin (PE) conjugated antibody was added and the cells incubated for 15 minutes at 4°C in the dark. 1 mL PBS was added to each tube, which was then centrifuged for 3 minutes at 2500×g. After removal of the supernatant, the cells were resuspended in 50-70 µL of PBS and analysed through flow cytometry.
The parameter which was evaluated was the mean fluorescence intensity (MFI). For each peptide to be analysed, the ratio between the MFI of the cells incubated with that peptide and the MFI of the cells incubated without that peptide was calculated. Only peptides with a ratio ≥ 1.2 were considered able to bind to the HLA-A0201 complex.

4.7.3 Expansion of peptide-specific T-cell populations

In order to assess the antigenicity of the candidate peptides, blood from healthy female volunteers was analysed for the presence of peptide-specific T-cells, able to recognize cells presenting those peptides.
In this study, irradiated T2 cells were used as antigen presenting cells; they were incubated with T-cells from healthy female donors and with two interleukins: IL-2, which stimulates growth and differentiation of T-cell during the immune response and

IL-7, which specifically stimulates the proliferation of cytotoxic and memory $CD8^+$ T-cells.

Donor T-cell preparation

50 mL of blood were collected from healthy female donors, mothers of one son, at least. Leukocytes were isolated by density gradient centrifugation, 2500×g for 20 minutes, using Biocoll 1.077 g/mL (Biochrom AG, UK), washed and resuspended in PBS. T-cells were separated from the leukocytes using the Pan T-cell Isolation Kit II, a MACS separator and LS separation columns (all from Miltenyi Biotec, Germany). This T-cell separation kit is an indirect magnetic labelling system for the isolation of untouched T-cells from human peripheral blood mononuclear cells (PBMCs). Non-T-cells, i.e. B-cells, NK cells, dendritic cells, monocytes, granulocytes, and erythroid cells, are indirectly magnetically labelled by using a cocktail of biotin-conjugated antibodies against CD14, CD16, CD19, CD36, CD56, CD123, and Glycophorin A, and Anti-Biotin MicroBeads. Isolation of highly pure T-cells is achieved by depletion of magnetically labelled cells.

Based on the manufacturer's protocol, the separation procedure was optimized. Total leukocytes were centrifuged at 900×g for 10 minutes and resuspended in 40 µL (per 10^7 cells) of ice-cold MACS buffer (BSA 0.5% in PBS). 10 µL (per 10^7 cells) of ice-cold Biotin-Antibody cocktail were added and the cells were incubated 15 minutes at 4°C. Afterwards, 30 µL (per 10^7 cells) of MACS buffer were added and the cells pipetted gently. 20 µL (per 10^7 cells) of ice-cold Anti-Biotin MicroBeads were added and the cells were incubated for 15 minutes at 4°C. Cells were then washed with 15-20 mL of MACS buffer, centrifuged at 900×g for 10 minutes and resuspended in 500 µL of MACS buffer. After this procedure, the cells were passed through a LS separation column placed into a strong magnetic field and previously rinsed with 3 mL of MACS buffer. The column was washed three times with 3 mL of MACS buffer. The entire eluate, representing the enriched T-cell populations, was collected, washed with PBS, centrifuged for 10 minutes at 900×g and resuspended in RPMI medium, supplemented with 2 mM L-glutamine, 100 U/mL penicillin, 100 µg/µL streptomycin and 5% heat-inactivated human AB serum (HS), up to a final concentration of 2×10^6 cells per mL.

T2 cell preparation

T2 cells were irradiated (100 Gy of γ-radiations) using ^{137}Cs for 30 minutes, in order to inhibit their possible interleukin-stimulated proliferation, centrifuged for 10 minutes at 900×g and, after discarding the supernatant, resuspended in RPMI medium (with L-glutamine, penicillin, streptomycin and 5% HS), up to a final concentration of 2×10^5 cells per mL. 8 ng/mL IL-7 and 10 µg/mL β_2-microglobulin were added to the cells.

Specific T-cell stimulation and expansion

On day 0, 1 mL of donor T-cells (2×10^6) and 1 mL of the previously prepared T2 cells (2×10^5) were pipetted into each well of a 24 well-plate. 10 mg/mL of the candidate peptide were added to each well. The so-prepared plate was incubated at 37°C with 5% CO_2 for 7 days (figure 4.6).

On day 7, 1 mL of medium was removed from each well. T2 cells were irradiated as previously described, washed and resuspended, at 2×10^5 cells per mL, in RPMI with 5% HS. 8 ng/mL IL-7 (Cell concepts, Umkirch, Germany), 80 U/mL IL-2 (Pan-Biotech, Aidenbach, Germany), and 10 µg/mL β_2-microglobulin were added to the cells, which were, then, added to each well of the plate. 10 mg/mL of the candidate peptide were also added to each well. The plate was incubated at 37°C with 5% CO_2 for 7 days.

On day 14, the same protocol of day 7 was repeated.

On day 21, the same procedure of day 7 and day 14 was repeated, but IL-2 only was added to the cells.

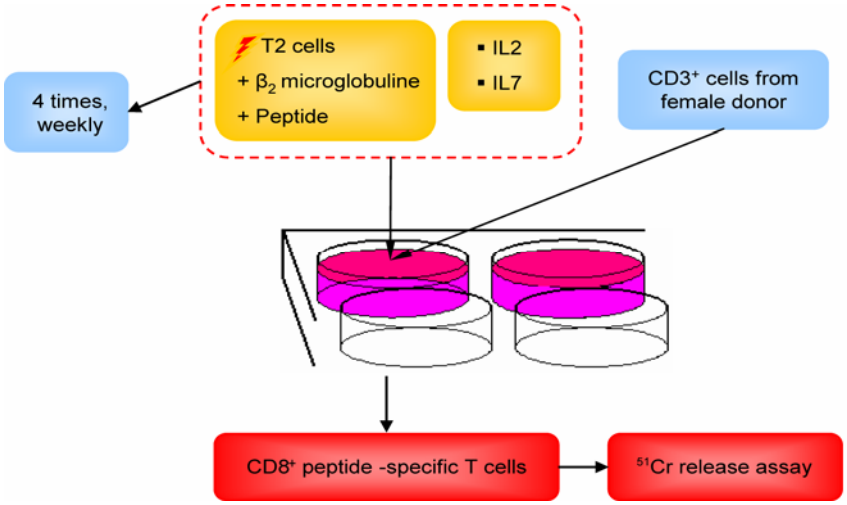

Figure 4.6. Expansion of peptide-specific $CD8^+$ cytotoxic cells. Irradiated T2 cells (target) were incubated with CD3 cells isolated from a healthy female, who gave birth to at least one son, in presence of the peptide of interest and of β_2 microglobuline, necessary to stabilize the bond between the peptide and the HLA-A0201 molecules. Fresh irradiated target cells, β microglobuline, and peptides were added weekly for a total of four times.

Chromium release assay

After the stimulations, the lytic activity of the peptide-specific T-cells was assessed in a well-established ^{51}Cr-release assay (Brunner, *et al* 1968, Hirohashi, *et al* 2002, Sato, *et al* 1986). T2 cells, the targets, were incubated overnight with the candidate peptide and β_2-microglobulin, both 10 µg/mL. The day after, the target cells were incubated with 100 µCi of ^{51}Cr in fetal bovine serum (FBS) for 1.5 hours at 37°C with 5% CO_2; washed twice with PBS, and resuspended in RPMI medium with 15% FBS. Then, ^{51}Cr-labelled target cells (2000 cells/well) were pipetted in a 96-well microtiter plate and the

previously stimulated T-cells were titrated in the plate, to get an initial ratio (effector cells:target cells) of 80:1 and a final one of 10:1. The plate was then incubated for 4 hours at 37°C. Radioactivity of the culture supernatant was measured by a γ-counter.
The cytotoxic effect of the sensitised lymphoid cells is expressed by the following formula:

$$\text{Specific lysis (\%)} = \frac{\text{Experimental release} - \text{Spontaneous release}}{\text{Maximum release} - \text{Spontaneous release}} \times 100$$

where "experimental release" means the ^{51}Chromium release in samples containing effector cells and target cells; "spontaneous release" refers to labelled target cells alone and "maximum release" to the total amount of label incorporated in the target cells. The term "specific lysis" means the lysis of target cells due to peptide-specific T-cells.

4.8 AML samples and healthy material

4.8.1 Samples characteristics

The study was conducted on AML M4 and M5 male patients' and healthy male volunteers' blood and bone marrow samples. In some cases the AML subtype was not determined at the diagnosis; in these cases the choice was limited to those samples in which the AML cells showed high expression of CD14 molecules. Informed consent was obtained from the patients before the blood was collected for experiments.
Some patient's samples were chosen between frozen samples stored in the clinic.
Table 4.1 illustrates the characteristics of the patients and the healthy volunteers.

Sample ID	Age	Diagnosis	Disease Status	Material	Cell count	$CD14^+$ cells
Patients						
P1	-	AML M5	fd	PB	9×10^7	ns
P2	-	AML M5	fd	PB	6×10^7	ns
P3	77	AML us	rel	BM	2×10^7	1.3×10^6
P4	63	AML M5	fd	BM	5×10^6	7×10^5
P5	68	CMML	fd	BM	2.7×10^7	6×10^6
P6	25	AML us	fd	PB	1×10^6	ns
P7*	47	AML M4 eo	fd	PB	2.3×10^6	1.3×10^5
P8*	32	AML M5	fd	PB	2.3×10^6	5×10^5
P9*	26	AML M5	fd	PB	9.7×10^6	4.7×10^6
P10*	57	AML us	fd	PB	1.4×10^7	2.5×10^6
P11	-	AML M4-M5	fd	PB	1.6×10^8	8.5×10^6
Volunteers						
D1	40	-	-	PB	1×10^8	6×10^6
D2	41	-	-	PB	1×10^8	6×10^6
D3	55	-	-	BM	3.1×10^7	2.8×10^6
D4	26	-	-	PB	7.1×10^8	2.3×10^7
D5	40	-	-	PB	7×10^8	2×10^7
D6	37	-	-	PB	1.7×10^8	4.5×10^7
D7	44	-	-	PB	8.1×10^7	2.1×10^7
D8	32	-	-	PB	5.3×10^7	8.6×10^6

Table 4.1. Characteristics of samples from AML patientsand healthy volunteers. PB: peripheral blood; BM: bone marrow; fd: first diagnosis; rel: relapse; us: unspecified; ns: not separated (when the $CD14^+$ cells in the blood were more than 90%). Starred samples were kindly provided from Prof. Ann Dickinson's group, from Newcastle University.

4.8.2 Cell separation

$CD14^+$ cell separation

The blood and the bone marrow samples were collected in heparinized tubes and the leukocytes isolated by density gradient centrifugation, 2500×g for 20 minutes, using Biocoll 1.077 g/mL (Biochrom AG, UK), washed and resuspended in PBS.

$CD14^+$ cells were separated from the leukocytes using CD14 MicroBeads, a MACS separator and LS separation columns (all from Miltenyi Biotec, Germany).

Based on the manufacturer's protocol, the separation procedure was optimized. Total leukocytes were centrifuged at 900×g for 10 minutes and resuspended in 80 µL (per 10^7 cells) of ice-cold MACS buffer (BSA 0.5% in PBS). 20 µL (per 10^7 cells) of ice-cold CD14 MicroBeads were added and the cells were incubated for 15 minutes at 4°C. Cells were washed with 15-20 mL of MACS buffer, centrifuged at 900×g for 10 minutes and resuspended in 500 µL of MACS buffer.

After labelling, the cells were passed twice through a LS separation column placed into a strong magnetic field and previously rinsed with 3 mL of MACS buffer. The columns were washed three times in each separation round with 3 mL of MACS buffer and finally eluted in 5 mL of the same buffer.
When possible, also the CD14 negative cell fraction of the AML patients was collected.

Separation verification
Before and after the magnetic separation, the samples were tested via flow cytometry to verify the enrichment of $CD14^+$ cell population. To do this, the cells were labelled with a combination of three fluorofor-conjugated antibodies. These antibodies were anti-CD14 fluorescein isothiocyanate (FITC)-conjugated, anti-CD33 phycoerythrin (PE)-conjugated, and anti-CD3 allophycocyanin (APC)-conjugated. The so-stained cells were subsequently analysed with a FACSCalibur (Becton Dickinson Biosciences) flow cytometer.
A small aliquot of cells was taken and incubated with 1.5 µL of each antibody for 15 minutes at 4°C.
Afterwards, 1 mL PBS was added to the cells and they were centrifuged at 2500×g for 3 minutes.
After resuspension in 100 µL PBS, the cells were analysed through flow cytometry. Figure 2.1 shows the typical result of flow cytometry of a healthy blood sample before and after $CD14^+$ cell separation (panels A and B, respectively).
The results of cytometry analysis after staining of the cells with the previously mentioned antibodies are shown in figure 2.1 (panels C and D). The cell samples which showed high purity of $CD14^+$ cells (not less than 90%) were centrifuged (10 minutes at 900×g) and stored at -80°C, as dry pellet of 3×10^6 cells, until use.

4.9 Bioinformatics

4.9.1 Gene expression evaluation

Y chromosome gene expression levels were evaluated using the ABI Prism 7900HT Sequence Detection System version (SDS) 2.2 software (Applied Biosystems). Real time-PCR and detectors parameters were set up according to the manufacturer's directions for the use of low density arrays. Threshold and baseline values were set up automatically from the software.

This software creates quantifiable relationships between test samples, AML and healthy samples in the present study, based on the number of PCR cycles elapsed before achieving detectable levels of fluorescence. Samples containing higher number of target RNA molecules cross the detection threshold at a lower cycle than samples containing lower initial template.
The SDS software uses arithmetic formulas to evaluate the relative amount of a target RNA molecule in a group of test samples, the AML patients in this study, compared with a group of so-called calibrator samples, the healthy volunteers' samples (see paragraph 4.5.3).
This expression fold-difference is given from the formula $2^{-\Delta\Delta C_T}$ (Livak and Schmittgen 2001), where the ΔC_T is the mean value of cycle difference between target RNA and endogenous control RNA in one of the samples' group. The difference of this value between the test group (AML samples) and the control group (healthy samples) represents the $\Delta\Delta C_T$.

4.9.2 DNA and protein sequence analysis

The software AlignX, which is a component of Vector NTI Advance 9.1.0 (Invitrogen, USA), was used to find and analyze similarities between several nucleotide or amino acid sequences.
This software aligns two or more sequences, starting from the first nucleotide or amino acid and ending with the last one, and highlights regions which share similarities. Moreover, the software calculates, in case of protein alignment, the amino acid substitution type at a given position, providing information about identical, conservative or non-similar substitutions (see also figure 2.9 and 2.11 for the graphical outcome of the software).
The Dot Matrix application provides the possibility to compare two sequences to find all possible nucleotide matches (figure 4.7). In the dot matrix, one sequence (A) is listed on the x-axis of the graph and the other sequence (B) is listed on the y-axis. Starting with the first positions in A and B, the program slides the window of n characters along the sequences performing a comparison of adjacent positions in the windows. If the similarity of residues in each position is above a given threshold, a dot is placed in the matrix in the position defined by the starting positions of the window for both sequences. A diagonal line segment indicates that the two sequences match consistently over an extended region.

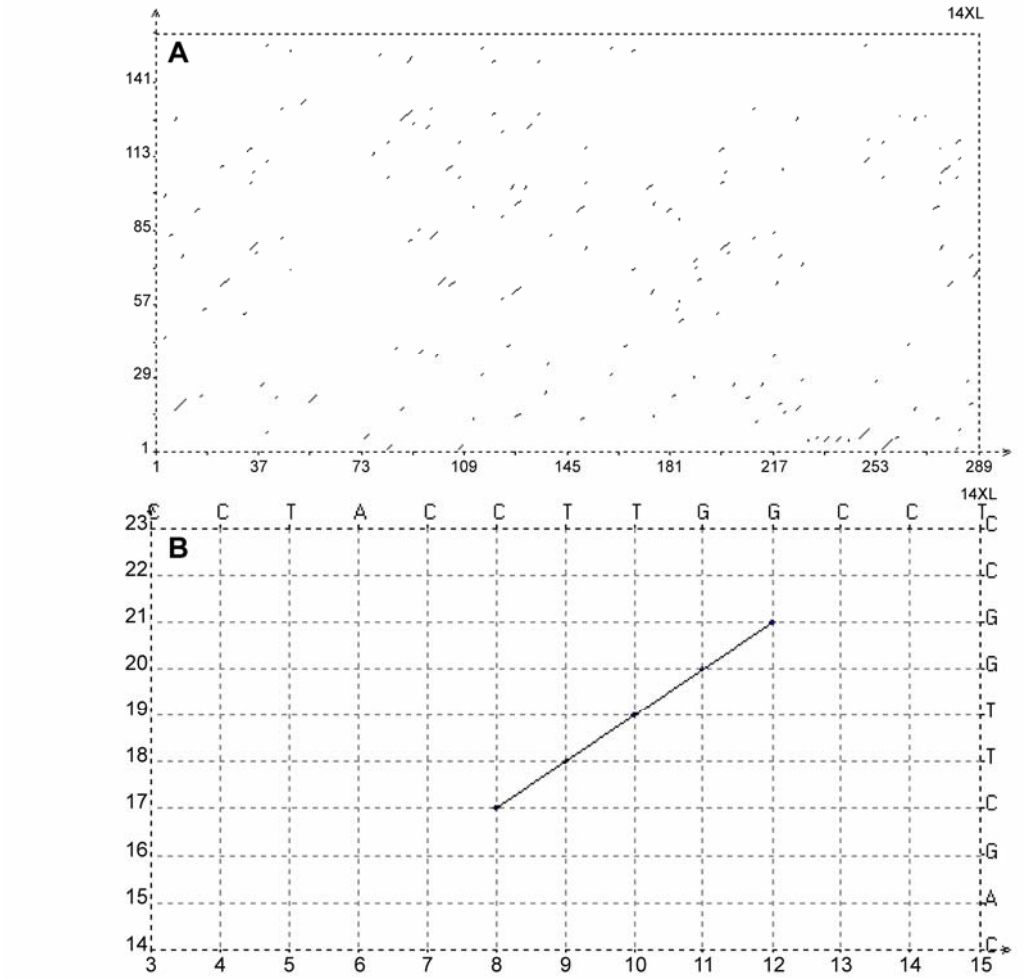

Figure 4.7. AlignX, dot matrix. on the two axis are the two DNA sequences to be compared. Diagonal lines represent a consistent match of the two sequences. Panel A: example of dot matrix in the original size. Panel B: particular.

4.9.3 3D modelling of proteins

3D structure of proteins was calculated using the online algorithm 3Dpro, a component of the publicly available software SCRATCH Protein Predictor (URL: http://scratch.proteomics.ics.uci.edu/). 3Dpro is a server that predicts protein tertiary structure. 3Dpro uses predicted amino acid features, derived from literature, to build the three dimensional structure of the submitted protein (Cheng*, et al* 2005).

4.9.4 HLA-A0201-restricted peptide prediction

Prediction of HLA-A0201-restricted peptide (see also paragraph 4.7.1) was accomplished using three different softwares: HLA Peptide Binding Predictions (Bimas), Syfpeithi, and Paproc.

HLA Peptide Binding Predictions (http://bimas.dcrt.nih.gov/molbio/hla_bind/)
This software is based on a predicted half-time of dissociation from HLA class I molecules of all the octameric or nonameric peptides contained in a given amino acid sequence. The analysis is based on coefficient tables deduced from published literature concerning HLA-binding peptides (Parker*, et al* 1994).
The scoring algorithm runs as follows: the initial running score is set to the value 1.0. For each residue position, the program examines which amino acid is appearing at that position. The running score is then multiplied by a coefficient specific for that amino acid type, at that position, for the chosen HLA molecule. Using nonamers, which are usually binding to HLA-A0201, nine multiplications are performed. The resulting running score is then multiplied by a final constant, specific for each HLA molecule, to yield an estimate of the half time of dissociation.

Syfpeithi (http://www.syfpeithi.de)
This software is also providing information about peptides which are predicted to bind HLA molecules. The algorithm evaluates a given amino acid sequence taking into account any possible nonamer, in the case of HLA-A0201 molecules, contained in it. The scoring system evaluates every amino acid within a given peptide. The arbitrary value of 1 is given to amino acids that are only slightly preferred in the respective position; the value of 15 is given to residues with predicted optimal anchoring features. Any value between these two is possible. Negative values are also possible for amino acids which are disadvantageous for the peptide's binding capacity at a certain sequence position. The allocation of values is based on the frequency of the respective amino acid in natural ligands, T-cell epitopes, or binding peptides (Rammensee*, et al* 1999).

Paproc (http://www.paproc.de)
This algorithm processes a given amino acid sequence, identifying all the possible 20S proteasomal cleavage sites within it, basing on published data concerning proteasomal cleavage. The software takes into account the possible proteasomal cleavage sites and their surrounding as well, evaluating, for each site, the probability that a cut really occurs, basing on the affinity of the surrounding of the cleavage site for the binding region of the proteasome (Kuttler*, et al* 2000).

4.9.5 Online publicly available servers and databases

Several databases have been used to retrieve general information about genes, transcripts, and proteins, as well as nucleotide and amino acid sequences. The databases with the respective URL are listed below in alphabetical order:

- Atlas of Genetics and Cytogenetics in Oncology and Haematology (http://atlasgeneticsoncology.org): provides information about involvement of genes in several types of cancer, basing on literature.
- Ensembl (http://www.ensembl.org): contains genome databases for several vertebrates and other eukaryotic species.
- Expasy (http://expasy.org): is a server containing several databases and tools for protein analysis.
- Geneatlas (http://genatlas.medecine.univ-paris5.fr): provides information especially concerning gene mapping and association between genes and genetic diseases.
- Genecards (http://www.genecards.org): provides genomic, proteomic, transcriptomic, genetic and functional information on human genes.
- NCBI (http://www.ncbi.nlm.nih.gov): provides molecular biology information about several eukaryotic and prokaryotic species. Among the tools provided from this server, there are Pubmed, literature database, and Blast, nucleotide and amino acid sequence database.

5 Bibliography

Ades, S.E. & Sauer, R.T. (1995) Specificity of minor-groove and major-groove interactions in a homeodomain-DNA complex. *Biochemistry,* **34,** 14601-14608.

Aguilella, C., Dubourg, C., Attia-Sobol, J., Vigneron, J., Blayau, M., Pasquier, L., Lazaro, L., Odent, S. & David, V. (2003) Molecular screening of the TGIF gene in holoprosencephaly: identification of two novel mutations. *Hum Genet,* **112,** 131-134.

Anasetti, C., Beatty, P.G., Storb, R., Martin, P.J., Mori, M., Sanders, J.E., Thomas, E.D. & Hansen, J.A. (1990) Effect of HLA incompatibility on graft-versus-host disease, relapse, and survival after marrow transplantation for patients with leukemia or lymphoma. *Hum Immunol,* **29,** 79-91.

Anderlini, P., Luna, M., Kantarjian, H.M., O'Brien, S., Pierce, S., Keating, M.J. & Estey, E.H. (1996) Causes of initial remission induction failure in patients with acute myeloid leukemia and myelodysplastic syndromes. *Leukemia,* **10,** 600-608.

Attisano, L. & Wrana, J.L. (2000) Smads as transcriptional co-modulators. *Curr Opin Cell Biol,* **12,** 235-243.

Austin, H., Delzell, E. & Cole, P. (1988) Benzene and leukemia. A review of the literature and a risk assessment. *Am J Epidemiol,* **127,** 419-439.

Aversa, F., Tabilio, A., Velardi, A., Cunningham, I., Terenzi, A., Falzetti, F., Ruggeri, L., Barbabietola, G., Aristei, C., Latini, P., Reisner, Y. & Martelli, M.F. (1998) Treatment of high-risk acute leukemia with T-cell-depleted stem cells from related donors with one fully mismatched HLA haplotype. *N Engl J Med,* **339,** 1186-1193.

Bafico, A., Gazit, A., Wu-Morgan, S.S., Yaniv, A. & Aaronson, S.A. (1998) Characterization of Wnt-1 and Wnt-2 induced growth alterations and signaling pathways in NIH3T3 fibroblasts. *Oncogene,* **16,** 2819-2825.

Barnes, D.W. & Loutit, J.F. (1957) Treatment of murine leukaemia with x-rays and homologous bone marrow. II. *Br J Haematol,* **3,** 241-252.

Beatty, P.G., Clift, R.A., Mickelson, E.M., Nisperos, B.B., Flournoy, N., Martin, P.J., Sanders, J.E., Stewart, P., Buckner, C.D., Storb, R. & et al. (1985) Marrow transplantation from related donors other than HLA-identical siblings. *N Engl J Med,* **313,** 765-771.

Bennett, J.M., Catovsky, D., Daniel, M.T., Flandrin, G., Galton, D.A., Gralnick, H.R. & Sultan, C. (1976) Proposals for the classification of the acute leukaemias. French-American-British (FAB) co-operative group. *Br J Haematol,* **33,** 451-458.

Bennett, J.M., Young, M.L., Andersen, J.W., Cassileth, P.A., Tallman, M.S., Paietta, E., Wiernik, P.H. & Rowe, J.M. (1997) Long-term survival in acute myeloid leukemia: the Eastern Cooperative Oncology Group experience. *Cancer,* **80,** 2205-2209.

Berges, C., Naujokat, C., Tinapp, S., Wieczorek, H., Hoh, A., Sadeghi, M., Opelz, G. & Daniel, V. (2005) A cell line model for the differentiation of human dendritic cells. *Biochem Biophys Res Commun,* **333,** 896-907.

Berman, E., Heller, G., Santorsa, J., McKenzie, S., Gee, T., Kempin, S., Gulati, S., Andreeff, M., Kolitz, J., Gabrilove, J. & et al. (1991) Results of a randomized trial comparing idarubicin and cytosine arabinoside with daunorubicin and cytosine arabinoside in adult patients with newly diagnosed acute myelogenous leukemia. *Blood,* **77,** 1666-1674.

Berman, E. & McBride, M. (1992) Comparative cellular pharmacology of daunorubicin and idarubicin in human multidrug-resistant leukemia cells. *Blood,* **79,** 3267-3273.

Bertani, G. (1951) Studies on lysogenesis. I. The mode of phage liberation by lysogenic Escherichia coli. *J Bacteriol,* **62,** 293-300.

Bertolino, E., Reimund, B., Wildt-Perinic, D. & Clerc, R.G. (1995) A novel homeobox protein which recognizes a TGT core and functionally interferes with a retinoid-responsive motif. *J Biol Chem,* **270,** 31178-31188.

Bishop, J.F., Lowenthal, R.M., Joshua, D., Matthews, J.P., Todd, D., Cobcroft, R., Whiteside, M.G., Kronenberg, H., Ma, D., Dodds, A. & et al. (1990) Etoposide in acute nonlymphocytic leukemia. Australian Leukemia Study Group. *Blood,* **75,** 27-32.

Blanco-Arias, P., Sargent, C.A. & Affara, N.A. (2002) The human-specific Yp11.2/Xq21.3 homology block encodes a potentially functional testis-specific TGIF-like retroposon. *Mamm Genome,* **13,** 463-468.

Blanco, P., Sargent, C.A., Boucher, C.A., Mitchell, M. & Affara, N.A. (2000) Conservation of PCDHX in mammals; expression of human X/Y genes predominantly in brain. *Mamm Genome,* **11,** 906-914.

Blank, J.L., Gerwins, P., Elliott, E.M., Sather, S. & Johnson, G.L. (1996) Molecular cloning of mitogen-activated protein/ERK kinase kinases (MEKK) 2 and 3. Regulation of sequential phosphorylation pathways involving mitogen-activated protein kinase and c-Jun kinase. *J Biol Chem,* **271,** 5361-5368.

Boehm, J., He, Y., Greiner, A., Staudt, L. & Wirth, T. (2001) Regulation of BOB.1/OBF.1 stability by SIAH. *Embo J,* **20,** 4153-4162.

Bonizzi, G. & Karin, M. (2004) The two NF-kappaB activation pathways and their role in innate and adaptive immunity. *Trends Immunol,* **25,** 280-288.

Bonnet, D., Warren, E.H., Greenberg, P.D., Dick, J.E. & Riddell, S.R. (1999) CD8(+) minor histocompatibility antigen-specific cytotoxic T lymphocyte clones eliminate human acute myeloid leukemia stem cells. *Proc Natl Acad Sci U S A,* **96,** 8639-8644.

Browne, G.J. & Proud, C.G. (2002) Regulation of peptide-chain elongation in mammalian cells. *Eur J Biochem,* **269,** 5360-5368.

Brunner, K.T., Mauel, J., Cerottini, J.C. & Chapuis, B. (1968) Quantitative assay of the lytic action of immune lymphoid cells on 51-Cr-labelled allogeneic target cells in vitro; inhibition by isoantibody and by drugs. *Immunology,* **14,** 181-196.

Buchner, T., Hiddemann, W., Schoch, C., Haferlach, T., Sauerland, M.C. & Heinecke, A. (2001) Acute myeloid leukaemia (AML): treatment of the older patient. *Best Pract Res Clin Haematol,* **14,** 139-151.

Buchner, T., Hiddemann, W., Wormann, B., Loffler, H., Gassmann, W., Haferlach, T., Fonatsch, C., Haase, D., Schoch, C., Hossfeld, D., Lengfelder, E., Aul, C., Heyll, A., Maschmeyer, G., Ludwig, W.D., Sauerland, M.C. & Heinecke, A.

(1999) Double induction strategy for acute myeloid leukemia: the effect of high-dose cytarabine with mitoxantrone instead of standard-dose cytarabine with daunorubicin and 6-thioguanine: a randomized trial by the German AML Cooperative Group. *Blood,* **93,** 4116-4124.

Burchenal, J.H., Oettgen, H.F., Holmberg, E.A., Hemphill, S.C. & Reppert, J.A. (1960) Effect of total-body irradiation on the transplantability of mouse leukemias. *Cancer Res,* **20,** 425-430.

Cardis, E., Gilbert, E.S., Carpenter, L., Howe, G., Kato, I., Armstrong, B.K., Beral, V., Cowper, G., Douglas, A., Fix, J. & et al. (1995) Effects of low doses and low dose rates of external ionizing radiation: cancer mortality among nuclear industry workers in three countries. *Radiat Res,* **142,** 117-132.

Cassileth, P.A., Lynch, E., Hines, J.D., Oken, M.M., Mazza, J.J., Bennett, J.M., McGlave, P.B., Edelstein, M., Harrington, D.P. & O'Connell, M.J. (1992) Varying intensity of postremission therapy in acute myeloid leukemia. *Blood,* **79,** 1924-1930.

Center, R., Lukeis, R., Vrazas, V. & Garson, O.M. (1993) Y chromosome loss and rearrangement in non-small-cell lung cancer. *Int J Cancer,* **55,** 390-393.

Chau, V., Tobias, J.W., Bachmair, A., Marriott, D., Ecker, D.J., Gonda, D.K. & Varshavsky, A. (1989) A multiubiquitin chain is confined to specific lysine in a targeted short-lived protein. *Science,* **243,** 1576-1583.

Chen, F., Ogawa, K., Nagarajan, R.P., Zhang, M., Kuang, C. & Chen, Y. (2003) Regulation of TG-interacting factor by transforming growth factor-beta. *Biochem J,* **371,** 257-263.

Chen, M.W., Vacherot, F., De La Taille, A., Gil-Diez-De-Medina, S., Shen, R., Friedman, R.A., Burchardt, M., Chopin, D.K. & Buttyan, R. (2002) The emergence of protocadherin-PC expression during the acquisition of apoptosis-resistance by prostate cancer cells. *Oncogene,* **21,** 7861-7871.

Cheng, J., Randall, A.Z., Sweredoski, M.J. & Baldi, P. (2005) SCRATCH: a protein structure and structural feature prediction server. *Nucleic Acids Res,* **33,** W72-76.

Cheson, B.D., Bennett, J.M., Kopecky, K.J., Buchner, T., Willman, C.L., Estey, E.H., Schiffer, C.A., Doehner, H., Tallman, M.S., Lister, T.A., Lo-Coco, F., Willemze, R., Biondi, A., Hiddemann, W., Larson, R.A., Lowenberg, B., Sanz, M.A., Head, D.R., Ohno, R. & Bloomfield, C.D. (2003) Revised recommendations of the International Working Group for Diagnosis, Standardization of Response Criteria, Treatment Outcomes, and Reporting Standards for Therapeutic Trials in Acute Myeloid Leukemia. *J Clin Oncol,* **21,** 4642-4649.

Cho, J.Y., Akbarali, Y., Zerbini, L.F., Gu, X., Boltax, J., Wang, Y., Oettgen, P., Zhang, D.E. & Libermann, T.A. (2004) Isoforms of the Ets transcription factor NERF/ELF-2 physically interact with AML1 and mediate opposing effects on AML1-mediated transcription of the B cell-specific blk gene. *J Biol Chem,* **279,** 19512-19522.

Chu, Z.L., Pio, F., Xie, Z., Welsh, K., Krajewska, M., Krajewski, S., Godzik, A. & Reed, J.C. (2001) A novel enhancer of the Apaf1 apoptosome involved in cytochrome c-dependent caspase activation and apoptosis. *J Biol Chem,* **276,** 9239-9245.

Court, E.L., Smith, M.A., Avent, N.D., Hancock, J.T., Morgan, L.M., Gray, A.G. & Smith, J.G. (2004) DNA microarray screening of differential gene expression in

bone marrow samples from AML, non-AML patients and AML cell lines. *Leuk Res,* **28,** 743-753.

Crane, M.M., Strom, S.S., Halabi, S., Berman, E.L., Fueger, J.J., Spitz, M.R. & Keating, M.J. (1996) Correlation between selected environmental exposures and karyotype in acute myelocytic leukemia. *Cancer Epidemiol Biomarkers Prev,* **5,** 639-644.

Crowley, K.S. & Payne, R.M. (1998) Ribosome binding to mitochondria is regulated by GTP and the transit peptide. *J Biol Chem,* **273,** 17278-17285.

Cryer, D.R., Eccleshall, R. & Marmur, J. (1975) Isolation of yeast DNA. *Methods Cell Biol,* **12,** 39-44.

Dahlback, B. & Stenflo, J. (1980) Inhibitory effect of activated protein C on activation of prothrombin by platelet-bound factor Xa. *Eur J Biochem,* **107,** 331-335.

Dahle, O., Andersen, T.O., Nordgard, O., Matre, V., Del Sal, G. & Gabrielsen, O.S. (2003) Transactivation properties of c-Myb are critically dependent on two SUMO-1 acceptor sites that are conjugated in a PIASy enhanced manner. *Eur J Biochem,* **270,** 1338-1348.

Damante, G., Pellizzari, L., Esposito, G., Fogolari, F., Viglino, P., Fabbro, D., Tell, G., Formisano, S. & Di Lauro, R. (1996) A molecular code dictates sequence-specific DNA recognition by homeodomains. *Embo J,* **15,** 4992-5000.

Dasari, V.K., Goharderakhshan, R.Z., Perinchery, G., Li, L.C., Tanaka, Y., Alonzo, J. & Dahiya, R. (2001) Expression analysis of Y chromosome genes in human prostate cancer. *J Urol,* **165,** 1335-1341.

Davies, S.M., Ruggieri, L., DeFor, T., Wagner, J.E., Weisdorf, D.J., Miller, J.S., Velardi, A. & Blazar, B.R. (2002) Evaluation of KIR ligand incompatibility in mismatched unrelated donor hematopoietic transplants. Killer immunoglobulin-like receptor. *Blood,* **100,** 3825-3827.

de La Coste, A., Romagnolo, B., Billuart, P., Renard, C.A., Buendia, M.A., Soubrane, O., Fabre, M., Chelly, J., Beldjord, C., Kahn, A. & Perret, C. (1998) Somatic mutations of the beta-catenin gene are frequent in mouse and human hepatocellular carcinomas. *Proc Natl Acad Sci U S A,* **95,** 8847-8851.

Dickinson, A.M., Wang, X.N., Sviland, L., Vyth-Dreese, F.A., Jackson, G.H., Schumacher, T.N., Haanen, J.B., Mutis, T. & Goulmy, E. (2002) In situ dissection of the graft-versus-host activities of cytotoxic T cells specific for minor histocompatibility antigens. *Nat Med,* **8,** 410-414.

Dillman, R.O., Davis, R.B., Green, M.R., Weiss, R.B., Gottlieb, A.J., Caplan, S., Kopel, S., Preisler, H., McIntyre, O.R. & Schiffer, C. (1991) A comparative study of two different doses of cytarabine for acute myeloid leukemia: a phase III trial of Cancer and Leukemia Group B. *Blood,* **78,** 2520-2526.

Duncan, C. & Roddie, H. (2008) Dendritic cell vaccines in acute leukaemia. *Best Pract Res Clin Haematol,* **21,** 521-541.

El-Jaick, K.B., Powers, S.E., Bartholin, L., Myers, K.R., Hahn, J., Orioli, I.M., Ouspenskaia, M., Lacbawan, F., Roessler, E., Wotton, D. & Muenke, M. (2007) Functional analysis of mutations in TGIF associated with holoprosencephaly. *Mol Genet Metab,* **90,** 97-111.

Elvin, J., Potter, C., Elliott, T., Cerundolo, V. & Townsend, A. (1993) A method to quantify binding of unlabeled peptides to class I MHC molecules and detect their allele specificity. *J Immunol Methods,* **158,** 161-171.

Emanuelsson, O., Nielsen, H., Brunak, S. & von Heijne, G. (2000) Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *J Mol Biol,* **300,** 1005-1016.

Estey, E., deLima, M., Strom, S., Pierce, S., Freireich, E.J. & Keating, M.J. (1997) Long-term follow-up of patients with newly diagnosed acute myeloid leukemia treated at the University of Texas M.D. Anderson Cancer Center. *Cancer,* **80,** 2176-2180.

Estey, E.H., Shen, Y. & Thall, P.F. (2000) Effect of time to complete remission on subsequent survival and disease-free survival time in AML, RAEB-t, and RAEB. *Blood,* **95,** 72-77.

Estey, E.H., Thall, P.F., Cortes, J.E., Giles, F.J., O'Brien, S., Pierce, S.A., Wang, X., Kantarjian, H.M. & Beran, M. (2001) Comparison of idarubicin + ara-C-, fludarabine + ara-C-, and topotecan + ara-C-based regimens in treatment of newly diagnosed acute myeloid leukemia, refractory anemia with excess blasts in transformation, or refractory anemia with excess blasts. *Blood,* **98,** 3575-3583.

Evans, D.I. & Steward, J.K. (1972) Down's syndrome and leukaemia. *Lancet,* **2,** 1322.

Faber, L.M., van der Hoeven, J., Goulmy, E., Hooftman-den Otter, A.L., van Luxemburg-Heijs, S.A., Willemze, R. & Falkenburg, J.H. (1995) Recognition of clonogenic leukemic cells, remission bone marrow and HLA-identical donor bone marrow by CD8+ or CD4+ minor histocompatibility antigen-specific cytotoxic T lymphocytes. *J Clin Invest,* **96,** 877-883.

Falkenburg, J.H., van de Corput, L., Marijt, E.W. & Willemze, R. (2003) Minor histocompatibility antigens in human stem cell transplantation. *Exp Hematol,* **31,** 743-751.

Faustin, B., Lartigue, L., Bruey, J.M., Luciano, F., Sergienko, E., Bailly-Maitre, B., Volkmann, N., Hanein, D., Rouiller, I. & Reed, J.C. (2007) Reconstituted NALP1 inflammasome reveals two-step mechanism of caspase-1 activation. *Mol Cell,* **25,** 713-724.

Fedi, P., Bafico, A., Nieto Soria, A., Burgess, W.H., Miki, T., Bottaro, D.P., Kraus, M.H. & Aaronson, S.A. (1999) Isolation and biochemical characterization of the human Dkk-1 homologue, a novel inhibitor of mammalian Wnt signaling. *J Biol Chem,* **274,** 19465-19472.

Fenaux, P., Chastang, C., Chevret, S., Sanz, M., Dombret, H., Archimbaud, E., Fey, M., Rayon, C., Huguet, F., Sotto, J.J., Gardin, C., Makhoul, P.C., Travade, P., Solary, E., Fegueux, N., Bordessoule, D., Miguel, J.S., Link, H., Desablens, B., Stamatoullas, A., Deconinck, E., Maloisel, F., Castaigne, S., Preudhomme, C. & Degos, L. (1999) A randomized comparison of all transretinoic acid (ATRA) followed by chemotherapy and ATRA plus chemotherapy and the role of maintenance therapy in newly diagnosed acute promyelocytic leukemia. The European APL Group. *Blood,* **94,** 1192-1200.

Ferrant, A., Labopin, M., Frassoni, F., Prentice, H.G., Cahn, J.Y., Blaise, D., Reiffers, J., Visani, G., Sanz, M.A., Boogaerts, M.A., Lowenberg, B. & Gorin, N.C. (1997) Karyotype in acute myeloblastic leukemia: prognostic significance for bone marrow transplantation in first remission: a European Group for Blood and Marrow Transplantation study. Acute Leukemia Working Party of the European Group for Blood and Marrow Transplantation (EBMT). *Blood,* **90,** 2931-2938.

Fontaine, P., Roy-Proulx, G., Knafo, L., Baron, C., Roy, D.C. & Perreault, C. (2001) Adoptive transfer of minor histocompatibility antigen-specific T lymphocytes

eradicates leukemia cells without causing graft-versus-host disease. *Nat Med,* **7,** 789-794.

Frohling, S., Scholl, C., Gilliland, D.G. & Levine, R.L. (2005) Genetics of myeloid malignancies: pathogenetic and clinical implications. *J Clin Oncol,* **23,** 6285-6295.

Gahrton, G. (2007) Risk assessment in haematopoietic stem cell transplantation: impact of donor-recipient sex combination in allogeneic transplantation. *Best Pract Res Clin Haematol,* **20,** 219-229.

Garred, P., Madsen, H.O., Hofmann, B. & Svejgaard, A. (1995) Increased frequency of homozygosity of abnormal mannan-binding-protein alleles in patients with suspected immunodeficiency. *Lancet,* **346,** 941-943.

Gavert, N., Ben-Shmuel, A., Raveh, S. & Ben-Ze'ev, A. (2008) L1-CAM in cancerous tissues. *Expert Opin Biol Ther,* **8,** 1749-1757.

Gibson, U.E., Heid, C.A. & Williams, P.M. (1996) A novel method for real time quantitative RT-PCR. *Genome Res,* **6,** 995-1001.

Godley, L.A. & Larson, R.A. (2008) Therapy-related myeloid leukemia. *Semin Oncol,* **35,** 418-429.

Greenlee, R.T., Hill-Harmon, M.B., Murray, T. & Thun, M. (2001) Cancer statistics, 2001. *CA Cancer J Clin,* **51,** 15-36.

Greiner, J., Dohner, H. & Schmitt, M. (2006) Cancer vaccines for patients with acute myeloid leukemia--definition of leukemia-associated antigens and current clinical protocols targeting these antigens. *Haematologica,* **91,** 1653-1661.

Grimwade, D., Walker, H., Harrison, G., Oliver, F., Chatters, S., Harrison, C.J., Wheatley, K., Burnett, A.K. & Goldstone, A.H. (2001) The predictive value of hierarchical cytogenetic classification in older adults with acute myeloid leukemia (AML): analysis of 1065 patients entered into the United Kingdom Medical Research Council AML11 trial. *Blood,* **98,** 1312-1320.

Grimwade, D., Walker, H., Oliver, F., Wheatley, K., Harrison, C., Harrison, G., Rees, J., Hann, I., Stevens, R., Burnett, A. & Goldstone, A. (1998) The importance of diagnostic cytogenetics on outcome in AML: analysis of 1,612 patients entered into the MRC AML 10 trial. The Medical Research Council Adult and Children's Leukaemia Working Parties. *Blood,* **92,** 2322-2333.

Grundy, C.B., Holding, S., Millar, D.S., Kakkar, V.V. & Cooper, D.N. (1992a) A novel missense mutation in the antithrombin III gene (Ser349----Pro) causing recurrent venous thrombosis. *Hum Genet,* **88,** 707-708.

Grundy, C.B., Schulman, S., Tengborn, L., Kakkar, V.V. & Cooper, D.N. (1992b) Two different missense mutations at Arg 178 of the protein C (PROC) gene causing recurrent venous thrombosis. *Hum Genet,* **89,** 685-686.

Guinn, B.A., Tobal, K. & Mills, K.I. (2007) Comparison of the survival implications of tumour-associated versus cancer-testis antigen expression in acute myeloid leukaemia. *Br J Haematol,* **136,** 510-512.

Habelhah, H., Frew, I.J., Laine, A., Janes, P.W., Relaix, F., Sassoon, D., Bowtell, D.D. & Ronai, Z. (2002) Stress-induced decrease in TRAF2 stability is mediated by Siah2. *Embo J,* **21,** 5756-5765.

Haferlach, T., Bacher, U., Kern, W., Schnittger, S. & Haferlach, C. (2007) Diagnostic pathways in acute leukemias: a proposal for a multimodal approach. *Ann Hematol,* **86,** 311-327.

Haferlach, T., Kern, W., Schnittger, S. & Schoch, C. (2005) Modern diagnostics in acute leukemias. *Crit Rev Oncol Hematol,* **56,** 223-234.

Hammaker, D.R., Boyle, D.L., Chabaud-Riou, M. & Firestein, G.S. (2004) Regulation of c-Jun N-terminal kinase by MEKK-2 and mitogen-activated protein kinase kinase kinases in rheumatoid arthritis. *J Immunol,* **172,** 1612-1618.

Heid, C.A., Stevens, J., Livak, K.J. & Williams, P.M. (1996) Real time quantitative PCR. *Genome Res,* **6,** 986-994.

Heim, S. & Mitelman, F. (1992) Cytogenetic analysis in the diagnosis of acute leukemia. *Cancer,* **70,** 1701-1709.

Hirohashi, Y., Torigoe, T., Maeda, A., Nabeta, Y., Kamiguchi, K., Sato, T., Yoda, J., Ikeda, H., Hirata, K., Yamanaka, N. & Sato, N. (2002) An HLA-A24-restricted cytotoxic T lymphocyte epitope of a tumor-associated protein, survivin. *Clin Cancer Res,* **8,** 1731-1739.

Hlaing, T., Guo, R.F., Dilley, K.A., Loussia, J.M., Morrish, T.A., Shi, M.M., Vincenz, C. & Ward, P.A. (2001) Molecular cloning and characterization of DEFCAP-L and -S, two isoforms of a novel member of the mammalian Ced-4 family of apoptosis proteins. *J Biol Chem,* **276,** 9230-9238.

Hu, G., Chung, Y.L., Glover, T., Valentine, V., Look, A.T. & Fearon, E.R. (1997a) Characterization of human homologs of the Drosophila seven in absentia (sina) gene. *Genomics,* **46,** 103-111.

Hu, G., Zhang, S., Vidal, M., Baer, J.L., Xu, T. & Fearon, E.R. (1997b) Mammalian homologs of seven in absentia regulate DCC via the ubiquitin-proteasome pathway. *Genes Dev,* **11,** 2701-2714.

Hua, S. & Sun, Z. (2001) Support vector machine approach for protein subcellular localization prediction. *Bioinformatics,* **17,** 721-728.

Huang, M.E., Ye, Y.C., Chen, S.R., Chai, J.R., Lu, J.X., Zhoa, L., Gu, L.J. & Wang, Z.Y. (1988) Use of all-trans retinoic acid in the treatment of acute promyelocytic leukemia. *Blood,* **72,** 567-572.

Imoto, I., Pimkhaokham, A., Watanabe, T., Saito-Ohara, F., Soeda, E. & Inazawa, J. (2000) Amplification and overexpression of TGIF2, a novel homeobox gene of the TALE superclass, in ovarian cancer cell lines. *Biochem Biophys Res Commun,* **276,** 264-270.

Imoto, S., Sugiyama, K., Muromoto, R., Sato, N., Yamamoto, T. & Matsuda, T. (2003) Regulation of transforming growth factor-beta signaling by protein inhibitor of activated STAT, PIASy through Smad3. *J Biol Chem,* **278,** 34253-34258.

Inohara, N. & Nunez, G. (2001) The NOD: a signaling module that regulates apoptosis and host defense against pathogens. *Oncogene,* **20,** 6473-6481.

Jemal, A., Siegel, R., Ward, E., Hao, Y., Xu, J., Murray, T. & Thun, M.J. (2008) Cancer statistics, 2008. *CA Cancer J Clin,* **58,** 71-96.

Jin, Y., Mailloux, C.M., Gowan, K., Riccardi, S.L., LaBerge, G., Bennett, D.C., Fain, P.R. & Spritz, R.A. (2007) NALP1 in vitiligo-associated multiple autoimmune disease. *N Engl J Med,* **356,** 1216-1225.

Kaiser, P., Flick, K., Wittenberg, C. & Reed, S.I. (2000) Regulation of transcription by ubiquitination without proteolysis: Cdc34/SCF(Met30)-mediated inactivation of the transcription factor Met4. *Cell,* **102,** 303-314.

Karre, K., Ljunggren, H.G., Piontek, G. & Kiessling, R. (1986) Selective rejection of H-2-deficient lymphoma variants suggests alternative immune defence strategy. *Nature,* **319,** 675-678.

Keating, M.J., Smith, T.L., Kantarjian, H., Cork, A., Walters, R., Trujillo, J.M., McCredie, K.B., Gehan, E.A. & Freireich, E.J. (1988) Cytogenetic pattern in

acute myelogenous leukemia: a major reproducible determinant of outcome. *Leukemia,* **2,** 403-412.

Kelly, L.M. & Gilliland, D.G. (2002) Genetics of myeloid leukemias. *Annu Rev Genomics Hum Genet,* **3,** 179-198.

Kern, W., Kohlmann, A., Schnittger, S., Hiddemann, W., Schoch, C. & Haferlach, T. (2004) Gene expression profiling as a diagnostic tool in acute myeloid leukemia. *Am J Pharmacogenomics,* **4,** 225-237.

Kohrt, H.E. & Coutre, S.E. (2008) Optimizing therapy for acute myeloid leukemia. *J Natl Compr Canc Netw,* **6,** 1003-1016.

Kolb, H.J. (2008) Graft-versus-leukemia effects of transplantation and donor lymphocytes. *Blood,* **112,** 4371-4383.

Kolb, H.J., Rieder, I., Rodt, H., Netzel, B., Grosse-Wilde, H., Scholz, S., Schaffer, E., Kolb, H. & Thierfelder, S. (1979) Antilymphocytic antibodies and marrow transplantation. VI. Graft-versus-host tolerance in DLA-incompatible dogs after in vitro treatment of bone marrow with absorbed antithymocyte globulin. *Transplantation,* **27,** 242-245.

Kummer, J.A., Broekhuizen, R., Everett, H., Agostini, L., Kuijk, L., Martinon, F., van Bruggen, R. & Tschopp, J. (2007) Inflammasome components NALP 1 and 3 show distinct but separate expression profiles in human tissues suggesting a site-specific role in the inflammatory response. *J Histochem Cytochem,* **55,** 443-452.

Kuras, L., Rouillon, A., Lee, T., Barbey, R., Tyers, M. & Thomas, D. (2002) Dual regulation of the met4 transcription factor by ubiquitin-dependent degradation and inhibition of promoter recruitment. *Mol Cell,* **10,** 69-80.

Kuttler, C., Nussbaum, A.K., Dick, T.P., Rammensee, H.G., Schild, H. & Hadeler, K.P. (2000) An algorithm for the prediction of proteasomal cleavages. *J Mol Biol,* **298,** 417-429.

Lahn, B.T. & Page, D.C. (2000) A human sex-chromosomal gene family expressed in male germ cells and encoding variably charged proteins. *Hum Mol Genet,* **9,** 311-319.

Lau, Y.F. & Zhang, J. (2000) Expression analysis of thirty one Y chromosome genes in human prostate cancer. *Mol Carcinog,* **27,** 308-321.

Le Beau, M.M., Albain, K.S., Larson, R.A., Vardiman, J.W., Davis, E.M., Blough, R.R., Golomb, H.M. & Rowley, J.D. (1986) Clinical and cytogenetic correlations in 63 patients with therapy-related myelodysplastic syndromes and acute nonlymphocytic leukemia: further evidence for characteristic abnormalities of chromosomes no. 5 and 7. *J Clin Oncol,* **4,** 325-345.

Lee, P.S., Chang, C., Liu, D. & Derynck, R. (2003) Sumoylation of Smad4, the common Smad mediator of transforming growth factor-beta family signaling. *J Biol Chem,* **278,** 27853-27863.

Leenders, F., Tesdorpf, J.G., Markus, M., Engel, T., Seedorf, U. & Adamski, J. (1996) Porcine 80-kDa protein reveals intrinsic 17 beta-hydroxysteroid dehydrogenase, fatty acyl-CoA-hydratase/dehydrogenase, and sterol transfer activities. *J Biol Chem,* **271,** 5438-5442.

Leith, C.P., Kopecky, K.J., Godwin, J., McConnell, T., Slovak, M.L., Chen, I.M., Head, D.R., Appelbaum, F.R. & Willman, C.L. (1997) Acute myeloid leukemia in the elderly: assessment of multidrug resistance (MDR1) and cytogenetics distinguishes biologic subgroups with remarkably distinct responses to standard chemotherapy. A Southwest Oncology Group study. *Blood,* **89,** 3323-3329.

Levy-Nissenbaum, O., Sagi-Assif, O., Raanani, P., Avigdor, A., Ben-Bassat, I. & Witz, I.P. (2003) cDNA microarray analysis reveals an overexpression of the dual-specificity MAPK phosphatase PYST2 in acute leukemia. *Methods Enzymol,* **366,** 103-113.

Lin, T.P., Guzman, R.C., Osborn, R.C., Thordarson, G. & Nandi, S. (1992) Role of endocrine, autocrine, and paracrine interactions in the development of mammary hyperplasia in Wnt-1 transgenic mice. *Cancer Res,* **52,** 4413-4419.

Linet, M.S. (1984) *The leukemias: Epidemiologic aspects.* Oxford University Press, New York, NY

Liu, B., Gross, M., ten Hoeve, J. & Shuai, K. (2001) A transcriptional corepressor of Stat1 with an essential LXXLL signature motif. *Proc Natl Acad Sci U S A,* **98,** 3203-3207.

Liu, D.W., Chen, S.T. & Liu, H.P. (2005) Choice of endogenous control for gene expression in nonsmall cell lung cancer. *Eur Respir J,* **26,** 1002-1008.

Livak, K.J. & Schmittgen, T.D. (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods,* **25,** 402-408.

Lu, W., van Eerde, A.M., Fan, X., Quintero-Rivera, F., Kulkarni, S., Ferguson, H., Kim, H.G., Fan, Y., Xi, Q., Li, Q.G., Sanlaville, D., Andrews, W., Sundaresan, V., Bi, W., Yan, J., Giltay, J.C., Wijmenga, C., de Jong, T.P., Feather, S.A., Woolf, A.S., Rao, Y., Lupski, J.R., Eccles, M.R., Quade, B.J., Gusella, J.F., Morton, C.C. & Maas, R.L. (2007) Disruption of ROBO2 is associated with urinary tract anomalies and confers risk of vesicoureteral reflux. *Am J Hum Genet,* **80,** 616-632.

Luo, L.Y., Herrera, I., Soosaipillai, A. & Diamandis, E.P. (2002) Identification of heat shock protein 90 and other proteins as tumour antigens by serological screening of an ovarian carcinoma expression library. *Br J Cancer,* **87,** 339-343.

Mao, S., Frank, R.C., Zhang, J., Miyazaki, Y. & Nimer, S.D. (1999) Functional and physical interactions between AML1 proteins and an ETS protein, MEF: implications for the pathogenesis of t(8;21)-positive leukemias. *Mol Cell Biol,* **19,** 3635-3644.

Marijt, W.A., Heemskerk, M.H., Kloosterboer, F.M., Goulmy, E., Kester, M.G., van der Hoorn, M.A., van Luxemburg-Heys, S.A., Hoogeboom, M., Mutis, T., Drijfhout, J.W., van Rood, J.J., Willemze, R. & Falkenburg, J.H. (2003) Hematopoiesis-restricted minor histocompatibility antigens HA-1- or HA-2-specific T cells can induce complete remissions of relapsed leukemia. *Proc Natl Acad Sci U S A,* **100,** 2742-2747.

Marlar, R.A., Kleiss, A.J. & Griffin, J.H. (1982) Mechanism of action of human activated protein C, a thrombin-dependent anticoagulant enzyme. *Blood,* **59,** 1067-1072.

Matsuzawa, S., Takayama, S., Froesch, B.A., Zapata, J.M. & Reed, J.C. (1998) p53-inducible human homologue of Drosophila seven in absentia (Siah) inhibits cell growth: suppression by BAG-1. *Embo J,* **17,** 2736-2747.

Mauritzson, N., Johansson, B., Albin, M., Billstrom, R., Ahlgren, T., Mikoczy, Z., Nilsson, P.G., Hagmar, L. & Mitelman, F. (1999) A single-center population-based consecutive series of 1500 cytogenetically investigated adult hematological malignancies: karyotypic features in relation to morphology, age and gender. *Eur J Haematol,* **62,** 95-102.

Mayer, R.J., Davis, R.B., Schiffer, C.A., Berg, D.T., Powell, B.L., Schulman, P., Omura, G.A., Moore, J.O., McIntyre, O.R. & Frei, E., 3rd (1994) Intensive postremission chemotherapy in adults with acute myeloid leukemia. Cancer and Leukemia Group B. *N Engl J Med,* **331,** 896-903.

Meadows, L., Wang, W., den Haan, J.M., Blokland, E., Reinhardus, C., Drijfhout, J.W., Shabanowitz, J., Pierce, R., Agulnik, A.I., Bishop, C.E., Hunt, D.F., Goulmy, E. & Engelhard, V.H. (1997) The HLA-A*0201-restricted H-Y antigen contains a posttranslationally modified cysteine that significantly affects T cell recognition. *Immunity,* **6,** 273-281.

Mehta, J., Singhal, S., Gee, A.P., Chiang, K.Y., Godder, K., Rhee Fv, F., DeRienzo, S., O'Neal, W., Lamb, L. & Henslee-Downey, P.J. (2004) Bone marrow transplantation from partially HLA-mismatched family donors for acute leukemia: single-center experience of 201 patients. *Bone Marrow Transplant,* **33,** 389-396.

Meklat, F., Li, Z., Wang, Z., Zhang, Y., Zhang, J., Jewell, A. & Lim, S.H. (2007) Cancer-testis antigens in haematological malignancies. *Br J Haematol,* **136,** 769-776.

Melhuish, T.A., Gallo, C.M. & Wotton, D. (2001) TGIF2 interacts with histone deacetylase 1 and represses transcription. *J Biol Chem,* **276,** 32109-32114.

Melloni, G., Ferreri, A.J., Russo, V., Gattinoni, L., Arrigoni, G., Ceresoli, G.L., Zannini, P. & Traversari, C. (2004) Prognostic significance of cancer-testis gene expression in resected non-small cell lung cancer patients. *Oncol Rep,* **12,** 145-151.

Miao, H., Chen, L., Riordan, S.M., Li, W., Juarez, S., Crabb, A.M., Lukas, T.J., Du, P., Lin, S.M., Wise, A., Agapova, O.A., Yang, P., Gu, C.C. & Hernandez, M.R. (2008) Gene expression and functional studies of the optic nerve head astrocyte transcriptome from normal African Americans and Caucasian Americans donors. *PLoS ONE,* **3,** e2847.

Middleton, D., Menchaca, L., Rood, H. & Komerofsky, R. (2003) New allele frequency database: http://www.allelefrequencies.net. *Tissue Antigens,* **61,** 403-407.

Miklos, D.B., Kim, H.T., Miller, K.H., Guo, L., Zorn, E., Lee, S.J., Hochberg, E.P., Wu, C.J., Alyea, E.P., Cutler, C., Ho, V., Soiffer, R.J., Antin, J.H. & Ritz, J. (2005) Antibody responses to H-Y minor histocompatibility antigens correlate with chronic graft-versus-host disease and disease remission. *Blood,* **105,** 2973-2978.

Mitus, A.J., Miller, K.B., Schenkein, D.P., Ryan, H.F., Parsons, S.K., Wheeler, C. & Antin, J.H. (1995) Improved survival for patients with acute myelogenous leukemia. *J Clin Oncol,* **13,** 560-569.

Montesinos, P., Lorenzo, I., Martin, G., Sanz, J., Perez-Sirvent, M.L., Martinez, D., Orti, G., Algarra, L., Martinez, J., Moscardo, F., de la Rubia, J., Jarque, I., Sanz, G. & Sanz, M.A. (2008) Tumor lysis syndrome in patients with acute myeloid leukemia: identification of risk factors and development of a predictive model. *Haematologica,* **93,** 67-74.

Moorman, A.V., Roman, E., Willett, E.V., Dovey, G.J., Cartwright, R.A. & Morgan, G.J. (2001) Karyotype and age in acute myeloid leukemia. Are they linked? *Cancer Genet Cytogenet,* **126,** 155-161.

Morin, P.J., Sparks, A.B., Korinek, V., Barker, N., Clevers, H., Vogelstein, B. & Kinzler, K.W. (1997) Activation of beta-catenin-Tcf signaling in colon cancer by mutations in beta-catenin or APC. *Science,* **275,** 1787-1790.

Mukhopadhyay, M., Shtrom, S., Rodriguez-Esteban, C., Chen, L., Tsukui, T., Gomer, L., Dorward, D.W., Glinka, A., Grinberg, A., Huang, S.P., Niehrs, C., Belmonte, J.C. & Westphal, H. (2001) Dickkopf1 is required for embryonic head induction and limb morphogenesis in the mouse. *Dev Cell,* **1,** 423-434.

Mutis, T., Gillespie, G., Schrama, E., Falkenburg, J.H., Moss, P. & Goulmy, E. (1999) Tetrameric HLA class I-minor histocompatibility antigen peptide complexes demonstrate minor histocompatibility antigen-specific cytotoxic T lymphocytes in patients with graft-versus-host disease. *Nat Med,* **5,** 839-842.

Mutis, T. & Goulmy, E. (2002) Hematopoietic system-specific antigens as targets for cellular immunotherapy of hematological malignancies. *Semin Hematol,* **39,** 23-31.

Nakakuki, K., Imoto, I., Pimkhaokham, A., Fukuda, Y., Shimada, Y., Imamura, M., Amagasa, T. & Inazawa, J. (2002) Novel targets for the 18p11.3 amplification frequently observed in esophageal squamous cell carcinomas. *Carcinogenesis,* **23,** 19-24.

Nakanishi, M., Tanaka, K., Shintani, T., Takahashi, T. & Kamada, N. (1999) Chromosomal instability in acute myelocytic leukemia and myelodysplastic syndrome patients among atomic bomb survivors. *J Radiat Res (Tokyo),* **40,** 159-167.

Newmark, H.L., Yang, K., Kurihara, N., Fan, K., Augenlicht, L.H. & Lipkin, M. (2009) Western-style diet-induced colonic tumors and their modulation by calcium and vitamin D in C57Bl/6 mice: a preclinical model for human sporadic colon cancer. *Carcinogenesis,* **30,** 88-92.

Nguyen, K.A., Cao, Y., Chen, J.R., Townsend, C.M., Jr. & Ko, T.C. (2006) Dietary fiber enhances a tumor suppressor signaling pathway in the gut. *Ann Surg,* **243,** 619-625; discussion 625-617.

Nijman, H.W., Houbiers, J.G., Vierboom, M.P., van der Burg, S.H., Drijfhout, J.W., D'Amaro, J., Kenemans, P., Melief, C.J. & Kast, W.M. (1993) Identification of peptide sequences that potentially trigger HLA-A2.1-restricted cytotoxic T lymphocytes. *Eur J Immunol,* **23,** 1215-1219.

Noyes, M.B., Christensen, R.G., Wakabayashi, A., Stormo, G.D., Brodsky, M.H. & Wolfe, S.A. (2008) Analysis of homeodomain specificities allows the family-wide prediction of preferred recognition sites. *Cell,* **133,** 1277-1289.

Oh, I.H. & Reddy, E.P. (1999) The myb gene family in cell growth, differentiation and apoptosis. *Oncogene,* **18,** 3017-3033.

Ohlenschlaeger, T., Garred, P., Madsen, H.O. & Jacobsen, S. (2004) Mannose-binding lectin variant alleles and the risk of arterial thrombosis in systemic lupus erythematosus. *N Engl J Med,* **351,** 260-267.

Okabe, H., Satoh, S., Furukawa, Y., Kato, T., Hasegawa, S., Nakajima, Y., Yamaoka, Y. & Nakamura, Y. (2003) Involvement of PEG10 in human hepatocellular carcinogenesis through interaction with SIAH1. *Cancer Res,* **63,** 3043-3048.

Ousati Ashtiani, Z., Ayati, M., Modarresi, M.H., Raoofian, R., Sabah Goulian, B., Greene, W.K. & Heidari, M. (2009) Association of TGIFLX/Y mRNA expression with prostate cancer. *Med Oncol,* **26,** 73-77.

Parham, P. & McQueen, K.L. (2003) Alloreactive killer cells: hindrance and help for haematopoietic transplants. *Nat Rev Immunol,* **3,** 108-122.

Park, M.H., Cho, S.A., Yoo, K.H., Yang, M.H., Ahn, J.Y., Lee, H.S., Lee, K.E., Mun, Y.C., Cho, D.H., Seong, C.M. & Park, J.H. (2007) Gene expression profile related to prognosis of acute myeloid leukemia. *Oncol Rep,* **18,** 1395-1402.

Parker, K.C., Bednarek, M.A. & Coligan, J.E. (1994) Scheme for ranking potential HLA-A2 binding peptides based on independent binding of individual peptide side-chains. *J Immunol,* **152,** 163-175.

Parker, K.C., Shields, M., DiBrino, M., Brooks, A. & Coligan, J.E. (1995) Peptide binding to MHC class I molecules: implications for antigenic peptide prediction. *Immunol Res,* **14,** 34-57.

Pasquini, M.C. (2008) Impact of graft-versus-host disease on survival. *Best Pract Res Clin Haematol,* **21,** 193-204.

Passner, J.M., Ryoo, H.D., Shen, L., Mann, R.S. & Aggarwal, A.K. (1999) Structure of a DNA-bound Ultrabithorax-Extradenticle homeodomain complex. *Nature,* **397,** 714-719.

Pelletier, N., Champagne, N., Stifani, S. & Yang, X.J. (2002) MOZ and MORF histone acetyltransferases interact with the Runt-domain transcription factor Runx2. *Oncogene,* **21,** 2729-2740.

Petersdorf, S.H., Rankin, C., Head, D.R., Terebelo, H.R., Willman, C.L., Balcerzak, S.P., Karnad, A.B., Dakhil, S.R. & Appelbaum, F.R. (2007) Phase II evaluation of an intensified induction therapy with standard daunomycin and cytarabine followed by high dose cytarabine for adults with previously untreated acute myeloid leukemia: a Southwest Oncology Group study (SWOG-9500). *Am J Hematol,* **82,** 1056-1062.

Piper, D.E., Batchelor, A.H., Chang, C.P., Cleary, M.L. & Wolberger, C. (1999) Structure of a HoxB1-Pbx1 heterodimer bound to DNA: role of the hexapeptide and a fourth homeodomain helix in complex formation. *Cell,* **96,** 587-597.

Plunkett, W., Liliemark, J.O., Adams, T.M., Nowak, B., Estey, E., Kantarjian, H. & Keating, M.J. (1987) Saturation of 1-beta-D-arabinofuranosylcytosine 5'-triphosphate accumulation in leukemia cells during high-dose 1-beta-D-arabinofuranosylcytosine therapy. *Cancer Res,* **47,** 3005-3011.

Ponting, C.P., Hofmann, K. & Bork, P. (1999) A latrophilin/CL-1-like GPS domain in polycystin-1. *Curr Biol,* **9,** R585-588.

Porter, D. & Levine, J.E. (2006) Graft-versus-host disease and graft-versus-leukemia after donor leukocyte infusion. *Semin Hematol,* **43,** 53-61.

Porter, D.L. & Antin, J.H. (2006) Donor leukocyte infusions in myeloid malignancies: new strategies. *Best Pract Res Clin Haematol,* **19,** 737-755.

Powles, R.L., Morgenstern, G.R., Kay, H.E., McElwain, T.J., Clink, H.M., Dady, P.J., Barrett, A., Jameson, B., Depledge, M.H., Watson, J.G., Sloane, J., Leigh, M., Lumley, H., Hedley, D., Lawler, S.D., Filshie, J. & Robinson, B. (1983) Mismatched family donors for bone-marrow transplantation as treatment for acute leukaemia. *Lancet,* **1,** 612-615.

Rammensee, H., Bachmann, J., Emmerich, N.P., Bachor, O.A. & Stevanovic, S. (1999) SYFPEITHI: database for MHC ligands and peptide motifs. *Immunogenetics,* **50,** 213-219.

Randolph, S., Gooley, T., Warren, E., Appelbaum, F. & Riddell, S. (2004) Female donors contribute to a selective graft-versus-leukemia effect in male recipients of HLA-matched, related hematopoietic stem cell transplants. *Blood,* **103,** 347-352.

Raspadori, D., Damiani, D., Michieli, M., Stocchi, R., Gentili, S., Gozzetti, A., Masolini, P., Michelutti, A., Geromin, A., Fanin, R. & Lauria, F. (2002) CD56 and PGP expression in acute myeloid leukemia: impact on clinical outcome. *Haematologica,* **87,** 1135-1140.

Redaelli, A., Lee, J.M., Stephens, J.M. & Pashos, C.L. (2003) Epidemiology and clinical burden of acute myeloid leukemia. *Expert Rev Anticancer Ther,* **3,** 695-710.

Reddy, P., Arora, M., Guimond, M. & Mackall, C.L. (2008) GVHD: a continuing barrier to the safety of allogeneic transplantation. *Biol Blood Marrow Transplant,* **15,** 162-168.

Rodt, H., Kolb, H.J., Netzel, B., Rieder, I., Janka, G., Belohradsky, B., Haas, R.J. & Thierfelder, S. (1979) GVHD suppression by incubation of bone marrow grafts with anti-T-cell globulin: effect in the canine model and application to clinical bone marrow transplantation. *Transplant Proc,* **11,** 962-966.

Romeo, G., Hassan, H.J., Staempfli, S., Roncuzzi, L., Cianetti, L., Leonardi, A., Vicente, V., Mannucci, P.M., Bertina, R., Peschle, C. & et al. (1987) Hereditary thrombophilia: identification of nonsense and missense mutations in the protein C gene. *Proc Natl Acad Sci U S A,* **84,** 2829-2832.

Rosinski, K.V., Fujii, N., Mito, J.K., Koo, K.K., Xuereb, S.M., Sala-Torra, O., Gibbs, J.S., Radich, J.P., Akatsuka, Y., Van den Eynde, B.J., Riddell, S.R. & Warren, E.H. (2008) DDX3Y encodes a class I MHC-restricted H-Y antigen that is expressed in leukemic stem cells. *Blood,* **111,** 4817-4826.

Rossi, G., Pelizzari, A.M., Bellotti, D., Tonelli, M. & Barlati, S. (2000) Cytogenetic analogy between myelodysplastic syndrome and acute myeloid leukemia of elderly patients. *Leukemia,* **14,** 636-641.

Rowe, J.M., Neuberg, D., Friedenberg, W., Bennett, J.M., Paietta, E., Makary, A.Z., Liesveld, J.L., Abboud, C.N., Dewald, G., Hayes, F.A., Tallman, M.S. & Wiernik, P.H. (2004) A phase 3 study of three induction regimens and of priming with GM-CSF in older adults with acute myeloid leukemia: a trial by the Eastern Cooperative Oncology Group. *Blood,* **103,** 479-485.

Rowe, J.M. & Tallman, M.S. (1997) Intensifying induction therapy in acute myeloid leukemia: has a new standard of care emerged? *Blood,* **90,** 2121-2126.

Rubinfeld, B., Robbins, P., El-Gamil, M., Albert, I., Porfiri, E. & Polakis, P. (1997) Stabilization of beta-catenin by genetic defects in melanoma cell lines. *Science,* **275,** 1790-1792.

Ruggeri, L., Capanni, M., Casucci, M., Volpi, I., Tosti, A., Perruccio, K., Urbani, E., Negrin, R.S., Martelli, M.F. & Velardi, A. (1999) Role of natural killer cell alloreactivity in HLA-mismatched hematopoietic stem cell transplantation. *Blood,* **94,** 333-339.

Ruggeri, L., Capanni, M., Urbani, E., Perruccio, K., Shlomchik, W.D., Tosti, A., Posati, S., Rogaia, D., Frassoni, F., Aversa, F., Martelli, M.F. & Velardi, A. (2002) Effectiveness of donor natural killer cell alloreactivity in mismatched hematopoietic transplants. *Science,* **295,** 2097-2100.

Sachdev, S., Bruhn, L., Sieber, H., Pichler, A., Melchior, F. & Grosschedl, R. (2001) PIASy, a nuclear matrix-associated SUMO E3 ligase, represses LEF1 activity by sequestration into nuclear bodies. *Genes Dev,* **15,** 3088-3103.

Salter, R.D. & Cresswell, P. (1986) Impaired assembly and transport of HLA-A and -B antigens in a mutant TxB cell hybrid. *Embo J,* **5,** 943-949.

Sambrook, J. & Russell, D.W. (2001) *Molecular Cloning: A Laboratory Manual.* Cold Spring Harbor Laboratory Press, Woodbury NY.

Sanderson, R.N., Johnson, P.R., Moorman, A.V., Roman, E., Willett, E., Taylor, P.R., Proctor, S.J., Bown, N., Ogston, S. & Bowen, D.T. (2006) Population-based

demographic study of karyotypes in 1709 patients with adult acute myeloid leukemia. *Leukemia,* **20,** 444-450.

Sandler, D.P. & Collman, G.W. (1987) Cytogenetic and environmental factors in the etiology of the acute leukemias in adults. *Am J Epidemiol,* **126,** 1017-1032.

Sandler, D.P. & Ross, J.A. (1997) Epidemiology of acute leukemia in children and adults. *Semin Oncol,* **24,** 3-16.

Sanz, C., Calasanz, M.J., Andreu, E., Richard, C., Prosper, F. & Fernandez-Luna, J.L. (2004) NALP1 is a transcriptional target for cAMP-response-element-binding protein (CREB) in myeloid leukaemia cells. *Biochem J,* **384,** 281-286.

Sanz, G.F., Sanz, M.A., Vallespi, T., Canizo, M.C., Torrabadella, M., Garcia, S., Irriguible, D. & San Miguel, J.F. (1989) Two regression models and a scoring system for predicting survival and planning treatment in myelodysplastic syndromes: a multivariate analysis of prognostic factors in 370 patients. *Blood,* **74,** 395-408.

Sato, T., Sato, N., Takahashi, S., Koshiba, H. & Kikuchi, K. (1986) Specific cytotoxicity of a long-term cultured T-cell clone on human autologous mammary cancer cells. *Cancer Res,* **46,** 4384-4389.

Schiffer, C.A., Lee, E.J., Tomiyasu, T., Wiernik, P.H. & Testa, J.R. (1989) Prognostic impact of cytogenetic abnormalities in patients with de novo acute nonlymphocytic leukemia. *Blood,* **73,** 263-270.

Schneider, G. & Fechner, U. (2004) Advances in the prediction of protein targeting signals. *Proteomics,* **4,** 1571-1580.

Schnittger, S., Schoch, C., Kern, W., Hiddemann, W. & Haferlach, T. (2004) FLT3 length mutations as marker for follow-up studies in acute myeloid leukaemia. *Acta Haematol,* **112,** 68-78.

Schoch, C., Kern, W., Schnittger, S., Hiddemann, W. & Haferlach, T. (2004) Karyotype is an independent prognostic parameter in therapy-related acute myeloid leukemia (t-AML): an analysis of 93 patients with t-AML in comparison to 1091 patients with de novo AML. *Leukemia,* **18,** 120-125.

Schuster, F.R., Buhmann, R., Reuther, S., Hubner, B., Grabrucker, C., Liepert, A., Reibke, R., Lichtner, P., Yang, T., Kroell, T., Kolb, H.J., Borkhardt, A. & Schmetzer, H. (2008) Improved effector function of leukemia-specific T-lymphocyte clones trained with AML-derived dendritic cells. *Cancer Genomics Proteomics,* **5,** 275-286.

Seymour, J.F., Pierce, S.A., Kantarjian, H.M., Keating, M.J. & Estey, E.H. (1994) Investigation of karyotypic, morphologic and clinical features in patients with acute myeloid leukemia blast cells expressing the neural cell adhesion molecule (CD56). *Leukemia,* **8,** 823-826.

Sharma, M.R., Polavarapu, R., Roseman, D., Patel, V., Eaton, E., Kishor, P.B. & Nanji, A.A. (2008) Increased severity of alcoholic liver injury in female verses male rats: a microarray analysis. *Exp Mol Pathol,* **84,** 46-58.

Shostak, Y., Van Gilst, M.R., Antebi, A. & Yamamoto, K.R. (2004) Identification of C. elegans DAF-12-binding sites, response elements, and target genes. *Genes Dev,* **18,** 2529-2544.

Simmons, W.A., Summerfield, S.G., Roopenian, D.C., Slaughter, C.A., Zuberi, A.R., Gaskell, S.J., Bordoli, R.S., Hoyes, J., Moomaw, C.R., Colbert, R.A., Leong, L.Y., Butcher, G.W., Hammer, R.E. & Taurog, J.D. (1997) Novel HY peptide antigens presented by HLA-B27. *J Immunol,* **159,** 2750-2759.

Skaletsky, H., Kuroda-Kawaguchi, T., Minx, P.J., Cordum, H.S., Hillier, L., Brown, L.G., Repping, S., Pyntikova, T., Ali, J., Bieri, T., Chinwalla, A., Delehaunty, A., Delehaunty, K., Du, H., Fewell, G., Fulton, L., Fulton, R., Graves, T., Hou, S.F., Latrielle, P., Leonard, S., Mardis, E., Maupin, R., McPherson, J., Miner, T., Nash, W., Nguyen, C., Ozersky, P., Pepin, K., Rock, S., Rohlfing, T., Scott, K., Schultz, B., Strong, C., Tin-Wollam, A., Yang, S.P., Waterston, R.H., Wilson, R.K., Rozen, S. & Page, D.C. (2003) The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. *Nature,* **423,** 825-837.

Sourisseau, T., Desbois, C., Debure, L., Bowtell, D.D., Cato, A.C., Schneikert, J., Moyse, E. & Michel, D. (2001) Alteration of the stability of Bag-1 protein in the control of olfactory neuronal apoptosis. *J Cell Sci,* **114,** 1409-1416.

Spierings, E., Vermeulen, C.J., Vogt, M.H., Doerner, L.E., Falkenburg, J.H., Mutis, T. & Goulmy, E. (2003) Identification of HLA class II-restricted H-Y-specific T-helper epitope evoking CD4+ T-helper cells in H-Y-mismatched transplantation. *Lancet,* **362,** 610-615.

Summerfield, J.A., Sumiya, M., Levin, M. & Turner, M.W. (1997) Association of mutations in mannose binding protein gene with childhood infection in consecutive hospital series. *BMJ,* **314,** 1229-1232.

Sun, C., Orozco, O., Olson, D.L., Choi, E., Garber, E., Tizard, R., Szak, S., Sanicola, M. & Carulli, J.P. (2008) CRIPTO3, a presumed pseudogene, is expressed in cancer. *Biochem Biophys Res Commun,* **377,** 215-220.

Suzuki, R., Onizuka, M., Kojima, M., Shimada, M., Fukagawa, S., Tsuboi, K., Kobayashi, H., Shintani, A., Ogawa, Y., Kawada, H., Hotta, T. & Ando, K. (2007) Preferential hypermethylation of the Dickkopf-1 promoter in core-binding factor leukaemia. *Br J Haematol,* **138,** 624-631.

Szydlo, R., Goldman, J.M., Klein, J.P., Gale, R.P., Ash, R.C., Bach, F.H., Bradley, B.A., Casper, J.T., Flomenberg, N., Gajewski, J.L., Gluckman, E., Henslee-Downey, P.J., Hows, J.M., Jacobsen, N., Kolb, H.J., Lowenberg, B., Masaoka, T., Rowlings, P.A., Sondel, P.M., van Bekkum, D.W., van Rood, J.J., Vowels, M.R., Zhang, M.J. & Horowitz, M.M. (1997) Results of allogeneic bone marrow transplants for leukemia using donors other than HLA-identical siblings. *J Clin Oncol,* **15,** 1767-1777.

Tadeu, A.M., Ribeiro, S., Johnston, J., Goldberg, I., Gerloff, D. & Earnshaw, W.C. (2008) CENP-V is required for centromere organization, chromosome alignment and cytokinesis. *Embo J,* **27,** 2510-2522.

Tallman, M.S., Andersen, J.W., Schiffer, C.A., Appelbaum, F.R., Feusner, J.H., Ogden, A., Shepherd, L., Willman, C., Bloomfield, C.D., Rowe, J.M. & Wiernik, P.H. (1997) All-trans-retinoic acid in acute promyelocytic leukemia. *N Engl J Med,* **337,** 1021-1028.

Tanaka, K. (2009) The proteasome: overview of structure and functions. *Proc Jpn Acad Ser B Phys Biol Sci,* **85,** 12-36.

Terry, S., Queires, L., Gil-Diez-de-Medina, S., Chen, M.W., de la Taille, A., Allory, Y., Tran, P.L., Abbou, C.C., Buttyan, R. & Vacherot, F. (2006) Protocadherin-PC promotes androgen-independent prostate cancer cell growth. *Prostate,* **66,** 1100-1113.

Thirman, M.J., Gill, H.J., Burnett, R.C., Mbangkollo, D., McCabe, N.R., Kobayashi, H., Ziemin-van der Poel, S., Kaneko, Y., Morgan, R., Sandberg, A.A. & et al. (1993) Rearrangement of the MLL gene in acute lymphoblastic and acute

myeloid leukemias with 11q23 chromosomal translocations. *N Engl J Med,* **329,** 909-914.

Thomas, E., Storb, R., Clift, R.A., Fefer, A., Johnson, F.L., Neiman, P.E., Lerner, K.G., Glucksberg, H. & Buckner, C.D. (1975) Bone-marrow transplantation (first of two parts). *N Engl J Med,* **292,** 832-843.

Thrower, J.S., Hoffman, L., Rechsteiner, M. & Pickart, C.M. (2000) Recognition of the polyubiquitin proteolytic signal. *Embo J,* **19,** 94-102.

Tsai, W.H., Shih, C.H., Lin, C.C., Ho, C.K., Hsu, F.C. & Hsu, H.C. (2008) Monocyte chemotactic protein-1 in the migration of differentiated leukaemic cells toward alveolar epithelial cells. *Eur Respir J,* **31,** 957-962.

Tsuchiya, S., Yamabe, M., Yamaguchi, Y., Kobayashi, Y., Konno, T. & Tada, K. (1980) Establishment and characterization of a human acute monocytic leukemia cell line (THP-1). *Int J Cancer,* **26,** 171-176.

van de Loosdrecht, A.A., van den Ancker, W., Houtenbos, I., Ossenkoppele, G.J. & Westers, T.M. (2009) Dendritic cell-based immunotherapy in myeloid leukaemia: translating fundamental mechanisms into clinical applications. *Handb Exp Pharmacol***,** 319-348.

van der Velden, V.H., Hochhaus, A., Cazzaniga, G., Szczepanski, T., Gabert, J. & van Dongen, J.J. (2003) Detection of minimal residual disease in hematologic malignancies by real-time quantitative PCR: principles, approaches, and laboratory aspects. *Leukemia,* **17,** 1013-1034.

Van Dyck, F., Delvaux, E.L., Van de Ven, W.J. & Chavez, M.V. (2004) Repression of the Transactivating Capacity of the Oncoprotein PLAG1 by SUMOylation. *J Biol Chem,* **279,** 36121-36131.

Vandesompele, J., De Preter, K., Pattyn, F., Poppe, B., Van Roy, N., De Paepe, A. & Speleman, F. (2002) Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol,* **3,** RESEARCH0034.

Vardiman, J.W., Harris, N.L. & Brunning, R.D. (2002) The World Health Organization (WHO) classification of the myeloid neoplasms. *Blood,* **100,** 2292-2302.

Vehar, G.A. & Davie, E.W. (1980) Preparation and properties of bovine factor VIII (antihemophilic factor). *Biochemistry,* **19,** 401-410.

Viatour, P., Merville, M.P., Bours, V. & Chariot, A. (2005) Phosphorylation of NF-kappaB and IkappaB proteins: implications in cancer and inflammation. *Trends Biochem Sci,* **30,** 43-52.

Vogler, W.R., Velez-Garcia, E., Weiner, R.S., Flaum, M.A., Bartolucci, A.A., Omura, G.A., Gerber, M.C. & Banks, P.L. (1992) A phase III trial comparing idarubicin and daunorubicin in combination with cytarabine in acute myelogenous leukemia: a Southeastern Cancer Study Group Study. *J Clin Oncol,* **10,** 1103-1111.

Vogt, M.H., van den Muijsenberg, J.W., Goulmy, E., Spierings, E., Kluck, P., Kester, M.G., van Soest, R.A., Drijfhout, J.W., Willemze, R. & Falkenburg, J.H. (2002) The DBY gene codes for an HLA-DQ5-restricted human male-specific minor histocompatibility antigen involved in graft-versus-host disease. *Blood,* **99,** 3027-3032.

von Muhlendahl, K.E. (1998) Chernobyl fallout, nuclear plants and leukaemia: review of recent literature. *Eur J Pediatr,* **157,** 602-604.

Voorter, C., Joos, S., Bringuier, P.P., Vallinga, M., Poddighe, P., Schalken, J., du Manoir, S., Ramaekers, F., Lichter, P. & Hopman, A. (1995) Detection of

chromosomal imbalances in transitional cell carcinoma of the bladder by comparative genomic hybridization. *Am J Pathol,* **146,** 1341-1354.

Wahlin, A., Hornsten, P. & Jonsson, H. (1991) Remission rate and survival in acute myeloid leukemia: impact of selection and chemotherapy. *Eur J Haematol,* **46,** 240-247.

Wallerath, T., Poleo, D., Li, H. & Forstermann, U. (2003) Red wine increases the expression of human endothelial nitric oxide synthase: a mechanism that may contribute to its beneficial cardiovascular effects. *J Am Coll Cardiol,* **41,** 471-478.

Walport, M.J. (2001) Complement. First of two parts. *N Engl J Med,* **344,** 1058-1066.

Wang, J. & Maldonado, M.A. (2006) The ubiquitin-proteasome system and its role in inflammatory and autoimmune diseases. *Cell Mol Immunol,* **3,** 255-261.

Wang, W., Meadows, L.R., den Haan, J.M., Sherman, N.E., Chen, Y., Blokland, E., Shabanowitz, J., Agulnik, A.I., Hendrickson, R.C., Bishop, C.E. & et al. (1995) Human H-Y: a male-specific histocompatibility antigen derived from the SMCY protein. *Science,* **269,** 1588-1590.

Wang, Z. & Mann, R.S. (2003) Requirement for two nearly identical TGIF-related homeobox genes in Drosophila spermatogenesis. *Development,* **130,** 2853-2865.

Warren, E.H., Gavin, M., Greenberg, P.D. & Riddell, S.R. (1998) Minor histocompatibility antigens as targets for T-cell therapy after bone marrow transplantation. *Curr Opin Hematol,* **5,** 429-433.

Warren, E.H., Gavin, M.A., Simpson, E., Chandler, P., Page, D.C., Disteche, C., Stankey, K.A., Greenberg, P.D. & Riddell, S.R. (2000) The human UTY gene encodes a novel HLA-B8-restricted H-Y antigen. *J Immunol,* **164,** 2807-2814.

Weick, J.K., Kopecky, K.J., Appelbaum, F.R., Head, D.R., Kingsbury, L.L., Balcerzak, S.P., Bickers, J.N., Hynes, H.E., Welborn, J.L., Simon, S.R. & Grever, M. (1996) A randomized investigation of high-dose versus standard-dose cytosine arabinoside with daunorubicin in patients with previously untreated acute myeloid leukemia: a Southwest Oncology Group study. *Blood,* **88,** 2841-2851.

Weiden, P.L., Flournoy, N., Thomas, E.D., Prentice, R., Fefer, A., Buckner, C.D. & Storb, R. (1979) Antileukemic effect of graft-versus-host disease in human recipients of allogeneic-marrow grafts. *N Engl J Med,* **300,** 1068-1073.

Weiden, P.L., Sullivan, K.M., Flournoy, N., Storb, R. & Thomas, E.D. (1981) Antileukemic effect of chronic graft-versus-host disease: contribution to improved survival after allogeneic marrow transplantation. *N Engl J Med,* **304,** 1529-1533.

Wolff, D., Knopp, A., Weirich, V., Steiner, B., Junghanss, C., Casper, J. & Freund, M. (2005) Loss of the GVL effect by loss of the Y-chromosome as putative mechanism of immune escape in ALL. *Bone Marrow Transplant,* **35,** 101-102.

Wong, A.K., Fang, B., Zhang, L., Guo, X., Lee, S. & Schreck, R. (2008) Loss of the Y chromosome: an age-related or clonal phenomenon in acute myelogenous leukemia/myelodysplastic syndrome? *Arch Pathol Lab Med,* **132,** 1329-1332.

Wong, G.T., Gavin, B.J. & McMahon, A.P. (1994) Differential transformation of mammary epithelial cells by Wnt genes. *Mol Cell Biol,* **14,** 6278-6286.

Wotton, D., Lo, R.S., Lee, S. & Massague, J. (1999) A Smad transcriptional corepressor. *Cell,* **97,** 29-39.

Yates, J., Glidewell, O., Wiernik, P., Cooper, M.R., Steinberg, D., Dosik, H., Levy, R., Hoagland, C., Henry, P., Gottlieb, A., Cornell, C., Berenberg, J., Hutchison, J.L., Raich, P., Nissen, N., Ellison, R.R., Frelick, R., James, G.W., Falkson, G.,

Silver, R.T., Haurani, F., Green, M., Henderson, E., Leone, L. & Holland, J.F. (1982) Cytosine arabinoside with daunorubicin or adriamycin for therapy of acute myelocytic leukemia: a CALGB study. *Blood,* **60,** 454-462.

Yingling, J.M., Blanchard, K.L. & Sawyer, J.S. (2004) Development of TGF-beta signalling inhibitors for cancer therapy. *Nat Rev Drug Discov,* **3,** 1011-1022.

Youn, B.S., Mantel, C. & Broxmeyer, H.E. (2000) Chemokines, chemokine receptors and hematopoiesis. *Immunol Rev,* **177,** 150-174.

Zhang, J., Guenther, M.G., Carthew, R.W. & Lazar, M.A. (1998) Proteasomal regulation of nuclear receptor corepressor-mediated repression. *Genes Dev,* **12,** 1775-1780.

Zhang, J., Xu, L.G., Han, K.J., Wei, X. & Shu, H.B. (2004) PIASy represses TRIF-induced ISRE and NF-kappaB activation but not apoptosis. *FEBS Lett,* **570,** 97-101.

Zhao, Q. & Lee, F.S. (1999) Mitogen-activated protein kinase/ERK kinase kinases 2 and 3 activate nuclear factor-kappaB through IkappaB kinase-alpha and IkappaB kinase-beta. *J Biol Chem,* **274,** 8355-8358.

Zou, M., Baitei, E.Y., Alzahrani, A.S., Al-Mohanna, F., Farid, N.R., Meyer, B. & Shi, Y. (2009) Oncogenic activation of MAP kinase by BRAF pseudogene in thyroid tumors. *Neoplasia,* **11,** 57-65.

Zou, S.W., Zhang, J.C., Zhang, X.D., Miao, S.Y., Zong, S.D., Sheng, Q. & Wang, L.F. (2003) Expression and localization of VCX/Y proteins and their possible involvement in regulation of ribosome assembly during spermatogenesis. *Cell Res,* **13,** 171-177.

Appendix

A.1 List of used primers

All the primers used in the present study are listed in table A.1. Forward (for) and reverse (rev) primers bind at the 5'- and at the 3'-side of the relative PCR products respectively.

Primer code	Sequence	Melting T	Product length	Purpose
TGIF2LY-pGEX-for-BamHI	TATAGGATCCATGGAGGCCGCTGCAGAC	55.7°C	578 bp	Cloning of TGIF2LY into pGEXΔBamHI, for heterologous expression and purification of the protein
TGIF2LY-pGEX-rev-HindIII	TTAAAAGCTTCTATTCCGGTGAGCTTCTGGCTAG	57.4°C		
TGIF2LY-C-SCL-for	TATAAAGCTTATGGAGGCCGCTGCAGAC	55.7°C	575 bp	Cloning of TGIF2LY into pcDNA4-C-Myc, for subcellular localisation
TGIF2LY-C-SCL-rev	TTAAGGATCCTTCCGGTGAGCTTCTGGCTAG	55.9°C		
TGIF2LY-N-SCL-for	TATAGAATTCATGGAGGCCGCTGCAGAC	55.7°C	578 bp	Cloning of TGIF2LY into pcDNA3-N-Myc, for subcellular localisation
TGIF2LY-N-SCL-rev	TTAAGAATTCCTATTCCGGTGAGCTTCTGGCTAG	57.4°C		
IGS-LINK-1	GGGGCGGCCGC	44.0°C	Various	Linkers for *in vitro* genomic selection. The primer is working both as forward and reverse. Linker 2 is phosphorylated at the 5'-end
IGS-LINK-2	Pho-GATCGCGGCCGCCCC	60.0°C		
IGS-PRIM	GGGGCGGCCGCGATC	60.4°C		
EMSA-1-for	Cy5-TGACATGACATGACATGACA	54.0°C	20 bp	Electromobility shift assay to determine TGIF2LY consensus sequence. All the forward primers are labelled with cyanine 5 at the 5'-end
EMSA-1-rev	TGTCATGTCATGTCATGTCA			
EMSA-2-for	Cy5-TGACCTGACCTGACCTGACC	63.0°C	20 bp	
EMSA-2-rev	GGTCAGGTCAGGTCAGGTCA			
EMSA-3-for	Cy5-TGACTTGACTTGACTTGACT	54.0°C	20 bp	
EMSA-3-rev	AGTCAAGTCAAGTCAAGTCA			
EMSA-4-for	Cy5-TGATATGATATGATATGATA	46.0°C	20 bp	
EMSA-4-rev	TATCATATCATATCATATCA			
EMSA-5-for	Cy5-TGATCTGATCTGATCTGATC	54.0°C	20 bp	
EMSA-5-rev	GATCAGATCAGATCAGATCA			
EMSA-6-for	Cy5-TGATTTGATTTGATTTGATT	46.0°C	20 bp	
EMSA-6-rev	AATCAAATCAAATCAAATCA			

Table A.1. List of primers used in the present study. Abbreviations: for: forward; rev: reverse; Pho: phosphorylation; Cy5: cyanine 5.

A.2 Genomic DNA fragments binding to TGIF2LY

Here are reported the sequences of the genomic DNA fragments isolated from THP-1 cells through in vitro genomic selection. The reverse complement sequence is reported as well for each fragment. The sequence 5'-TGA(C/T)(A/C/T)-3', which is similar to the consensus sequence of the human TGIF1 and TGIF2, and of the vismay protein

from *Drosophila melanogaster* is highlighted in the fragments. 5'-TGATA-3' and 5'-TGATC-3', for which TGIF2LY showed high affinity, are marked in red. 5'-TGACA-3' and 5'-TGACC-3', for which TGIF2LY showed low affinity, are marked in blue. The sequences where TGIF2LY was not binding, 5'-TGACT-3' and 5'-TGATT-3', are indicated by bold black letters.

Fragment "XL4"

CCTGAGGAATCACCACAC**TGACT**TCCACAATGGTTGAACTAGTTTACAGTC
CCACCAACAGTGTAAAAGTGTTCCTATTTCTCCACTTCCTCTCCAGCACCTG
TTGTTTCC**TGACT**TTTTAAT

Reverse complement

ATTAAAAAGTCAGGAAACAACAGGTGCTGGAGAGGAAGTGGAGAAATAGG
AACACTTTTACACTGTTGGTGGGACTGTAAACTAGTTCAACCATTGTGGAAG
TCAGTGTGG**TGATT**CCTCAGG

Fragment "XL8"

CCACTAGGGAGCAGT**TGACT**GATGCCGAACCTACTCTCCCAAAGCCCCACA
CCCGTGTTGCCAGGGTAGGCAGAGAGCTCCAGGGAGACAGAAACAGAGCT
GGGGAGCCTACACCAGCAGAAGCTGCAATTGGATAGGGTCAGAAACAGAA
ACCTTGGTTTTGCCACAGC**TGACC**AAAATCTCCACCAGTTCCTTGGTAAGGA
AAAAAGCAAAGCAGGAATTTGCTCGGGAACCTCTGTGTA

Reverse complement

TACACAGAGGTTCCCGTGCAAATTCCTGCTTTGCTTTTTTCCTTACCAAGGA
ACTGGTGGAGATTTTGGTCAGCTGTGGCAAAACCAAGGTTTCTGTTTC**TGAC**
CCTATCCAATTGCAGCTTCTGCTGGTGTAGGCTCCCCAGCTCTGTTTCTGTCT
CCCTGGAGCTCTCTGCCTACCCTGGCAACACGGGTGTGGGGCTTTGGGAGA
GTAGGTTCGGCATCAGTCAACTGCTCCCTAGTGG

Fragment "XL14"

TGGGG**TGACC**AACCATTCTGGTTTGCCTGGGACCAATGGGGCTTCCTTAGA
TAGAGGACTCAGTGCTAAAACCGAAAATCCAGGCAAACTGGACCGTGTTGG
TCATCCTATCCTGAACCTTTTGCTCTGTGCCTGAAGTCCCGCAGCTCCCTGGT
AGCCCACCCTGGTGTCTCTTCTGGCTCTGTTGCAGGTCTCAGCCCACCTATC
AGAATTTGCAGCTGTTTAATCTTAAACTCCCTGTTCC**TGATT**GATGGTTGGG
ACTG**TGACT**GGGAGTTGCTTAAATAGCTCCCTCCCGTTTCTCAGCTAAACTC
AATTTCACTATCAGCCCCCGCAGCTGGGGTAGATCTCGGTTCACTGCAACCT
CTGCCTCCCTGGGTTCAAGCAGTTCTTCTGCCTCAGCCTCCCGAGTAGCTGG

GATTACAGGCGCCCACCACCATGCCTGGTTAATTTTTTTGTATTTTCAGTAG
AGACAGAGTTTCACCGTGTTAGCCAGGATGGTCTG

Reverse complement

CAGACCATCCTGGCTAACACGGTGAAACTCTGTCTCTACTGAAAATACAAA
AAAATTAACCAGGCATGGTGGTGGGCGCCTGTAATCCCAGCTACTCGGGAG
GCTGAGGCAGAAGAACTGCTTGAACCCAGGGAGGCAGAGGTTGCAGTGAA
CCGAGATCTACCCCAGCTGCGGGGGC**TGATA**GTGAAATTGAGTTTAGCTGA
GAAACGGGAGGGAGCTATTTAAGCAACTCCCAGTCACAGTCCCAACCATCA
ATCAGGAACAGGGAGTTTAAGATTAAACAGCTGCAAATTC**TGATA**GGTGGG
CTGAGACCTGCAACAGAGCCAGAAGAGACACCAGGGTGGGCTACCAGGGA
GCTGCGGGACTTCAGGCACAGAGCAAAAGGTTCAGGATAGGA**TGACC**AAC
ACGGTCCAGTTTGCCTGGATTTTCGGTTTTAGCACTGAGTCCTCTATCTAAG
GAAGCCCCATTGGTCCCAGGCAAACCAGAATGGTTGGTCACCCCA

Fragment “XL15”

CAGCTGCTTGGCATTTTGAAGGCTGCTTTAGTGTTTTCCAGGTGTACTCTGTA
TTGGCAGACCATCGCCAGTGAGGCTCTTTGAGAATCACCAGAGGAGCTTGG
CTCACACGAACATTCCTGGTCCTTGTTTTAGTTAAAGCACTAGTATGTAGTC
TGCCTGCCAGGACTCAAACCCAGCTCTACCTTCTGCTAGCTGTGTGAACTCA
CTATGTGTTCATTTCCTTGGTAAATAGGAACAATGGATAGTACATGTCTTTA
TAGGGTTTAGGGACAGGACAAAATGAGACAATACCTAAATACAAGGGGTAT
TAGCAGAATTCTTGGCACATAGTTAAGAATCACTAATTATTATAATTAATTT
CTAAAGTGCCAAGGATATTGCTAACAGCATATTTTGTAATTGAAAATTTCCA
GTGCAACTGCTGGGAAAAAAGTTGAGAACTGCTGCTG**TGATA**CCTCTTTTC
ACAGATGCA

Reverse complement

TGCATCTGTGAAAAGAGGTATCACAGCAGCAGTTCTCAACTTTTTTCCCAGC
AGTTGCACTGGAAATTTTCAATTACAAAATATGCTGTTAGCAATATCCTTGG
CACTTTAGAAATTAATTATAATAATTAG**TGATT**CTTAACTATGTGCCAAGAA
TTCTGCTAATACCCCTTGTATTTAGGTATTGTCTCATTTTGTCCTGTCCCTAA
ACCCTATAAAGACATGTACTATCCATTGTTCCTATTTACCAAGGAAATGAAC
ACATAGTGAGTTCACACAGCTAGCAGAAGGTAGAGCTGGGTTTGAGTCCTG
GCAGGCAGACTACATACTAGTGCTTTAACTAAAACAAGGACCAGGAATGTT
CGTGTGAGCCAAGCTCCTCTGG**TGATT**CTCAAAGAGCCTCACTGGCGATGG
TCTGCCAATACAGAGTACACCTGGAAAACACTAAAGCAGCCTTCAAAATGC
CAAGCAGCTG

Fragment "XL17"

ACCTGGGAACGTGTCTGAGTACAGACTCTCAGGGCCCCAATGCAGATTTAC
TGAAAC**TGATA**CTCCAAGCAGGGCCCAGCAATCTGTGCTGTAATAGGCCCT
CCAGG**TGACT**CTCATGCCCGCTTGAATTTGAGAACCACTCAGCTATTCAGTG
CAAGTGAATTTGCCCCCAACCAATTCTCCAAATGGCCAATTTACTGCAAATC
AGTATGTCAAATGAAAATTTGAAATTAAAAAGACAAAAACT**TGACA**AATTA
AAACAAATGCATTTTAAAAGGAAAAAAAAATACAACCCCATGCTAATTTGT
GCACCTTTAATCTT**TGACT**TAAGAAATATAAAGAGAATTCAATTTTGTTTAC
TGATAATGAGAGAAGTTTTGTGGGAACTGTTCCCTGAGATTCTGCAGTTTAG
CTAACTTTATAGTCATAAAAACAAAACCAAAAAAACCTCACTGTGCTATGT
GGTATTTAGTTCATGCTGAATTTTTTTAATTTGAAAAATG**TGATT**TTTTTTTT
ACTTTTGCCAATGTATCAATATTCATCTTTTAGCAATTTTATTTATTTATTTA
AATTTTATCAATGCCTAAGGAATGTACATATTTCAGGGGTACATA**TGATA**AT
TTAATATAGTCTTAATTTTTAATTGCCAAATCAGTGTATTTGAGATATCTATC
ACCTTAAAATATCTTTTCTTTATGCTTAAAACATTCAAATTATTCTCTTCGAG
CGATTTTCTTTTCTTTCTTTGTTTTTTGTTTGTTTGTTTGTTTGTTTTTTTG
AGATGGAGTCTTACTCTGTCACCCAGGCTGGAGTGCAGTGGCACAATCTCA
GCTCACTGCAACCTTTGCCTCCCAGGTTCAAG**TGATT**CTCCTACCTCAGCTT
CCCGAGTAGCTGGGACTATAGGTGCCCGCCACCACTCCTGGCTAATTTTTGT
ATTTTTAGTAGAGATGGGGTTTCACCATATTGGCCAGGCTGGTCTTGAACTC
C**TGACC**TTGT

Reverse complement

ACAAGGTCAGGAGTTCAAGACCAGCCTGGCCAATATGGTGAAACCCCATCT
CTACTAAAAATACAAAAATTAGCCAGGAGTGGTGGCGGGCACCTATAGTCC
CAGCTACTCGGGAAGCTGAGGTAGGAGAATCACTTGAACCTGGGAGGCAAA
GGTTGCAGTGAGCTGAGATTGTGCCACTGCACTCCAGCCTGGG**TGACA**GAG
TAAGACTCCATCTCAAAAAACAAACAAACAAACAAACAAACAAAAAACAA
AGAAAGAAAAGAAAATCGCTCGAAGAGAATAATTTGAATGTTTTAAGCATA
AAGAAAAGATATTTTAAGG**TGATA**GATATCTCAAATACAC**TGATT**TGGCAA
TTAAAAATTAAGACTATATTAAATTATCATATGTACCCCTGAAATATGTACA
TTCCTTAGGCAT**TGATA**AAATTTAAATAAATAAATAAAATTGCTAAAAGAT
GAATAT**TGATA**CATTGGCAAAAGTAAAAAAAAAATCACATTTTTCAAATTA
AAAAAATTCAGCATGAACTAAATACCACATAGCACAGTGAGGTTTTTTTGG
TTTTGTTTTTA**TGACT**ATAAAGTTAGCTAAACTGCAGAATCTCAGGGAACAG
TTCCCACAAAACTTCTCTCATTATCAGTAAACAAAATTGAATTCTCTTTATAT
TTCTTAAGTCAAAGATTAAAGGTGCACAAATTAGCATGGGGTTGTATTTTTT
TTTCCTTTTAAAATGCATTTGTTTTAATTTGTCAAGTTTTTGTCTTTTTAATTT
CAAATTTTCATT**TGACA**TAC**TGATT**TGCAGTAAATTGGCCATTTGGAGAATT
GGTTGGGGGCAAATTCACTTGCACTGAATAGCTGAGTGGTTCTCAAATTCAA
GCGGGCATGAGAGTCACCTGGAGGGCCTATTACAGCACAGATTGCTGGGCC

CTGCTTGGAGTATCAGTTTCAGTAAATCTGCATTGGGGCCCTGAGAGTCTGT
ACTCAGACACGTTCCCAGGT

Fragment "2"

TGCCCGCCTCAGCCTCCCAAAGTGCTGGGACTACAGGCGTGAGCCATTGCG
CCTGTTCTTACTACACAGACTCACCAGGCCTCCCTGAACCCATTTGTTTAAT
CATCTAAAAGACAAAATCCACTTTTTTTTTTTTTTTTGAGACAGAGTCTTGG
TCTGTTGCCCAGGCTAGAGTGCAATGGTGCCATCTCAGCTCACTGCAACCTC
CACCTCCCGGGTTCCAGCAATTCTCCTGCCTCAGTCTCTCGAGTAGCTGGAA
CTACAGGTGTCTGCCACCATGCCCAGCTAATTTTTGTATTTTTAGTAGAGAT
GGGGTTTCATCATATTGGCCAGGCTGGTCCCGAACTCC**TGACC**TCAT

Reverse complement

ATGAGGTCAGGAGTTCGGGACCAGCCTGGCCAATATGATGAAACCCCATCT
CTACTAAAAATACAAAAATTAGCTGGGCATGGTGGCAGACACCTGTAGTTC
CAGCTACTCGAGAGACTGAGGCAGGAGAATTGCTGGAACCCGGGAGGTGG
AGGTTGCAGTGAGCTGAGATGGCACCATTGCACTCTAGCCTGGGCAACAGA
CCAAGACTCTGTCTCAAAAAAAAAAAAAAAAAGTGGATTTTGTCTTTTAGA
TGATTAAACAAATGGGTTCAGGGAGGCCTGGTGAGTCTGTGTAGTAAGAAC
AGGCGCAATGGCTCACGCCTGTAGTCCCAGCACTTTGGGAGGCTGAGGCGG
GCA

Fragment "3"

TGCCCTCCTTGGCCTCCCAAGGTGCTGGGATTACAGGCATGAGCCACTGTGC
CCAGCCAGCACATGCCTTTTTAATTGCGTTCTTGAGTTTTGGGTTTGCAATA
ACTATTCAATTATTCTCTTCACCTCAAACCTACTTAA**TGATA**TAATGTCTTTA
AAATAATAGCCTAACTTATGTTGTTCTTCTTCTTCTTTTTTTTTTTTTTTTTT**T**
GATACAGAGTTTTGTTCTTGTTGCCCAGGCTGGAGTGCAATGGCCGATCTCA
GCACACTGCAACCTCCACCTCCCGGGTTCAAG**TGATT**CTCCTGCCTCAGCCT
CCTGAATAGCTGGGATTACAGGCATGTGCTACCACGCCTGGCTAATTTTGTA
TTTTTATTAGAGATGAGGTTTCTCCACGTTGGTCAGGCTGGTCTCAAACTCT
TGACCTCAGGT

Reverse complement

ACCTGAGGTCAAGAGTTTGAGACCAGCC**TGACC**AACGTGGAGAAACCTCAT
CTCTAATAAAAATACAAAATTAGCCAGGCGTGGTAGCACATGCCTGTAATC
CCAGCTATTCAGGAGGCTGAGGCAGGAGAATCACTTGAACCCGGGAGGTGG
AGGTTGCAGTGTGCTGAGATCGGCCATTGCACTCCAGCCTGGGCAACAAGA
ACAAAACTCTGTATCAAAAAAAAAAAAAAAAAAAAGAAGAAGAAGAACAA
CATAAGTTAGGCTATTATTTTAAAGACATTATATCATTAAGTAGGTTTGAGG
TGAAGAGAATAATTGAATAGTTATTGCAAACCCAAAACTCAAGAACGCAAT

TAAAAAGGCATGTGCTGGCTGGGCACAGTGGCTCATGCCTGTAATCCCAGC
ACCTTGGGAGGCCAAGGAGGGCA

Fragment “6”

TGCATCCACCCTGGGCCATCTCCCGGCTCCTCAGGTCCTCCTGCTACCACTC
ACCAGATTCAGGTTTGGCTTGAGTTATCGATATTAGGGTCCTACTAGGATTC
TCTAAGTTTATATCTAAGTTCAGCCTCTCTTCAAACTCTCC**TGATT**CCTTTTG
CCTCTGTGAAAGAATAGGTAATTTTGTTTAGCCTACCCATATGAGTCAACTT
GTACCAGACAAAACAACAGTTTTCTGCTCTGGGCTGGTCAGGAATTTTTTTT
TTTTTTAACTACAACTTAGATACAATCAAGTGCGCAAATTTTTTTTTTTGAG
ACGGARTCTCACTCTGTCGCCCAGGCTGGAGTGCAGTGGCGT

Reverse complement

ACGCCACTGCACTCCAGCCTGGGCGACAGAGTGAGAYTCCGTCTCAAAAAA
AAAAAATTTGCGCACT**TGATT**GTATCTAAGTTGTAGTTAAAAAAAAAAAAA
ATTCC**TGACC**AGCCCAGAGCAGAAAACTGTTGTTTTGTCTGGTACAAGT**TG**
ACTCATATGGGTAGGCTAAACAAAATTACCTATTCTTTCACAGAGGCAAAA
GGAATCAGGAGAGTTTGAAGAGAGGCTGAACTTAGATATAAACTTAGAGAA
TCCTAGTAGGACCCTAATATCGATAACTCAAGCCAAACCTGAATCTGGTGA
GTGGTAGCAGGAGGACCTGAGGAGCCGGGAGATGGCCCAGGGTGGATGCA

Fragment “8”

GATCACCTGAGGTAAAGGAGTCTGAGCCAGCC**TGACC**AATATAGTGAAACC
CCATCTCTACCAAAAATACAAAAATTCACCAGGTATGGTGGCTGGTTTCTGT
AGTCCCAGCTACTTGGGAGGCTGAGACAGGAGAATTGGTTGAACCTGGGAG
GCGGAGGGTGCAG**TGACC**CAAGATGGCTCCACTGCACTCTGAGACTCTGTC
TCAAAAAACAAACAAACAAAAAAAAGAACCTTATATCCAATGGATAGGAA
AGAGAGGAAAGCAAACAAGTTAGCCAACAGTCCTGGTGTGGTGGTAGCAC
ATTCTCTACCATGTGCTAGAGAGGCCTAGTGGACTTGTTGCAAGAAGAGGA
GGGACAGAGAGAGG**TGACA**GAGAGGAAGGCACCCAAGCCAGACTGGGGG
ATTGAAAAGGGTGTCTCAGCTGCTCCA

Reverse complement

TGGAGCAGCTGAGACACCCTTTTCAATCCCCCAGTCTGGCTTGGGTGCCTTC
CTCTCTGTCACCTCTCTCTGTCCCTCCTCTTCTTGCAACAAGTCCACTAGGCC
TCTCTAGCACATGGTAGAGAATGTGCTACCACCACACCAGGACTGTTGGCT
AACTTGTTTGCTTTCCTCTCTTTCCTATCCATTGGATATAAGGTTCTTTTTTTT
GTTTGTTTGTTTTTTGAGACAGAGTCTCAGAGTGCAGTGGAGCCATCTTGGG
TCACTGCACCCTCCGCCTCCCAGGTTCAACCAATTCTCCTGTCTCAGCCTCCC
AAGTAGCTGGGACTACAGAAACCAGCCACCATACCTGGTGAATTTTTGTATT

TTTGGTAGAGATGGGGTTTCACTATATTGGTCAGGCTGGCTCAGACTCCTTT
ACCTCAGG**TGATC**

Fragment "10"

GTGCCACTGCACTCCAGCCTGGG**TGACA**GAGCAAGACTCTGTCTCAAAAAA
AAAAAAAAAAAAAAGAAAAGAAAAGCTGGTCGGTACACATCCAGCCATGC
TCAGATACAGGGCATCACGCTGCTCTATCCTTGAAGTTAGGAC**TGACC**AGC
AGGAAGTGGGGG**TGACA**AGCTTCTAAGCTATTCTTTCTGAACTCATAGTCC
CCAGCCCAGGGCCAGAGCGACAGCCCCATTGTGCTGCTCCGAGATAAGCAT
ACCCTTCAAAAAACTCTCACTGCTTTGGGCTTGGATCGCAAGCCAGAGACC
ATCCAGCTCATCACCCGGGACATGGTCCGAGAACTCATGTGAGTACACTGT
CCCCTGCCCTTCTATGTCCAGTCCTCATCCTTTTGCTAAG

Reverse complement

CTTAGCAAAAGGATGAGGACTGGACATAGAAGGGCAGGGGACAGTGTACT
CACATGAGTTCTCGGACCATGTCCCGGGTGATGAGCTGGATGGTCTCTGGCT
TGCGATCCAAGCCCAAAGCAGTGAGAGTTTTTTGAAGGGTATGCTTATCTCG
GAGCAGCACAATGGGGCTGTCGCTCTGGCCCTGGGCTGGGGACTATGAGTT
CAGAAAGAATAGCTTAGAAGCTTGTCACCCCCACTTCCTGCTGGTCAGTCCT
AACTTCAAGGATAGAGCAGCGTGATGCCCTGTATCTGAGCATGGCTGGATG
TGTACCGACCAGCTTTTCTTTTCTTTTTTTTTTTTTTTTTTTGAGACAGAGTC
TTGCTCTGTCACCCAGGCTGGAGTGCAGTGGCAC

Fragment "19"

GCGGCGCACTGCAACCTCCGCCTCCCCAGCTTCAAG**TGATT**CTCCTGCCTCA
GCCTCCCAAGTAGCTGGGACTACAGGCATGTGCCACCATGCCCAGCTAATTT
TCCTATTTTTTTTTTTTTTATTTTTTTTAGCAGAGTTGGGGCTTCACCACTTTG
GCCAGGG**TGATC**TCATGAGACTTATTCACTATCATGAGAACAGCATGGGAA
AGACTTGCCTCCATCATTCAATTACCTCCCACCGGGTCCCTCCCACAACACG
TGGGAATTAAAGATGAGATTTTGTTGGGGACACAGCCAAACCATATCACCA
TACCTCTCTAATCTTCAATAACAGGTCCCATGTAACCACTAAACCTTGAAGT
TCAGCACTGCTGCCCAGGTTGGAGTGCAGTGGCGCAATCTTGGCTCACTGCA
ACCTCTGCCTCCCAGGTTCAAG**TGATT**CTTGTACCTCAGCATCCTGAGTAGG
AGGCTGGGACTACAGGCACACGCCACCA**TGACT**GGCTAATTATTATTTTTCT
TGCATTTCTTTTTTTTTTGGTAGAGACGGGGTTTCACCATGTCAGCCAGGCTG
GTCTCAAACTCCTGGCCTCAAGT

Reverse complement

ACTTGAGGCCAGGAGTTTGAGACCAGCCTGGC**TGACA**TGGTGAAACCCCGT
CTCTACCAAAAAAAAAAGAAATGCAAGAAAAATAATAATTAGCCAGTCATG
GTGGCGTGTGCCTGTAGTCCCAGCCTCCTACTCAGGATGCTGAGGTACAAG

AATCACTTGAACCTGGGAGGCAGAGGTTGCAGTGAGCCAAGATTGCGCCAC
TGCACTCCAACCTGGGCAGCAGTGCTGAACTTCAAGGTTTAGTGGTTACATG
GGACCTGTTATTGAAGATTAGAGAGGTATGG**TGATA**TGGTTTGGCTGTGTC
CCCAACAAAATCTCATCTTTAATTCCCACGTGTTGTGGGAGGGACCCGGTGG
GAGGTAATTGAATGATGGAGGCAAGTCTTTCCCATGCTGTTCTCA**TGATA**GT
GAATAAGTCTCATGAGATCACCCTGGCCAAAGTGGTGAAGCCCCAACTCTG
CTAAAAAAAATAAAAAAAAAAAAAATAGGAAAATTAGCTGGGCATGGTGG
CACATGCCTGTAGTCCCAGCTACTTGGGAGGCTGAGGCAGGAGAATCACTT
GAAGCTGGGGAGGCGGAGGTTGCAGTGCGCCGC

Fragment “23”

TGTGAAGTTTCCCATCACCGTGGCAGTCCACAATGTGAAGCCCCTGAGGCC
ACCCTTCTCCATGTGTCCCCCACACTGTCCC**TGACC**CTCCCACCATCCTAGA
AAACCAGCCTCCCAGGTCCTCTGTCATGGAAGTAATTTACAGCAAGGCAGA
AAGAGCTGAAAATTTATAAATGTGTAAAATTTACAAAGGTGTTTCTTTCTTT
CTTTCTTTCTTTTTTTTTTTTTTTTTTTGAGGCAGAGTCTCACTCTGTTGCCC
AGGCTGGAGTGCAGTGGCGT

Reverse complement

ACGCCACTGCACTCCAGCCTGGGCAACAGAGTGAGACTCTGCCTCAAAAAA
AAAAAAAAAAAAAAGAAAGAAAGAAAGAAAGAAACACCTTTGTAAATTTT
ACACATTTATAAATTTTCAGCTCTTTCTGCCTTGCTGTAAATTACTTCCA**TGA**
CAGAGGACCTGGGAGGCTGGTTTTCTAGGATGGTGGGAGGGTCAGGGACAG
TGTGGGGGACACATGGAGAAGGGTGGCCTCAGGGGCTTCACATTGTGGACT
GCCACGGTGATGGGAAACTTCACA

Fragment “25”

GCACCACTGCACTCCAGCCTGGGCAACAGAGCAAGACTCCATCTCAAAAAA
AAAAAAAAAAAAAAAAAGGAAGATATGAGCACTCCCTTATATTTTGTACCC
TTCTCATCCTTCCACTTTAAGGTTG**TGATA**TTCTGGTGTTTTGGCCCTGTCAG
TTTTCTTATCCTCCCTCCCATTTCAACCCATGTATCTTC**TGATA**ATCTGTGTT
CCTTTTTTTTTGTTTTTTTGTTTTTTTGCTTACCTACTTCTGTGGCCACCACTT
GTACCCGATATGGACTTCCGGCTTCTCTGTCCAATGGAGCCACACTAAGG

Reverse complement

CCTTAGTGTGGCTCCATTGGACAGAGAAGCCGGAAGTCCATATCGGGTACA
AGTGGTGGCCACAGAAGTAGGTAAGCAAAAAAACAAAAAAACAAAAAAAA
AGGAACACAGATTATCAGAAGATACATGGGTTGAAATGGGAGGGAGGATA
AGAAAAC**TGACA**GGGCCAAAACACCAGAATATCACAACCTTAAAGTGGAA
GGATGAGAAGGGTACAAAATATAAGGGAGTGCTCATATCTTCCTTTTTTTTT

TTTTTTTTTTTTTGAGATGGAGTCTTGCTCTGTTGCCCAGGCTGGAGTGCAG
TGGTGC

Fragment “5B”

ATCCCTCCTCTCCCCTACTCCCCTAC**TGACC**ATCCCTTCTACCTGCTTTATTT
TTCTCCATAGCAT**TGACT**ATCTTCAATGTATTATACTGTATAAAATCTTTGC
ATATATTTTGTTTATTGCCTTCTTCTCTCCA**TGACA**TGTAAGCAACACAAAG
GCTAGAGGTTTTTTGCTTGTTTGTTTTATTTTGTTTTGAAGACAGGGTCTCAC
TCTGTCACCAGGCTGGAGTTGCAGTGGTGAAATCTCAGCTCATTGCAACCTC
CACCTCCTGGGTTCAAGCAATTCTTGTGCCTCAGCCTCCTGAGTAGCTGGGA
CTACAGGCTTGTACCACCACGCATGGCTAATTTTTTTTTGTAGTTTTAGTTGA
GACTGGGTTTCACCATGTTGGCCAGTCTGGTCTCAAACTCCTGGCTTCAAG**T**
GACCTGCCAGCCTTGGCCTGCCAGAATGCTG**TGATT**ACAGGTATGAGCCAC
CATGCCCAGCCAAGGGCTAGAGTTTTTGTACATTTTGCTG

Reverse complement

CAGCAAAATGTACAAAAACTCTAGCCCTTGGCTGGGCATGGTGGCTCATAC
CTGTAATCACAGCATTCTGGCAGGCCAAGGCTGGCAGGTCACTTGAAGCCA
GGAGTTTGAGACCAGACTGGCCAACATGGTGAAACCCAGTCTCAACTAAAA
CTACAAAAAAAAATTAGCCATGCGTGGTGGTACAAGCCTGTAGTCCCAGCT
ACTCAGGAGGCTGAGGCACAAGAATTGCTTGAACCCAGGAGGTGGAGGTTG
CAATGAGCTGAGATTTCACCACTGCAACTCCAGCCTGG**TGACA**GAGTGAGA
CCCTGTCTTCAAAACAAAATAAAACAAACAAGCAAAAAACCTCTAGCCTTT
GTGTTGCTTACATGTCATGGAGAGAAGAAGGCAATAAACAAAATATAT

Fragment “19B”

GAGCCACCCTGGCAGACGCGTCAGAAAGAGCTTTGTAGGCAGACAAGGGCT
TGGCTTATTCCAGGACTGGAAAGGAGGCCAGTGAGGCTGGAATATTGTGAA
CAGGGGAGAAGA**TGATA**CCTGATGTGGTTGGAGAGGTGGACAGGGGCCAG
A**TGATA**CAGCACTTTGTGGCTGTGGCAGGGAGTTTTATTCTAAGAATGATG
GAAAAAAGAGCTAATAGAGGGTTTAACCAGGGAGTGATGTATACCAGATCA
GTCTGAAGTCTAAGGGGGCTCTTGTTACGTCACACTGCCTTAAAGAAATTCT
TTTTTTTTTTTTTTTTTTTTTTTGAGACGGAGTCTCACTCTGTCGCCAGGCTG
GAGTGCAGTGGTGT

Reverse complement

ACACCACTGCACTCCAGCCTGGCGACAGAGTGAGACTCCGTCTCAAAAAAA
AAAAAAAAAAAAAAAAAAAGAATTTCTTTAAGGCAGTGTGACGTAACAAGA
GCCCCCTTAGACTTCAGAC**TGATC**TGGTATACATCACTCCCTGGTTAAACCC
TCTATTAGCTCTTTTTTCCATCATTCTTAGAATAAAACTCCCTGCCACAGCCA
CAAAGTGCTGTATCATCTGGCCCCTGTCCACCTCTCCAACCACATCAGGTAT

CATCTTCTCCCCTGTTCACAATATTCCAGCCTCACTGGCCTCCTTTCCAGTCC
TGGAATAAGCCAAGCCCTTGTCTGCCTACAAAGCTCTTTCTGACGCGTCTGC
CAGGGTGGCTC

Fragment “37B”

CACCCGCCTTGGCCTCCCAAAGTGCTGGGCGTACAGGCGTGAGCCACCACG
CCCAGCCTAACATGTCCATTTTAAGGAATAGGTAAAACAGTTTGCAGTTTAC
TTAGTCATAATGTGGAATAAAAATGTGTGCCCTGGGGAATTAAACAAACAA
ACATAGTGTGGAAATGTGTCCAAACTGTCTGTTACAGT**TGACA**GAAAGTGA
ATTGGGTCATTTGGGATTCCTGGCCCTGCCGGGCCAGAGAACAGCCCTCCCC
TGCCCACTGGCGGTGGCCCC**TGACA**GCCATGTCTCCCTCTCTTTGCTCTAAG
CCCCCAAACCATGGCAGAGGCCCC**TGACA**GCCGTGTCTCTCTCTCTGCTCTA
AGCCCCCAAACCACAAAGGGCAGAGGCTGGGTGCCAGAGGCTCTG**TGACA**
GCCCCACTCTGTTTTTGAACTTTCTGCTTCCCTTCCACTCGTTACAGGAGGAA
GATGTT

Reverse complement

AACATCTTCCTCCTGTAACGAGTGGAAGGGAAGCAGAAAGTTCAAAAACAG
AGTGGGGCTGTCACAGAGCCTCTGGCACCCAGCCTCTGCCCTTTGTGGTTTG
GGGGCTTAGAGCAGAGAGAGAGACACGGCTGTCAGGGGCCTCTGCCATGGT
TTGGGGGCTTAGAGCAAAGAGAGGGAGACATGGCTGTCAGGGGCCACCGC
CAGTGGGCAGGGGAGGGCTGTTCTCTGGCCCGGCAGGGCCAGGAATCCCAA
A**TGACC**CAATTCACTTTCTGTCAACTGTAACAGACAGTTTGGACACATTTCC
ACACTATGTTTGTTTGTTTAATTCCCCAGGGCACACATTTTTATTCCACATTA
TGACTAAGTAAACTGCAAACTGTTTTACCTATTCCTTAAAATGGACATGTTA
GGCTGGGCGTGGTGGCTCACGCCTGTACGCCCAGCACTTTGGGAGGCCAAG
GCGGGTG

Fragment “45B”

ATGGGAATGGCTGTCAATTCTTTTTTTTTTTTTTTTTTTTGAGACAGAGTA
TTGCTCTTGTTGCCCAGGCTGGACTGCAGTGGTGCAATCTCGGCTCACTGCA
ACCTCAACCTCTGGGTCTCAAGCTATCCTCCCACTTCAGTCTCCCGAGTAGC
TGGGGCTACAGGCATGCACCACCATGCCCAGCTAATTTTTTAATTTTTTGTA
GAGATAGGGTTTCCCTTGTTGCCCAGACTGGTCTCAAACTCTTGGGCTCAAG
CAGTCCTCTTGCCTCAGACTCCCAAAGTGCTGGGATTAGAGGCACGAACTA
CAGCGCCTGGTCAA**TGAC**TGATATATTTAGATTGAAAGCTAGCATCTTACT
TGGTTTGTACTTATCCAACCTCCCTGTGTTTCTTTTTCTCTCCTTTCTTGCCTT
TTTCTGGAATAATCACGTTTTTTTTTTTTTTTTTTTGAGATTTCATTTCCCCG
CTCTATTAGCTTGTCATTGTTACATTATTTTACAGTTCTAA**TGATC**GCTCTAG
TAATTACAACATGCATCCT**TGATT**TATTAGCATC**TGATA**TAAATAATTACAT

TGACCACTTTCTTTCTTTTTCTTTTTTTTTTGGAGATGGAGTCT**TGATC**TGTT
GCCAGGCTGGAGTGCAGTGGTGC

Reverse complement

GCACCACTGCACTCCAGCCTGGCAACAGATCAAGACTCCATCTCCAAAAAA
AAAAGAAAAAGAAAGAAAGTGGTCAATGTAATTATTTATATCAGATGCTAA
TAAATCAAGGATGCATGTTGTAATTACTAGAGCGATCATTAGAACTGTAAA
ATAATGTAACAA**TGACA**AGCTAATAGAGCGGGGAAATGAAATCTCAAAAA
AAAAAAAAAAAAAAAACG**TGATT**ATTCCAGAAAAAGGCAAGAAAGGAGAG
AAAAAGAAACACAGGGAGGTTGGATAAGTACAAACCAAGTAAGATGCTAG
CTTTCAATCTAAATATATCAGTCAT**TGACC**AGGCGCTGTAGTTCGTGCCTCT
AATCCCAGCACTTTGGGAGTCTGAGGCAAGAGGACTGCTTGAGCCCAAGAG
TTTGAGACCAGTCTGGGCAACAAGGGAAACCCTATCTCTACAAAAAATTAA
AAAATTAGCTGGGCATGGTGGTGCATGCCTGTAGCCCCAGCTACTCGGGAG
ACTGAAGTGGGAGGATAGCTTGAGACCCAGAGGTTGAGGTTGCAGTGAGCC
GAGATTGCACCACTGCAGTCCAGCCTGGGCAACAAGAGCAATACTCTGTCT
CAAAAAAAAAAAAAAAAAAAAAAGAAT**TGACA**GCCATTCCCAT

Acknowledgements

It is probably not possible to mention here all the persons whom I have to thank. The first should be Prof. J. Adamski and Prof. H.J. Kolb, who gave me the possibility to work on my PhD in Munich. I would like to thank them both for the help, in and outside the lab, and for the formative and stimulating discussions that we have had.
Of course the acknowledgements have to be extended to all the present and former members of their groups, in particular to Gabi, Rebekka, Raymund, Dagmar, and Marc. It would have been for me not possible to achieve this goal without their precious help both in my scientific and in my personal maturation.
I owe many thanks to the TRANSNET consortium, with all its participants. Special thanks to Prof. Anne M. Dickinson, who coordinated the consortium in the best possible way. It is not possible to name here all the people who deserve my gratitude for the fantastic time I had during our meetings and our late-night discussions. To all the TRANSNET fellows: thank you.
Of course, I wouldn't have even thought about doing a PhD without the support of the people in my former lab in Rome, among which I would specially like to thank Ombretta, and of my family, Mariano, Gabriella, Eduardo, Gabriele, and Elena. They deserve probably the biggest "thank you".
I have also to thank all the people, in and outside the lab, who have helped me to get integrated in this city and in this country. I cannot name all of them here, but I would like to thank particularly my cousin Francesca, Alessandra, Antonio, Damiano, Edoardo, Federico, Silvia, and Victor.
I am very thankful to my friends in Italy, Antonello, Claudio, Francesca, Pasquale, Raffaele, who were always close to me, although being far away. The short time we spent together during your visits is unforgettable.
I am enormously grateful to Olga, who was always present with her help and support. Thanks to her I never felt alone when I needed somebody.
My last thanks go to all the people who have helped me becoming who and what I am now, but unfortunately are not anymore here.
Grazie a tutti voi.