UNIVERSITÀ DEGLI STUDI DI TRIESTE

XXXI CICLO DEL DOTTORATO DI RICERCA IN

BIOMEDICINA MOLECOLARE

APPLICATION OF A NEW MOLECULAR APPROACH BASED ON NEXT GENERATION SEQUENCING TO MONITOR THE CHIMERISM POST ALLOGENEIC HSCT

Settore scientifico-disciplinare: **MED/04 Patologia Generale**

DOTTORANDO
Michelangelo Aloisio

COORDINATORE
PROF. Germana Meroni

SUPERVISORE DI TESI
PROF. Roberta Bulla

SUPERVISORE DI TESI
DOTT. Giovanni Maria Severini

ANNO ACCADEMICO 2017/2018

ABSTRACT

After the hematopoietic stem cell transplant, the chimerism analysis is routinely evaluated to monitor donor immune system engraftment. Today, the most common genetic diagnostic method for chimerism is the low sensitivity STR analysis by capillary electrophoresis. Recent studies have shown the application of the chimerism analysis in the risk of relapse prediction when high sensitivity technique is used and in the HLA-loss relapse recognition when genetic markers outside and inside the HLA locus are evaluated. As first objective of the PhD, a Next Generation Sequencing (NGS) approach to detect the chimerism was developed to meet high informativity and sensitivity. Based on the technology of Ion AmpliSeq, a 44-amplicon custom chimerism SNP panel was designed, and a bioinformatics pipeline dedicated to the genotyping and quantification of NGS data was coded. The SNP panel informativity was high: 16 informative recipient alleles in unrelated pairs and 9 in sibling ones. The method sensitivity was fixed at 1% due to the NGS background (<1%). The protocol followed the standard Ion AmpliSeq library preparation and Ion Torrent Personal Genome Machine guidelines. The NGS method sensitivity tested on several artificial chimeras and on 20 clinical samples, resulted higher than the STRs one. The NGS method passed the UKNEQAS external quality program validation, resulting in suitable for diagnostic use. To decrease the NGS background with a view to reach a sensitivity of 0.1%, an upgrade of the NGS method was proposed as second objective of the thesis. In addition, as third objective, with a view to recognize the HLA Loss relapse in all the transplant couples, 20 haplotype blocks were designed inside the HLA locus and 12 outside. For these purposes, the previous custom bioinformatics pipeline was implemented to recognize the NGS background. A cheaper custom NGS library protocol based on 100 ng of input DNA, was developed and successfully tested on four DNA samples. The panel informativity was higher for both outside and inside the HLA blocks. The NGS background detected (0,01%) was ten times lower than the desired sensitivity (0,1%). However, a full work on artificial chimera will be done to experimentally measure the sensitivity. Since alignment errors in some HLA blocks were found, the bioinformatics tools will be improved. Moreover, the method performance will be validated on clinical patient samples. After further investigations, our NGS haplotype block method could promptly recognize the graft rejection, the risk of relapse and the HLA-loss relapse together.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

A

ACCh	Ampliseq Custom Chimerism panel
ACCh_hb	Ampliseq Custom Chimerism panel haplo-block
aCh	artificial Chimera
aGvHD	acute Graft versus Host Disease
AIBT	Italian Association of the Immuno-genetic and Transplant Biology
ALL	Acute Lymphoblastic Leukaemia
ALL_p	All the 1000genomes database populations
allo-HSCT	allogeneic Hematopoietic Stem Cells Transplant
AMELX	X variant of the amelogenin gene
AMELY	Y variant of the amelogenin gene
AMGX	X genetic marker
AMGY	Y genetic marker
AML	Acute Myeloid Leukemia
APC	Antigen Presenting Cells
ASO	Allele Specific Oligonucleotide
auto-HSCT	autologous Hematopoietic Stem Cells Transplant

B

BM	Bone Marrow
Bp	Base Pair

C

CA	Chimerism Analysis
CAR-T	Chimeric Antigen Receptor-modified T Cells
CB	Cord Blood
CC	Complete Chimerism
CD	Cluster Designation
cGvHD	chronic Graft versus Host Disease
Chim	Chimera sample
CI	Congenic Immunodeficiency
CIBMTR	Centre for International Blood and Marrow Transplant Research
CMV	Cytomegalovirus
Ct	Cycle of threshold

D

DLI	Donor Lymphocyte Infusion
Don	Donor sample

E

EAS	East Asian 1000genomes population
EBMT	European Society for Blood and Marrow Transplantation

EQA	external quality Assessment
EUR	European 1000genomes population

F

FISH	Fluorescence In Situ Hybridization
FITC	fluorescein iso-thiocyanate

G

G-CSF	Granulocyte Colony Stimulating Factor
GITMO	Italian Group of Bone Marrow Transplantation
GvHD	Graft versus Host Disease
GVL	Graft Versus Leukemia

H

Haplo-block	Haplotype Block Marker
HLA	Human Leukocyte Antigen
HSC	Hematopoietic Stem Cells
HSCT	Hematopoietic Stem Cells Transplant

I

IBMRD	Italian Bone Marrow Donor Registry
Ig	Immunoglobulin
IMC	Increased Mixed Chimerism
In/Del	Insertion Deletion polymorphism
IR	Immune Reconstruction
IRA	Informative Recipient Allele
IRR	Informative Recipient Read
ISP	Ione Sphere Particles

K

Kb	Kilobase
kDA	kiloDalton

L

LD	Limit of Detection

M

MAC	Myeloablative Conditioning
MAF	Minor Allelic Frequency
MC	Mixed Chimerism
MDS	Myelodysplastic Syndrome
MHC	Major Histocompatibility Complex
MRD	Minimal Residual Disease
mSNP	main SNP
MMUD	HLA mismatch unrelated donor
MUD	Match Unrelated Donor

N

nd	not defined
ng	nanogram
NGS	Next Generation Sequencing
NI	Not Informative
NK	Natural Killer Cells
NRM	Non Relapse Mortality

O

OS	Overall Survival

P

PB	Peripheral Blood
PBSC	Peripheral Blood Stem Cell
pCh	patient Chimera sample
PCR	Polymerase Chain Reaction
pD	patient Donor sample
PGM	Personal Genome Machine
Pool_HLA	pool primers HLA locus
Pool_N_HLA	pool primers non HLA locus
pR	patient Recipient sample

Q

qPCR	quantitative PCR

R

R%	Recipient percentage
R_0,1%	Recipient 0,1% DNA sample
R_0%	Recipient 0,0% DNA sample
R_1%	Recipient 1,0% DNA sample
R_10%	Recipient 10% DNA sample
R_100%	Recipient 100% DNA sample
Rec	Recipient sample
RIC	Reduced Intensity Conditioning
Rn value	Normalized report value

S

SAA	Severe Aplastic Anaemia
SBT	Sequence Based Typing
SCA	Sickle Cell Anaemia
SE	Standard Error
SNP	Single Nucleotide Polymorphism
SSOP	Sequence Specific Oligonucleotide priming
SSP	Sequence Specific Priming
St.Dev.	Standard Deviation
STR	Short Tandem Repeat
STRs-CE	STRs Capillary Electrophoresis

T

T cell	Lymphocytes T cell
Ta	Annealing Temperature
TD-PCR	Touch-Down PCR
Tm	Melting Temperature
TrP1	Truncate p1 adapter

W

WHO	World Health Organization

1. INTRODUCTION

1.1. Hematopoietic Stem Cells Transplant

The **Hematopoietic Stem Cells Transplant** (HSCT) represents a potential therapy for treatment of hematological neoplastic disorders or congenital metabolism defects. The Hematopoietic Stem Cells (HSCs) are characterized by a membrane glycoprotein $CD34^{+}$ (*Cluster Designation*) that it is the marker for their identification and selection. The HSCs are the precursor of all the mature blood cells in the differentiation process called Hematopoiesis (Fig 1) [1].

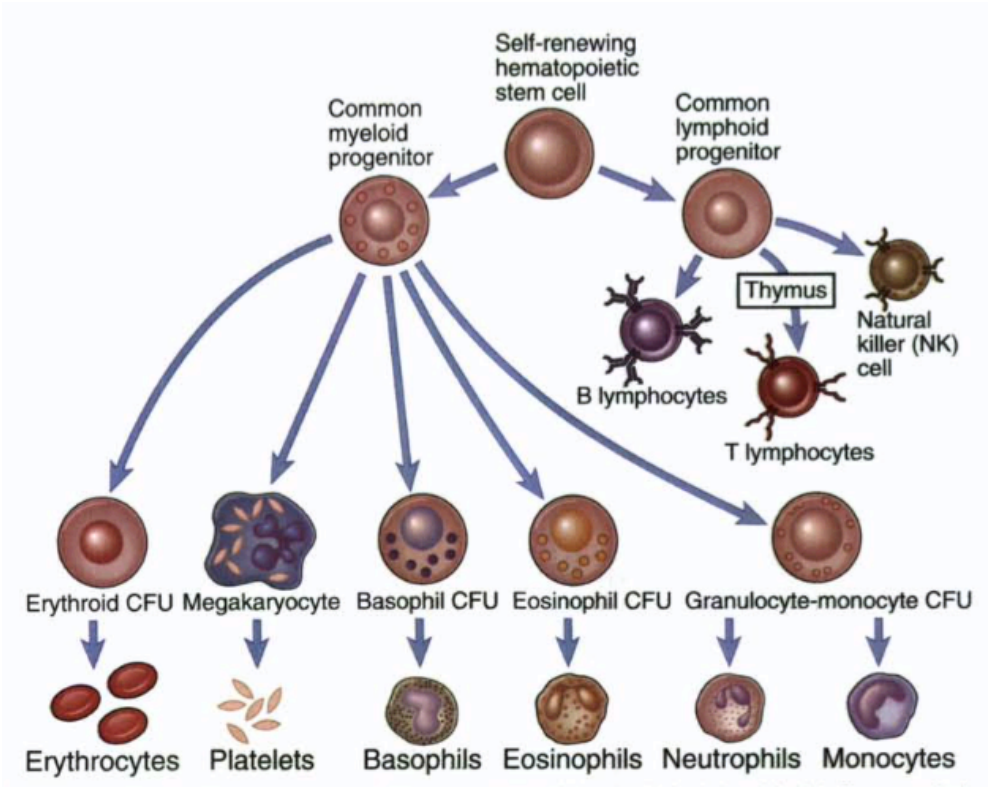

Figure 1. Hematopoiesis [1].

Two kinds of transplant are used in the clinical practice: **autologous** and **allogeneic**. In the **autologous HSCT (auto-HSCT**), the graft derives from the patients themselves and there are no problems with the adverse reaction caused by the

transplanted immune system. However, when the auto-HSCT is applied to malignant disorders, the disease relapse risk is higher [2]. Conversely, in the **allogeneic HSCT** (**allo-HSCT**), the CD-34^{+} are harvested from a healthy donor and infused in a different patient [1]. The focus of the thesis is on the allo-HSCT, therefore we will not further describe the auto-HSCT.

1.2. Allo-HSCT from the past to the present

In 1957, Donnall Thomas performed the first human allogeneic BMT [3]. At that time, the histocompatibility was largely unexplained and the donors were not selected on the basis of the compatibility. Indeed, the first transplanted patients died within 100 days of the transplant [4]. In 1972, after the discovery of the human Major Histocompatibility Complex (**MHC**), or Human Leucocyte Antigen (HLA)**,** Donnall Thomas performed the first successfully allogeneic Bone Marrow transplants [5, 6]. In the late 70's, it was remarked that the graft versus leukemia effect is the key of allo-HSCT success (see paragraph 1.3.4.) [7]. Since 1990's the number of allogeneic and autologous HSCTs increased worldwide (Fig 2) through the introduction in the clinical protocols of the Granulocyte Colony Stimulating Factor (G-CSF) (see paragraph 1.3.3.) the Reduced Intensity Conditioning (see paragraph 1.3.2.), and the HLA-molecular typing methods (see paragraph 1.3.1.2) [8, 9].

In the last ten years, the Non Relapse Mortality (NRM) that represents the death not caused by the disease relapse, has decreased. It has depended by the development of technical protocols to control of GvHD and so infectious complications [11]. As a result of the advances in the technologies and supportive patient care, the allo-HSCT has become safer and the number of allo-HSCT has increased continuously across the world. Indeed, the European Society for Blood and Marrow Transplantation (EBMT) has registered an increase of about 52% from 2006 to 2016, with more than 43000 HSCTs performed in 2016. Among these, about 17641 are allo-HSCTs, out of which 2647 are Haploidentical (see paragraph 1.3.1.2.) [12]. The "Italian Group of Bone Marrow Transplantation" (GITMO) registered a total of 4701 HSCTs performed in Italy during 2016. Among these, 1796 (38%) are allo-HSCTs and 573 are haploidentical [13]. Both in Europe [12] and in Italy [13] the majority of allo-HSCT is

performed for the treatment of patients affected by Acute Myeloid Leukemia (AML), acute Lymphoblastic Leukemia (ALL) and Myelodysplastic Syndrome (MDS).

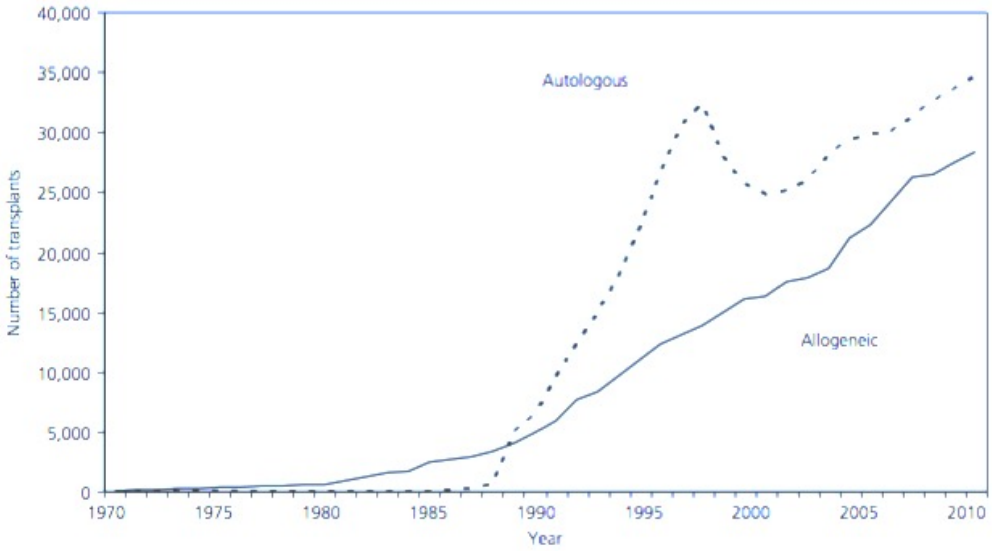

Figure 2. Number of Allogeneic and Autologous HSCT per year worldwide. Data from Centre for International Blood and Marrow Transplant Research (CIBMTR) [10].

Although an obvious progress in allo-HSCT treatments have led to an increased overall survival (OS), the relapse remains the most important cause of death after the first 100 days (Fig 3) [14, 15].

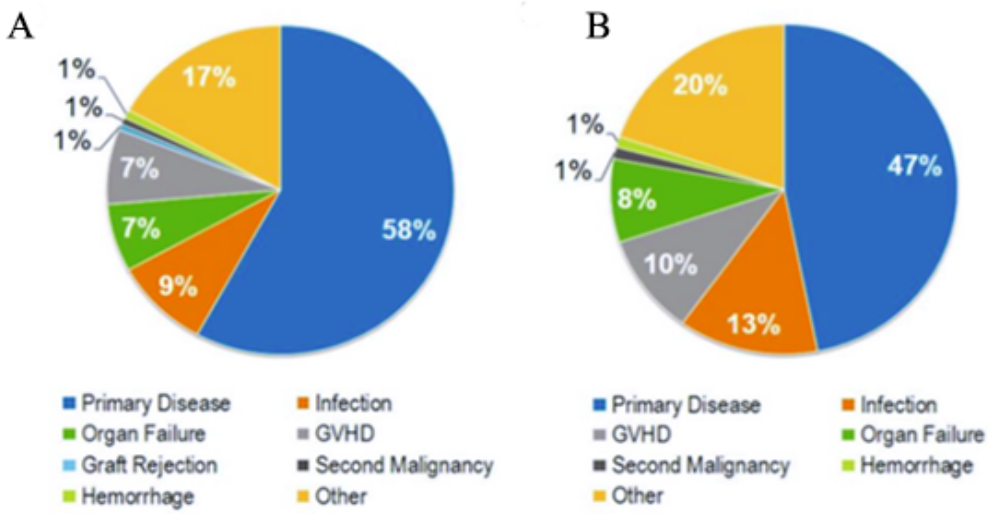

Figure 3. Cause of death at or beyond the 100 days after allo-HSCT (2014-2015) A) in HLA-matched sibling B) in Unrelated Donor [16].

1.3. HSCT procedure

The choice to perform a HSCT is made after evaluating of the individual patient condition. Since the HSCT is a therapy with high risks, the ratio between the risk and the benefit of the treatment must be evaluated before proceeding to the transplant. Therefore, different aspects must be taken into account, as patient overall health, disease status at the time of transplantation, prior therapies, age, comorbidity, degree of HLA-match, donor age and gender or other donor characteristics, including any donor health disorders [17, 18]. The allo-HSCT procedure consists of four phases:

- Identification of HLA compatible donor;
- Recipient conditioning;
- Collection and infusion of the CD-34^{+};
- Immune reconstruction.

1.3.1. Identification of HLA compatible donor

In the first phase of the transplant, the selection of the suitable donor is determined mainly by the donor-recipient HLA matching. The level of HLA matching is fundamental for the outcome of allo-HSCT. The risks associated with the HLA mismatches are the graft failure, Graft versus Host Disease (GvHD) (see paragraph 1.3.4.) and delayed immune reconstruction [19-21]. When more than one donor is available, the non-HLA factors may be considered during the donor selection. Among these, we have the donor age [22, 23], sex [24] and cytomegalovirus serostatus [23].
Before describing how to identify a suitable donor for the HSCT, considering the grade of HLA homology between donor and recipient and defining their possible combinations (see paragraph 1.3.1.2), it is crucial to introduce the immune-genetic aspects of the HLA.

1.3.1.1. HLA overview

The HLA genomic region spans about 4000 kilobases (Kb) and is located on the short arm of the chromosome 6 (6p21.3). This region contains high polymorphic genes (more than 220) clustered in three classes (Class III, Class II and Class I) (Fig 4).

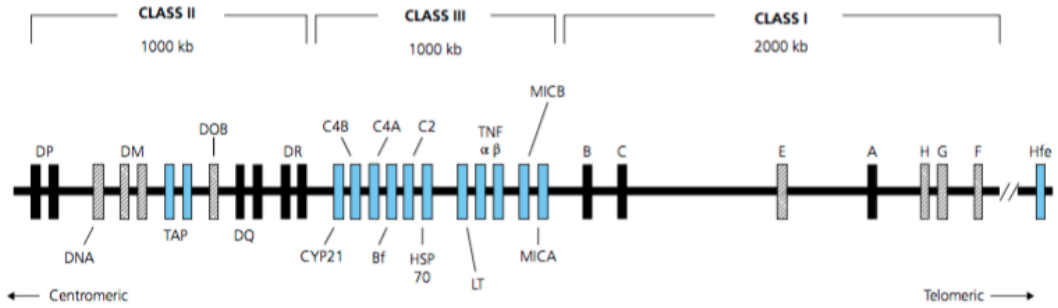

Figure 4 The HLA gene organization on short arm of the chr6. The classical class I genes: HLA-A, -B, -C and class II genes: HLA-DP, -DQ, -DR are shown in black [10].

The **class III** genes are involved in complement cascade. The **HLA class II and class I molecules** are cell surface glycoproteins that present the peptides (antigens) to the T Lymphocytes Cells (T cells). These molecules take part in the system (adaptive immunity) that controls the infections, tumors and the allo-reactivity.

The **allo-reactivity** is a process mediated by the immune system to distinguish the **self** antigens from the **non-self** ones. The T cells recognize and react against the HLA non-self antigens. This mechanism is behind transplant complications, such as **graft-rejection** and **GvHD**. There are two kinds of allo-reactivity mechanisms: direct an indirect. In the **direct** mechanism, the recipient T-cells recognize the self-peptides bound to donor (foreign) HLA antigens. Instead, in the **indirect** mechanism, the recipient T cells recognize the peptides derived from the donor HLA, presented by the self Antigen Presenting Cells **(APC)** [1].

The genes of the **HLA class I and II** are inherited according to the mendelian rules. These genes show a **codominant** expression levels, indeed each individual shows the alleles that are inherited from each of the two parents. A fundamental characteristic of the HLA region is the positive **Linkage Disequilibrium,** indeed two or more alleles

that belong to the same chromosome strand are inherited in a block **(haplotype).** If each HLA allele was inherited randomly, there would be more than 5,7 x 10^{31} unique HLA-A, -B, -C, -DRB1, -DQB1 phenotypes and then the identification of a matched unrelated donor would be insurmountable. There is a strong Linkage Disequilibrium between the locus HLA-B and -C and between the HLA-DR and –DQ, indeed, if the couple patient-donor matched for the HLA-A, -B and -DR then with a good probability they will match also for the –DQ and -C gene [10].

The **HLA class I molecules** are expressed on the membranes of virtually all the nucleated cells and they have the function of presenting the peptides derived from intracellular proteins to the cytotoxic CD8+ T cells. These molecules are composed of light β-chain (12 kDA) and a heavy α-chain (45 kDA) (Fig 5). The β-chain or β_2 micro-globulin, is expressed by a low polymorphic gene located outside the HLA (chromosome 15). The α-chain is composed by three Ig domains (α1, α2 and α3) and can be expressed by the gene HLA-A or -B or -C (Fig 5). These genes are composed of 8 exons, showing a high number of polymorphisms clustered in the hyper-variable regions (HvR) (exons 2 and 3) that encode for the peptides (antigen) binding groove (protein domains α1 and α2) [25]. Today, 4340 alleles are known for the HLA-A, 5212 for HLA-B and 3939 for HLA-C (release 3.33.0, 2018-07-11) [26].

The **HLA class II molecules** are expressed only on dendritic cells, B lymphocytes, macrophages and in other few cell types. They have the function of presenting the extracellular peptides to the helper CD4+ T cells and are composed of two chains with similar size: the α chain (34 kDa) and β chain (30 kDA) (Fig 5). The α-chains are composed of two Ig domains encoded by HLA -DRA, -DQA1 and -DPA1 genes [25]. These genes, composed of 5 exons each, do not show a high polymorphic level. Indeed today only 7 alleles are described for DRA, 95 for DQA1 and 67 for DPA1 (release 3.33.0, 2018-07-11) [26]. The β-chains are composed of two Ig domains encoded by HLA -DRB, -DQB1 and -DPB1 genes [25]. These genes are constituted of 5 exons each and present a high number of polymorphisms clustered in the exon 2 encoding for the hypervariable β1 domain. Today, 2593 alleles are described for DRB, 1257 for DQB1 and 1014 for DPB1 (release 3.33.0, 2018-07-11) [26].

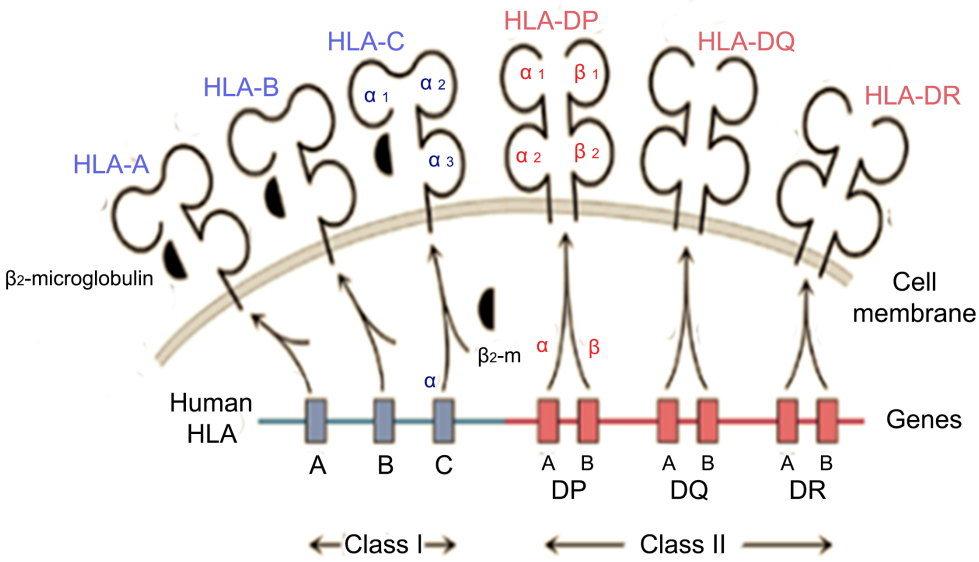

Figure 5. Schematic depiction of gene coding for the HLA class I (A, B, C) and class II (DP, DQ, DR) molecules. The β2-microglobulin encoded outside HLA.

1.3.1.2. HLA typing methods and donor selection

To select a suitable donor for the transplant, we need to establish the grade of homology between donor and recipient. To this end, different techniques for HLA typing have been developed during the years. The serological method was used until 1990s. Subsequently, after the discovery of the Polymerase Chain Reaction (PCR), the HLA typing was performed using **molecular methods.** Among these we have: the PCR Sequence Specific Priming (SSP), the PCR Sequence Specific Oligonucleotide probing (SSOP) and the Sanger Sequence Based Typing (SBT). These molecular methods are able to perform the HLA typing with a superior grade of resolution compared to the serological method. This major resolution has led to the discovery of new HLA alleles and for this reason it was necessary to develop a new nomenclature system [27] (Fig 6). The new nomenclature method was established by the World Health Organization (WHO). This is composed of the name of the HLA gene separated with an asterisk from the four Fields (Fig 6). The Field 1 describes the allele group and corresponds to the low resolution typing reached by the serological method. The Field 2 describes the HLA polymorphic peptide antigen binding-groove

detected by the molecular methods. The HSCT accreditation standards, defines that the identification of the first two fields and the description of the Null alleles (suffix N in the Field 4) are generally sufficient to define high resolution typing (Fig 6) [27, 29].

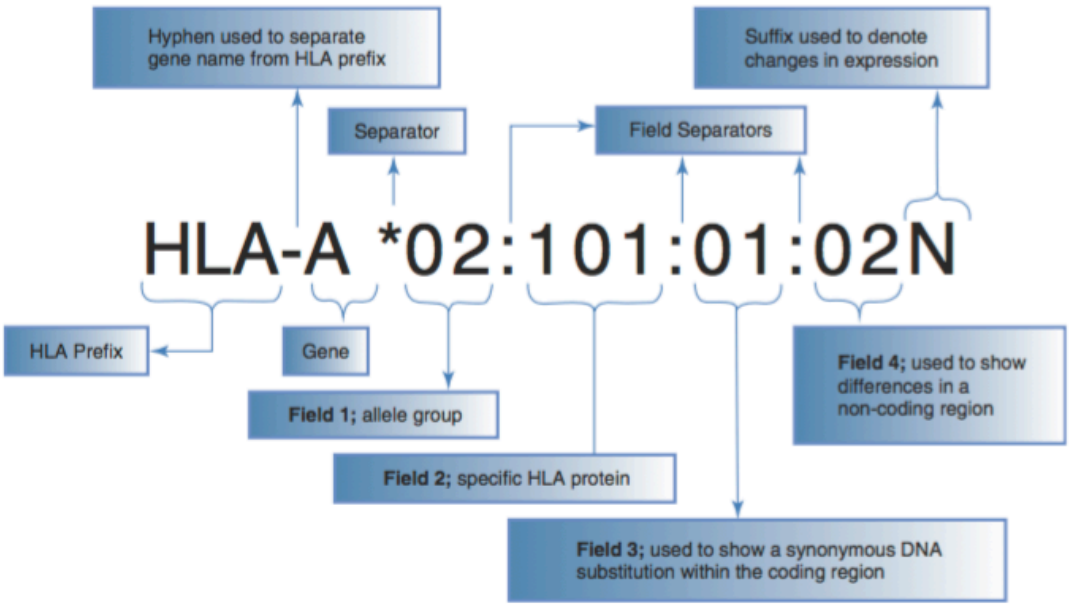

Figure 6. The structure of the HLA nomenclature with the meaning of all the Fields. Figure from [28].

Nowadays, the use of the NGS typing is spreading in clinical laboratories and has greatly increased our knowledge on HLA polymorphisms. Indeed the number of HLA alleles reported to the IMGT-HLA databases keeps always increasing by time [30, 31].

The definition of HLA matching between the donor and the recipient depends not only on the HLA typing resolution level, but also from which loci are tested. In most of the European centers, the loci tested to find a donor are: HLA-A, -B, -C, -DRB1 and -DQB1 [32]. During the donor selection phase, the compatible donors are first searched among the patient's family and if nobody results HLA compatible then the unrelated donor registries are consulted (Italian Bone Marrow Donor Registry IBMDR) [33].

The high-resolution typing methods have allowed the extension of the HSCT also in the cases in which only the unrelated donor is available [34]. Furthermore, the graft manipulation strategies, such as the extensively T-cell–depletion [35] and pharmacological strategies developed to control the GvHD [36] allowed the extension of the HSCT also when HLA partially Miss-Matched donors were available. Today, thanks to the improvement of the HSCT knowledge, based on the **homology between donor-recipient,** it is possible to distinguish 5 types of transplants:

1. **Syngeneic**: when the donor is a monozygotic twin.
2. **HLA match related donor (sibling):** when the donor is related and shares both the HLA haplotypes with the recipient. This is the "gold standard" HSCT homology, but its probability is 25% between siblings.
3. **HLA mismatch related donor (Haploidentical):** when the donor is related and shares exactly 1 HLA haplotype with the recipient and differs by a variable number of HLA genes on the unshared haplotype.
4. **HLA match unrelated donor (MUD)**: when the donor is unrelated (from registry) and it is compatible for all the single HLA alleles evaluated.
5. **HLA mismatch unrelated donor (MMUD):** when donor is unrelated (from registry) and it is compatible only for 50% of the HLA alleles [37].

1.3.2. Recipient conditioning

Before HSCT, the conditioning regimen administered to the patient consists in radiotherapy chemotherapy protocols with the purpose of decreasing the number of patient's own cells and suppressing the immune system in order to avoid the rejection and eradicate the malignancy. Also, it is thought that the donor HSCs need a space in the patient BM for the engraftment. This thought is still controversial [38].

The are many variables in the conditioning regimen such as the intensity, the duration the drug and the irradiation used. The ideal conditioning regimen should permit a fast donor HSCs engraftment and present low grade of toxicity. Moreover, it should be able to eliminate the maximum number of cancer cells in the case of malignancy. According to the intensity, the conditioning regimes can be divided in two classes: **Myeloablative Conditioning** (MAC) and the **Reduced Intensity Conditioning** (RIC) [39].

The **MAC** consists in high dose of chemotherapy and radiotherapy and it is ordinarily used to eradicate the disease in patients affected by malignant disorders. After the MAC, the patients are severely immune-compromised and they will not survive if they are not transplanted. To avoid the graft rejection that often affects the RIC regimen, the MAC can be applied also to treat patients that are not affected by neoplastic disease, for example in the case in which the patients are young and fit [40].

The **RICs**, as mentioned above (paragraph 1.2), were introduced in the clinical protocols during the 1990s and they are considered less harsh than the MAC. The RIC is characterized by reduced doses of chemotherapy drugs with partial or no radiotherapy treatments. This less invasive approach has enabled the elderly patients and those affected by co-morbidities to receive the allo-HSCT. The advantage of the RIC is its broad range of application because of the minor toxicity level compared to the MAC. When the RIC is applied to the malignant disease it needs the graft versus leukemia effect (see paragraph 1.3.4) to eliminate the cancer cells [40].

1.3.3. Collection and infusion of the CD-34^{+}

In the allogeneic transplants, CD34^{+} cells may be collected from three different sources: the bone marrow, the peripheral and the cord blood.
In the past, the unique graft source was the **bone marrow** (BM), in which the CD34+ cells are highly enriched. At that time, the HSCT was called Bone Marrow

Transplantation (BMT) because the HSCs were aspired directly from the BM. This way to collect the graft is painful and for this reason the donor is sedated [41].

Today the **Peripheral Blood (PB)** is the most used graft source, principally because the collection of the HSCs from the PB has become a minimally invasive procedure. In the first step of this procedure the donor is treated for several days with the mobilization factor (G-CSF) in order to induce the proliferation and migration of the $CD34^+$ cells from the BM to the PB. Subsequently, since the donor blood is extremely enriched of the Peripheral Blood Stem Cells (PBSCs), these are collected, purified by the apheresis and infused into the patient [42, 43].

A third HSCs graft source used in the HSCT today is the **Cord Blood** (CB). This represents the volume of blood that remains in the discarded umbilical cord during the delivery. Contrary to the transplants performed with BM or PB, in those based on CB graft, the phase of graft collection and graft infusion are temporarily separated. Indeed today there are massive banks in which the CBs graft are stored. The advantage of the CB as graft source is its prompt availability and the disadvantage is its limited content of HSCs. The CB contains about 1/10 of the HSCs compared to those contained in the BM or PB. This reduced content of $CD34^+$ may lead to a slower engraftment and may cause serious post-transplant complications as infections. Due to the lower level of HSC in the CB graft, this can be used to treat only children and adolescents, since the HSCs amount required for the engraftment is directly proportional to the recipient body weight. To overcame this problem in the clinical practice it is possible to use two different CB grafts to treat a single patient. [44, 29].

The day in which the **stem cells** are **infused** in the patient is called “day 0”. The CD34+ cells are introduced in the bloodstream trough a common intravenous administration. Then they move to colonize the recipient Bone Marrow and begin to grow and to produce new blood cells (this process is termed engraftment). Before the infusion, the graft is manipulated *in vitro*. The purification of the erythrocytes and of the plasma is the minimal essential manipulation to be done in order to overcome the donor-recipient AB0 incompatibility. Usually, also the number of T cells in the graft is reduced especially in pairs with high risk to develop a severe GvHD [29].

1.3.4. Immune reconstruction

At the "day 0", the recipient immune system is harshly compromised (immune deficiency state) due to conditioning regimen and the fact that the donor HSCs infused need time to engraft and expand. The allo-HSCT will achieve the success when the donor HSCs carry out the complete **immune reconstruction (IR)**. This is a process in which the physiological hematopoiesis is restored. The time to reach the IR can vary from months to years and it depends on several factors, as the cell type, the HSCs source, conditioning regimen (MAC or RIC), age of patient and also on viral infections [45-47]. Cells appertaining to the innate immune system are restored faster than those of the adaptive immune system because the formers do not require time for receptor rearrangement and cell instructions. The Neutrophils are restored first and their engraftment is usually used as a marker of overall engraftment. They reach the normal peripheral blood levels in about 2-4 week after allo-HSCT. Also the Natural Killer cells (NK) are fast enough, indeed they take up in about 1-3 months. On the contrary, the T and B cells are restored in several months or years (Fig 7) [48].

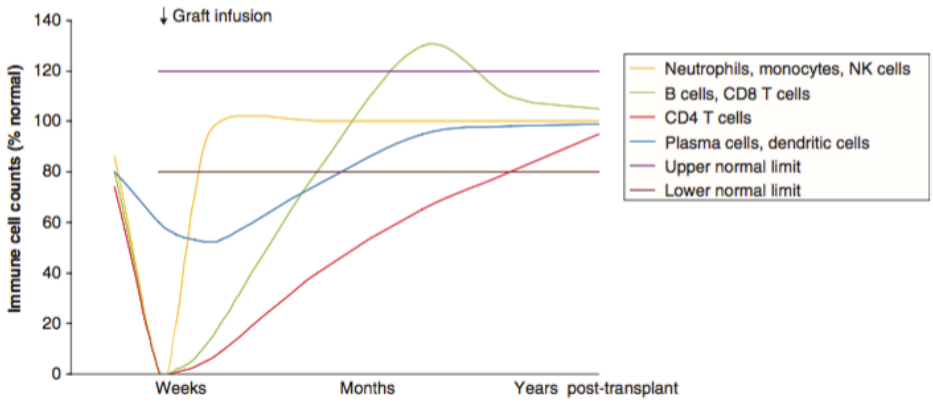

Figure 7. Immune reconstruction after HSCT. Approximate immune cell counts (expressed as percentages of normal counts) peri- and post- myeloablative hematopoietic cell transplantation [48].

The clinical benefit of the allo-HSCT is given by the **Graft-Versus-Leukemia (GVL)**, a mechanism in which the donor immune system eliminates the malignant cells survived after the conditioning. In the GVL, the cancer cells are recognized as **non-self** and non-normal principally by donor $CD8^+$ and $CD4^+$ T cells. The GVL

effect is fundamental to eradicate the cancer cells especially in the patient when the RIC regimen is used. As mentioned before, the latter is a weak conditioning regimen and it is not able to eliminate all the malignant cells [49, 50]. The **donor lymphocyte infusion (DLI)** is a treatment used to boost the GVL effect. It is applied to treat the cases of overt relapse, to control and prevent the relapse when an increase mixed chimerism or positive **minimal residual disease** marker are detected and also to treat patients presenting a high-risk of relapse [51]. It consists in the infusion of the active donor T cells into the patient with a view to eliminating the cancer cells [52].

1.4. Allogeneic HSCT complications

The allo-HSCT is the only therapeutic option able to resolve malignant and severe disorders. However, this treatment can present complications that can lead to morbidity and also to mortality. Among these complications we have infections, acute and chronic Graft versus Host disease, graft rejection, disease relapse and the HLA-loss relapse.

The **Infections** are complications associated to the post transplant recipient immune deficiency state evolved from the conditioning regimen. These complications represent the most frequent causes of death in the 100 days after HSCT (Fig 8) [53]. The common infections are of viral origin, but there are also those derived from bacteria or fungi. The viruses were usually present in the patient before the transplant in a state of latency because they are controlled by the full active immune system. After the transplant the patients are treated with antiviral drugs. The infections represented by the Epstein-Barr virus, Cytomegalovirus (CMV), varicella zoster virus and herpes simplex virus are the most widespread [54-56].

The acute Graft versus Host Disease (**aGvHD)** is a frequent cause of death within the 100 days after HSCT (Fig 8) [53]. The aGvHD usually takes place within two months after the allo-HSCT [57], but it may occur also later [58]. It is a syndrome in which the donor T cells recognize as **non-self** the antigens presented by the **healthy** cells of the immune-compromised recipient and react against them [59]. Acute GvHD is

correlated with the presence of HLA mismatches between recipient and donor. However, aGvHD can also be caused by mismatches in minor histocompatibility antigens which are usually not typed before transplantation [24].

Considering that the donor immune system during its maturation has become able to recognize and attack the cells that show non-self antigens, in every transplant there is a risk the that donor graft will react against the patient cells. This complication shows four different grades: grades I and II are weak and patients do not need any systemic treatment. Instead the more severe grades III and IV may be lethal, therefore the patient requires immune suppression treatment and the organs affected are the skin, liver and gut [60]. Acute GVHD has the same mechanism of non-self recognition that is present in the GVL effect, indeed the GVL and GvHD are strictly related. To limit the damages in the patients, the clinicians try to stimulate a low-grade GvHD in order to activate the beneficial effect of GVL [61]. Moreover, the aGvHD represents the major complications that affect the DLI treatment [62]. The chronic graft versus host disease **cGVHD** occurs later than aGVHD, usually 100 days after the HSCT. The symptoms of chronic GvHD have been expressed within 3 years after allo-HSCT and they are often preceded by the aGVHD [63].

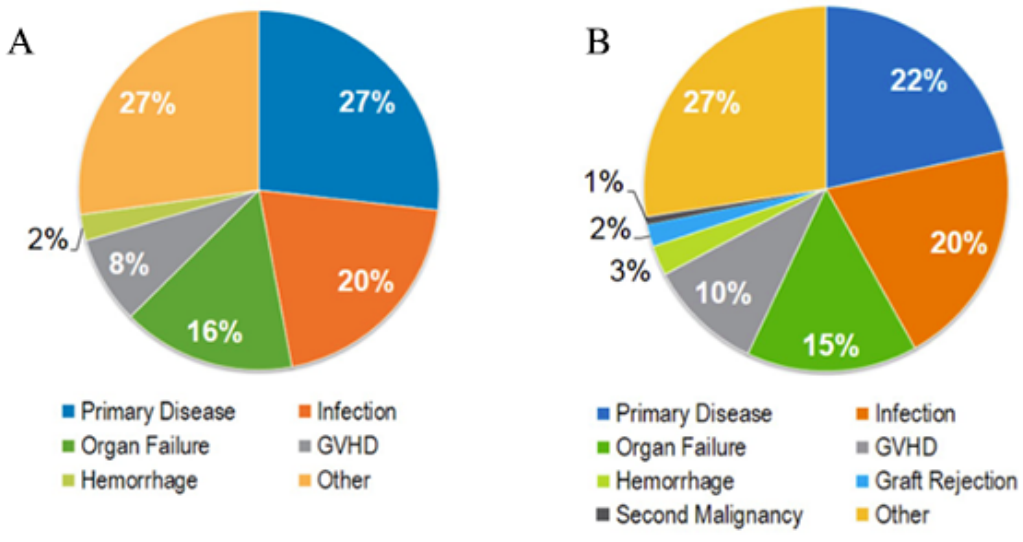

Figure 8. Cause of death at within the 100 days after allo-HSCT (2014-2015). A) in HLA-matched sibling B) in Unrelated Donor [16].

The **graft rejection** represents another complication of the allo-HSCT. In this complication the recipient immune cells that are survived at the conditioning regimen, react against the donor HSCs infused. The risk of graft failure is high in patients that are treated with RIC regimens, that have been transplanted with a HLA mismatched graft or with donor T cell depleted graft. The graft rejection may be achieved in two different ways. In the first one there is no evidence of donor engraftment, indeed, the patients do not recover from the first phase of the neutropenia. They result in a phase of pancytopenia (low number of blood red cells, white cells and platelets) and need a second transplant [64]. In the second way, there is a loss of the donor cells after the initial engraftment that brings to an autologous recovery or to a BM aplasia and pancytopenia.

The **engraftment** is measured through the classical "blood count" and it is achieved when the number of the granulocyte neutrophils in recipient blood reach a value $>0,5*10^9$/L and the number of platelets $> 20*10^9$/L [65]. Usually the cause of the graft failure consists in an immunological reaction mediated by the recipient immune cells against the donor cells [64]. During the graft rejection, the patient remains without the functionally immune system and he/she is subject at risks of infections and of disease relapse [66,67].

The disease **Relapse** derived from the original malignancy is the most common cause of treatment failure and mortality beyond 100 days after-HSCT (Fig 3).
Allogeneic HSCT is designed as a curative treatment for hematological malignancies, used when other less-intensive treatments are no longer effective. HSCT makes use of two therapeutic modalities to eradicate hematological malignancies:

- The anti-leukemia treatment with chemotherapy, radiotherapy or other modalities (e.g., tyrosine kinase inhibitors [TKIs] and antibodies),
- The powerful GVL effect of the allograft, mediated through donor-derived T cells, NK cells and possibly antibodies [68].

Causes for an early disease relapse after transplant are the impossibility to bring the malignancies to a level at which a GVL effect can prevent recurrence or the lack of an effective GVL effect. However, even with an effective GVL, relapse can occur if the

immune system becomes weak or tolerant to the residual disease, or if the disease undergoes immune escape through clonal selection of immune-resistant progenitors.
It should also be borne in mind, occasionally, leukemia can recur also in donor cells as a de novo event, masquerading as a relapse [69, 70].

The **HLA-loss relapse** is the result of the cancer immune escape mechanism and it occurs after the haploidentical HSCT in 33% of the leukemia relapses [71]. The number of haploidentical transplants are increasing worldwide [37, 72]. Moreover, the HLA-loss Relapse is described sporadically in cases of HLA mismatch unrelated donor and of HLA match related donor transplants [71, 73]. The immune escape is a mechanism of the cancer clonal evolution, in which the cancer cells, under the pressure of the immune system, acquire somatic mutations that lead them to interfere with the immune surveillance. The HLA-loss is caused by a somatic mutation called uniparental disomy that consists in a deletion of an HLA haplotype (loss of heterozygosity) and in the duplication of the other haplotype [74,75]. In this acquired mechanism, the cancer cells lose the donor mismatched HLA haplotype allele to escape from the donor immune system [75]. The risk of the HLA loss is linked to the presence of the GvHD before the relapse and also to the amount of the T cells infused. The “classical” relapse occurs in median time of 94 days after the HSCT, whereas, the HLA loss relapse in 300 days. The outcome of the HLA loss is poor. The knowledge about the presence of the HLA loss relapse is not only biological curiosity, but it is important as it allows clinicians to choose the best therapy. Indeed, when the cancer cells lose the donor HLA mismatched, they are less susceptible to the DLI [75]. This picture suggests the clinicians to consider a different strategy, like a second transplantation with another donor, or a cellular therapy based on Chimeric Antigen Receptor-modified T Cells (CAR-T) because the latter are based on unconventional target recognition mechanism. Therefore, before submitting the patient to a DLI treatment after the haploidentical transplant, it is recommended to investigate the HLA loss relapse because of its high frequency [76].

1.5. Chimerism after allogeneic HSCT

In modern medicine, the term **chimerism** after allogeneic HSCT refers to the coexistence in the same individual of cells derived from two different persons, where the recipient became a genetic chimera (term derived from the classical mythology to define a creature composed of anatomic parts derived from different animals). The aim of the allo-HSCT is to reach the status of **Full chimerism** or Complete Chimerism, in which the recipient hematopoietic cells derive only from donor HSCs. The term **Mixed Chimerism** refers to a status in which the recipient post-transplant blood cells derived both from donor and recipient HSCs. Moreover, the **"Split chimerism"** represents a particular mixed chimerism status in which only some hematopoietic subpopulation cells derive from recipient HSCs and the others from donor HSCs.

The **Chimerism analysis (CA)** allows defining the **donor/recipient** cells ratio in post-HSCT PB or BM samples. For this purpose, the techniques utilized are based on the evaluation of the genomic differences between recipient and donor. These differences are called **genetic markers**. The **informative** genetic markers are those useful to distinguish the donor from the recipient. The probability of finding informative markers is called **informativity** and it depends on the number of loci tested, the number of alleles at each locus, the allele distribution and on the relationship between donor-recipient. There is a higher probability, indeed, to find informative markers in pairs of unrelated that in related [10].

1.5.1. First techniques for the evaluation of chimerism

The techniques and the genetic markers used to evaluate the chimerism have evolved very much during the HSCT history. Among the first markers, there were the **Erythrocyte antigens** (AB0, MN, Rhesus system and those of Kell, Kidd, Duffy, Lutheran, Ss and P system) evaluated by mixed agglutination test and flow cytometer. Their detection reached high informativity and sensitivity (0,5-0,1%) [77] but the

erythrocytes relatively long life affected the CA results. For this reason, a transfusion performed before the transplant may lead to a mistake in the CA detection [10].
The classical **cytogenetic analysis** of the sex chromosomes in metaphases was also widely applied for the chimerism detection. The principal limitation of this technique is the low informativity, indeed it is applicable only in the cases in which the donor and the recipient are of different gender. Its sensitivity is also limited due to the low number of metaphases that could be evaluated at each test.

At the end of the 80s, the development of molecular biology techniques has revolutionized the use of the genetic markers, introducing the fluorescence "***in situ"*** **hybridization (FISH)** (Fig 9). As the cytogenetic analysis, the latter shows a limited informativity, as it is based on fluorescent probes specific for the sexual chromosomes [78]. However, conversely to cytogenetic analysis, the FISH could be applied not only in the metaphases but also in the interphases [79-81] increasing the number of the cells tested and therefore the sensitivity [82].

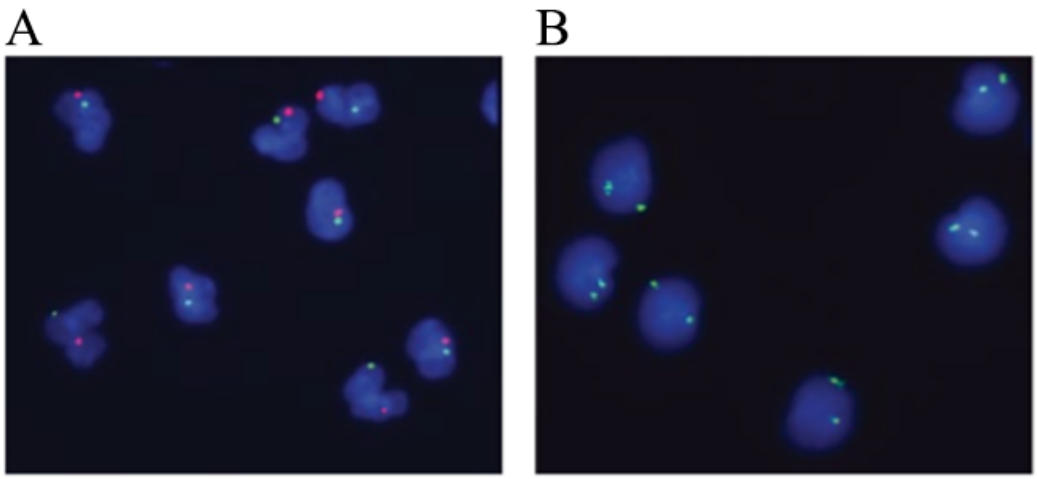

Figure 9. Chimerism evaluation using the dual-color FISH. A) The male cells are colored with red fluorescent and green fluorescent spot. **B)** The female cells contain two green fluorescent spots [10].

1.5.2. Chimerism analysis based on PCR

CA is currently performed by using molecular biology techniques that are based on the amplification with the PCR. With PCR, it is possible to use as genetic marker small differences of DNA sequence between donor and recipient. Among these new markers there are: the Short Tandem Repeats (STRs), Insertion or Deletion (In/Del) and Single Nucleotide Polymorphisms (SNPs) [83,84].

The **STRs** are a class of DNA polymorphisms composed of a core and a constant flanking regions (Fig 10). The core is composed by blocks or units that are repeated in tandem. Each unit is composed of 2-6 base pairs (bp) and, based on the number of base per block, the STRs are classified in bi-, tri-, -tetra, -penta and esa- nucleotide. This markers are spread on the whole genome and their frequency is of one locus every 6-10 kilo bases [85]. They are high polymorphic, reaching about 20-30 alleles per locus [86]. The number of repeats can be highly variable among individuals making, the STR markers effective for human identification purposes [87].

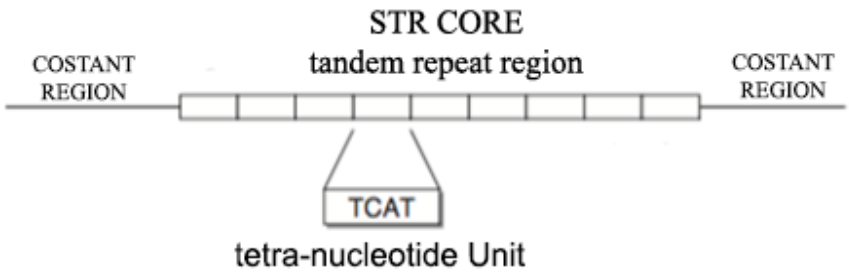

Figure 10. Short Tandem Repeat marker.

The **In/Dels** are natural insertions or deletions, characterized by a variable range of length from 1 to 10 Kilo-bases (Kb). These markers are located both in coding and in the non-coding regions and are spread in all the genome with a frequency of about one In/Del every Kb of DNA [88].

The **SNPs** are the most abundant genetic variation present in the human genome. They are usually bi-allelic variations characterized by the substitution of a single base with another one. They show a frequency in the population higher than 1%. Between

two individuals there are about 1 million different SNPs, with a frequency of about 1 SNP every 1000 bases. Most of them are located in the non-coding region and about 1% in the coding one [89].

Today, the techniques utilized to assess the CA are based on the PCR amplification. The most used technique in diagnostic laboratories is the analysis of STRs by fluorescent PCR. Moreover, another fairly used technique is the analysis of In/Del or SNP by the Real-Time quantitative PCR (qPCR).
Both these techniques are composed of two phases: the **genotyping** and the **quantification**. In the first phase, the DNA of the **donor** and the **recipient pre-transplant** are genotyped in order to find the informative markers. In the second phase, the chimerism quantification is performed on post-transplant DNA samples. For each informative marker found during the genotyping phase, a chimerism value is calculated. However, the results of the CA is calculated from their average. In the diagnostic routine, at least three informative genetic markers are recommended to obtain a robust CA result [90].

1.5.3. Analysis of STRs by Fluorescent PCR

During the CA assessed by fluorescent STRs amplification, the genotyping and the quantification phases are concurrently performed. Therefore, the DNA of the donor, of the recipient pre- and post- transplant are amplified in parallel PCRs. For each STR marker, the primers are designed in order to bind the constant flanking regions (Fig 11-A). The discrimination between donor and recipient in the CA is based on the length differences (size) of the amplified region. The amplicon size depends on the number of the tandem repetition motif for the selected marker [91]. One of the primers utilized to amplify each STRs locus, is linked to a fluorescent dye. The capillary electrophoresis is able to discriminate the STRs allele derived from the donor and the recipient by separating the amplicons on different size and detecting their fluorescence intensity. The alleles are graphically represented as separated peaks (Fig 11-B). The area under the peak derived by the fluorescence intensity of the amplicon represents its quantity. The amount of the generated amplicons is

proportional to the number of initial recipient or donor cells. Therefore, it is possible to determine the proportion between the cells of the different genotypes by analysing the area under the peaks derived by the post transplant sample.

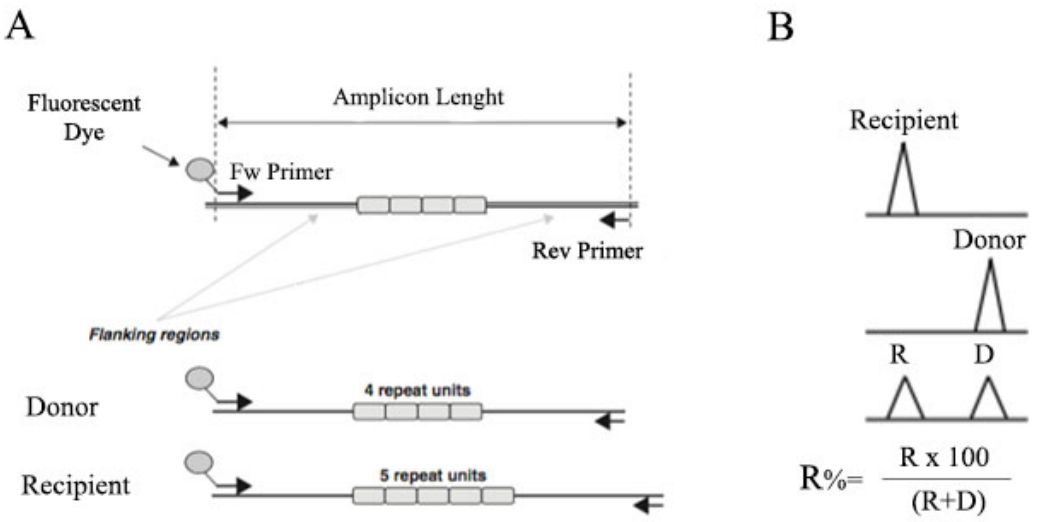

Figure 11. Chimerism analysis performed with the Fluorescent amplification of the STRs A) Fluorescent PCR Amplification scheme. **B)** Peaks result by capillary electrophoresis separation. The Homozygous recipient and donor peaks represent the genotyping phase. The R and D peaks represent the quantification phase in the post-transplant sample. The formula defines the recipient percentage (R%). R and D in the formula represent the area under the peaks.

Nowadays, commercial forensic kits based on the fluorescent amplification of the STRs are used in laboratories to perform the CA. These kits are easy to use because of the multiplex amplification of several STRs loci [92-96]. The capillary electrophoresis instruments are capable to distinguish different fluorescent wavelengths. The multiplex forensic kits are able, indeed, to distinguish also 17 different STRs loci using four dyes with different fluorescent wavelengths (6-FAMT^{M} dye emits the shortest wavelength and it is displayed as blue, followed by the VIC$^{®}$ dye (green), NEDTM dye (yellow), PET$^{®}$ dye (red)). The markers having the same dye are distinguished between them because characterized by different amplicon lengths. (Fig 12).

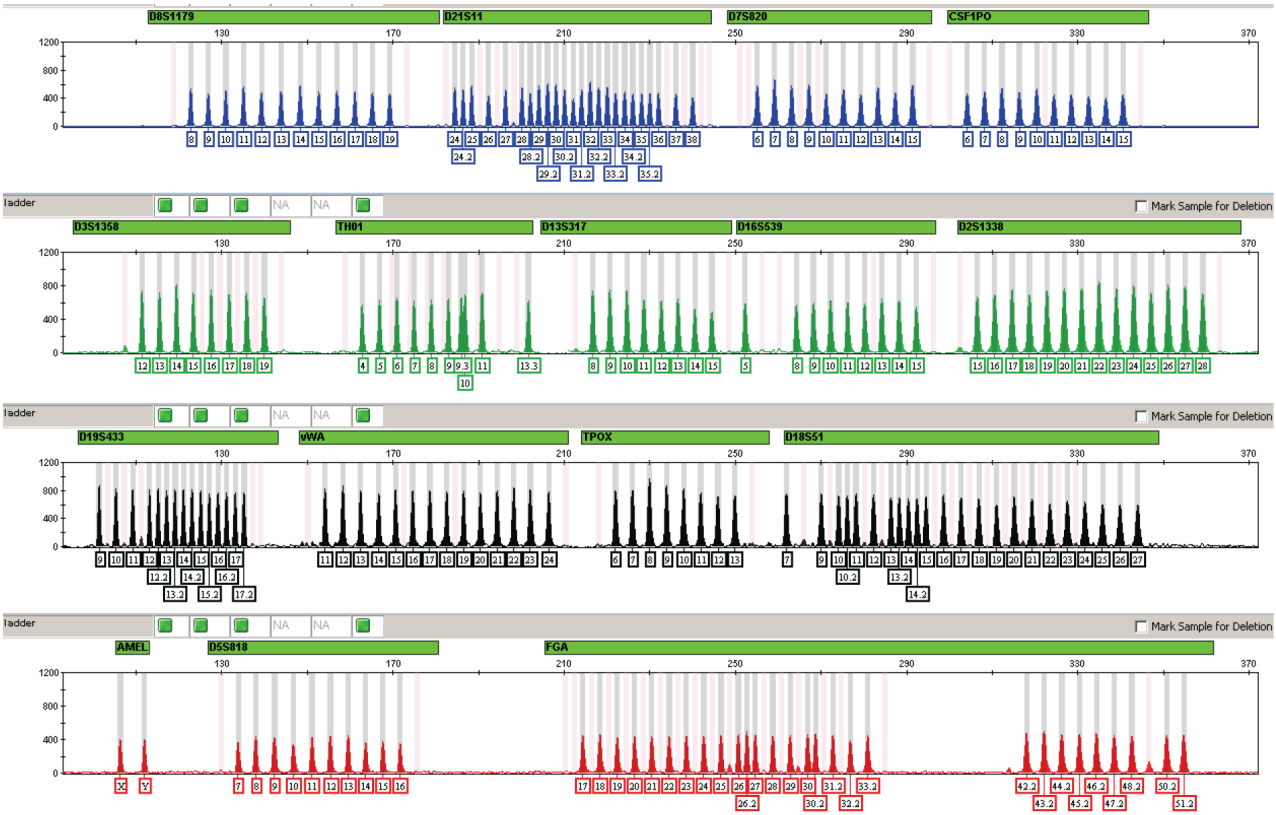

Figure 12. GeneMapper® ID-X Software (Thermo Fisher Scientific, Waltham, MA, USA) plot of the AmpFlSTR® Identifiler® Plus Kit Allelic Ladder (Thermo Fisher Scientific, Waltham, MA, USA). The x- axis represents the fragment size in bp and on the y-axis the fluorescence intensity. In blue dye the loci D8S1179, D21S11, D7S820, CSF1PO; in green dye D3S1358, TH01, D13S317, D16S539, D2S1338; in yellow dye D19S433, vWA, TPOX, D18S51 and in red dye the Amelogenin, D5S818, FGA. For each marker are showed the most common alleles called also allelic ladder.

The principal **advantages** of the STRs amplification that led to the application of this technique in various diagnostic laboratories are the high informativity and the technical simplicity achieved thanks to the development of commercial forensic kits. Specifically, the high informativity is fundamental because it allows to analyse the chimerism in all transplant couples.

On the other hand, the principal **disadvantages** of the STRs amplification affect its quantitative aspect. It is considered a semi-quantitative technique because of:

- A low **sensitivity,** indeed the commercial forensic kits provide a good estimate of the minor component in the chimeric sample when its amount is higher than 5% [93]. In different scientific papers and also in the clinical practice, the input DNA charge in the STRs PCR reaction is 0,5-2-ng (corresponding to 75-300 diploid

cells) according to the kits user guide instructions [92-96]. Such low amount of input DNA affects the sensitivity because very likely, the DNA content of the cells less represented in the chimera sample will not be charged in the PCR reaction [90].

- An **unbalanced amplification** caused by a primer competition in the amplification of the alleles of the donor and the recipient having different size. Indeed the PCR is more efficient in the amplification of the short fragments than the longer ones shifting the quantitative results [97].

- The **stutter peaks** caused by a slippage of the DNA polymerase during the PCR amplification reduce the quantity of the main peak.

- The **pull-up peaks** are artefacts generated during the multiplex PCR amplification when a specific peak in a channel wavelength result off-scale and it appears in its specific position also in other colour channels [98].

1.5.4. Analysis of In/Del or SNP by the Real-Time qPCR

The real time qPCR is used to quantify the initial amount of the DNA tracking the PCR exponential phase. There is a linear relationship between the signal detected during the PCR exponential phase and the initial DNA target amount. In order to track this phase the fluorescence dyes have been applied. The most used dyes in the CA assays are the SybrGreen and the TaqMan probes [97, 99-101]. The former is a non-specifically dye that binds the newly double-stranded DNA products at every cycle increasing the reaction fluorescence. Moreover, the TaqMan probes are composed by an oligonucleotide that binds specifically the target DNA and is linked to a quencher and a reporter. The quencher adsorbs the fluorescence of the reporter when they are close together. Instead, during every PCR extension step, when the polymerase hits the probe, its exo-nucleases activity cuts the quencher that moves away from the reporter allowing it to start its fluorescent activity that increases at every PCR cycle. In the qPCR the **Cycle of threshold (Ct)** value represents the number of PCR cycles

in which the reaction fluorescence intersects the **threshold** line and it is inversely proportional to the initial target DNA amount. The threshold represents the lowest level from which the reaction fluorescence intensity began to increase exponentially (Fig 13) while the full interval is called exponential or logarithmic phase.

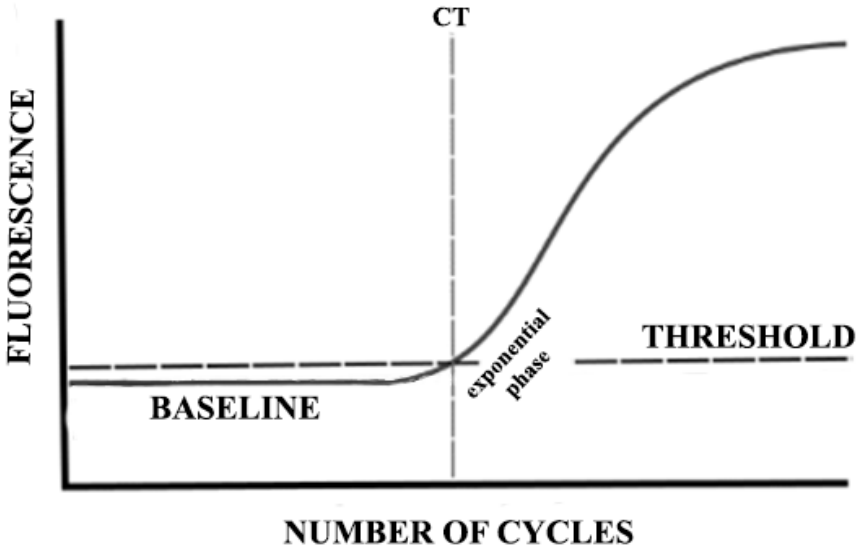

Figure 13. Real-Time PCR amplification curve. The x-axis represents the number of the PCR cycles, the y-axis represents the amount of the fluorescence detected, the baseline represents the basal reaction fluorescence.

During the CA with the amplification of the SNPs or In/Dels by qPCR, the phase of the genotyping and of the quantification are performed in two different steps. Indeed in the first phase, in order to find the informative markers, the DNA of the donor and the recipient are genotyped for a panel of SNPs or INDEL. It is reported that a panel composed by 16-25 polymorphisms derived by 8-15 loci is able to discriminate the 80-98% of the couple donor recipient [97, 100-102]. The genotyping phase is carried out amplifying the DNA of donor and recipient in different tubes, each one with an allele specific oligonucleotide (ASO) in qPCR platform. Each ASO primer of the panel is able to discriminate and amplify one allele at each locus. From Cts values obtained by the ASO amplifications it is possible to define the informative markers. An allele was considered informative when it was positive on recipient DNA and negative on donor DNA, or at contrary negative on recipient DNA and positive on

donor DNA. The allele is defined positive for example when the Ct value of the ASO amplification is minor than 23, and negative for value greater than 35 [97, 103]. The quantification phase of CA could be performed by the classical qPCR absolute quantification comparing the Ct of the post-transplant sample with a calibration curve [100, 104] or by the relative quantification comparing the Ct of the transplant sample with a calibrator DNA using the mathematical formula ΔΔCt [97, 105]. From 2002, different in house qPCR methods [97, 99, 100, 103, 105-108] as well as commercial kit, "AlleleSEQR chimerism assay" (Abbott Molecular, Santa Clara, CA, USA), based SNPs and on In/Dels amplification, were developed. The main advantage of the qPCR amplification is the **high sensitivity** that reaches the 0,001% to 0,1% of the recipient cells detected (Tab 1).

Genetic Markers	Detection method	DNA input (ng)	LD (%)	Reference
INDEL	TaqMan	250	0,1	[97]
INDEL	SYBR Green	100	0,1	[99]
SNP	TaqMan	50 - 500	0,1 -0,01	[100]
INDEL	Hybridization	100	0,01	[105]
SNP	TaqMan	100	0,1	[106]
SNP -INDEL	TaqMan	250	0,003 -0,006	[107]
SNP	SYBR Green	100	0,001	[103]
SNP-INDEL	TaqMan	50 -100	0,1 - 0,01	[108]
INDEL	TaqMan	25 -250	0,1 - 0,024	[109]

Table 1. List of the qPCR methods developed to assess the chimerism. The detection method column indicates the kind of dye used: TaqMan probe, Hybridization probe or the SYBR green. The amount of the DNA charged in the PCR reaction is indicated as ng of Input DNA. The limit of detection (LD) is expressed as percentage of the recipient cells amount.

Figure 14, taken from Alizadeh M (2002) paper [97], shows the higher sensitivity of the qPCR compared with the STRs PCR that is tested on artificial chimera. Nowadays, qPCR is not the most used technique, despite the high sensitivity reached in chimerism detection, due to the low informativity of the biallelic SNP and INDEL markers. Therefore, different markers should be tested to find the three markers recommended for the clinical routine application [90]. Moreover, the qPCR is more

accurate for the detection of chimerism amount between 0,1-30%, but it is inaccurate for higher amounts [99].

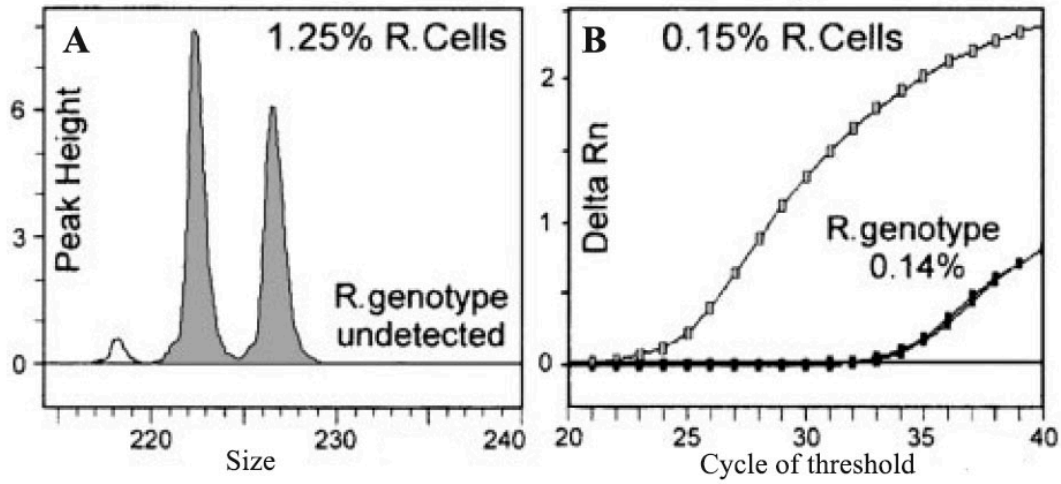

Figure 14. Comparison of STR PCR and qPCR sensitivity in CA. A) The artificial chimera composed by the 1,25% of recipient cells is not detected by the STR PCR method. **B)** The artificial chimera composed by the 0,15% of recipient cells is detected as 0,14% by the qPCR [97].

1.5.5. Chimerism analysis as clinical marker

The CA analysis is an instrument to monitor the outcome of the allogeneic stem cells transplant. It is used, indeed, as a marker to monitor immunological complications like the **engraftment-rejection**, **risk of relapse** and **HLA-loss relapse** (see paragraph 1.4.). Before analysing how to use the CA value as predictive marker for the aforementioned complications, we will define the CA sampling time points. In general the samples to perform the CA in the patient follow-up are collected on day +20-30 (useful to monitor the state of the engraftment), on day +60 (fundamental if the patient does not have the minimal residual disease markers) and on day +90-100 (to confirm the engraftment). These sampling time points are recommended by the "Italian association of the immunogenetic and transplant biology" **(AIBT)**, but they may vary according to the kind of pathology, to the immune-suppressor therapy and to the DLI treatment [110].

The CA is mainly used as a marker to monitor the **engraftment/graft rejection**. The CA is performed after the transplant, subsequently to engraftment evaluated by the number of neutrophils and platelets that is tested by a routine blood count (see paragraph 1.4.). The condition of Complete Chimerism **(CC)** is desirable, because it confirms the engraftment of the donor cells. Moreover, a transient mixed chimerism **(MC)** in the early post transplant period, followed by a CC confirms the IR [111]. On the contrary, an Increased Mixed Chimerism **(IMC)** predicts a graft rejection [112].
CA after the allo-HSCT, may be used also as a marker to monitor the **risk of relapse** in patients affected by a malignant disease, when a high sensitive method is used [113]. The association between CA and risk of relapse, has been proven by different authors [114-116].The CA has high predictive relapse value especially if it is monitored in the tumor specific subpopulation cells [117-119] or in the CD34+ cells [120]. However, when a high sensitivity technique is used, a basal MC not correlated with the relapse is detected [121-122]. Therefore, we must underline that the CA is a specific marker to detect the recipient cells and not the Minimal Residual Disease (MRD) cells. The latter are defined as leukemic cells at levels below the morphologic detection and their persistence is related to the disease relapse [123]. To monitor the MRD, genetic markers as gene mutations, translocations, B-cell receptor and T-cell receptor rearrangement, that are specific for the cancer cells, are used [124-126]. However, despite the fact that the CA is not completely specific for the cancer cells, it gives clinicians a fundamental indication about the disease relapse risk when the patient does not have the specific MRD markers [113].

The CA can also be used as a marker to recognize the **HLA-loss relapse**. In the paragraph 1.4 we had described why it is important to distinguish between HLA-loss relapse and classical relapse. Routine evaluation of the presence or absence of the former and its early detection in a status of molecular rather than overt hematological relapse is clinically desirable. In order to detect this condition, a comparative quantification of patient-specific genomic markers outside and inside the HLA locus is necessary. To do this, there are separate techniques with different sensitivities, such as STR for polymorphisms outside HLA and SSOP or SNP arrays for polymorphisms inside the HLA [75]. The sensitivity of these methods, however, is limited to 5% and reliable information on the presence or absence of HLA loss can only be obtained from purified leukemic blasts, which can be sorted only in the presence of overt

hematologic relapse. In 2017, a Real Time qPCR tool, based on chimerism evaluation, was developed to apply the HLA loss relapse in the clinical routine. This tool is able to distinguish the classical from the HLA loss relapse by evaluating the difference between the values of the recipient chimerism on SNP marker located inside and outside the HLA. Compared to the STRs-SSOP method mentioned before, the high sensitivity reached by this method (0,16%) permits an early detection of the HLA loss relapse. Its application, however, is limited only to 70% of the haploidentical transplants, due to the low informativity of the genetic markers inside the HLA [127].

1.6. Next Generation Sequencing

Next generation sequencing (NGS) technologies, developed in 2000s, are an innovation in human and animal genomics research, as they are capable of producing 100-fold more data than the most powerful Sanger based capillary sequencers. In this way, researchers are able to investigate the large number of queries still to be addressed [128]. NGS generates hundreds of giga-bases of nucleotide sequences per instrument run and at a lower cost, thus motivating researchers to use NGS for various purposes: to identify rare variations on the whole genome or on a target sequence, to analyze transcriptome profiling of cells, tissues and organisms and to identify epigenetic markers for disease diagnosis. Progress in the optimization of procedures, in addition to further reduction of costs, are the key factors that will lead to a more extensive uptake of this technique both in diagnosis and for practical clinical applications. NGS provides qualitative and quantitative data. Quantitative data depends on the depth of sequence data collected on each sample. For samples with a lower abundance target, many more sequence reads are required to achieve accurate quantification [129]. A previous study demonstrated that NGS exhibits sensitivity comparable to that of qPCR in the evaluation of MRD in B cell disorders [130]. In this thesis, the NGS platform used is the Ion Torrent Personal Genome Machine (PGM) of the Thermo Fisher Scientific. This is a "sequencing by synthesis" method which combines the semiconductor technology with the biochemical DNA synthesis reaction. The sequencing chip is composed by several micro-wells, each containing a DNA single strand template to be sequenced. The chip is flooded with one single species of dNTP at a time. When the latter is incorporated in the growing complementary strand it releases hydrogen ions that are detected by sensors.

2. OBJECTIVES

The increasing measure of the recipient percentage, with a high sensitivity technique, can be used to predict the disease relapse risk. It is fundamental when patient specific MRD markers are lost [113]. Today, in the vast majority of laboratories, the chimerism is evaluated using the fluorescent PCR amplification of STRs. The key advantage offered by this method is the highly polymorphic nature of the STR markers, which allows a high informativity. However, its inadequate sensitivity (≥ 5,0%), caused by the intrinsically semi-quantitative platform [93], reduces the risk of relapse predictive value. Moreover, the allele specific amplification of biallelic markers, such as SNP or In/Del, by qPCR has been used in a great number of laboratories. Despite the fact that this method reaches an optimal sensitivity (0,01-0,1%), it lacks in the informativity, because of the SNPs biallelic nature [97].

The **first objective** of this thesis is:

- To develop a new diagnostic system to analyse the chimerism on NGS platform in order to combine a high informativity with an optimal sensitivity.

- To validate the NGS method, comparing its performance both with the STRs method (*"gold standard"* in the diagnostic laboratories) and with an in-house real time PCR protocol, published during the period of my PhD [131].

- To validate the NGS method by participating in the "UK NEQAS" chimerism external quality Assessment (EQA) program.

The NGS protocol development and part of validation with the STRs method was published in a peer-reviewed journal [132].

The **second objective** of this thesis is to perform an upgrade of the method in order to decrease the Ion Torren NGS background with a view to increase the sensitivity and to compare its performance with the previous version.

The **third objective** is to make the system able to recognize the HLA loss relapse in all the transplant pairs.

3. MATERIALS AND METHODS

3.1. FIRST OBJECTIVE

3.1.1. DNA sample preparation

The Ethics Committee of the Institute for Maternal and Child Health, IRCCS 'Burlo Garofolo' approved the present study (approval number: Prot. 18-2015, Cl. M-11). Written informed consent was obtained from all the participants.

For the NGS method development a total peripheral blood sample was collected from 10 volunteer donors (V01-V10; 4 males and 6 females) with ages ranging from 20-50 years. All DNA samples were isolated using QIAamp DNA Blood kit according to manufacturer's protocol (Qiagen GmbH, Hilden, Germany). The DNA status was evaluated using the NanoDrop 1000 spectrophotometer (Thermo Fisher Scientific, Inc., Waltham, MA, USA) and the Qubit dsDNA HS Assay kit and the Qubit fluorometer (Thermo Fisher Scientific, Inc., Waltham, MA, USA). A DNA stock solution was prepared for all DNA samples at 10,0 ng/µl. DNA samples from the 10 volunteer donors were randomly paired. A total of 5 artificial chimeric DNA mixtures, were created by diluting a DNA with its paired DNA at several percentages for each artificial DNA mixture (aCh) (Tab 2). The informativity was also tested on a couple of siblings and for this purpose, the pR_1 sample from a male recipient and the pD_1 from a female donor were collected.

DNA ID	CHIMERA ID	NOTES
V01, V02	aCh_1	Chimera: 1,0% of V01; 99,0% of V02
V03, V04	aCh_2	Chimera: 1,25% of V03; 98,75% of V04
V03, V04	aCh_3	Chimera: 2,5% of V03; 97,5% of V04
V05, V06	aCh_4	Chimera: 5,0% of V05; 95,0% of V06
V05, V06	aCh_5	Chimera: 10,0% of V05; 90,0% of V06
V07, V08	$aCh_{6,7,8,9,10,11}$	Chimeras: 0,5%, 1,0%, 4,0%, 8,0%, 12,0%, 20,0% of V07; 99,5%, 99,0%, 96,0%, 92,0%, 88%, 80,0% of V08
V09, V10	$aCh_{12,13}$	Chimeras: 40,0% of V09 and 100% of V09; 60,0% of V10 and 0% of V10

Table 2. List of the 13 artificial chimeric mixtures with the corresponding percentage.

To perform the validation of the NGS method, different samples of the peripheral whole blood was collected. Three types of validation were done:

A. The first validation was made on **patient follow-up**, comparing the NGS results with the STRs ones. For this validation, the blood samples of a male recipient (pR_2), his donor (pD_2) and six time points (+1, +2, +3, +4, +6, +10 months) of the patient follow up ($pCh\text{-}_{1\text{-}6}$) were collected (Tab 3).

B. The second validation, called **blind test,** was made comparing the NGS results with the STRs ones performed in an external laboratory (Ospedale Santa Maria della Misericordia, Udine). During the clinical routine activity performed in our Institute, 8 patients PB samples ($pR\text{-}bt\text{-}_{1\text{-}8}$), their respective donors ($pD\text{-}bt_{1\text{-}8}$) and post-transplant blood chimeric samples ($pCh\text{-}bt\text{-}_{1\text{-}8}$) were collected. Later, CD3+, CD15+ and CD19+ were respectively depleted from the pCh-bt-4-6, pCh-bt-5-6 and pCh-bt-5 samples by using magnetic beads (Miltenyi Biotech, Bergisch-Gladbach, Germany) performed according to the manufacturer's instructions. The cells depletion purity was tested using flow cytometer (FACS) (BD FACSCalibur, BD Biosciences, USA); all the three sub-population cells were labeled using anti-CD45 (R-phycoerythrin, PE) (Invitrogen, Thermo Fisher Scientific, Waltham, MA, USA) and anti-CD2 (fluorescein iso-thiocyanate, FITC) (Invitrogen, Thermo Fisher Scientific, Waltham, MA, USA) for T lymphocytes, anti-CD15 (FITC) (Immunostep, Salamanca, Spain) for the granulocytes and anti-CD20 (FITC) (Invitrogen, Thermo Fisher Scientific, Waltham, MA, USA) for the B Lymphocytes (Tab 3).

C. As a third validation, we participated to the **UKNEQAS** [144] Chimerism EQA. During this validation 4 patient samples (183, 188, 192 and 196), their 4 respective donors (182, 187, 191 and 195) and two post-transplant chimeric samples (185-186, 189-190, 193-194 and 197-198) for each patient were collected (Tab 3).

The DNA extraction and the quality status were performed as described above. The DNA was stored at -20 ° C.

VALIDATION TYPE	RECIPIENT				DONOR		CHIMERA
	ID	SEX	AGE	DISEASE	ID	SEX	ID
Follow-up	pR2	M	12 y	ALL	pD2	F	pCh1 pCh2 pCh3 pCh4 pCh5 pCh6
Blind Test	pR-bt-1	M	6y 3m	ALL	pD-bt-1	F	pCh-bt-1
	pR-bt-2	M	5y 6m	SCA	pD-bt-2	M	pCh-bt-2
	pR-bt-3	M	16 y	ALL	pD-bt-3	M	pCh-bt-3
	pR-bt-4	F	3y 1 m	ALL	pD-bt-4	M	pCh-bt-4
	pR-bt-5	M	10 y 10 m	ALL	pD-bt-5	F	pCh-bt-5
	pR-bt-6	M	2 y 11 m	CI	pD-bt-6	M	pCh-bt-6
	pR-bt-7	M	7 y 1m	SAA	pD-bt-7	M	pCh-bt-7
	pR-bt-8	M	5y 3m	ALL	pD-bt-8	M	pCh-bt-8
UKNEQAS	183	nd	nd	nd	182	nd	185 186
	188	nd	nd	nd	187	nd	189 190
	192	nd	nd	nd	191	nd	193 194
	196	nd	nd	nd	195	nd	197 198

Table 3. List of the patient DNA samples used for the validation of the NGS method. The samples ID (recipient, donor and chimera) and the corresponding type validation are shown. Nd: not defined Age: age of the patient at the time of the chimerism analysis, y: year;m: months . SCA: Sickle Cell Anemia; ALL: Acute Lymphoblastic Leukemia; CI: Congenic Immunodeficiency; SAA: Severe Aplastic Anemia.

3.1.2. Ion AmpliSeq Custom Chimerism panel design.

A multi-phase strategy was employed to evaluate the main characteristics to develop the Ampliseq Custom Chimerism panel (**ACCh)** composed by SNPs as genetic markers to evaluate the chimerism with NGS:

- The panel average heterozygosity was assessed around 0,5 for the European population (HapMap Phase 3 CEU population);

- Two SNPs per somatic chromosome, termed ‘main SNPs’ (mSNPs), were selected and located in two different regions of the same chromosome;
- The amplicon composition was evaluated according to the following requirements:

    a) The GC percentage ranging between 40,0% and 60,0%;

    b) Presence of one mSNP inside each amplicon;

    c) mSNP location preferably in the centre of the amplicon;

    d) Absence of In/Del SNPs;

    e) Absence of homopolymers and potential homopolymer generation from SNP variants and their flanking regions;

    f) Absence of flanking SNPs to the mSNPs.

In total, 44 single-nucleotide, biallelic, polymorphisms were selected from the NCBI dbSNPs database [136] (Tab 4). A total of 4 base sequences including mSNPs were used as target regions for the primer design. The primer pool, intended for DNA library construction through multiplex PCR, was defined automatically by Ion AmpliSeq Designer software, version 3.0.1 (Thermo Fisher Scientific, Waltham, MA, USA). A single-tube, containing the 44 primer pair pool, was purchased from Life Technologies (Thermo Fisher Scientific, Waltham, MA, USA).

SNP ID	Genome Position	Alleles	European Heterozygosity
rs12070036	chr1:g.227819514	A-G	0,407
rs1234315	chr1:g.173178463	C-T	0,513
rs10496711	chr2:g.134516742	C-G	0,407
rs12612347	chr2:g.219057338	A-G	0,442
rs1984630	chr3:g.134414219	G-T	0,522
rs9831477	chr3:g.30693522	A-T	0,483
rs10033900	chr4:g.110659067	C-T	0,496
rs5335	chr4:g.148463840	C-G	0,492
rs983889	chr5:g.15555486	A-C	0,487

rs10038113	chr5:g.25902342	C-T	0,469
rs552655	chr6:g.13370488	A-G	0,504
rs2077163	chr6:g.33636907	C-T	0,460
rs39395	chr7:g.103489729	A-G	0,425
rs2270188	chr7:g.116140524	G-T	0,496
rs10505477	chr8:g.128407443	C-T	0,531
rs532841	chr8:g.12957475	C-T	0,549
rs2297313	chr9:g.91669362	A-G	0,960
rs424539	chr9:g.14442595	C-G	0,467
rs1561570	chr10:g.13155726	C-T	0,522
rs619824	chr10:g.104581288	G-T	0,407
rs198464	chr11:g.61521621	C-T	0,504
rs178503	chr11:g.44082931	A-G	0,442
rs1126758	chr12:g.103248924	A-G	0,416
rs8608	chr12:g.53294381	A-G	0,522
rs1061472	chr13:g.52524488	A-G	0,504
rs504544	chr13:g.19735891	A-T	0,508
rs10143250	chr14:g.104723433	C-T	0,434
rs1957779	chr14:g.63669647	C-T	0,449
rs634990	chr15:g.35006073	A-G	0,492
rs2117215	chr15:g.94879684	C-T	0,603
rs121893	chr16:g.66183995	C-T	0,414
rs2191125	chr16:g.7720923	T-C	0,550
rs6808	chr17:g.62400575	C-G	0,450
rs744166	chr17:g.40514201	T-C	0,441
rs620898	chr18:g.48509148	A-T	0,467
rs633265	chr18:g.57831468	A-C	0,496
rs108295	chr19:g.34224816	A-G	0,496
rs892086	chr19:g.10837677	C-T	0,451
rs753381	chr20:g.39797465	T-C	0,451
rs715147	chr20:g.50055350	G-A	0,367
rs225436	chr21:g.43729034	A-G	0,517
rs8128316	chr21:g.35721560	C-T	0,542
rs4444	chr22:g.31205334	A-G	0,483
rs132985	chr22:g.38563471	C-T	0,517

Table 4. List of all the 44 main SNPs included in the Ion AmpliSeq custom chimerism panel. SNP, single nucleotide polymorphism; A, adenine; G, guanine; C, cytosine; T, thymine.

3.1.3. Ion torrent library preparation.

In each Chimerism analysis test, the DNA of the donor, recipient and the chimeric

sample were tested to perform the genotyping phase and the quantification ones in parallel. The DNA sample library preparation was performed according to the AmpliSeq Library Preparation protocol (Life Technologies; Thermo Fisher Scientific, Waltham, MA, USA). For each DNA sample, a library was constructed using 30,0 ng of genomic DNA through the Ion AmpliSeq Library kit, version 2.0. The library preparation consists of five phases:

a. Amplification of the DNA target regions.

The **target regions** were **amplified** through a multiplex-PCR according to the following reaction and thermal protocols:

REACTION PROTOCOL			**THERMAL PROTOCOL**		
2,0	µL	Master Mix 5x	99°C	2'	
5,0	µL	Pool primer 2x	99°C	15"	21 cycles
3,0	µL	gDNA 5,0 ng/µL	60°C	4'	
10,0		µL Total	8°C	+∞	

The standard amplification cycles indicated in the Ion torrent guide were 18, but other 3 cycles were added in order to increase the multiplex final yield and facilitate the amplification in the case of difficult DNA samples.

b. Enzymatic digestion of the amplicons

The protocol of the **enzymatic digestion** was performed adding 2 µL of the FuPa enzyme to each library reaction and the final volume was 12 µL. The thermal cycler protocol is:

THERMAL PROTOCOL	
50°C	20'
55°C	20'
60°C	20'
8°C	+∞

c. Adapter ligation.

In the **ligation step** the universal p1 adaptor and the barcode specific for each library

(donor, recipient and chimera) were added to each digested library. The ligation protocol is composed by the preparation of two pre-mixes:

LIGATION PRE-MIX 1			LIGATION PRE-MIX 2		
0,5	µL	p1 adapter			
0,5	µL	Barcode	2,0	µL	Switch solution
1,0	µL	H_2O	1,0	µL	Ligase enzyme
2,0		µL Total	3,0		µL Total

First 1 µL of the first the pre-mix and after 2 µL of the second pre-mix were added to the digested library, that reaches a final volume of 15 µL. The ligation thermal protocol was:

THERMAL PROTOCOL

22°C	30'
68°C	5'
72°C	5'
8°C	+∞

d. Library purification

Each library was **purified** from the: dNTPs, salts, primers, primer dimers with 1,5x of AMPure XP magnetic beads (Beckman Coulter, Inc., Brea, CA, USA) according to the Ion AmpliSeq Library kit version 2.0 instructions.

e. Library quantification by the real-time PCR

The purified libraries were **quantified** by qPCR to define their concentration on the thermo-cycler 7900HT Fast Real-Time PCR system using KAPA Library Quantification Kits-Ion Torrent-Universal (Kapa Biosystems Inc., Massachusetts, USA). The libraries with a concentration minor than 30 pM were re-amplified from the multiplex amplification step. Instead, the libraries having a concentration greater than 30 pM, were diluted at 8 pM in low TE buffer (Thermo Fisher Scientific, Waltham, MA, USA) and after pooled using the following rates: Donor-recipient 1:1; and chimera-chimera, 1:1. The recipient-chimera rate was fixed at 1:40 in order to obtain an average coverage of the above libraries around 250x :10.000x.

3.1.4. NGS sequencing

During the **emulsion PCR,** the library fragments were amplified on the Ione Sphere Particles (ISPs). The reaction was performed using the PGM™ Hi-Q™ OT2 Kit (Thermo Fisher Scientific, Waltham, MA, USA) on the Ion OneTouch™ 2 instrument (Thermo Fisher Scientific, Waltham, MA, USA). The amplification reaction occurs in microscopic drops which are generated during the reaction itself. To obtain a monoclonal amplification each drop should contain a single sphere and a single amplicon. The reaction of amplification is composed by 60 PCR cycles performed at two different temperatures: 95°C and 64°C. After the amplification the template positive ISPs were **enriched** using the Ion OneTouch™ ES (Thermo Fisher Scientific, Waltham, MA, USA). The template positive ISPs are those in which the amplicon had been cloned. The selection takes place through the use of streptavidin magnetic beads that capture the template-positive ISP because the high affinity for the biotinylated ends of the amplicons. Subsequently the washing with NaOH was performed to remove the bond streptavidin-ISP. Template Ion Sphere Particles were arranged using the Ion Personal Genome Machine (PGM) Template OT2 200 kit (Thermo Fisher Scientific, Waltham, MA, USA) and a single end 200 base-read **sequencing** run was conducted using the Ion Torrent PGM system. Library pools were sequenced on ion-314 and 316 chip. The Ion torrent sequencer is based on the semiconductor technology. The machine sequentially floods the chip with one nucleotide after another. The chip is composed of millions of micro-wells in which the positive template ISPs are charged and pH variations detected by a sensor. If in a micro-well, nucleotides will be incorporate because complementary to the DNA sequence, a hydrogen ions are released.

3.1.5. Tool for the Chimerism Analysis: Chimeval

A tool called **Chimeval** to automatically perform the Genotyping of all DNA together with the quantification of all the chimera samples was developed [132]. The tool was

written using the Shiny package in R, a web framework to build interactive web applications [133]. A functional diagram of the code is presented in (Fig 15). The custom tool get as input the sequencing **Bam files** (see below for the definition) of 'Donor', 'Recipient' and 'Chimeric' patients. The tool uses readGAlignments and pileLettersAt functions, from the GenomicAlignments package [134], to read Bam files and extracts the letters-nucleotides into a set of individual genomic positions defined from the **Bed file** (see below the definition). Thresholds for 'Donor' and 'Recipient', homozygous and heterozygous genotyping calls are settled in base counts frequency ranges of 90,0 - 100,0% and 30,0 - 60,0%, respectively. Genotyping calls not included in the thresholds ranges were excluded as unreliable. Users can modify the thresholds according to their needs from the user interface. Genotypes from each library were crosschecked to select only SNPs comparable in all conditions. Selected SNPs from donors and recipients were labelled as informative recipient alleles **(IRA)** according to the following schema:

- Recipient heterozygous and Donor homozygous:
    - Recipient (Aa) and Donor (AA);
    - Recipient (Aa) and Donor (aa);
- Recipient homozygous and Donor homozygous:
    - Recipient (aa) and Donor (AA);
    - Recipient (AA) and Donor (aa);

Only the IRA SNPs tagged as informative were used to calculate the chimera's donor-recipient ratio as median of the allele frequency ratio, while **standard error** was used to calculate confidence intervals of prediction at 95,0%.

To cross validate the tool, genotyping of all donor and recipient samples was also performed manually obtaining the variant data from the Ion Torrent plugin Variant Caller, version 4.4 using the 'Generic - PGM - Germ Line - Low Stringency' configuration coupled by the HP286SNPs hotspot Bed file.

Bam file definition: it is a tab delimited text file that contains sequence alignment data downloaded directly by the Ion Torrent PGM system after the sequencing.

Bed file definition: it consists in a list containing the genomic position of the all the

SNPs presents in the panel. The Bed file was created using the UCSC Genome Browser [135] and the NCBI dbSNPs database [136]. All the INDELs present across the amplicons, and the SNPs near the 5' and 3' ends of the amplicons were excluded from the file. Instead, all the SNPs located in the central region of each amplicon were included in the file. They result in a total of 286 SNPs of which the 44 selected SNPs (paragraph 8.1) were marked as 'mSNP' and the SNPs belonging to the same amplicon were indexed with the same chromosome-amplicon ID number. Finally, a hotspot panel Bed file was created: 'HP286SNPs'.

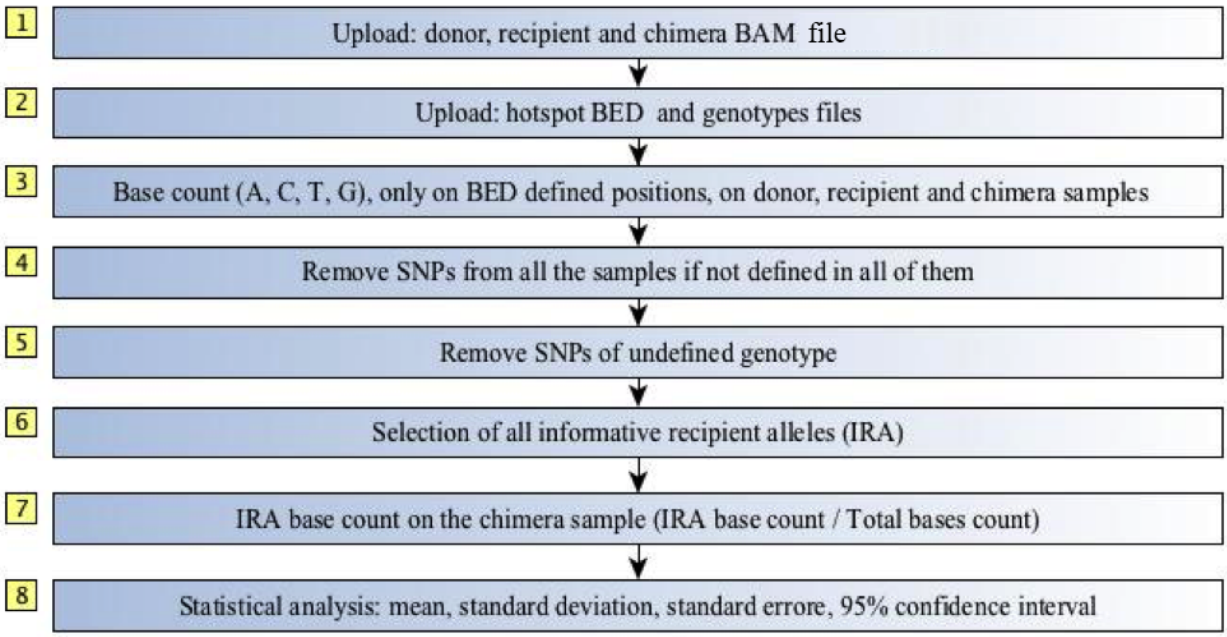

Figure 15. Chimeval Code functional diagram. The Bam, and reference files were uploaded to the custom pipeline. A quality control filtering step per target base was performed and the genotypes for donor and recipient samples were defined. Finally, subsequent to identification of the informative recipient allele, statistical analysis was conducted to determine the chimerism percentage. A, adenosine; C, cytosine; T, thymine; G, guanine; SNP, single nucleotide polymorphism; IRA, informative recipient allele.

3.1.6. STRs Analysis for NGS method validation

Multiplex PCR amplification of V01-V10 and aCh1-13 samples, in addition to the patient samples pD2, pR2 and pCh1-6, was performed according to the manufacturer's

instructions of the AmpFlSTR Identifiler Plus PCR Amplification kit (Thermo Fisher Scientific, Waltham, MA, USA). Amplicons were resolved on a Genetic Analyzer 3130 and analyzed with GeneMapper software, version 4.1 (Life Technologies; Thermo Fisher Scientific, Waltham, MA, USA). This validation was performed in the external laboratory of Department of Molecular Medicine, University of Padova, Padova, Italy. The Multiplex PCR amplification of the blind test validation on pCh-bt-1-8 samples was performed according to the manufacturer's instructions STR-PCR PowerPlex 16HS (Promega, Madison, USA). Amplicons were resolved on a Genetic Analyzer 3500 and evaluated with GeneMapper software, version 1.2 (Life Technologies; Thermo Fisher Scientific, Waltham, MA, USA). This validation was performed in the external laboratory of "Ospedale Santa Maria della Misericordia, Udine".

3.1.7. Absolute real time PCR for NGS method validation

An additional validation method developed by our research group was used, where a discrepancy between NGS and STR data was assessed. It is based on allele specific absolute qPCR amplification with SybrGreen [131]. This in-house method was applied in the cases in which the recipient is a male and donor female. For each test a patient specific standard curve was built with 3 serial dilutions of the recipient pre-transplant DNA that represent the recipient 100% (R_100%) with the donor DNA that represent the donor 100% or the recipient (R_0%). The standard curve is composed by four standard points R_100%, R_10%, R_1% and R_0,1%. Before the serial dilution both the DNAs R_100% and R_0% were diluted at the same concentration (Fig 16).

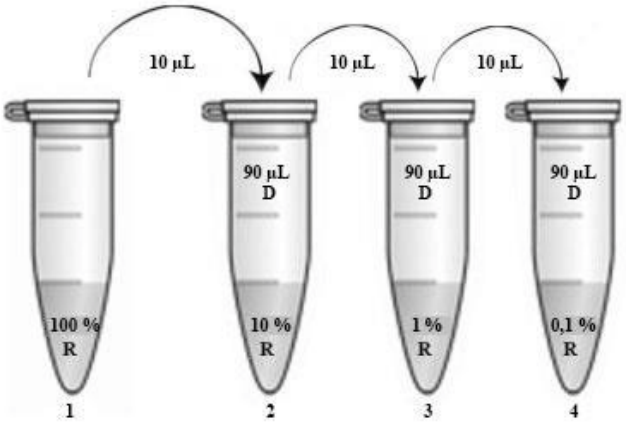

Figure 16. Schematic representation of the standard curve preparation. Ten microliters of the Recipient DNA R_100% were serial diluted (three times) with 90 µL of Donor 100% DNA (D). The four standard points (R_100%, R_10%, R_1% and R_0,1%) were produced.

For each patient, to assess the chimerism with the qPCR, the four points of the standard curve, the donor DNA (R_0%) and the chimeric samples were amplified with the semi-nested real-time PCR developed [131] This protocol permits the quantification of the chimerism of a male recipient with high sensitivity (0,1%) through the allele specific amplification of a genetic marker called AMGY. This marker is composed by four nucleotides (GATA) located inside the Y chromosome variant (AMELY) of the amelogenin gene and it differs from the marker AMGX (AAGG) located inside the X chromosome variant AMELX.

The semi-nested real time protocol is composed by two subsequent PCR reactions.

In the **first PCR reaction**, the genomic DNA of the standards, patient chimeric samples and the water control, were amplified using a pair of primers that bind the nucleotides externally located to the marker permitting the amplification of both the alleles AMGY and AMGX. In this way, the initial recipient-donor proportions are maintained at the plateau phase. The reaction was performed on the GeneAmp PCR System 9700 (Life Technologies, Foster City, CA, USA) term cycler, following a Touch Down PCR (TD-PCR) protocol. In TD-PCR the initial annealing Temperature (Ta) was set over the primers melting temperatures (Tm) but then decreased at each PCR cycle of -0,5°C until it reached a value lower then the Tm.

The sequence of primers used in the first PCR are:

- Primer-FB: (5'- CCTTTGAAGTGGTACCAGAGCA-3 ')

- Primer-Rev: (5'- TGACTCCAACCAGAGAAGCAG-3 ')

The first PCR protocol was:

7,50	µL	K2G Fast Hot Start Ready mix 2X (KAPA Biosystems, Wilmington, MA, USA)
1,25	µL	Primer FB [5µM]
1,25	µL	Primer Rev [5µM]
5,00	µL	gDNA 20ng/µL
15,00		µL Total

The TD-PCR thermal protocol was:

96°C	3'	
96°C	15"	x 10 cycles
63°C*	10"	
72°C	1"	
96°C	15"	x 30 cycles
58°C	15"	
72°C	1"	
8°C	+∞	*=-0,5°C x cycle

In the Second PCR, the product of the first PCR diluted 1:8000 was amplified in duplicate reactions with the ASO primers and its respective reverse primer (Primer-Rev), the same that was utilized in the first PCR amplification. The reaction was done with ABI 7900HT Fast Real Time PCR System (Life Technologies, Foster City, CA, USA) in order to quantify the recipient chimerism. The ASO primer sequence was:

• primer ASO-Y: (5'- GAAGTGGTACCAGAGCATGATA- 3').

The second PCR protocol is the following:

7,50	µL	SYBR Green PCR Master Mix 2X (Life Technologies, Foster City, CA, USA)
0,75	µL	Primer ASO-Y [5µM]
0,75	µL	Primer Rev [5µM]
6,00	µL	DNA [1:8000]
15,00		µL Total

The thermal protocol was:

95°C	10'	
95°C 65°C	15" 10"	X40 cycles
95°C 60°C 95°C		Dissociation curve

The ΔCt obtained from the R_100% and the R_0% was used to test the ASO-Y primer specificity. Moreover, the Ct's of the four standards were plotted against the $\log_{10}$ of the standards concentration. The slope, R^2 and y-intercept of the standard curve were obtained with the software SDS v-2.3 (Thermo Fisher Scientific, Waltham, MA, USA) and the amplification efficiency was calculated by the formula: $\text{Efficiency} = (10^{\frac{-1}{\text{slope}}})\text{-}1$.

3.2. SECOND and THIRD OBJECTIVES

To improve our NGS method sensitivity, an upgrade of the **ACCh** panel and of the tool **Chimeval** was done. The new panel is called ACCh haplo-block **(ACCh_hb).** It is based on markers that are composed by different SNPs, sequenced in a single amplicon (sequencing read). These markers are called haplotype blocks (**haplo-blocks**) and are described as useful to perform an ultrasensitive NGS counting [137]. Moreover, to identify the HLA-loss relapse, several markers located in the HLA locus were added to ACCh_hb panel. Therefore, the panel is composed by two groups of markers, one inside and the other outside the HLA locus.

3.2.1. Haplo-block counting approach

To genotype the recipient pre-transplant DNA and the donor, two NGS libraries were sequenced at low coverage (100x) (Fig 17 A-B). The Informative haplo-blocks are those in which the recipient sample is homozygous with two Informative Recipient Reads (IRR) (case 1 described in the paragraphs 3.2.2. and 3.4.1.) or heterozygous with one IRR (case 2 described in the paragraph 3.2.2. and 3.4.1.). In the example below (Fig 17 A-B) we can see the genotyping of a theoretical haplo-block composed by three SNPs. Both the reads of the recipient have the adenine (AAA) as allele in all three SNPs, instead both of the donors have three guanine (GGG). Therefore, both the recipient reads result informative (IRR) and useful for the chimerism quantification. A third library was constructed for the quantification of the post-transplanted sample. The latter was deeply sequenced (about 10.000x) and the IRR were used to define the recipient percentage. The quantification consists in the counting of the sequenced reads that match the IRR and in the definition of their percentage on the total reads counted. The reads that do not match completely the recipient reads, due to the occurrence of a single base error (background), can be recognized and interpreted as sequencing error by the other 2 SNPs. The example (Fig17-C1) shows a post transplant chimera sample containing the 1% of the recipient. A total of 9850 reads

that match completely the IRR (AAA) were counted. A total of 50 reads affected by the single base error (AGA) in the SNP2 were recognized as mistake by the other two SNPs (SNP1 and SNP2) and properly assigned to the recipient. The correct chimerism amount was calculated counting the amount of the total IRR (IRR corrected and uncorrected) over the total reads. Moreover, in the example (Fig 17-C2) the donor 100% sample (0% recipient) was used as chimera to measure the recipient background (NGS single base error). The 10 reads (x10) with adenine (red) at first SNP and the others ten with adenine (red) at second SNP were assigned to the donor, interpreting correctly the sample as 100% donor.

Figure 17. NGS haplotype block counting vs NGS based on SNP counting. A-B) Genotyping of the Haplo-block composed by 3 SNPs in Recipient pre-transplant and Donor DNA. **C1)** Counting strategy correction in Recipient 1% chimeric sample. **C2)** Background correction during the quantification of the Donor 100% samples using the haplo-block counting strategy. **D)** NGS background in Donor 100% quantification using a SNP marker.

Instead, the first NGS method developed, that used the single SNP as marker, was affected by this error. Indeed in the example (Fig 17-D), 50 reads (with adenine in red) are wrongly assigned to the recipient and its percentage (0,5%) is wrongly calculated despite the fact that the sample is composed by 100% donor.

3.2.2. Haplotype blocks selection

A custom bioinformatics script was written using python and bash programming languages to select the haplo-block **outside the HLA** from 1000genomes phase 3 data. The script algorithm executes the following steps:

a) It selects blocks of 200 bases that contain at least 5 bi-allelic SNPs with an allele frequency >0,1 in a specific European (EUR) population.

b) It compares the blocks of all the individuals present in the specific population to define the **possible haplotypes** and their frequency (Fig 18).

CHROMOSOME	POSITION	rs_ID	POSSIBLE HAPLOTYPES															
			1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
			1|2	1|2	1|2	1|2	1|2	1|2	1|2	1|2	1|2	1|2	1|2	1|2	1|2	1|2	1|2	1|2
Chr1	188247616	rs75818240	T|T	T|T	T|T	A|T	T|T	T|T	T|A	A|A	T|A	T|T	A|T	T|T	T|T	T|T	T|A	A|T
Chr1	188247654	rs3010658	T|C	T|T	C|C	C|T	C|T	T|T	C|C	C|C	T|C	T|C	C|C	T|T	C|T	T|T	T|C	C|T
Chr1	188247674	rs2931598	A|G	G|G	G|G	G|A	G|G	A|G	G|G	G|G	A|G	G|G	G|G	A|A	G|A	G|A	G|G	G|G
Chr1	188247734	rs2931599	G|A	A|A	A|A	A|G	A|A	G|A	A|A	A|A	G|A	A|A	A|A	G|G	A|G	A|G	A|A	A|A
Chr1	188247741	rs61811374	C|C	T|T	C|C	C|C	C|T	C|T	C|C	C|C	C|C	T|C	C|C	C|C	C|C	T|C	T|C	C|T
Chr1	188247742	rs3011972	G|A	G|G	A|A	A|G	A|G	G|G	A|A	A|A	G|A	G|A	A|A	G|G	A|G	G|G	G|A	A|G
Chr1	188247809	rs3010657	A|T	T|T	T|T	T|A	T|T	A|T	T|T	T|T	A|T	T|T	T|T	A|A	T|A	T|A	T|T	T|T
			N° INDIVIDUAL															
			32	61	24	24	32	61	12	1	24	51	16	63	29	36	26	12
			FREQUENCY															
			0,06	0,12	0,05	0,05	0,06	0,12	0,02	0,00	0,05	0,10	0,03	0,13	0,06	0,07	0,05	0,02

Figure 18. Example of the 619 block (193 bases) composed by 7 SNPs. The 16 "possible haplotypes" in the EUR (composed by 503 individuals) each composed by 2 reads (1|2) are shown. For each "possible haplotypes" the numbers of individual and the frequency are showed. The frequency is calculated dividing the N° of individuals of "possible haplotype" with the total number of the individual in the population (503).

c) It defines the **informative couples** comparing "the possible haplotypes" between them. A couple results informative in two cases:

1. When both the reads of a haplotype (homozygous) show at least 3 different SNPs from both the reads of another haplotype (Fig 19-A).

2. When one read of one haplotype (heterozygous) shows at least 3 different SNPs from its other reads and from both the reads of another haplotype (Fig 19-B).

A **CASE 1**

Haplotype 2 1|2	Haplotype 3 1|2
T|T	T|T
T|T	C|C
G|G	G|G
A|A	A|A
T|T	C|C
G|G	A|A
T|T	T|T

B **CASE 2**

Haplotype 1 1|2	Haplotype 2 1|2
T|T	T|T
T|C	T|T
A|G	G|G
G|A	A|A
C|C	T|T
G|A	G|G
A|T	T|T

Figure 19. A) Example of informativity in homozygosity, both the reads of the haplotype 2 are informative because they have 3 different SNPs from both the reads (red colour) of the haplotype 3 (blue colour). **B)** Example of informativity in heterozygosity, only the read 1 of the haplotype 1 is informative because it has 3 different SNPs (red colour) from its read 2 (blue colour) and from both the reads of the haplotype 2 (blue colour).

d) The probability of the informative couple is determined by multiplying the frequencies of the haplotypes belonging to an informative couple. The sum of all the probabilities gives the **block informativity**. The latter is defined as the probability that the block results informative when two individuals in a specific population is randomly chosen.

From the total number of the blocks selected in the population EUR, it was carried out a **ranking** based on the block informativity value. Moreover, the positions of the blocks found in the EUR population have been used to calculate the block

informativity also in Asiatic (EAS) and in all the populations present 1000genomes database (ALL_p).

3.2.3. Haplotype blocks selection and primer design

The **ACCh_hb panel** is composed by two primer pools: the HLA pool (**pool_HLA)** to amplify the haplo-blocks inside the HLA locus and the non-HLA pool (pool_N_HLA), to amplify the blocks outside the HLA. The primers of the **pool_HLA** used in this study were described by Lange [138]. To obtain the primers sequences, a non-disclosure agreement with the DKMS Life Science Lab GmbH Germany Dresden was signed (see data paragraph 8.2.). The pool_HLA is composed by **12 blocks,** amplified by a total of 73 degenerate primers. The 12 HLA blocks correspond to the exons 2 and 3 of the genes HLA-A, -B, -C, -DPB1, -DQB1 and –DRB1. To design the primers for the pool_N_HLA, a total of **29 blocks** with high informativity were selected on the 22 autosomes. A set of 58 primers was designed by hand and the *in silico* **parameters** were chosen in order to work in multiplex with the pool_HLA primers. The tools and the parameters used to design the primers are described below.

1. The primers were designed at more than 30 bp from the beginning or end of the corresponding block.
2. The amplicon lengths ranged from 230-290 bp.
3. The amplicon GC percentage ranged between 40,0 and 60,0%.
4. The primers' specificity was tested using the tool blat [139]
5. The absence of the SNPs within the primers was tested using the database Common SNPs build 150 of UCSC Genome Browser.
6. The primers selected had a melting temperature (Tm) similar between them and it was evaluated using the Multiple Primer Analyzer [140]
7. Values of $\Delta G > -16$ were considered acceptable in order to avoid the secondary structure formation as harpin, Self-dimer and hetero-dimers. The Tm was evaluated using OligoAnalyzer tool [141].

A tail sequence was added at the 5'-end of all the primers of the ACCh_hb panel**.** A sequence called UNI-linker (5'-CAGGACCAGGGTACGGTG-3'), kindly given by

Prof. Florian Fiorella and by prof. Pallavicini Alberto of the University of Trieste, was added to the target specific sequence of the forward primers. Moreover, the sequence of the Ion truncated p1 adapter (**TrP1)** (5' CCTCTCTATGGGCAGTCGGTGAT-3') was added to each reverse primer (Fig 20-A). The primers sequences have not been attached to the thesis because they could be patented.

3.2.4. Custom Ion torrent library protocol

A custom protocol composed by two successive PCRs was developed to prepare the DNA sample libraries. In the first PCR, the genomic DNA was amplified using the targets **ACCh_hb panel** primers. During the annealing phase, in the first cycle of the PCR amplification, the target specific portion of the primers bound the genomic DNA and the primer-tails remained free in a single strand status (Fig 20-A).

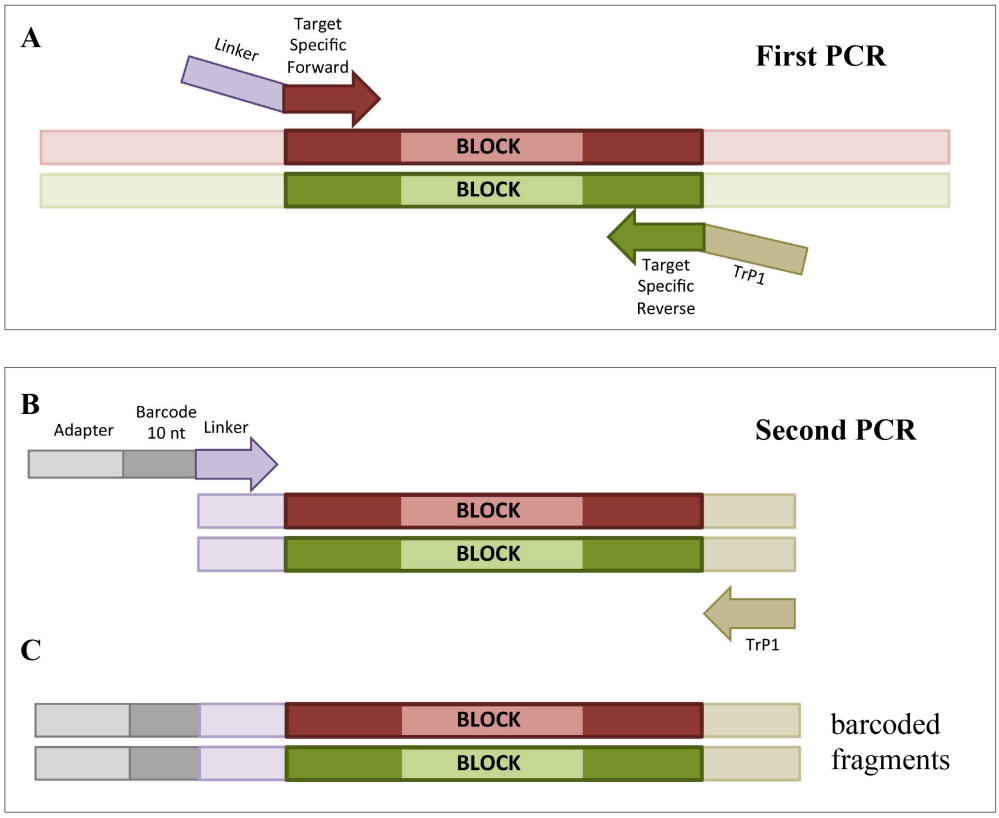

Figure 20. Custom Ion library protocol strategy. A) Schematic representation of the amplification of one Haplo-block during the first PCR. **B)** Schematic representation of the barcoding strategy with the barcode and TrP1 primers during the second PCR. **C)** Schematic representation of the barcoded fragment at the end of the second PCR.

Instead from the second to the last one cycle, the tails were incorporated in the generated amplicons. In this way, at the end of the first PCR amplification, both sides of the amplified fragments were provided with the linker and the Trp1 sequences.

With the aim of developing a multiplex PCR protocol to amplify the haplo-blocks of the ACCh_hb panel, a singleplex PCR protocol was set. The V01 DNA was amplified in 41 singleplex PCRs, 12 for the HLA blocks and 29 for the N-HLA ones. Each of them was performed with a gradient of four annealing temperatures (61°, 63°, 65° and 67°C). The common annealing temperature at which the majority of the haplo-blocks showed a specific band on 3% agarose gel was chosen. The non-HLA haplo-blocks characterized by non specific bands were discarded from the ACCh_hb panel. Moreover, the binomial analysis was used to estimate the theoretical number of informative markers in the ACCh_hb panel for unrelated and related pairs (see paragraph 8.4).

The reactions were performed on CFX96 real-time PCR detection system (Bio-Rad, California, USA). The Ct values were used as indirect parameter of the amplification efficiencies. For each block, a gradient of primer concentration was tested to standardize the Ct values between them. The reaction and the thermal protocols of the first PCR were respectively:

SINGLEPLEX PROTOCOL

12,50	µL	KAPA2G Fast Multiplex PCR Kit[1]
2,00	µL	PRIMER Forward [concentration][2]
2,00	µL	PRIMER Reverse [concentration][2]
1,00	µL	DNA [100 ng/µL]
5,00	µL	H2O
1,25	µL	Eva Green 20x [3]
1,25	µL	DMSO ≈5%
25,00	µL	TOTAL

MULTIPLEX PROTOCOL

12,50	µL	KAPA2G Fast Multiplex PCR Kit[1]
4,00	µL	PRIMER pool_HLA[4]
4,00	µL	PRIMER pool_N_HLA[5]
1,00	µL	DNA [100 ng/µL]
1,00	µL	H2O
1,25	µL	Eva Green 20x [3]
1,25	µL	DMSO ≈5%
25,00	µL	TOTAL

Reaction Protocols legend: 1) KAPA2G Fast Multiplex PCR Kit (KapaBiosystems, Massachusetts, USA). 2) Concentration variable, in order to optimize the amplification efficiency. 3) Eva Green (Biotium, Inc. Hayward, California, USA). 4) In the pool_HLA the optimized primer concentration for the blocks DQ-Exon-2 and DR-Exon-2 were [2,5 µM] each, for the block A-exon-E2 it was [3,125 µM] and for the others block it was [1,25 µM]. 5) In the pool_HLA each block had an optimized primer concentration of [1,25 µM].

SINGLE- and MULTIPLEX THERMAL PROTOCOL

95°C	5'	
95°C	15"	x20-25 cycles
61°C	10"	
72°C	1'	
72°C	10'	

Four NGS libraries with 100 ng of the V01 and V02 DNAs in duplicate reactions were constructed to verify the custom protocol performances and to measure the haplo-block background. The first PCR of the library protocol was performed with the multiplex protocol. Subsequently, the fragments were purified using 0,7x of the Mag-Bind Total Pure NGS Bead (omega Bio-Teck, Georgia, USA) according to the manufacturer's instructions.

In the Second PCR an aliquot of the first PCR was amplified by the TrP1 sequence (5'-CCTCTCTATGGGCAGTCGGTGAT-3') as reverse primer and by a **barcode primer** (5'-CCATCTCATCCCTGCGTGTCTCCGACTCAGNNNNNNNNNNCAG GACCAGGGTACGGTG-3') as the forward one. The latter is composed by three parts: the Ion Torrent adapter A (30 nt) in green colour, the **barcode index** (10 nt) specific for each library (donor, recipient and chimera) in black colour and the **linker sequence** (18 nt), useful to bind the first PCR fragments, in red colour (Fig 20-B). A different barcode index was used for each DNA. At the end of the second PCR, all the library fragments were barcoded (Fig 20-C). The second PCR was performed on CFX96 real-time PCR detection system (Bio-Rad, California, USA). The total length of the barcoded fragments for the pool_N_HLA ranged from 307 to 380 bp.

The reaction and the thermal protocols were respectively:

REACTION PROTOCOL

12,5	µL	KAPA2G Fast Multiplex PCR Kit[1]
0,50	µL	PRIMER Barcode [10 µM]
0,50	µL	PRIMER TrP1 [10 µM]
0,50	µL	Template of theFirst PCR
8,50	µL	H2O
1,25	µL	Eva Green 20x [2]
1,25	µL	DMSO ≈5%
25,00	µL	TOTAL

THERMAL PROTOCOL

95°C	5'	
95°C	15"	x5 cycles
61°C	10"	
72°C	1'	
72°C	10'	

Also after the second PCR, the barcoded fragments were purified using 0,7x of the Mag-Bind Total Pure NGS Bead (omega Bio-Teck, Georgia, USA), according to the manufacturer's instructions.
The libraries were quantified using the real-time PCR quantification protocol described in paragraph 3.1.3. Subsequently, they were diluted at 30 pM in low TE buffer (Thermo Fisher Scientific, Waltham, MA, USA) and pooled using the following rates: V01-V02 1:1; and chimera-chimera, 1:1. The fragments pool were sequenced following the protocol described in paragraph 3.1.4.

3.2.5. Chimeval haplo-block

An upgrade of the tool Chimeval, called **Chimeval haplo-block** (Chimeval_hb), was developed to automatically perform the genotyping and quantification of all the DNA and chimera samples. A functional diagram of the code is presented in (Fig 21). As the previous version, the Chimeval_hb requires the sequencing **Bam files** of 'Donor', 'Recipient' and 'Chimeric' patients as input. The first six steps of the Chimeval_hb useful to define the IRA SNPs are the same as in the Chimeval tool. Subsequently, the tool reads the Bam files and associates the SNPs belonging to the same block (read) as defined in the **bed BLOCK file** (see the definition below). During the recipient and donor Bam file genotyping, the tool discarded the blocks having less than 3 IRA SNPs. The thresholds for 'Donor' and 'Recipient', homozygous-read and heterozygous-read genotyping calls were settled in reads counts frequency ranging from 90,0 to 100,0% and from 30,0 to 60,0%, respectively. Selected reads from donors and recipients were labelled as informative recipient reads **(IRR)** according to the following scheme:

- **Case 1:** when both the reads of a recipient (homozygous-read) showed at least 3 different SNPs from both the reads of the donor.
- **Case 2:** when one read of the recipient (heterozygous-read) showed at least 3 different SNPs from the other read and from both the reads of the donor.

Only the IRR tagged as informative were used to calculate the chimera's donor-recipient ratio as median of the IRR frequency ratio, while standard error was used to

calculate confidence intervals of prediction at 95,0%.

As the Bed file described in paragraph 3.1.2, **the Bed BLOCK file** consists in a list containing the genomic position of all the SNPs present in the panel, but with the difference that they are associated in different blocks. The Bed file was created using the UCSC Genome Browser [135] and the NCBI dbSNPs database [136]. The SNPs result in a total of 763 associated in 32 blocks. The SNPs belonging to the same amplicon (block) were indexed with the same Tag. Finally, a hotspot panel Bed BLOCK file (HP763SNPs) was created (see paragraph 8.3).

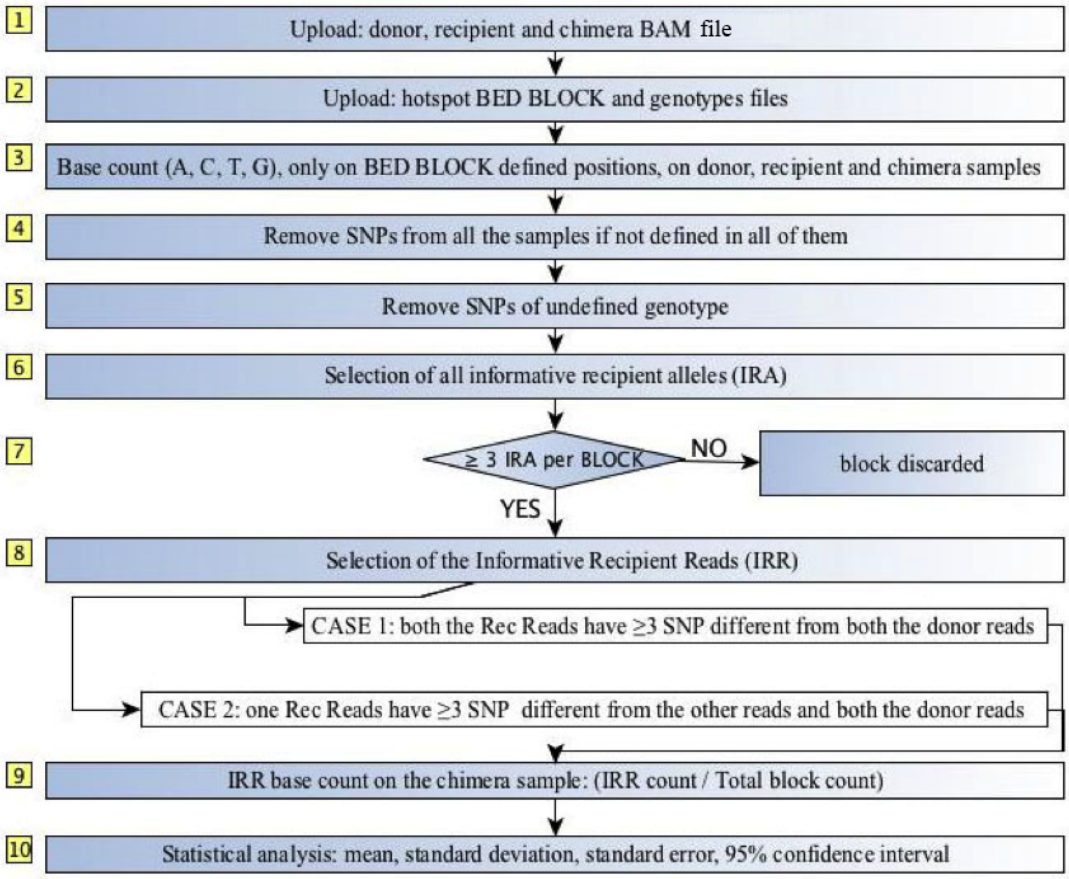

Figure 21. Chimeval Haploblock Code functional diagram. All output next generation sequencing files and reference files were uploaded to the custom pipeline. The first six points of the algorithm correspond to the Chimeval tool. Finally, subsequent to identification of the informative recipient reads, statistical analysis was conducted to determine the chimerism percentage. A, adenosine; C, cytosine; T, thymine; G, guanine; SNP, single nucleotide polymorphism; IRA, informative recipient allele; IRR, informative recipient reads.

4. RESULTS

4.1. FIRST OBJECTIVE

4.1.1. Ion chips and ACCh panel performance

To set up and to validate the NGS method, a total of 12 library pools were loaded and sequenced on ion chip 314 and other 10 library pools on ion chip 316. For each sequencing run, a donor, a recipient and one chimera samples were loaded and sequenced on ion 314 chips. Instead, on 316 chip, up to four chimeras libraries at high coverage were sequenced. The mean performance values of the 22 runs and of all samples are summarized in Table 5.

In-house observed performance	**Ion 314 chip**	**Ion 316 chip**
Ion sphere particles loading	75,65	72,2
Total bases (Mb)	99,05	606
Total reads	489.522	3.043.969
Reads on target	99,45	99,63
Panel uniformity	97,45	96,03
Coverage performance	**Ion 314 chip**	**Ion 316 chip**
Chimera samples	43-44	43-44
amplicons over 2,500X	range (3.600X-25.634X)	range (3.000X-44.000X)
Donor & recipient samples	44-44	44-44
amplicons over 50X	range (90X-1200X)	range (64X-1300X)

Table 5. In-house Ion Torrent Personal Genome Machine analysis observed average performances using the Ion AmpliSeq custom chimerism panel on ion 314 and 316 chips.

4.1.2. NGS genotyping performance using the ACCh panel

A total of 38 DNA samples were genotyped on Ion Torrent PGM, using the ACCh panel with the HP286SNPs Bed file. Regarding the mSNPs, the Variant Caller output assigned 2 mSNPs (rs121893 and rs12612347) as 'No Call' in over 50,0% of the

genotyping runs due to low quality. The remaining 42 mSNPs were successfully genotyped. Out of the remaining 242 SNPs, 27 SNPs were assigned as 'No Call', with an average of 10 SNPs per patient. The call of these SNPs failed in two Variant Caller filtering steps: 'Maximum common signal shift' and 'minimum coverage on either strand'. Genotyping of all samples was additionally performed using our tool with the HP286SNPs Bed file. To identify the IRAs, the data of the donor was compared with the recipient using the tool (Tab 6) and manually cross-validated with the Variant Caller genotypes. Inside the genotyping calls of 242 SNPs, a small bias was present between these 2 tools; this is due to the high conserved filters of variant caller, dedicated predominantly for standard sequencing applications, and to the absence of these filters in our custom tool. An average of 18 IRA SNPs was found in 14 pairs of unrelated and MUD (Tab 6), 15 of them were mSNPs. Moreover, an average of 9 IRA SNPs were founded in 6 sibling pairs and 8 of them were mSNPs (Tab 6).

DNA pair	donor	Custom pipeline		IRA genotypes	
'Donor/Recipient'	type	44 mSNPs	242 SNPs	Heterozygous	Homozygous
V01/V02	unrelated	15	3	10	8
V03/V04	unrelated	18	4	19	3
V05/V06	unrelated	14	4	14	4
V07/V08	unrelated	13	5	10	8
V09/V10	unrelated	8	3	8	3
V02/V01	unrelated	15	5	12	8
V04/V03	unrelated	7	2	6	3
V06/V05	unrelated	14	2	12	4
V08/V07	unrelated	16	4	12	8
V10/V09	unrelated	19	3	19	3
pD1/pR1	sibling	9	2	9	2
pD2/pR2	MUD	17	4	18	3
pR-bt-1/pD-bt-1	sibling	8	1	5	4
pR-bt-2/pD-bt-2	sibling	5	1	4	2
pR-bt-3/pD-bt-3	sibling	12	0	10	2
pR-bt-4/pD-bt-4	sibling	8	1	7	2
pR-bt-5/pD-bt-5	MUD	14	1	8	7
pR-bt-6/pD-bt-6	sibling	5	2	5	2
pR-bt-7/pD-bt-7	MUD	16	1	15	2
pR-bt-8/pD-bt-8	MUD	20	1	12	9
184/183	nd	18	5	17	6
188/187	nd	17	2	12	7
192/191	nd	20	3	13	10
196/195	nd	16	3	16	3

Table 6. List of IRA numbers identified in each donor-recipient pair using the custom pipeline. The donor type is described as Match Unrelated Donor (MUD) or sibling for the transplanted patients and as Unrelated for the artificial chimera samples. For the UKNEQAS samples, the donor type was not defined (nd).

4.1.3. ACCh panel: linearity, detection limit and accuracy

In order to test the linearity of Ion Torrent PGM with the ACCh panel in a fixed detection range (0,5 - 100,0%), a series of DNA mixtures was developed, diluting a DNA with its paired DNA at several percentages of the original. In order to increase the genetic marker variability in addition to the biological variability, a total of 13 artificial chimeras (aCh_{1-13}) were prepared from 5 different DNA pairs. Subsequent to Ion Torrent sequencing, using the custom tool, quantitative data for all IRAs of each artificial chimera were obtained (Tab 7).

Rec-Don ID	Chimera ID	REFERENCE %	NGS %
V09-V10	aCh13	100,0%	99,4%
V04-V03	aCh3	97,5%	96,4%
V06-V05	aCh4	95,0%	97,0%
V06-V05	aCh5	90,0%	90,4%
V08-V07	aCh11	80,0%	81,9%
V10-V09	aCh12	60,0%	59,3%
V09-V10	aCh12	40,0%	41,4%
V07-V08	aCh11	20,0%	19,0%
V07-V08	aCh10	12,0%	11,6%
V05-V06	aCh5	10,0%	9,5%
V07-V08	aCh9	8,0%	7,5%
V05-V06	aCh4	5,0%	5,0%
V07-V08	aCh8	4,0%	3,4%
V03-V04	aCh3	2,5%	2,6%
V03-V04	aCh2	1,3%	1,6%
V01-V02	aCh1	1,0%	1,0%
V07-V08	aCh7	1,0%	1,0%
V07-V08	aCh6	0,5%	0,5%

Table 7. List of the volunteer DNA coupled, the identity of the artificial chimera, the reference amount and the recipient chimerism detected by the NGS method.

Least-squares analysis of the above putative points identified a clear linearity (R^2=0,999; Y=1,006X - 0,0017) between NGS and the reference values (Fig. 22).

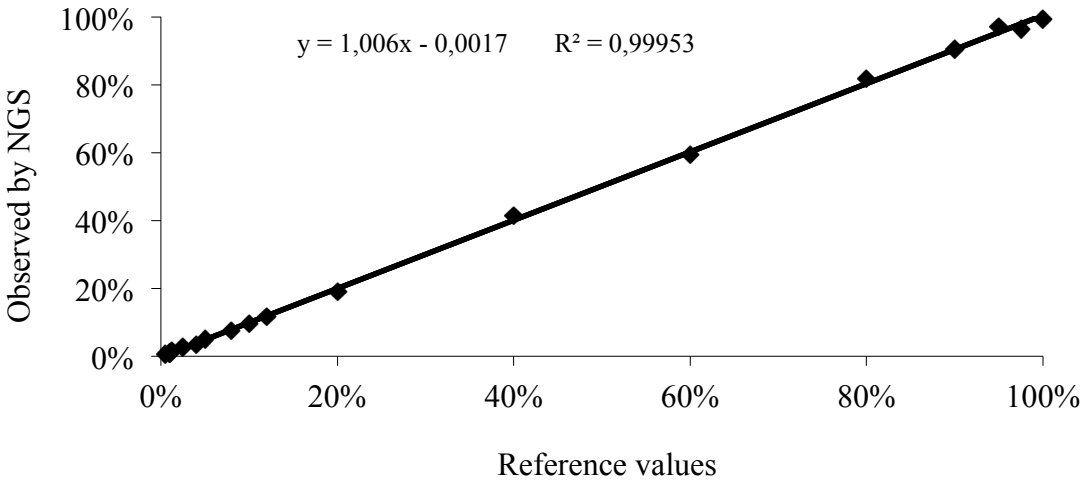

Figure 22. Linearity of Ion AmpliSeq custom chimerism panel: NGS against references values. A series of artificial chimeras, ranging between 0,5-100,0%, were quantified by NGS and plotted against their reference values.

Analyzing the artificial chimeras aCh_{4-5} and aCh_{-8-13} by capillary electrophoresis using the STRs markers (Tab 8), chimerism ranging from 4,0 - 100,0% was detected. Least-squares analysis identified a clear linearity (R^2=0,999; Y=1,005X - 0,0032) between NGS and STRs values (Fig 23).

Rec-Don ID	Chimera ID	STRs-CE	NGS
V09-V10	aCh13	100,0%	99,4%
V08-V07	aCh11	80,8%	81,9%
V10-V09	aCh12	60,6%	59,3%
V09-V10	aCh12	39,5%	41,4%
V07-V08	aCh11	19,2%	19,0%
V07-V08	aCh10	11,2%	11,6%
V05-V06	aCh5	10,4%	9,5%
V07-V08	aCh9	8,4%	7,5%
V05-V06	aCh4	5,4%	5,0%
V07-V08	aCh8	3,8%	3,4%

Table 8. List of the volunteer DNA coupled, the identity of the artificial chimera and the corresponding

recipient amount detected by the STRs capillary electrophoresis (STRs-CE) and by the NGS method.

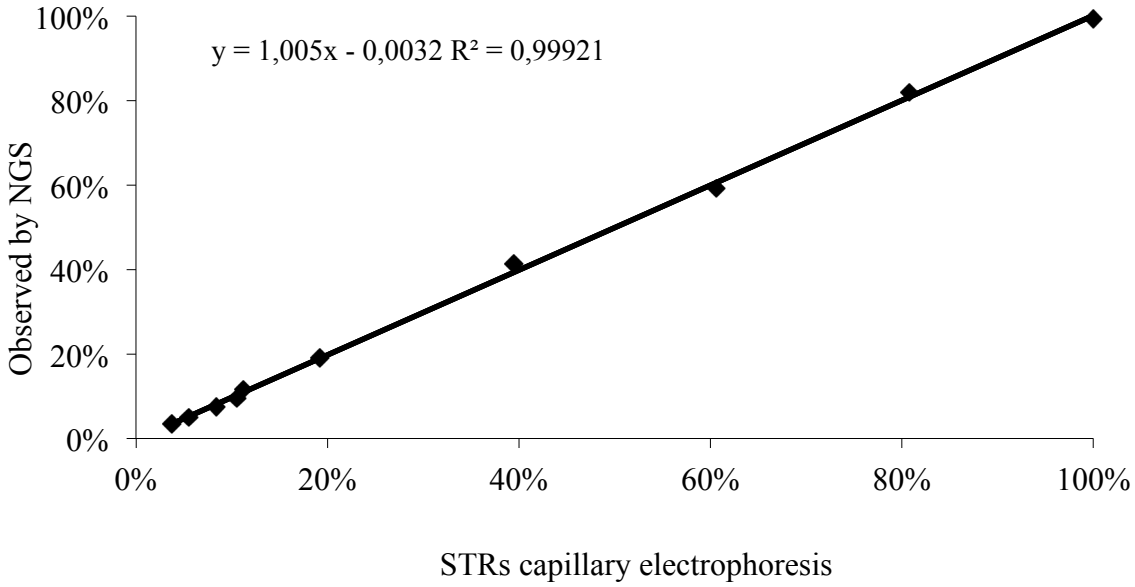

Figure 23. Linearity of Ion AmpliSeq custom chimerism panel: NGS against STRs Capillary electrophoresis values. A series of artificial chimeras was quantified by standard STRs capillary electrophoresis method and the data were plotted compared to the NGS values. The chimeras ranging between 0,5-2,5% could not be quantified by the short tandem repeats due to the detection limit of the informative markers.

In addition, the background of Ion Torrent generated by the ACCh panel was estimated. In this case, the custom tool was used, considering the 'donor' samples as chimera. The background was calculated on 130 IRA SNPs. The average background value at each SNP was estimated at 0,3% ranging from 0,0% to 0,8% and with a 95,0% confidence interval between 0,1% and 0,5% (Tab 9). Considering the background values and the reported literature on the error rates at each base of NGS technologies (range from 0,04-1,0%) [142], the detection limit of the NGS protocol with the ACCh panel was set at 1,0%, although an artificial chimera was detected at 0,5%.

Rec-Don ID	Don 100% as Chimera ID	IRA	Recipient %	95% CI	
V02-V03	V03	22	0,40%	0,70%	0,10%
V04-V05	V05	19	0,30%	0,60%	0,00%
V06-V10	V10	18	0,10%	0,10%	0,00%
V07-V10	V10	11	0,10%	0,40%	0,00%
V03-V02	V02	10	0,30%	0,50%	0,10%
V05-V04	V04	18	0,80%	1,70%	0,00%
V10-V06	V06	20	0,30%	0,40%	0,20%
V10-V07	V07	12	0,00%	0,00%	0,00%
		130	**0,30%**		

Table 9. Background determination. Eight couples of recipient-donor were used to evaluate the NGS background. The latter was assessed as the recipient amount detected in 100% donor samples. The average of the recipient percentage (background) tested on 130 IRA SNPs was 0,3% (in bold character) with 95% of CI 0,5-0,1%. CI: confidence Interval.

Finally, considering that the method determined each chimera, calculating the average value of all IRA, the average Standard Error (SE) and the *Accuracy value* were used as markers of the accuracy, using the data of all artificial chimeras ranging from 1,0 to 99,0% (excluding 0,5 and 100,0%). An $Accuracy_{value} = \frac{|NGS_{value} - Reference_{value}|}{Standard\ Error} < 2$ was considered acceptable, with high accuracy for values < 1,5. For the dynamic range of 1,0-20,0%, the average standard error was calculated at 0,3% with the $Accuracy_{value} < 1{,}5$ (Tab 10). For higher values of chimeras, up to 99,0%, the standard error increased up to 1,8% with a maximum $Accuracy_{value}$ of 2,0 (Tab 10).

Rec-Don ID	Chim ID	IRA	Reference	NGS	SE	$\|NGS - Reference\|$	$Accuracy_{value}$
V04-V03	aCh3	9	97,50%	96,40%	1,5%	1,1%	0,7
V06-V05	aCh4	16	95,00%	97,00%	1,0%	2,0%	2,0
V06-V05	aCh5	16	90,00%	90,40%	1,5%	0,4%	0,3
V08-V07	aCh11	20	80,00%	81,90%	1,2%	1,9%	1,6
V10-V09	aCh12	22	60,00%	59,30%	1,8%	0,7%	0,4
V09-V10	aCh12	11	40,00%	41,40%	0,7%	1,4%	2,0
V07-V08	aCh11	18	20,00%	19,00%	0,7%	1,0%	1,4
V07-V08	aCh10	18	12,00%	11,60%	0,3%	0,4%	1,3
V05-V06	aCh5	18	10,00%	9,50%	0,4%	0,5%	1,3
V07-V08	aCh9	18	8,00%	7,50%	0,4%	0,5%	1,3
V05-V06	aCh4	18	5,00%	5,00%	0,2%	0,0%	0,0
V07-V08	aCh8	18	4,00%	3,40%	0,4%	0,6%	1,5
V03-V04	aCh3	22	2,50%	2,60%	0,3%	0,1%	0,3
V03-V04	aCh2	22	1,25%	1,60%	0,3%	0,4%	1,2
V01-V02	aCh1	18	1,00%	1,00%	0,1%	0,0%	0,0

V07-V08	aCh7	18	1,00%	1,00%	0,1%	0,0%	0,0

Table 10. Accuracy determination. The chimerism value in the NGS column and the Standard Error (SE) were calculated with the NGS method for each chimera sample (Chim ID column) on the IRA. The Reference percentage represents the true chimerism value that was used to calculate the $Accuracy_{value}$. The dynamic range 1-20% is represented with a white background and the range >20% in grey.

4.1.4. NGS method validation

After the development of the NGS system and the evaluation of the technical performance on the artificial chimers, three types of validation on the clinical samples were carried out.

4.1.4.1. Patient follow-up validation

The first validation was made on six **patient follow-up** DNA samples, comparing the NGS results with the STRs ones. The chimerism quantification of six samples derived from the same patient (pCh_{1-6}), was previously performed by STR-CE analysis in an external laboratory (Department of Molecular Medicine, University of Padua, Padua, Italy). The samples presented at least one CC between two MCs (Table 11). The NGS analysis detected a mixed chimerism in all samples (pC_1, pCh_2, pCh_3, pCh_4, pCh_5 and pCh_6), while the STR-CE analysis only in 4 of them (pCh_1, pCh_3, pCh_4 and pCh_5). For these 4 MC samples, the percentage of predicted chimerism was equal in both the methods of analysis (Fig. 24). As to the pCh_2 and pCh_6 samples, NGS analysis evaluated a mixed chimerism at 3,0 and 2,0%, respectively (Tab 11) (Fig 24). To confirm the obtained NGS data for the pCh_2 and pCh_6 samples, the qPCR was additionally performed (see paragraph 4.1.4.3).

Rec-Don ID	Chimera ID	sample collection time	STRs-CE		NGS
pR2-pD2	pCh1	+1 month post-HSCT	MC[a]	9%	9%
	pCh2	+2 months post-HSCT	CC[a]	0%	3%
	pCh3	+3 months post-HSCT	MC[a]	7%	7%
	pCh4	+4 months post-HSCT	MC[a]	6%	6%
	pCh5	+6 months post-HSCT	MC[a]	4%	4%
	pCh6	+10 months post-HSCT	CC[a]	0%	2%

Table 11. Patient follow-up validation. [a]Chimerism evaluation of all patient samples (pCh1-6) was performed by STR-CE analysis in an external laboratory (Department of Molecular Medicine, University of Padua, Padua, Italy). HSCT, hematopoietic stem cell transplantation; MC, mixed chimerism; CC, complete chimerism; STR-CE, short tandem repeat capillary electrophoresis.

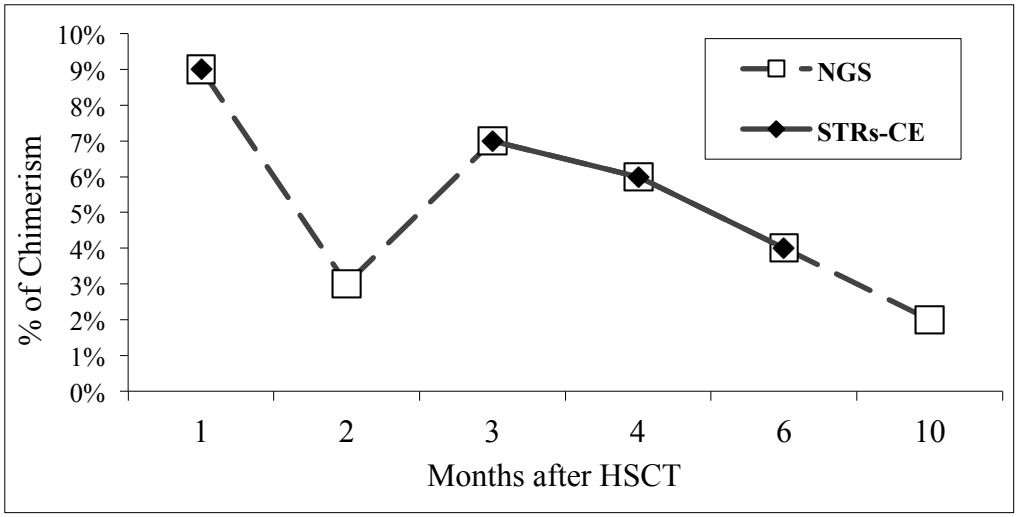

Figure 24. Patient sample evaluation with NGS and STRs-CE methods. Samples (n=6) taken from the same patient at different time points, ranging from 1-10 months subsequent to HSCT, were compared using the two methods. At +2 (pCh_2) and +10 (pCh_6) months, the STRs-CE method was not able to detect the recipient DNA.

4.1.4.2. Blind test validation

The second validation, called blind test, was made comparing the CA results obtained from the analysis of 14 DNA samples with the NGS method and with the STRs ones performed in an external laboratory (Ospedale Santa Maria della Misericordia, Udine). Eight DNAs were extracted from the whole blood of 8 patients (pCh-bt-1-8) and the other 6 DNAs from the subpopulation cells of three of them (pCh-bt-4-6) (Tab 12). The purity of the subpopulation cells depleted was expressed as percentage on the total blood cells and it was assessed by the flow cytometer. The purity values of the two laboratories for the corresponding sample are shown in the Table 12. The cell depletion reached a high level of purity, indeed it was >97% in 6/7 samples and only in the CD19+ sample it was 75%. One sample (pCh-bt-8) was detected as CC by the NGS method, whereas 6 samples (pCh-bt-8; pCh-bt-4-WB, pCh-bt-4-CD3+; pCh-bt-

5-WB, pCh-bt-5-CD13+, pCh-bt-5-CD 15+) were detected as CC by the STR method. The 5 discordant samples were detected as mixed chimerism by the NGS method in a range 1-3% (Tab 12). To confirm the obtained NGS data for pCh-bt-5-WB, pCh-bt-5-CD13+ and pCh-bt-5-CD 15+ samples, the CA was additionally performed with the qPCR (see paragraph 4.1.4.3.).

Rec-Don ID		Chimera blood ID	NGS method							STRs-CE						
			WB	CD3		CD15		CD19		WB	CD3		CD15		CD19	
				%	FC	%	FC	%	FC		%	FC	%	FC	%	FC
pR-bt-1	pD-bt-1	pCh-bt-1	12	-	-	-	-	-	-	12	-	-	-	-	-	-
pR-bt-2	pD-bt-2	pCh-bt-2	70	-	-	-	-	-	-	71	-	-	-	-	-	-
pR-bt-3	pD-bt-3	pCh-bt-3	2	-	-	-	-	-	-	3	-	-	-	-	-	-
pR-bt-4	pD-bt-4	pCh-bt-4	1	3	98	-	-	-	-	0	0	95	-	-	-	-
pR-bt-5	pD-bt-5	pCh-bt-5	1	2	97	1	97	7	75	0	0	98	0	96	6	79
pR-bt-6	pD-bt-6	pCh-bt-6	60	2	97	95	98	-	-	53	7	90	93	95	-	-
pR-bt-7	pD-bt-7	pCh-bt-7	8	-	-	-	-	-	-	10	-	-	-	-	-	-
pR-bt-8	pD-bt-8	pCh-bt-8	0	-	-	-	-	-	-	0	-	-	-	-	-	-

Table 12. Blind test validation. Chimerism evaluation of all patient samples was performed by STR-CE analysis in an external laboratory (Ospedale Santa Maria della Misericordia, UDINE). FC: Flow cytometer purity value is expressed as percentage of the subpopulation cells depleted on the total blood cells; WB: whole blood; %: recipient percentage detected; CD3: T-lymphocytes; CD15: granulocytes; CD19: B-lymphocytes. In red color the DNA samples assessed as CC by the STR-CE and as MC by the NGS method.

4.1.4.3. qPCR validation

The qPCR was used as a third technique to evaluate the chimerism status in the samples detected as MC with the NGS method and as CC with the STR-CE. The in-house qPCR protocol developed in our laboratory [131] was used to quantify the chimerism in samples derived from a male recipient and a female donor. It could be applied, indeed, for the analysis of the samples pCh_2 and pCh_6 and in the DNA samples pCh-bt-5-WB, pCh-bt-5-CD13+ and pCh-bt-5-CD 15+ (Tab 3). As to the samples pCh-bt-4-WB and pCh-bt-4-CD3+, despite the discordant results between NGS and STR method, they were not analyzed with qPCR because the recipients were females. Our in-house method was based on the semi-nested absolute amplification with the Syber green. A patient specific standard curve composed by

four points (R_100%, R_10%, R_1% and R_0,1%) was built diluting the recipient DNA with the corresponding donor DNA. In each qPCR assay, the four points of the standard curve, the donor DNA (R_0%) and the chimeric samples were amplified in duplicate. The CTs median values obtained by the qPCR amplification are shown (Tab 13). The reaction specificity was calculated performing the ΔCt between the R_0% and R_100% samples. For patient pR_2, it was 20,4 and for patient pR-bt-5, it was 20,6. A ΔCt of about 20 results in an error of approximately 0,0001%, therefore the semi-nested qPCR protocol showed a high amplification specificity .

Patient pR2 Threshold: 0,623		**Patient pR-bt-5** Threshold: 0,288	
Sample ID	Ct	Sample ID	Ct
Standard 1 (R_100%)	14,0	Standard 1 (R_100%)	14,3
Standard 2 (R_10%)	17,3	Standard 2 (R_10%)	17,3
Standard 3 (R_1%)	20,6	Standard 3 (R_1%)	21,5
Standard 4 (R_0,1%)	23,4	Standard 4 (R_0,1%)	24,0
Donor sample (R_0%)	34,5	Donor sample (R_0%)	34,9
pCh2	18,6	pCh-bt-5-WB	20,8
pCh6	19,7	pCh-bt-5-CD3+	21,3
-	-	pCh-bt-5-CD 15+	21,1

Table 13. qPCR cycles of threshold. The Cts obtained by the intersections of the threshold line with the amplification curves for each patients.

As example, the amplification curves of the four standards for patient pR-bt-5 are shown (Fig 25).

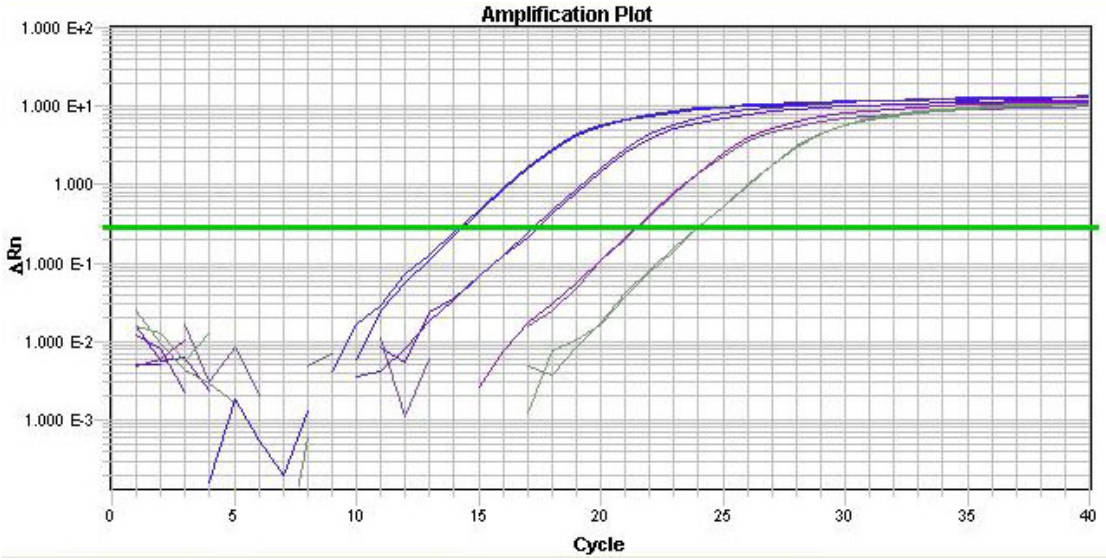

Figure 25. SDS v-2.3 (Thermo Fisher Scientific, USA) amplification plot. The four standard curves for patient pR-bt-5 are represented. On the x-axis the CT values and on the y-axis the ΔRn values are shown. The Rn value, or normalized reporter value, is the fluorescent signal from SYBR Green normalized to (divided by) the signal of the passive reference dye (ROX dye) for a given reaction. The ΔRn value is the Rn value of an experimental reaction minus the Rn value of the baseline signal generated by the instrument. This parameter reliably calculates the magnitude of the specific signal generated from a given set of PCR conditions. The Threshold line (green color) intersects the four standard curves generating the CT values that are plotted to generate the equation of the standard curve.

For patient pR_2, the standard curve equation resulted to be $y = -3{,}10\,x + 20{,}39$, the R^2 =0,999, the standard deviation of the duplicate points was <0,160 and the amplification efficiency 110%. For patient pR-bt-5, the standard curve equation resulted to be $y = -3{,}34\,x + 21{,}00$, the R^2 =0,991, the standard deviation of the duplicate points was <0,160 and the amplification efficiency resulted 99% (Fig 26). The amplification efficiency in both the amplification curve ranged acceptably from 90-110%. A value of the determination coefficient (R^2) >0,985 indicates a high level of correlation between the Ct values and the recipient amount. Moreover, the standard deviation of the duplicate points < 167 indicates a high level of PCR amplification repeatability [143].

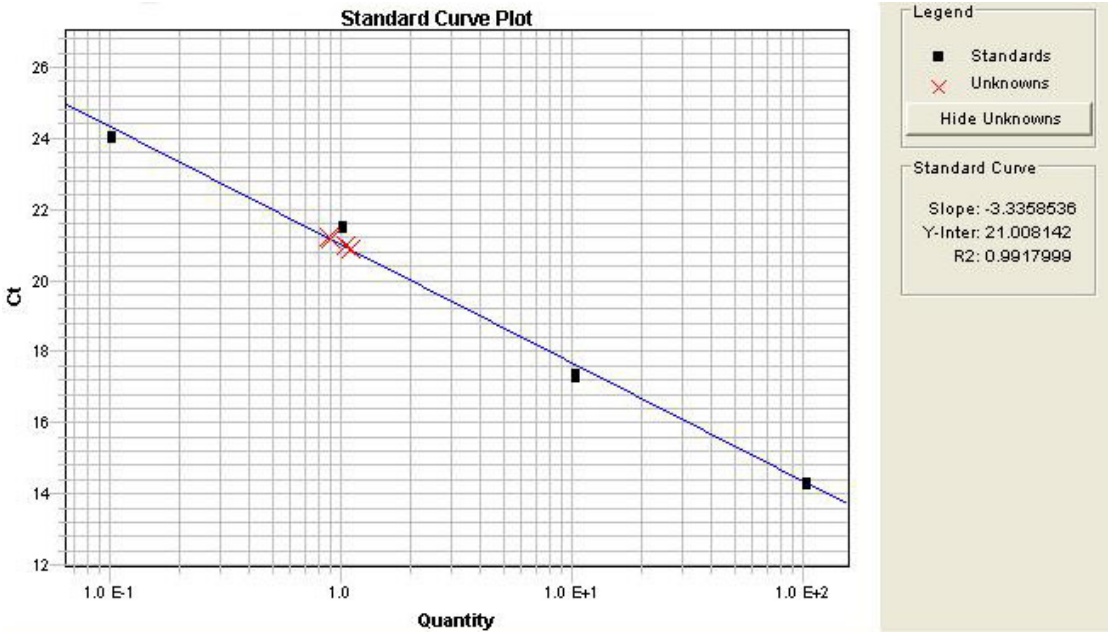

Figure 26. SDS v-2.3 (Thermo Fisher Scientific, USA) equation plot. The equation for patient pR-bt-5 is represented as example. On the x-axis the $\log_{10}[standard\ point\ concentration]$ and on the y-axis the CT values derived from the amplification plot, are shown. The Slope, Y-intercept and R^2 are automatically calculated. The black points represent the Standard amplification and the red crosses the duplicates of the three post-transplant samples (pCh-bt-5-WB, pCh-bt-5-CD13+ and pCh-bt-5-CD 15+).

The result of the qPCR confirmed the MC in the samples analyzed. A total of 3,7% of recipient cells were detected in the samples pCh_2 and 1,6% in pCh_6. Moreover, 1,1% of recipient cells were detected in sample pCh-bt-5-WB, 0,9 % in sample pCh-bt-5-CD3+ and 1% in sample pCh-bt-5-CD 15+.

4.1.4.4. UKNEQAS validation

The third validation was made through the analysis of 8 chimeras samples provided by the **UKNEQAS,** a consortium of external quality assessment laboratories. The latter sent the same chimeric samples to the laboratories involved in the program in order to compare their results and accredited the laboratory for CA. During this validation phase, a total of four patient samples (183, 188, 192 and 196) and their 4 respective donors (182, 187, 191 and 195) were genotyped (table 6). Moreover, 2 post-transplant chimeric samples for each patient (185-186, 189-190, 193-194 and 197-198) were quantified (Tab 14). For each sample, the chimerism amount detected was uploaded to the UKNEQAS website [144]. Subsequently, UKNEQAS provided us with a statistical report, comparing the results of about 107 laboratories (Tab 14). Ninety percent of the laboratories analyzed the samples with the STRs-CE technique, 9% with the real-time PCR, 0,5% with the Droplet Digital PCR and 0,5% with the FISH. Table 14 shows, the UKNEQAS comparison for the 8 chimeric samples. The NGS results are expressed as mean of the recipient amount derived by different informative markers. Instead, the UKNEQAS results are expressed as the mean of the recipient amount detected in all the laboratories (Tab 14). Moreover, the **z-score** was calculated for each chimeric sample using the formula $z = \frac{(x-X)}{\sigma}$, where x is the NGS mean, X is the UKNEQAS mean and σ is the standard deviation.
The interpretation of z-scores is as follows:

- A result between 2,0 and -2,0 would be classified as satisfactory;
- A result between 3,0 and 2,0 or -2,0 and -3,0 is seen as an 'action' result, highlighting a potential issue to the laboratory. Two 'action' results in successive rounds would be classified 'critical';
- A result above 3,0 or below -3,0 is considered to be a 'critical' result, requiring immediate investigation by the laboratory (source from PDF manual on the website [144]).

Rec-Don ID		Chimera ID	NGS	UKNEQAS			Z-Score
			Average	Average	N° lab	St.Dev	
184	183	185	19	20,2	106	1,6	0,75
		186	18	20,2	106	1,5	1,47
188	187	189	9	10,2	109	1,7	0,71
		190	15	15,7	109	2,6	0,27
192	191	193	90	91,5	110	1,5	1
		194	39	37,8	110	3,3	-0,36
196	195	197	11	12,5	106	2,7	0,56
		198	10	10,7	106	1,8	0,39

Table 14. UKNEQAS validation results. NGS average derived by the mean of all the informative SNPs. UKNEQAS average: is the mean of the results reported of UKNEQAS laboratory. N° lab is the number of the laboratories that analyzed the specific chimera sample. St.Dev: is the standard deviation of the UKNEQAS results. Z-score is the UKNEQUAS qualitative statistical parameter.

All the 8 samples analyzed by the NGS method were in a z-score satisfactory interval. The original UKNEQAS report is shown in Figure 27. Moreover, as an example, the frequency distribution histogram taken from UKNEQAS report is shown in Figure 28. This graphic shows the chimerism percentage engraftment for the post-HSCT sample 198.

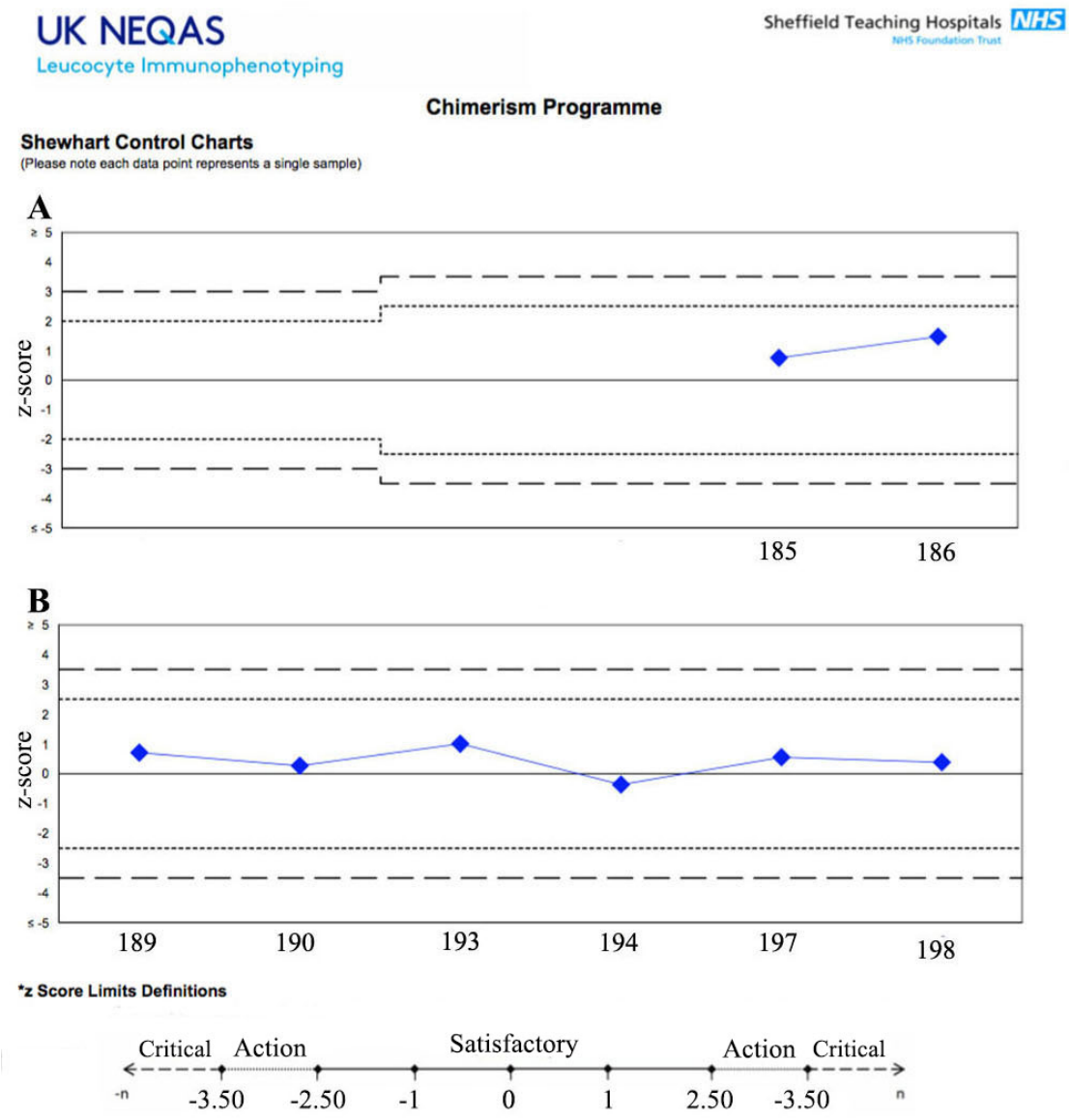

Figure 27. Z-score UKNEQAS report. The sample identities are reported on x-axis. Z-score values are reported on y-axis. The blue points represent the chimeric samples.

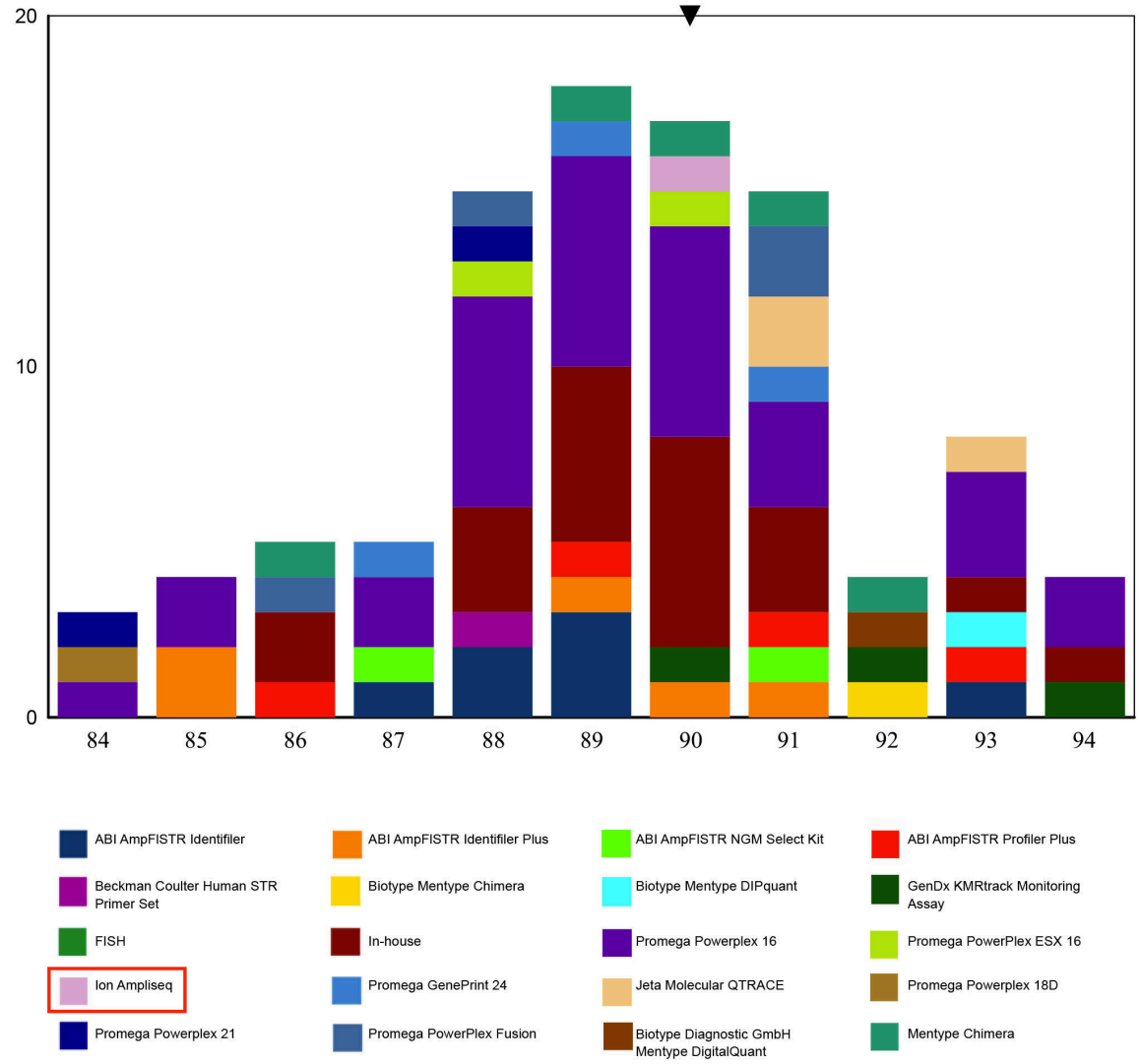

Figure 28. An example of frequency distribution histogram for the sample 198 (from the UKNEQAS report). The different chimerism amounts detected by the 106 laboratories for the sample 198 are reported on x-axis. The frequency of the laboratories identified as technique/kit used is reported on y-axis. Each color indicates a technique/kit used to perform the CA. The NGS method result is indicated as black arrow in the column corresponding to 90% of donor (or 10% of recipient) and it is reported as Ion Ampliseq in pink color (highlighted with the red square)

4.2. SECOND and THIRD OBJECTIVE

4.2.1. Haplo-block selection and primer design

To design the pool_N_HLA primers, 16288 blocks spread on the 22 autosomes were selected by the 1000 genomes database using the custom script (described in paragraph 3.2.2.). The mean of the blocks informativity for the EUR population was 0,381, with the minimum value at 0,002 and the maximum at 0,972. A ranking based on blocks informativity was done. Twenty-nine blocks with high informativity value and meeting the primer design criteria (paragraph 3.3), were selected. A total of 29 primer pairs were designed and identified using the abbreviations: PCR-1 to PCR-29 (Tab 15). The block informativity was subsequently verified in the EAS and ALL_p (Tab 15). The mean of the block informativity was 0,644 for the EUR population, 0,536 for the EAS and 0,618 for ALL_p.

BLOCK ID	Chr	Start Block	End Block	length	n° SNPs	EUR	EAS	ALL_p
PCR-1	17	41401199	41401285	86	21	0,814	0,819	0,821
PCR-2	2	2726421	2726581	160	14	0,643	0,697	0,734
PCR-3	4	114018992	114019101	109	8	0,681	0,598	0,634
PCR-4	7	157149908	157150024	116	22	0,649	0,823	0,749
PCR-5	7	7902123	7902258	135	9	0,656	0,260	0,595
PCR-6	9	31000648	31000803	155	12	0,637	0,337	0,578
PCR-7	10	53679978	53680124	146	15	0,616	0,513	0,598
PCR-8	11	118558333	118558455	122	11	0,610	0,374	0,466
PCR-9	11	5906048	5906205	157	11	0,622	0,603	0,633
PCR-10	12	3194564	3194668	104	11	0,648	0,590	0,598
PCR-11	12	29829639	29829794	155	16	0,601	0,628	0,647
PCR-12	14	63141952	63142026	74	8	0,606	0,427	0,540
PCR-13	16	81204066	81204188	122	11	0,617	0,390	0,587
PCR-14	16	78383509	78383645	136	12	0,610	0,658	0,649
PCR-15	16	6713387	6713563	176	16	0,616	0,573	0,617
PCR-16	17	6165696	6165863	167	15	0,640	0,512	0,588
PCR-17	19	21063039	21063169	130	15	0,657	0,625	0,638
PCR-18	21	11126185	11126289	104	9	0,632	0,597	0,626
PCR-19	21	25405328	25405455	127	9	0,603	0,587	0,570
PCR-20	22	19581946	19582066	120	9	0,614	0,484	0,504
PCR-21	1	188247616	188247809	193	9	0,665	0,536	0,592
PCR-22	3	11955892	11955986	94	13	0,718	0,719	0,780
PCR-23	10	127833998	127834142	144	7	0,656	0,507	0,592
PCR-24	11	247905	248077	172	25	0,643	0,308	0,544
PCR-25	15	31405278	31405442	164	24	0,627	0,355	0,604
PCR-26	16	83740033	83740179	146	11	0,665	0,465	0,640
PCR-27	18	1960439	1960590	151	12	0,632	0,530	0,617
PCR-28	19	30071237	30071417	180	12	0,656	0,485	0,566
PCR-29	20	59309997	59310162	165	8	0,642	0,609	0,608

Table 15. Pool_N_HLA block informativity. The pool_N_HLA blocks are identified (ID) by the acronym (PCR-1-PCR-29). The chromosome, the start-end position, the length and the number of SNPs for each block are described. Table shows the informativity calculated by the custom pipeline on the 1000genomes database for each population.

4.2.2. Libraries custom protocol development

To prepare the NGS libraries, a custom protocol composed by two subsequent PCR reactions was developed. The first PCR was useful to amplify the genomic DNA samples (recipient, donor and chimera) and the second one to barcode the resulted fragments. In order to develop a multiplex PCR useful to amplify all the blocks of the ACCh_hb panel during the first PCR, a singleplex PCR was used to test the haplo-block amplification specificity and efficiency. The DNA V01 was used to amplify the 29 N-HLA and the 12 HLA blocks in different reactions in order to develop a common singleplex PCR protocol. The amplification specificity was tested on gel electrophoresis. For the N-HLA pool, 20 blocks, showing the specific band (Fig 29) at 61°C of the annealing temperature, were selected and included in the ACCh_hb panel. Instead, the other 9 blocks were discarded. The average informativity for the 20 N-HLA pool was calculated by the value in Table 15 and it was 0,639 in EUR, 0,496 in EAS and 0,596 in ALL_p. According to our binomial analysis (Supplementary table 3), the 20 non-HLA markers ensure a 100% of probability to find at least the three recommended informative markers in each population. This value is reduced when relates pairs are taken into account. We found 97,7% for EUR, 90,9% for EAS and 96,5% for ALL_p. However, the probability to find at least one informative marker in these cases is 100%. Moreover, 12/12 HLA blocks HLA blocks were included in the ACCh_hb panel, despite the fact that the HLA-B exon 2 and 3 blocks did not show a specific single band (Fig 29). This choice was made on the hypothesis that the amplification specificity in the HLA locus may vary depending on the DNA tested due to the locus hyper-variability. The final ACCh_hb panel was composed by 32 haplo-blocks (12 HLA and 20 N-HLA) that showed a specific amplification with the same PCR conditions. They potentially could work in multiplex.

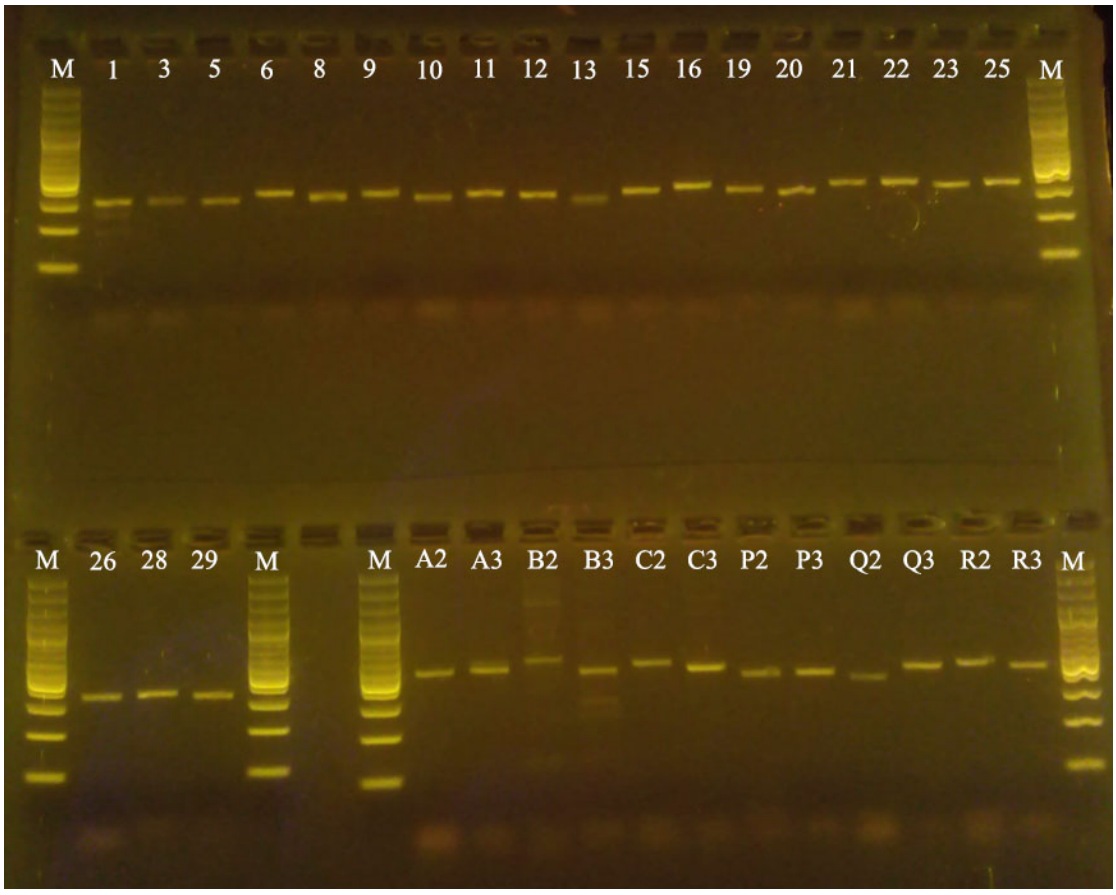

Figure 29. Singleplex PCR specificity. Agarose gel 3% of a total of 21/29 singleplex N-HLA PCRs and of the 12/12 HLA PCR. The 21 N-HLA block are marked by a corresponding number (e.g.: 1 for PCR-1; 3 for PCR-3; 5 for PCR-5; etc). The PCR-1 was discarded because it showed non specific bands. Instead, the other 20 N-HLA PCRs were validated because they showed a single specific band. The 12 HLA blocks are marked by the corresponding gene name (full or abbreviated) and by the exon number (e.g.: HLA-A exon 2= A2 (full); HLA DP B1 exon 2 = P2 (abbreviated); HLA DQ B1 exon 3 = Q3 (abbreviated); etc..). The blocks HLA-B exon 2 and 3 showed non specific bands. M=marker 100bp.

Subsequently, the primer concentration was optimized in order to standardise the amplification efficiency of the 32 selected blocks to obtain a similar amplicon coverage during the NGS sequencing. For each of them, a gradient of primer concentration was tested and the ones in which the Cts were closest between them were selected. The optimized primer concentrations in the final PCR volume are reported (Tab 16). The haplo-blocks of the pool N-HLA (Fig 30) and of pool HLA (Fig 30) were tested in separate runs. To compare their Cts, the same threshold value between the two runs was settled (Tab 16). The PCR-28 was the block with the lowest Ct (20,37), the PCR-13 was the one with the highest Ct (23,35) (Tab 16) and their ΔCt was 2,98.

Pool ID	Block ID	Ct value	Primer pair concentration (nM)
pool_N_HLA	PCR-3	22,4	200
	PCR-5	21,79	200
	PCR-6	21,31	200
	PCR-8	21,65	200
	PCR-9	21,88	200
	PCR-10	22,26	200
	PCR-11	21,75	200
	PCR-12	21,86	200
	PCR-13	23,35	200
	PCR-15	21,79	200
	PCR-16	21,48	200
	PCR-19	22,17	200
	PCR-20	21,46	200
	PCR-21	21,65	200
	PCR-22	21,35	200
	PCR-23	22,01	200
	PCR-25	21,59	200
	PCR-26	21,26	200
	PCR-28	20,37	200
	PCR-29	21,3	200
pool_HLA	HLA-A-E2	21,56	500
	HLA-A-E3	21,2	200
	HLA-B-E2	20,64	200
	HLA-B-E3	20,5	200
	HLA-C-E2	21,78	200
	HLA-C-E3	20,57	200
	HLA-DP-E2	21,2	200
	HLA-DP-E3	21,22	200
	HLA-DQ-E2	22,64	400
	HLA-DQ-E3	21,52	200
	HLA-DR-E2	21,17	400
	HLA-DR-E3	21,33	200

Table 16. Optimization of the primer concentration in singleplex real time PCR. The Ct of the 32 blocks belonging to the ACCh_hb panel are shown. The optimized concentration calculated in the final PCR reaction volume is shown.

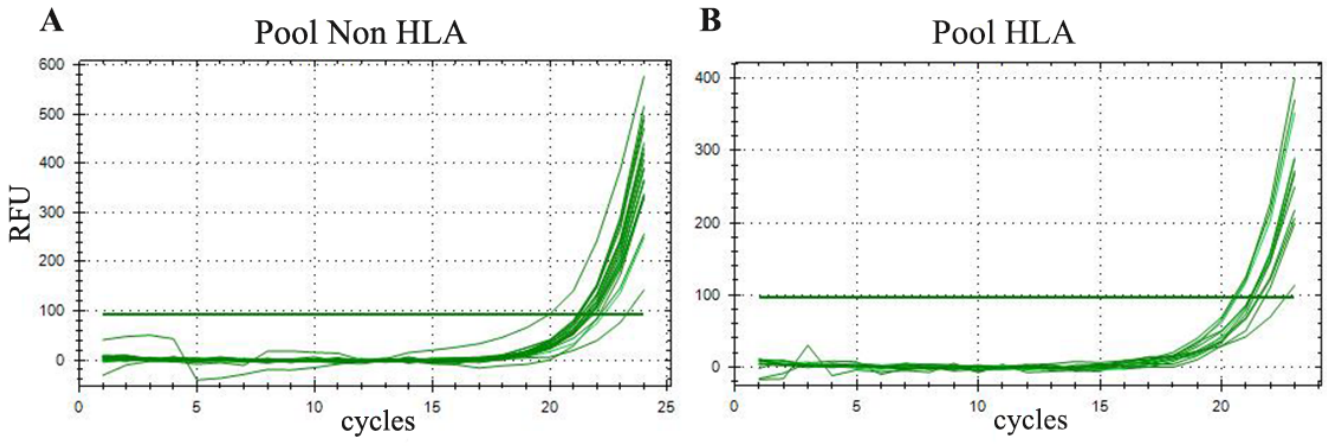

Figure 30. Optimization of the primer concentration in singleplex real time PCR. A) The amplification curves of the 20 N-HLA blocks are shown. **B)** The amplification curves of the 12 HLA blocks are shown. The Threshold value was the same between the two runs.

4.2.3. Library custom protocol performance

Four NGS libraries in duplicate for the V01 and V02 DNAs (V01a, V01b, V02a, V02b) were built following the two steps of the Custom Ion torrent library protocol. In the first PCR, the DNAs were amplified by the multiplex protocol settled. In the second PCR, the amplification fragments were barcoded. The libraries were sequenced on Ion Chip 316. The sequencing performance were the following: the Ion sphere particles loaded resulted to be 84% of the chip wells, the total bases sequenced were 1,07 Giga Bases, 3.864.412 of total reads with 96,7% of aligned reads were obtained (Fig 31), the reads on target were 90,4% and the panel uniformity was 95,98%.

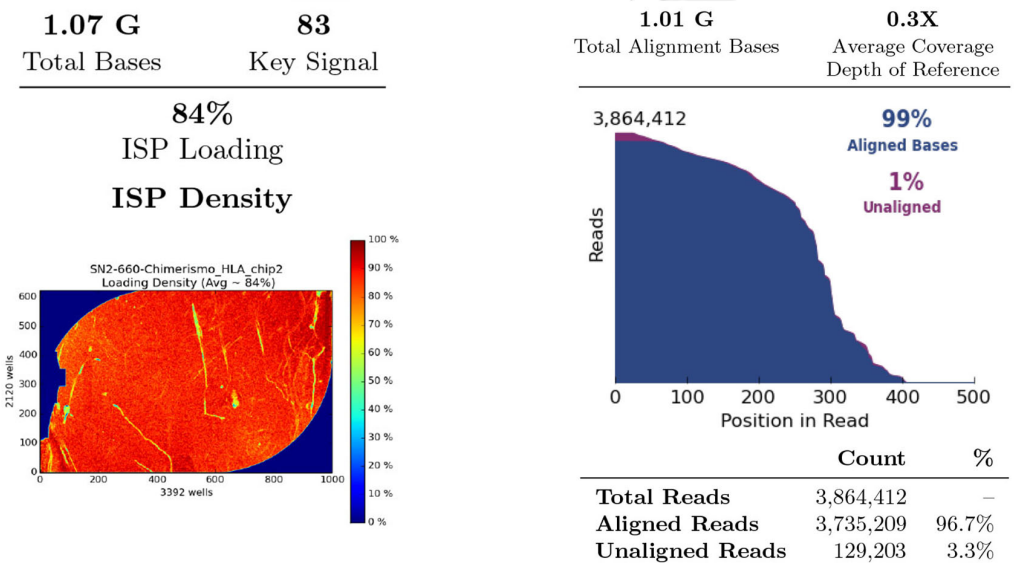

Figure 31. PGM report of the Ion custom library protocol sequenced.

The four libraries were analyzed with the Ion Torrent coverage analysis tool. Table 17 shows the average coverage for each block/amplicon. A total of 954900 reads were obtained for V01a library, 980204 for the V01b library, 848591 for the V02a library and 577440 for the V02b. All the 32 blocks both outside and inside the HLA were successfully sequenced. The least covered amplicon for all the four libraries was the PCR-3 and the most covered amplicon was the PCR-26. The PCR-3 was approximately 7 times less covered then the PCR-26 for the V01a library, 8,4 for the V01b library, 10,0 for the V02a library and 8,6 for the V02b library.

BLOCK ID	V01a	V01b	V02a	V02b
PCR-3	7087	6316	4692	3675
PCR-5	23049	23890	19426	14733
PCR-6	27294	23643	21519	13335
PCR-8	42872	46215	38286	25293
PCR-9	35488	36496	31744	21646
PCR-10	31949	36975	29644	19157
PCR-11	46484	52125	43942	30136
PCR-12	31940	35201	30469	21717
PCR-13	12240	13262	10149	7220
PCR-15	37162	39889	33434	24097
PCR-16	38391	40014	35023	24430
PCR-19	20737	23267	20508	13391
PCR-20	40467	44301	38789	24892
PCR-21	25872	27032	22426	16072
PCR-22	43216	31316	27731	18514
PCR-23	24974	26105	21778	15830
PCR-25	45931	46500	40815	26841
PCR-26	51332	53400	47136	31923
PCR-28	47192	50415	43059	29244
PCR-29	38986	28928	25238	17135
HLA-A-E2	33371	31513	27739	19071
HLA-A-E3	21595	19590	16774	11433
HLA-B-E2	23294	19040	16921	11586
HLA-B-E3	9302	8988	7969	5441
HLA-C-E2	7287	7493	7017	4332
HLA-C-E3	18306	18639	16816	11273
HLA-DP-E2	34572	35736	31809	22396
HLA-DP-E3	41399	42550	37646	25844
HLA-DQ-E2	28894	30179	26866	18070
HLA-DQ-E3	21300	22306	19515	12965
HLA-DR-E2	13587	28085	25753	16770
HLA-DR-E3	29330	30795	27958	18978

Table 17. Amplicon average coverage: output of the Ion Torrent coverage analysis tool.

4.2.4. Haplo-block NGS method performance

The haplo-block marker background generated by the ACCh_hb panel was estimated considering the 'donor' samples as chimera. Moreover, the Haplo-block NGS performance was compared with the first version of the method. For these purposes, the V01a and V02a Bam file was analysed with both Chimeval and Chimeval_hb tools.

For the pair Rec-V01a:Don-V02a:Chim-V02a, 67 informative SNPs were found by Chimeval. Moreover, 3 informative blocks inside the HLA (HLA-AE3, -BE2 and -BE3) and other 8 blocks outside the HLA (PCR-6, -9, -11, -15, -19, -20, -28 and -29) by Chimeval_hb were found (Tab 18).

For the pair Rec-V02a:Don-V01a:Chim-V01a, 49 informative SNPs were found by Chimeval. Moreover, 3 informative blocks inside the HLA (HLA-AE2, -AE3 and -BE3) and other 6 blocks outside the HLA (PCR-6, -11, -12, -19, -22, and -29) were found by Chimeval_hb (Tab 19).

For the pair **Rec**_V01a: **Don**_V02a: **Chim**_V02a, the average background calculated by Chimeval was 0,50 with 95% CI ranging from 0,62 to 0,39. Instead, the average background calculated by Chimeval_hb was 0,016 with 95% CI ranging from 0,007 to 0,023. The SNPs belonging to the block HLA-BE2 were excluded from the background calculation due to an alignment error of the sequencing reads against the hg19 template was found.

For the pair Rec-V02a:Don-V01a:Chim-V01a, the average background calculated by Chimeval was 0,355 with 95% CI ranging from 0,46 to 0,24. Instead, the average background calculated by Chimeval_hb was 0,006 with 95% CI ranging from 0,012 to 0,001. Also in this case, during the background evaluation performed by both the tools, the SNPs belonging to the block HLA-BE3 were excluded because of the reads alignment error.

The total background calculated on the informative SNPs derived from both the pairs with Chimeval was in average 0,445 with 95% CI ranging from 0,53 to 0,36.

The background calculated on 18 informative blocks derived from both the pairs with Chimeval_hb was in average 0,011 with 95% of IC ranging from 0,017 to 0,006.

Rec-V01a: **Don**-V02a: **Chim**-V02a				
BLOCK ID	**SNP ID**	**CHIMEVAL**	**AVERAGE**	**Chimeval_hb**
HLA-AE2	rs3173429	0,2		NI
HLA-AE3	rs199474485	0,2	0,46	0,016
	rs1059509	0,1		
	rs1059514	0,7		
	rs1136695	0,6		
	rs3173420	1,0		
	rs1059542	0,1		
	rs1059488	0,9		
	rs1059506	0,1		
	rs1059517	0,3		
	rs1059536	0,6		
HLA-BE2	rs1050388	1,4	4,51	1,583
	rs1131212	5,6		
	rs1131215	2,4		
	rs1050556	2,9		
	rs1050543	2,2		
	rs1050486	16,6		
	rs147324178	0,5		
HLA-BE3	rs2596493	0,2	0,78	0,032
	rs1131285	0,7		
	rs1131279	0,1		
	rs1131275	1,2		
	rs709053	1,5		
	rs3179865	1,0		
HLA-CE2	rs1050420	1,5		NI
HLA-DRB1E2	rs17879702	0,3		NI
	rs17884945	1,0		
PCR-6	rs10813408	0,5	0,77	0,012
	rs10970034	1,4		
	rs11521418	0,2		
	rs10970035	0,8		
	rs11534129	0,2		
	rs11532869	1,7		
	rs11548860	0,6		
PCR-8	rs149509653	0,2		NI
	rs532425127	0,3		
PCR-9	rs4758168	1,0	0,83	0,009
	rs4757986	1,2		
	rs1453435	0,3		
PCR-11	rs759888	0,2	0,22	0,005
	rs759890	0,0		
	rs2033041	0,4		
	rs1079365	0,0		
	rs759889	0,0		
	rs10843476	0,7		
PCR-13	rs1901817	0,1		NI
	rs1901818	0,1		
	rs1901819	0,4		

PCR-15	rs12922407	1,0	0,43	0
	rs12921979	0,3		
	rs12922020	0,0		
PCR-19	rs2251998	0,2	0,2	0,04
	rs2828739	0,2		
	rs34737404	0,2		
PCR-20	rs5748329	0,1	0,2	0,021
	rs5992433	0,5		
	rs5993715	0,0		
PCR-22	rs303846	0,1		NI
PCR-25	rs79259282	0,2		NI
PCR-28	rs12608877	0,2	0,46	0,005
	rs10853818	0,7		
	rs10853820	0,4		
	rs10853821	0,5		
	rs12608899	0,5		
PCR-29	rs2865944	0,7	0,9	0,017
	rs1202013	1,7		
	rs1202014	0,3		
		Average 0,507 95% CI:0,62-0,39		Average 0,016 95% CI:0,007-0,023

Table 18. Comparison of Chimeval and Chimeval_hb results on the DNA pair Rec-V01a: Don-V02a and V02a DNA as chimera. In the CHIMEVAL column, the Chimeval background result is shown for each SNP. In the Chimeval_hb column, the Chimeval_hb background result is shown for each Informative block. Moreover, the non informative blocks are indicated as Not Informative (NI). In the column AVERAGE, the average of the background detected by Chimeval is shown for each informative block. In red, the SNPs excluded from the background calculation.

Rec-V02a: **Don**-V01a: **Chim**-V01a				
BLOCK ID	**SNP ID**	**CHIMEVAL**	**AVERAGE**	**Chimeval_hb**
HLA-AE2	rs1136683	0,0	0,03	0
	rs1059455	0,0		
	rs199474424	0,0		
	rs199474436	0,1		
HLA-AE3	rs1136702	0,0	0,08	0
	rs199474485	0,1		
	rs1059509	0,0		
	rs1059514	0,2		
HLA-BE2	rs1050388	0,1		NI
	rs1131212	0,7		
HLA-BE3	rs2596493	1,0	0,57	0,179
	rs1131285	0,3		
	rs1131279	0,5		
	rs1131275	0,3		
	rs709053	0,2		
	rs3179865	1,5		
	rs1131235	0,2		
PCR-5	rs6463729	0,6		NI
	rs6463730	0,5		
PCR-6	rs11532869	0,3	0,27	0,016
	rs11534129	0,0		
	rs11548860	0,5		
PCR-9	rs4757987	0,2		NI
PCR-10	rs6489433	0,7		NI
PCR-11	rs759888	0,4	0,27	0
	rs759890	0,2		
	rs2033041	0,2		
PCR-12	rs4899091	0,6	0,43	0
	rs4899092	0,4		
	rs4899093	0,4		
	rs8009247	0,2		
	rs8009249	0,2		
	rs8009109	0,8		
PCR-13	rs1901819	0,1		NI
PCR-15	rs8053333	0,0		NI
PCR-19	rs2251998	0,0	0,3	0,015
	rs2828739	0,4		
	rs34737404	0,5		
PCR-22	rs303846	0,5	0,73	0,006
	rs303847	1,7		
	rs2600246	1,4		
	rs62247767	0,1		
	rs55955151	0,1		
	rs9878282	0,6		
PCR-25	rs12907809	0,3		NI
	rs7170825	0,9		
PCR-29	rs2865944	0,3	0,3	0,016
	rs1202013	0,2		
	rs1202014	0,4		
		Average 0,335		Average 0,007
		95% CI:0,46-0,24		95% CI:0,012-0,001

Table 19. Comparison of Chimeval and Chimeval_hb results on the DNA pair Rec-V02a: Don-V01a and V01a DNA as chimera.

5. DISCUSSION

During the allogeneic HSCT procedure, the patient is treated with the conditioning regimen (RIC or MAC) to suppress his/her own immune system with the purpose of avoiding the graft rejection and to eradicate the cancer cells in cases of malignancy. Therefore, at the day 0, when CD34+ cells are infused, the patient immune system results harshly compromised and the blood count test is routinely performed to verify the neutrophils and platelet engraftment [65]. Subsequently, since the allo-HSCT is successful when the donor HSCs carries out the complete IR, the chimerism analysis is used to monitor the donor immune system engraftment. Therefore there is no need for a high sensitivity technique. Although the STRs fluorescent amplification is a low sensitive (>5%) technique, due to the STR marker informativity and the availability of user-friendly commercial kits, it is widely used in diagnostic laboratories [92-96]. Today, despite the progress in allo-HSCT treatments which has led to a growth of the overall survival, the disease relapse remains the principal challenge [15, 53]. Small increases of the recipient chimerism give clinicians information about the disease relapse risk. Therefore, the use of high sensitive techniques, as qPCR, is fundamental when the patient specific MRD marker is lost [113, 145]. However, the amplification of the SNPs and of the INDELs by the qPCR lacks in informativity because of the marker biallelic nature [97].

Nowadays, the NGS technology have revolutionized the field of genomics and its application has been extended to different fields, such as clinical diagnostics and forensic science [142, 146, 147] . As a result of the continuous development of NGS, several applications previously performed on Sanger sequencing with capillary electrophoresis have been transferred onto the NGS platform, enabling fast and cost-effective generation sequence data with high resolution and accuracy. For this reason, different panels are being developed for the sequencing of genetic mutations involved in human diseases (e.g. MiSeqDx Cystic Fibrosis Clinical Sequencing Assay; Illumina, Inc., San Diego, CA, USA) or in cancer (Ion AmpliSeq BRCA1 and BRCA2 Panel; Life Technologies; Thermo Fisher Scientific, Inc.). In the field of

chimerism, Kim *et al* [148] briefly reported a relative quantification analysis of SNP markers by NGS in one human bone marrow chimerism sample. However, the study by Kim *et al* [148] was conducted without any detail concerning the limit of detection, the technical error or additional important technical information and validation data of NGS application in chimerism quantification.

As the first objective of this PhD thesis, a new diagnostic system to analyze the chimerism on NGS platform was proposed in order to combine a high informativity and an optimal sensitivity. A full workflow and a custom chimerism panel based on Ion AmpliSeq technology were designed. In addition, a bioinformatics tool dedicated to the genotyping and quantification of NGS data was coded. These resources were created in order to provide a novel tool for the evaluation of the chimerism following allogenic HSCT, thus potentially increasing the number of clinical analyses supported on NGS platforms. The ACCh proposed panel is composed of 44 amplicons, containing 44 selected mSNPs, of which 2 mSNPs are located in different regions of each somatic chromosome. It is suggested that the different mSNP locations in all somatic chromosome may be useful to avoid predominantly false negative results caused by chromosomal deletions, specific of certain malignancies [149]. In addition, the Bed file uploaded in the custom tool, containing all targeted SNPs (a total of 286 SNPs), can be modified in order to exclude the SNPs present in chromosome target regions subjected to deletions in a specific patient. For a more precise and robust quantification, the panel average heterozygosity was assessed around 0,5 for the European population in order to obtain different informative markers for each transplanting pair. The theoretical informativity of the panel for unrelated donor:recipient values, computed according to the data present on the NCBI dbSNPs database, was estimated to be approximately 16/42 mSNPs, while for siblings the informativity was estimated to be 50,0% (approximately 8/42 mSNPs). It was confirmed experimentally on a total of 14 unrelated and 6 sibling couples. Moreover, the ACCh panel informativity was improved by adding to the Bed file other 242 SNPs present in the targeted regions (Tab 6). The ACCh panel showed linearity with the artificial chimera reference values ranging from 0,5 to 100%. The achieved limit of detection on the Ion Torrent PGM platform was 0,5%. However, this number was updated to the conservative value of 1,0% for two reasons: i) The Ion Torrent error is defined to be between 0,04 and 1,0% [150]; and ii) the background of the ACCh

panel, based on the IRA data of our experiment, was ranging from 0,1 to 0,5%. The CA accuracy detected for chimeric samples in the range 1-20% resulted optimal and for those up to 99% it was acceptable. A comparison between the NGS method and the gold standard STR method was tested on artificial chimeras ranging from 0,5 to 100%. The linearity between the two methods was detected in the range from 4 to 100%. The chimeric samples under 4% were detected only by the NGS method, proving its better sensitivity. Furthermore, the NGS and STRs methods performance were compared on a total of 20 clinical samples. Six of them were collected at different points of a patient follow-up and the other 14 from 8 patients during the "blind test validation". The results obtained in 7/20 samples were discordant between the methods. These samples were detected as mixed chimerism by the NGS method and as complete chimerism by the STRs one. The discordant samples were analyzed by a third technique to evaluate their chimeric status. For this purpose, an in house protocol based on the detection of the AMGY marker by Real Time PCR was used [131]. It was applied only on 5/7 samples because the recipient was male and the donor female. A mixed chimerism status was detected in all the 5 samples, confirming the better NGS sensitivity. Moreover, the NGS method passed the UKNEQAS external quality program validation, resulting suitable for diagnostic use. Indeed, all the 8/8 samples analyzed were in the satisfactory phase.

In 2014, Debeljak *et al*. [137] described, for the first time, the haplo-block marker as proof of concept to detect the chimerism post allogeneic-HSCT with high sensitivity (0,01%). This marker is composed by different SNPs closed within the same amplicon and it is useful to recognize the single base error (background). Subsequently, in 2017, the same authors developed a panel composed by 8 haplo-blocks, extending the application of the ultrasensitive chimerism detection to CEU, JPT and YRI populations [151]. In this study, they located two of the 8 haplo-blocks in the HLA locus (HLA-A and -B genes) with the sole aim of increasing the panel informativity. However, in the discussion, they mentioned the potential application of two HLA markers in the HLA-loss relapse recognition [151]. Nowadays, the HLA-loss relapse could be detected by the comparative quantification of patient specific-genomic markers located outside and inside the HLA locus. Separated techniques were used to this end. Specifically, the chimerism performed with the STR analysis was applied to the outsider HLA markers and the SSOP or the SNP arrays to the insider one. The

sensitivity of these techniques is limited to 5% and the HLA-loss relapse could be detected only on purified leukemic blasts that could be enriched only in cases of overt relapse [75].

Instead, to early recognize the HLA loss relapse, Ahci *et al* [127] developed a tool based on high sensitivity qPCR, useful to perform a comparative chimerism quantification of markers located inside and outside the HLA. They designed 10 qPCR reactions specific for 27 HLA-A, -C, -DPB1 allele groups. These three loci were selected because they are often mismatched in haplo- and matched unrelated HSCT, cases in which the HLA-loss relapse is reported [73, 152-154]. However, despite the high sensitivity reached by the Ahci tool (0,16% of the recipient cells), the informativity tested on 454-haplo-HSCT and on 113 MUD HSCT, was respectively, only 70 % and 66,4%.

As second objective of the thesis, an upgrade of the NGS method to decrease the NGS background with a view to 0,1% of sensitivity was proposed. To this end, 29 haplo-blocks with high informativity in the EUR, EAS, ALL_p 1000genomes populations were selected to build the ACCh haplo-block panel. Moreover, since the third thesis objective was to extend the application of the NGS method to the recognition of the HLA-loss relapse in all the transplant pairs, 12 hyper-variable HLA blocks were added to the panel. The added HLA blocks were the exons 2 and 3 of the genes HLA-A, -B, -C, -DPB1, -DQB1 and -DRB1. For their amplification, the primers described by Lange *et al* [138] were used. Moreover, about the non-HLA blocks, 29 primer pairs meeting the requirements to work in multiplex PCR with the HLA primers were designed. Subsequently, a custom library protocol composed by two PCR steps was developed. In order to set-up a multiplex PCR protocol to amplify all the ACCh haplo-blocks during the first PCR step, the specificity of the 41 haplo-blocks was preliminary tested by the singleplex PCR protocol. Nine non-HLA haplo-blocks were discarded from the panel due to their non-specific amplification. According to the binomial analysis, the remaining 20 non-HLA markers ensure 100% probability to find at least three informative markers in unrelated pairs for all the population considered. Instead, when related pairs are taken in account, this probability is reduced to 97,7% for EUR, 90,9% for EAS and 96,5% for ALL_p. Anyway, the 20 non-HLA blocks assured the informativity of at least one marker in 100% of the

related pairs. However, since the UKNEQAS consortium recommends a minimum of 3 informative markers for a reliable chimerism analysis [95] and since the HLA loss relapse was described in 33% of the haplo-identical (related) transplants [71], we need to increase the number of the non HLA blocks in the ACCh_hb panel to ensure at least the 3 informative markers outside the HLA in 100% of the related pairs. The first PCR step was settled with 100 ng of input DNA with a view to reach 0,1% of sensitivity. One hundred Nano-grams of genomic DNA, that correspond to approximately 16700 diploid cells (about 6 pg/diploid genome) [151], were used to largely exceed (≈16x) the number of cells required to reach 0,1% of sensitivity (1:1000 cells). Four libraries (V01a, V01b, V02a and V02b) were sequenced with the custom library protocol. The sequencing performance in term of Ion sphere particles loaded, total bases sequenced and total reads obtained were similar between the custom library protocol and the classical AmpliSeq one. The huge advantage of the custom protocol, compared to the classical one, is the lower cost. The cost per NGS library with the standard AmpliSeq Library Preparation kit is of 70-90 Euros; instead, it drops to 2-3 Euros when the custom library protocol is used. The developed multiplex protocol was successfully settled; indeed, all the 32 haplo-blocks (inside and outside the HLA) were covered in all the four sequenced libraries. The V01a and V02a Bam file was analyzed with Chimeval and Chimeval_hb. About the informativity, 67 and 49 informative SNPs were found, respectively, in the two unrelated pairs analyzed with the ACCh_hb panel. Instead, an average of 16 informative SNPs was found in unrelated pairs with the ACCh panel. Therefore, despite the fact that the ACCh panel_hb was composed by a lower number of amplicons (32) compared to the ACCh one (44), it was characterized by a higher informativity. The advantage of a reduced number of amplicons per library is the possibility of sequencing more samples per chip. About the ACCh_hb panel informativity, more than 3 blocks inside and outside the HLA were informative in both the pairs analyzed. Therefore, it could potentially be used to recognize the HLA loss-relapse, but further pairs will be analyzed to confirm the data. The average total ACCh_hb background calculated on SNPs by Chimeval was 0,445 with 95% of IC 0,53-0,36. Moreover, it calculated on Haplo-blocks by Chimeval_hb was in average 0,011 with 95% of IC 0,017-0,006. Therefore, in the new NGS method, it is decreased 40 times from the previous version. This was about 10 ten times lower than the desired sensitivity (0,1%), but a full work on artificial chimera will be done to

experimentally measure the sensitivity. During the background evaluation, the Haplo-block HLA-BE2 and HLA-BE3 were discarded due to alignment errors of the raw reads against the hg19 template. For this reason, we are investigating how implement the Chimeval_hb with a new alignment strategy for the HLA blocks in order to recognize the HLA-loss relapse.

Subsequently, we will validate the method sensitivity by analyzing post-HSCT clinical samples and comparing the data with the gold standard technique for the sensitivity: the Real-Time PCR. A further validation will be done to test the ability of the new method in HLA-relapse recognition by analysing several post-transplant DNA samples derived from haplo-identical transplanted patients.

Regarding the timing of chimerism analysis, the UKNEQAS Consortium has recommended that results should be assessed in 5 working days from the reception of the sample and in 3 working days for urgent requests [90]. Both the methods suggested are feasible in 2 days. The first day for library preparation and quantification and the second for template preparation, run sequence and data analysis. Concerning the custom bioinformatics tools, any Bam file coupled by a Bed file, generated from any platform could be used. At present, the cost of NGS analysis, compared with microsatellite methods, remains high. However considering the continuously reducing cost per NGS run, an NGS-based method for chimerism quantification could be adopted in laboratories with a high volume of activity and with NGS platforms already in use for other purposes.

6. CONCLUSIONS

The overall objective of this PhD thesis was the technical development of a novel method based on NGS technology to evaluate the chimerism with a view to improve the hematopoietic stem cell transplant success.

Firstly, a full workflow of the NGS method composed of the ACCh SNPs panel and of the Chimeval tool to analyze the data was described. The technical performance in term of sensitivity (1%) resulted higher then the gold standard STRs method. The NGS method was also accredited for diagnostic use as it satisfactorily passed the UKNEQAS chimerism external quality program.

Subsequently, an upgrade of the method was done to decrease the NGS background with a view to reach a 0,1% of sensitivity and to extend its application in the HLA-relapse recognition. The new method, composed of the ACCh haplotype block panel and of the Chimeval_hb tool, will be validated first on artificial chimera and then on post-HSCT samples to fully describe its sensitivity and its application in HLA-loss relapse recognition. Therefore, we can conclude that after further investigations, our NGS haplotype block method could promptly recognize the graft rejection, the risk of relapse and the HLA-loss relapse together.

7. BIBLIOGRAPHY

1. Abbas A, Lichtman A, and Pillai S. Cellular and molecular immunology. 6a Ed, (2007), W. B. Saunders Company, Philadelphia, USA. pp 57; 97-111; 379-393.
2. Greco R, Bondanza A, Oliveira MC, Badoglio M, Burman J, Piehl F, Hagglund H, Krasulova E, Simões BP, Carlson K, Pohlreich D, Labopin M, Saccardi R, Comi G, Mancardi GL, Bacigalupo A, Ciceri F, Farge D. Autologous hematopoietic stem cell transplantation in neuromyelitis optica: a registry study of the EBMT Autoimmune Diseases Working Party. Mult Scler. 2015;21(2):189-197.
3. Thomas ED, Lochte HL, Jr., Lu WC, Ferrebee JW. Intravenous infusion of bone marrow in patients receiving radiation and chemotherapy. N Engl J Med. 1957;257(11):491-496.
4. Appelbaum FR. Hematopoietic-cell transplantation at 50. N Engl J Med. 2007;357(15):1472- 1475.
5. Thomas ED, Storb R, Clift RA, Fefer A, Johnson L, Neiman PE, Lerner KG, Glucksberg H, Buckner CD. Bone-marrow transplantation (second of two parts). N Engl J Med. 1975;292(17):895-902.
6. Thomas E, Storb R, Clift RA, Fefer A, Johnson FL, Neiman PE, Lerner KG, Glucksberg H, Buckner CD Bone-marrow transplantation (first of two parts). N Engl J Med. 1975;292(16):832-843.
7. Weiden PL, Flournoy N, Thomas ED, Prentice R, Fefer A, Buckner CD, Storb R. Antileukemic effect of graft-versus-host disease in human recipients of allogeneic-marrow grafts. N Engl J Med. 1979;300(19):1068-1073.
8. Sheridan WP, Begley CG, Juttner CA, Szer J , A LB , Maher D , McGrath KM , Morstyn G , Fox RM. Effect of peripheral-blood progenitor cells mobilised by filgrastim (G-CSF) on platelet recovery after high-dose chemotherapy. Lancet. 1992;339(8794):640-644.
9. Welte K, Gabrilove J, Bronchud MH, Platzer E, Morstyn G. Filgrastim (r-metHuG-CSF): the first 10 years. Blood. 1996;88(6):1907-1929.
10. Appelbaum FR, Forman SJ, Negrin RS, and Blume KG. Thomas' Hematopoietic Cell Transplantation: Stem Cell Transplantation. 5a Ed. (2016), Blackwell Publishing Ltd, UK. pp 9, 112-125; 272-280.
11. Majhail NS, Chitphakdithai P, Logan B, King R, Devine S, Rossmann SN, Hale G, Hartzman RJ, Karanes C, Laport GG, Nemecek E, Snyder EL, Switzer GE, Miller J, Navarro W, Confer DL, Levine JE. Significant improvement in survival after unrelated donor hematopoietic cell transplantation in the recent era. Biol Blood Marrow Transplant. 2015;21(1):142-150.
12. Passweg JR, Baldomero H, Bader P, Basak GW, Bonini C, Duarte R, Dufour C, Kröger N, Kuball J, Lankester A, Montoto S, Nagler A, Snowden JA, Styczynski J, Mohty M; European Society for Blood and Marrow Transplantation (EBMT). Is the use of unrelated donor transplantation leveling off in Europe? The 2016 European Society for Blood and Marrow Transplant activity survey report. Bone Marrow Transplant. 2018;53(9):1139-1148.
13. http://www.gitmo.it/images/pdf/segreteria/presentazioni/Report2016.pdf

14. Scott EC, Hari P, Sharma M,Scott EC, Hari P, Sharma M, Le-Rademacher J , Huang J , Vogl D , Abidi M , Beitinjane A, Fung H, Ganguly S, Hildebrandt G, Holmberg L, Kalaycio M, Kumar S, Kyle R, Lazarus H, Lee C , Maziarz RT, Meehan K, Mikhael J, Nishihori T, Ramanathan M, Usmani S, Tay J , Vesole D , Wirk B , Yared J, Savani BN, Gasparetto C, Krishnan A, Mark T, Nieto Y, D'Souza A. Post-Transplant Outcomes in High-Risk Compared with Non-High-Risk Multiple Myeloma: A CIBMTR Analysis. Biol Blood Marrow Transplant. 2016;22(10):1893-1899.
15. Gibson CJ, Lindsley RC, Tchekmedyian V, Mar BG, Shi J, Jaiswal S, Bosworth A, Francisco L, He J, Bansal A, Morgan EA, Lacasce AS, Freedman AS, Fisher DC, Jacobsen E, Armand P, Alyea EP, Koreth J, Ho V, Soiffer RJ, Antin JH, Ritz J, Nikiforow S, Forman SJ, Michor F, Neuberg D, Bhatia R, Bhatia S, Ebert BL. Clonal Hematopoiesis Associated With Adverse Outcomes After Autologous Stem-Cell Transplantation for Lymphoma. J Clin Oncol. 2017;35(14):1598-1605.
16. www.cibmtr.org
17. Okamoto S. Current indication for hematopoietic cell transplantation in adults. Hematol Oncol Stem Cell Ther. 2017;10(4):178-183.
18. Armand P, Gibson CJ, Cutler C, Ho VT, Koreth J, Alyea EP, Ritz J, Sorror ML, Lee SJ, Deeg HJ, Storer BE, Appelbaum FR, Antin JH, Soiffer RJ, and Kim HT. A disease risk index for patients undergoing allogeneic stem cell transplantation. Blood. 2012; 120(4): 905–913.
19. Anasetti C, Hansen JA. Effect of HLA incompatibility in marrow transplantation from unrelated and HLA-mismatched related donors. Transfus Sci. 1994;15:221–230.
20. O'Donnell PV, Luznik L, Jones RJ , Vogelsang GB, Leffell MS, Phelps M, Rhubart P, Cowan K, Piantados S, Fuchs EJ. Nonmyeloablative bone marrow trans- plantation from partially HLA-mismatched related donors using posttransplanta- tion cyclophosphamide. Biol Blood Marrow Transplant. 2002;8:377–386.
21. Kernan NA, Flomenberg N, Dupont B, O'Reilly RJ. Graft rejection in recipients of T-cell-depleted HLA-nonidentical marrow transplants for leukemia. Identification of host-derived antidonor allocytotoxic T lymphocytes. Transplantation. 1987;43:842–847.
22. Finke J, Schmoor C, Bethge WA, Ottinger HD, Stelljes M, Zander AR, Volin L, Heim DA, Schwerdtfeger R, Kolbe K, Mayer J, Maertens JA, Linkesch W, Holler E, Koza V, Bornhäuser M, Einsele H, Bertz H, Grishina O, Socié G; ATG-Fresenius Trial Group. Prognostic factors affecting outcome after allogeneic transplantation for hematological malignancies from unrelated donors: results from a randomized trial. Biol Blood Marrow Transplant. 2012;18(11):1716-1726.
23. Kollman C, Howe CW, Anasetti C, Antin JH, Davies SM, Filipovich AH, Hegland J, Kamani N, Kernan NA, King R, Ratanatharathorn V, Weisdorf D, Confer DL. Donor characteristics as risk factors in recipients after transplantation of bone marrow from unrelated donors: the effect of donor age. Blood. 2001;98(7):2043-2051.
24. Flowers ME, Inamoto Y, Carpenter PA, Lee SJ, Kiem HP, Petersdorf EW, Pereira SE, Nash RA, Mielcarek M, Fero ML, Warren EH, Sanders JE, Storb RF, Appelbaum FR, Storer BE, Martin PJ. Comparative analysis of risk factors for acute graft-versus-host disease and for chronic graft-versus-host disease according to National Institutes of Health consensus criteria. Blood. 2011;117(11):3214-3219.

25. Rammensee HG. Chemistry of peptides associated with MHC class I and class II molecules. Curr Opin Immunol. 1995;7:85–96.
26. https:www.ebi.ac.uk-ipd-imgt-hla
27. Marsh SG, Albert ED, Bodmer WF, Bontrop RE, Dupont B, Erlich HA, Fernández-Viña M, Geraghty DE, Holdsworth R, Hurley CK, Lau M, Lee KW, Mach B, Maiers M, Mayr WR, Müller CR, Parham P, Petersdorf EW, Sasazuki T, Strominger JL, Svejgaard A, Terasaki PI, Tiercy JM, Trowsdale J. Nomenclature for factors of the HLA system, 2010. Tissue Antigens. 2010;75(4):291-455.
28. http://www.hla.alleles.org/nomenclature/naming.html
29. Gluckman É, Niederwieser D, Aljurf M. Establishing a Hematopoietic Stem Cell Transplantation Unit. A pratical Guide. 1a Ed, (2018), Springer International Publishing, CH. pp 111-128; 129-139.
30. Robinson J, Halliwell JA, James D. Hayhurst JD, Flicek P, Parham P, E. Marsh SGE. The IPD and IMGT/HLA database: allele variant databases. Nucleic Acids Res. 2015;43(Database issue): D423–D431.
31. Lind C, Ferriola D, Mackiewicz K, Heron S, Rogers M, Slavich L, Walker R, Hsiao T, McLaughlin L, D'Arcy M, Gai X, Goodridge D, Sayer D, Monos D. Next-generation sequencing: the solution for high-resolution, unambiguous human leukocyte antigen typing. Hum Immunol. 2010;71(10):1033-1042.
32. Tiercy JM. How to select the best available related or unrelated donor of hematopoietic stem cells? Haematologica. 2016;101(6):680–687.
33. https:--ibmdr.it
34. Lee SJ, Klein J, Haagenson M, Baxter-Lowe LA, Confer DL, Eapen M, Fernandez-Vina M, Flomenberg N, Horowitz M, Hurley CK, Noreen H, Oudshoorn M, Petersdorf E, Setterholm M, Spellman S, Weisdorf D, Williams TM, Anasetti C. High-resolution donor-recipient HLA matching contributes to the success of unrelated donor marrow transplantation. Blood. 2007;110(13):4576-4583.
35. Aversa F, Terenzi A, Tabilio A, Falzetti F, Carotti A, Ballanti S, Felicini R, Falcinelli F, Velardi A, Ruggeri L, Aloisi T, Saab JP, Santucci A, Perruccio K, Martelli MP, Mecucci C, Reisner Y , Martelli MF. Full haplotype-mismatched hematopoiet- ic stem-cell transplantation: a phase II study in patients with acute leukemia at high risk of relapse. J Clin Oncol. 2005;23:3447-3454.
36. Koh LP, Rizzieri DA, Chao NJ. Allogeneic hematopoietic stem cell transplant using mismatched/haploidentical donors. Biol Blood Marrow Transplant. 2007;13:1249-1267.
37. Apperley J, Niederwieser D, Huang X-J, Nagler A, Fuchs E, Szer J, Kodera Y. Haploidentical Hematopoietic Stem Cell Transplantation: A Global Overview Comparing Asia, the European Union, and the United States. Biol. Blood Marrow Transplant. J. Am. Soc. Blood Marrow Transplant. 2016;22(1):23– 26.
38. Ruiz-Argüelles GJ, Gómez-Almaguer D. Breaking dogmata to help patients: non-myeloablative haematopoietic stem cell transplantation. Expert Opin Biol Ther. 2004;4(10):1693-1699.
39. Gyurkocza B, Sandmaier BM. Conditioning regimens for hematopoietic cell transplantation: one size does not fit all. Blood. 2014;124(3):344-353.

40. Reshef R, Porter DL. Reduced-intensity conditioned allogeneic SCT in adults with AML. Bone Marrow Transplant. 2015;50(6):759-769.
41. Bosi A, Bartolozzi B. Safety of bone marrow stem cell donation: a review. Transplant Proc. 2010;42(6):2192-2194.
42. Baldomero H, Gratwohl M, Gratwohl A, Tichelli A, Niederwieser D, Madrigal A Frauendorfer K .The EBMT activity survey 2009: trends over the past 5 years. Bone Marrow Transplant. 2011;46(4):485-501.
43. Stroncek DF, Clay ME, Petzoldt ML, Smith J, Jaszcz W, Oldham FB McCullough J. Treatment of normal individuals with granulocyte-colony-stimulating factor: donor experiences and the effects on peripheral blood CD34+ cell counts and on the collection of peripheral blood stem cells. Transfusion. 1996;36(7):601-610.
44. Gertow J, Berglund S, Okas M, Uzunel M, Berg L, Karre K Mattsson J, Uhlin M. Characterization of long-term mixed donor-donor chimerism after double cord blood transplantation. Clin Exp Immunol. 2010;162(1):146-155.
45. Ogonek J, Kralj Juric M, Ghimire S, Varanasi PR, Holler E, Greinix H, Weissinger E. Immune Reconstitution after Allogeneic Hematopoietic Stem Cell Transplantation. Front Immunol. 2016;7:507.
46. Van den Brink MR, Velardi E, Perales MA. Immune reconstitution following stem cell transplantation. Hematology Am Soc Hematol Educ Program. 2015;2015:215-219.
47. Chaudhry MS, Velardi E, Malard F, van den Brink MR. Immune Reconstitution after Allogeneic Hematopoietic Stem Cell Transplantation: Time To T Up the Thymus. J Immunol. 2017;198(1):40-46.
48. Storek J. Immunological reconstitution after hematopoietic cell transplantation - its relation to the contents of the graft. Expert Opin Biol Ther. 2008;8(5):583-597.
49. Ringden O, Karlsson H, Olsson R, Omazic B, Uhlin M. The allogeneic graft-versus-cancer effect. Br J Haematol 2009;147(5): 614-633.
50. Dickinson AM, Norden J, Li S, Hromadnikova I, Schmid C, Schmetzer H Jochem-Kolb H. Graft-versus-Leukemia Effect Following Hematopoietic Stem Cell Transplantation for Leukemia. Front Immunol 2017;8:496.
51. Castagna L, Sarina B, Bramanti S, Perseghin P, Mariotti J, Morabito L. Donor lymphocyte infusion after allogeneic stem cell transplantation. Transfus Apher Sci 2016;54:345 – 355.
52. Roddie C, Peggs KS. Donor lymphocyte infusion following allogeneic hematopoietic stem cell transplantation. Expert Opin Biol Ther. 2011;11(4):473-487.
53. Scott EC, Hari P, SharmaM, Le-Rademacher J, Huang J, Vogl D, Abidi M, Beitinjaneh A, Fung H, Ganguly S, Hildebrandt G, Holmberg L, Kalaycio M, Kumar S, Kyle R, Lazarus H, Lee C, Maziarz RT, Meehan K, Mikhael J, Nishihori T, Ramanathan M, Usmani S, Tay J, Vesole D, Wirk B, Yared J, Savani BN, Gasparetto C, Krishnan A, Mark T, Nieto Y, D'Souza A. Post-Transplant Outcomes in High-Risk Compared with Non-High-Risk Multiple Myeloma: A CIBMTR Analysis. Biol Blood Marrow Transplant. 2016;22(10):1893-1899.

54. Ljungman P, Perez-Bercoff L, Jonsson J, Avetisyan G, Sparrelid E, Aschan J, Barkholt L, Larsson K, Winiarski J, Yun Z, Ringdén O. Risk factors for the development of cytomegalovirus disease after allogeneic stem cell transplantation. Haematologica. 2006;91(1):78-83.
55. Uhlin M, Wikell H, Sundin M, Blennow O, Maeurer M, Ringden O, Winiarski J, Ljungman P, Remberger M, Mattsson J. Risk factors for Epstein-Barr virus-related post-transplant lymphoproliferative disease after allogeneic hematopoietic stem cell transplantation. Haematologica 2014;99(2):346-352.
56. Fuji S, Loffler J, Einsele H, Kapp M. Immunotherapy for opportunistic infections: Current status and future perspectives. Virulence 2016;7(8): 939-949.
57. Ferrara JL, Levine JE, Reddy P, Holler E. Graft-versus-host disease. Lancet. 2009;373:1550–1561.
58. Jagasia M, Arora M, Flowers ME, Chao NJ, McCarthy PL, Cutler CS, Urbano-Ispizua A, Pavletic SZ, Haagenson MD, Zhang MJ, Antin JH, Bolwell BJ, Bredeson C, Cahn JY, Cairo M, Gale RP, Gupta V, Lee SJ, Litzow M, Weisdorf DJ, Horowitz MM, Hahn T. Risk factors for acute GVHD and survival after hematopoietic cell transplantation. Blood. 2012;119:296–307.
59. Billingham RE. The biology of graft-versus-host reactions. Harvey Lect 1966; 62: 21-78.
60. Zeiser R, Blazar BR. Acute Graft-versus-Host Disease - Biologic Process, Prevention, and Therapy. N Engl J Med. 2017;377(22):2167-2179.
61. Weiden PL, Flournoy N, Thomas ED, Prentice R, Fefer A, Buckner CD Storb R. Antileukemic effect of graft-versus-host disease in human recipients of allogeneic-marrow grafts. N Engl J Med 1979;300(19):1068-1073.
62. Frey NV, Porter DL. Graft-versus-host disease after donor leukocyte infusions: presentation and management. Best Pract Res Clin Haematol. 2008;21(2):205-222.
63. Filipovich AH, Weisdorf D, Pavletic S, Socie G, Wingard JR, Lee SJ, Martin P, Chien J, Przepiorka D, Couriel D, Cowen EW, Dinndorf P, Farrell A, Hartzman R, Henslee-Downey J, Jacobsohn D, McDonald G, Mittleman B, Rizzo JD, Robinson M, Schubert M, Schultz K, Shulman H, Turner M, Vogelsang G, Flowers ME. National Institutes of Health consensus development project on criteria for clinical trials in chronic graft-versus-host disease: I. Diagnosis and staging working group report. Biol Blood Marrow Transplant 2005;11:945–956.
64. Locatelli F, Lucarelli B, Merli P. Current and future approaches to treat graft failure after allogeneic hematopoietic stem cell transplantation. Expert Opin Pharmacother. 2014;15(1):23-36.
65. Kim HT, Armand P. Clinical endpoints in allogeneic hematopoietic stem cell transplantation studies: the cost of freedom. Biol Blood Marrow Transplant. 2013;19(6):860-866.
66. Olsson R, Remberger M, Schaffer M, Berggren DM, Svahn BM, Mattsson J Olsson R1, Remberger M, Schaffer M, Berggren DM, Svahn BM, Mattsson J, Ringden O.. Graft failure in the modern era of allogeneic hematopoietic SCT. Bone Marrow Transplant 2013;48(4):537-543.
67. Van den Brink MR, Porter DL, Giralt S, Lu SX, Jenq RR, Hanash A, Bishop MR. Relapse after allogeneic hematopoietic cell therapy. Biol Blood Marrow Transplant 2010;16(1 Suppl):S138-145.
68. Barrett AJ. Understanding and harnessing the graft-versus-leukaemia effect. Br J Haematol. 2008; 142(6):877–888.

69. Murata M, Ishikawa Y, Ohashi H, Terakura S, Ozeki K, Kiyoi H, Naoe T. Donor cell leukemia after allogeneic peripheral blood stem cell transplantation: a case report and literature review. Int J Hematol. 2008;88(1):111–115.
70. Ruiz-Arguelles GJ, Ruiz-Arguelles A, Garces-Eisele J. Donor cell leukemia: a critical review. Leuk Lymphoma. 2007;48(1):25–38.
71. Crucitti L, Crocchiolo R, Toffalori C, Mazzi B, Greco R, Signori A, Sizzano F, Chiesa L, Zino E, Lupo Stanghellini MT, Assanelli A, Carrabba MG, Marktel S, Marcatti M, Bordignon C, Corti C, Bernardi M, Peccatori J, Bonini C, Fleischhauer K, Ciceri F, Vago L. Incidence, risk factors and clinical outcome of leukemia relapses with loss of the mismatched HLA after partially incompatible hematopoietic stem cell transplantation. Leukemia. 2015;29(5):1143-1152.
72. Passweg JR, Baldomero H, Bader P, Bonini C, Duarte RF, Dufour C, Gennery A, Kröger N, Kuball J, Lanza F, Montoto S, Nagler A, Snowden JA, Styczynski J13, Mohty M. Use of haploidentical stem cell transplantation continues to increase: the 2015 European Society for Blood and Marrow Transplant activity survey report. Bone Marrow Transplant. 2017;52(6):811-817.
73. Cieri N, Greco R, Crucitti L, Morelli M, Giglio F, Levati G, Assanelli A, Carrabba MG, Bellio L, Milani R, Lorentino F, Stanghellini MT, De Freitas T, Marktel S, Bernardi M, Corti C, Vago L, Bonini C, Ciceri F, Peccatori J. Post-transplantation Cyclophosphamide and Sirolimus after Haploidentical Hematopoietic Stem Cell Transplantation Using a Treosulfan-based Myeloablative Conditioning and Peripheral Blood Stem Cells. Biol Blood Marrow Transplant. 2015;21(8):1506-1514.
74. Raghavan M, Smith LL, Lillington DM, , Chaplin T, Kakkas I, Molloy G, Chelala C, Cazier JB, Cavenagh JD, Fitzgibbon J, Lister TA, Young BD. Segmental uniparental disomy is a commonly acquired genetic event in relapsed acute myeloid leukemia. Blood. 2008;112(3):814-821.
75. Vago L, Perna SK, Zanussi M, Mazzi B, Barlassina C, Stanghellini MT, Perrelli NF, Cosentino C, Torri F, Angius A, Forno B, Casucci M, Bernardi M, Peccatori J, Corti C, Bondanza A, Ferrari M, Rossini S, Roncarolo MG, Bordignon C, Bonini C, Ciceri F, Fleischhauer K. Loss of mismatched HLA in leukemia after stem-cell transplantation. N Engl J Med. 2009;361(5):478-488.
76. Horowitz M, Schreiber H, Elder A, Heidenreich O, Vormoor J, Toffalori C, Vago L, Kröger N. Epidemiology and biology of relapse after stem cell transplantation. Bone Marrow Transplant. 2018.
77. David B, Bernard D, Navenot JM, Muller JY, Blanchard D. Flow cytometric monitoring of red blood cell chimerism after bone marrow transplantation. Transfus Med. 1999;9(3):209-217.
78. Buño I, Díez-Martín JL, López-Fernández C, Fernández JL, Gosálvez J. Polymorphisms for the size of heterochromatic regions allow sex-independent quantification of post-BMT chimerism targeting metaphase and interphase cells. Haematologica. 1999;84(2):138-141.
79. Van den Berg H, Vossen JM, Langlois van den Bergh R, Bayer J, van Tol MJ. Detection of Y chromosome by in situ hybridization in combination with membrane antigens by two-color immunofluorescence. Lab Invest. 1991;64(5):623-628.
80. Khan F, Agarwal A, Agrawal S. Significance of chimerism in hematopoietic stem cell transplantation: new variations on an old theme. Bone Marrow Transplant. 2004;34(1):1-12.

81. Thiede C. Diagnostic chimerism analysis after allogeneic stem cell transplantation: new methods and markers. Am J Pharmacogenomics. 2004;4(3):177-187.
82. Morisaki H, Morisaki T, Nakahori Y, Ogura H, Kanno H, Tani K, Kodo H, Fujii H, Asano S, Miwa S. Genotypic analysis using a Y-chromosome-specific probe following bone marrow transplantation. Am J Hematol. 1988;27(1):30-33.
83. Feuk L, Carson AR, and Scherer SW. Structural variation in the human genome. Nat. Review. (2006) 7:85-97.
84. Scherer SW, Lee C, Birney E, Altshuler DM, Eichler EE, Carter NP, Hurles ME, and Feuk L. Challenges and standards in integrating surveys of structural variation. Nat. Genetics. (2007) 39:S7-S15.
85. Beckman JS, Weber JL. Survey of human and rat microsatellites. Genomics. 1992;12(4):627-631.
86. Thiede C, Bornhäuser M, Ehninger G. Evaluation of STR informativity for chimerism testing--comparative analysis of 27 STR systems in 203 matched related donor recipient pairs. Leukemia. 2004;18(2):248-254.
87. Romsos EL, Vallone PM. Rapid PCR of STR markers: Applications to human identification. Forensic Sci Int Genet. 2015;18:90-99.
88. Mills RE, Luttig CT, Larkins CE, Beauchamp A, Tsui C, Pittard WS, Devine SE. An initial map of insertion and deletion (INDEL) variation in the human genome. Genome Res. 2006;16(9):1182-1190.
89. Stoneking M. Single nucleotide polymorphisms. From the evolutionary past... Nature. 2001;409(6822):821-822.
90. Clark JR, Scott SD, Jack AL, Lee H, Mason J, Carter GI, Pearce L, Jackson T, Clouston H, Sproul A, Keen L, Molloy K, Folarin N, Whitby L, Snowden JA, Reilly JT, Barnett D; United Kingdom National External Quality Assessment Service for Leucocyte Immunophenotyping Chimerism Working Group. Monitoring of chimerism following allogeneic haematopoietic stem cell transplantation (HSCT): technical recommendations for the use of short tandem repeat (STR) based techniques, on behalf of the United Kingdom National External Quality Assessment Service for Leucocyte Immunophenotyping Chimerism Working Group. Br J Haematol. 2015;168(1):26-37.
91. Agrawal S, Khan F, Talwar S, Nityanand S. Short tandem repeat technology has diverse applications: individual identification, phylogenetic reconstruction and chimerism based post haematopoietic stem cell transplantation graft monitoring. Indian J Med Sci. 2004;58(7):297-304.
92. Santurtún A, Riancho JA, Yañez L, Santurtún M, Zarrabeitia MT. Analysis of post-transplant chimerism by using a single amplification reaction of 38 Indel polymorphic loci. Bone Marrow Transplant. 2014;49(11):1432-1435.
93. Odriozola A, Riancho JA, Colorado M, Zarrabeitia MT. Evaluation of the sensitivity of two recently developed STR multiplexes for the analysis of chimerism after haematopoietic stem cell transplantation. Int J Immunogenet. 2013;40(2):88-92.
94. Thyagarajan B, Young S, Floodman S, Peterson R, Wang X. Systematic analysis of interference due to stutter in estimating chimerism following hematopoietic cell transplantation. J Clin Lab Anal. 2009;23(5):308-313.

95. Pepinski W, Skawronska M, Janica J, Niemcunowicz-Janica A. Polymorphism of 11 non-CODIS STRs in a population sample of religious minority of Old Believers residing in northeastern Poland. Adv Med Sci. 2010;55(2):328-332.
96. Borrill V, Schlaphoff T, du Toit E, Marx M, Wood L, Jacobs P. The use of short tandem repeat polymorphisms for monitoring chimerism following bone marrow transplantation: a short report. Hematology. 2008;13(4):210-214.
97. Alizadeh M, Bernard M, Danic B, Dauriac C, Birebent B, Lapart C, Lamy T, Le Prisé PY, Beauplet A, Bories D, Semana G, Quelvennec E. Quantitative assessment of hematopoietic chimerism after bone marrow transplantation by real-time quantitative polymerase chain reaction. Blood. 2002;99(12):4618-4625.
98. Schraml E, Lion T. Interference of dye-associated fluorescence signals with quantitative analysis of chimerism by capillary electrophoresis. Leukemia. 2003;17(1):221-223.
99. Bai L, Deng YM, Dodds AJ, Milliken S, Moore J, Ma DD. A SYBR green-based real-time PCR method for detection of haemopoietic chimerism in allogeneic haemopoietic stem cell transplant recipients. Eur J Haematol. 2006;77(5):425-431.
100. Maas F, Schaap N, Kolen S, Zoetbrood A, Buño I, Dolstra H, de Witte T, Schattenberg A, van de Wiel-van Kemenade E. Quantification of donor and recipient hemopoietic cells by real-time PCR of single nucleotide polymorphisms. Leukemia. 2003;17(3):630-633.
101. Koldehoff M, Steckel NK, Hlinka M, Beelen DW, Elmaagacli AH. Quantitative analysis of chimerism after allogeneic stem cell transplantation by real-time polymerase chain reaction with single nucleotide polymorphisms, standard tandem repeats, and Y-chromosome-specific sequences. Am J Hematol. 2006;81(10):735-746.
102. Willasch A, Eing S, Weber G, Kuçi S, Schneider G, Soerensen J, Jarisch A, Rettinger E, Koehl U, Klingebiel T, Kreyenberg H, Bader P. Enrichment of cell subpopulations applying automated MACS technique: purity, recovery and applicability for PCR-based chimerism analysis. Bone Marrow Transplant. 2010;45(1):181-189.
103. Almeida CA, Dreyfuss JL, Azevedo-Shimmoto MM, Figueiredo MS, de Oliveira JS. Evaluation of 16 SNPs allele-specific to quantify post hSCT chimerism by SYBR green-based qRT-PCR. J Clin Pathol. 2013;66(3):238-242.
104. Harries LW, Wickham CL, Evans JC, Rule SA, Joyner MV, Ellard S. Analysis of haematopoietic chimaerism by quantitative real-time polymerase chain reaction. Bone Marrow Transplant. 2005;35(3):283-290.
105. Jiménez-Velasco A, Barrios M, Román-Gómez J, Navarro G, Buño I, Castillejo JA, Rodríguez AI, García-Gemar G, Torres A, Heiniger AI. Reliable quantification of hematopoietic chimerism after allogeneic transplantation for acute leukemia using amplification by real-time PCR of null alleles and insertion/deletion polymorphisms. Leukemia. 2005;19(3):336-343.
106. Gineikiene E, Stoskus M, Griskevicius L. Single nucleotide polymorphism-based system improves the applicability of quantitative PCR for chimerism monitoring. J Mol Diagn. 2009;11(1):66-74.

107.Chen DP, Tseng CP, Wang WT, Wang MC, Tsai SH, Sun CF. Real-time biallelic polymorphism-polymerase chain reaction for chimerism monitoring of hematopoietic stem cell transplantation relapsed patients.Clin Chim Acta. 2011;412(7-8):625-630.

108.Qin XY, Li GX, Qin YZ, Wang Y, Wang FR, Liu DH, Xu LP, Chen H, Han W, Wang JZ, Zhang XH, Li JL, Li LD, Liu KY, Huang XJ. Quantitative assessment of hematopoietic chimerism by quantitative real-time polymerase chain reaction of sequence polymorphism systems after hematopoietic stem cell transplantation. Chin Med J (Engl). 2011;124(15):2301-2308.

109.Kim SY, Jeong MH, Park N, Ra E, Park H, Seo SH, Kim JY, Seong MW, Park SS. Chimerism monitoring after allogeneic hematopoietic stem cell transplantation using quantitative real-time PCR of biallelic insertion/deletion polymorphisms. J Mol Diagn. 2014;16(6):679-688.

110. https://aibt.it/it/library/linee-guida-aibt

111. Khan F, Agarwal A, Agrawal S.Significance of chimerism in hematopoietic stem cell transplantation: new variations on an old theme. Bone Marrow Transplant. 2004;34(1):1-12.

112. Maury S, Jouault H, Kuentz M, Vernant JP, Tulliez M, Cordonnier C, Bories D. Chimerism analysis by lineage-specific fluorescent polymerase chain reaction in secondary graft failure after allogeneic stem cell transplantation. Transplantation. 2001;71(3):374-380.

113. Jacque N, Nguyen S, Golmard JL, Uzunov M, Garnier A, Leblond V, Vernant JP, Bories D, Dhédin N. Chimerism analysis in peripheral blood using indel quantitative real-time PCR is a useful tool to predict post-transplant relapse in acute leukemia. Bone Marrow Transplant. 2015;50(2):259-265.

114. Zeiser R, Spyridonidis A, Wäsch R, Ihorst G, Grüllich C, Bertz H, Finke J. Evaluation of immunomodulatory treatment based on conventional and lineage-specific chimerism analysis in patients with myeloid malignancies after myeloablative allogeneic hematopoietic cell transplantation. Leukemia. 2005;19(5):814-821.

115. Barrios M, Jiménez-Velasco A, Román-Gómez J, Madrigal ME, Castillejo JA, Torres A, Heiniger A. Chimerism status is a useful predictor of relapse after allogeneic stem cell transplantation for acute leukemia. Haematologica. 2003;88(7):801-810.

116. Horky O, Mayer J, Kablaskova L, Razga F, Krejci M, Kissova J, Borsky M, Jeziskova I, Dvorakova D. Increasing hematopoietic microchimerism is a reliable indicator of incipient AML relapse. Int J Lab Hematol. 2011;33(1):57-66.

117. Broglie L, Helenowski I, Jennings LJ, Schafernak K, Duerst R, Schneiderman J, Tse W, Kletzel M, Chaudhury S. Early mixed T- cell chimerism is predictive of pediatric AML or MDS relapse after hematopoietic stem cell transplant. Pediatr Blood Cancer 2017;64(9).

118. Lamba R, Abella E, Kukuruga D, Klein J, Savasan S, Abidi MH, Mohamed A, Peres E. Mixed hematopoietic chimerism at day 90 following allogenic myeloablative stem cell transplantation is a predictor of relapse and survival. Leukemia 2004;18(10): 1681-1686.

119. Mattsson J, Uzunel M, Tammik L, Aschan J, Ringden O. Leukemia lineage-specific chimerism analysis is a sensitive predictor of relapse in patients with acute myeloid leukemia and myelodysplastic syndrome after allogeneic stem cell transplantation. Leukemia 2001; 15(12): 1976-1985.

120. Bacher U, Haferlach T, Fehse B, Schnittger S, Kröger N. Minimal residual disease diagnostics and chimerism in the post-transplant period in acute myeloid leukemia. ScientificWorldJournal. 2011;11:310-319.

121. Serrano J, Roman J, Sanchez J, Jimenez A, Castillejo JA, Herrera C, Gonzalez MG, Reina L, Rodriguez MC, Alvarez MA, Maldonado J, Torres A. Molecular analysis of lineage-specific chimerism and minimal residual disease by RT-PCR of p210(BCR-ABL) and p190(BCR-ABL) after allogeneic bone marrow transplantation for chronic myeloid leukemia: increasing mixed myeloid chimerism and p190(BCR-ABL) detection precede cytogenetic relapse. Blood. 2000;95(8):2659-2665.

122. Wiedemann B, Klyuchnikov E, Kröger N, Zabelina T, Stahl T, Zeschke S, Badbaran A, Ayuk F, Alchalby H, Wolschke C, Bokemeyer C, Fehse B, Zander AR, Bacher U. Chimerism studies with quantitative real-time PCR in stem cell recipients with acute myeloid leukemia. Exp Hematol. 2010;38(12):1261-1271.

123. Ossenkoppele G, Schuurhuis GJ. MRD in AML: time for redefinition of CR? Blood. 2013;121(12):2166-2168.

124. Kröger N, Miyamura K, Bishop MR. Minimal residual disease following allogeneic hematopoietic stem cell transplantation. Biol Blood Marrow Transplant. 2011;17(1 Suppl):S94-100.

125. Ravandi F, Walter RB, Freeman SD. Evaluating measurable residual disease in acute myeloid leukemia. Blood Adv. 2018;2(11):1356-1366.

126. Agathangelidis A, Rosenquist R, Davi F, Ghia P, Belessi C, Hadzidimitriou A, Stamatopoulos K. Immunoglobulin Gene Analysis in Chronic Lymphocytic Leukemia. Methods Mol Biol. 2019;1881:51-62.

127. Ahci M, Toffalori C, Bouwmans E, Crivello P, Brambati C, Pultrone C, Stempelmann K, Bost D, Mazzi B, Beelen DW, Ciceri F, Mulder W, Fleischhauer K, Vago L. A new tool for rapid and reliable diagnosis of HLA loss relapses after HSCT. Blood. 2017;130(10):1270-1273.

128. Pareek CS, Smoczynski R, Tretyn A. Sequencing technologies and genome sequencing. J Appl Genet. 2011;52(4):413-435.

129. Xuan J, Yu Y, Qing T, Guo L, Shi L. Next-generation sequencing in the clinic: promises and challenges. Cancer Lett. 2013;340(2):284-295.

130. Ladetto M, Brüggemann M, Monitillo L, Ferrero S, Pepin F, Drandi D, Barbero D, Palumbo A, Passera R, Boccadoro M, Ritgen M, Gökbuget N, Zheng J, Carlton V3, Trautmann H, Faham M, Pott C. Next-generation sequencing and real-time quantitative PCR for minimal residual disease detection in B-cell disorders. Leukemia. 2014;28(6):1299-1307.

131. Aloisio M, Bortot B, Gandin I, Severini GM, Athanasakis E. A semi-nested real-time PCR method to detect low chimerism percentage in small quantity of hematopoietic stem cell transplant DNA samples. Genome. 2017;60(2):183-192.

132. Aloisio M, Licastro D, Caenazzo L, Torboli V, D'Eustacchio A, Severini GM, Athanasakis E. A technical application of quantitative next generation sequencing for chimerism evaluation. Mol Med Rep. 2016;14(4):2967-2974.

133. https:--cran.r-project.org-web-packages-shiny-index.html

134. Lawrence M, Huber W, Pagès H, Aboyoun P, Carlson M, Gentleman R, Morgan MT, Carey VJ. Software for computing and annotating genomic ranges. PLoS Comput Biol. 2013;9(8):e1003118.
135. https:--genome.ucsc.edu
136. http:--www.ncbi. nlm.nih.gov-SNP-, build 138
137. Debeljak M, Freed DN, Welch JA, Haley L, Beierl K, Iglehart BS,Pallavajjala A, Gocke CD, Leffell MS, Lin MT, Pevsner J, Wheelan SJ, Eshleman JR. Haplotype counting by next-generation sequencing for ultrasensitive human DNA detection. J Mol Diagn. 2014;16(5):495-503.
138. Lange V, Böhme I, Hofmann J, Lang K, Sauter J, Schöne B, Paul P, Albrecht V, Andreas JM, Baier DM, Nething J, Ehninger U, Schwarzelt C, Pingel J, Ehninger G, Schmidt AH. Cost-efficient high-throughput HLA typing by MiSeq amplicon sequencing. BMC Genomics. 2014;15:63.
139. https:--genome.ucsc.edu-cgi-bin-hgBlat
140. https:--www.thermofisher.com
141. https:--eu.idtdna.com-calc-analyzer
142. Mwaigwisya S, Assiri RA and O'Grady J: Emerging commercial molecular tests for the diagnosis of bloodstream infection. Expert Rev Mol Diagn. 2015;15(5):681-692.
143.https://www.thermofisher.com/it/en/home/life-science/pcr/real-time-pcr/real-time-pcr-learning-center/real-time-pcr-basics/real-time-pcr-understanding-ct.html
144. www.ukneqasli.co.uk
145. Ahci M, Stempelmann K, Buttkereit U, Crivello P, Trilling M, Heinold A, Steckel NK, Koldehoff M, Horn PA, Beelen DW, Fleischhauer K. Clinical Utility of Quantitative PCR for Chimerism and Engraftment Monitoring after Allogeneic Stem Cell Transplantation for Hematologic Malignancies. Biol Blood Marrow Transplant. 2017;23(10):1658-1668.
146. D'Argenio V, Esposito MV, Telese A, Precone V, Starnone F, Nunziato M, Cantiello P, Iorio M, Evangelista E, D'Aiuto M, Calabrese A, Frisso G, D'Aiuto G, Salvatore F. The molecular analysis of BRCA1 and BRCA2: Next-generation sequencing supersedes conventional approaches. Clin Chim Acta. 2015;446:221-225.
147. Børsting C and Morling N: Next generation sequencing and its applications in forensic genetics. Forensic Sci Int Genet. 2015;18:78-89.
148. Kim J, Hwang IS, Kim HS, Joo DJ, Hong KR and Choi JR: Bone marrow chimerism detection using next generation sequencing based on single nucleotide polymorphisms following liver transplantation: Comparison with short tandem repeat-PCR. Ann Lab Med 36: 82-84, 2016;36(1): 82-84.
149. Chen DP, Tsai SH, Tseng CP, Wu TL, Chang PY and Sun CF: Bone marrow transplant relapse with loss of an allele. Clin Chim Acta. 2008;387(1-2):161-164.
150. Bragg LM, Stone G, Butler MK, Hugenholtz P and Tyson GW: Shining a light on dark sequencing: Characterizing errors in Ion Torrent PGM data. PLoS Comput Biol. 2013;9(4):e1003031.
151. Debeljak M, Mocci E, Morrison MC, Pallavajjalla A, Beierl K, Amiel M, Noë M, Wood LD, Lin MT, Gocke CD, Klein AP, Fuchs EJ, Jones RJ, Eshleman JR. Haplotype Counting for Sensitive

Chimerism Testing: Potential for Early Leukemia Relapse Detection. J Mol Diagn. 2017;19(3):427-436.

152. Toffalori C, Cavattoni I, Deola S, Mastaglio S, Giglio F, Mazzi B, Assanelli A, Peccatori J, Bordignon C, Bonini C, Cortelazzo S, Ciceri F, Fleischhauer K, Vago L. Genomic loss of patient-specific HLA in acute myeloid leukemia relapse after well-matched unrelated donor HSCT. Blood. 2012;119(20):4813–4815.

153. Waterhouse M, Pfeifer D, Pantic M, Emmerich F, Bertz H, Finke J. Genome-wide profiling in AML patients relapsing after allogeneic hematopoietic cell transplantation. Biol. Blood Marrow Transplant. J. Am. Soc. Blood Marrow Transplant. 2011;17(10):1450–1459.

154. Hamdi A, Cao K, Poon LM, Aung F, Kornblau S, Fernandez Vina MA, Champlin RE, Ciurea SO. Are changes in HLA Ags responsible for leukemia relapse after HLA-matched allogeneic hematopoietic SCT? Bone Marrow Transplant. 2015;50(3):411– 413.

8. SUPPLEMENTARY DATA

8.1. HP286SNPs Bed file

SNP ID	Genome position	Variation	Type of SNP
rs10489266	chr1:g.173178453	A-G	-
rs12070002	chr1:g.227819490	G-A	-
rs12070036	chr1:g.227819514	G-A	mSNP
rs1234315	chr1:g.173178463	C-T	mSNP
rs180908510	chr1:g.227819487	G-T	-
rs186139921	chr1:g.227819603	C-T	-
rs188508819	chr1:g.227819449	A-G	-
rs192073945	chr1:g.173178435	G-C	-
rs60397612	chr1:g.227819482	G-A	-
rs10496711	chr2:g.134516742	C-G	mSNP
rs12612347	chr2:g.219057338	A-G	mSNP
rs143618768	chr2:g.219057377	G-C	-
rs148872993	chr2:g.219057313	A-G	-
rs187665855	chr2:g.219057358	T-A	-
rs79182081	chr2:g.134516803	T-C	-
rs118065079	chr3:g.30693481	T-C	-
rs140843574	chr3:g.30693570	T-C	-
rs145079088	chr3:g.30693489	T-G	-
rs1466079	chr3:g.134414273	T-C	-
rs185422938	chr3:g.30693478	A-T	-
rs187381520	chr3:g.134414105	G-T	-
rs190180628	chr3:g.30693494	C-T	-
rs192327182	chr3:g.134414300	C-A	-
rs1984630	chr3:g.134414219	C-A	mSNP
rs72849396	chr3:g.30693562	T-G	-
rs9831477	chr3:g.30693522	T-A	mSNP
rs10033900	chr4:g.110659067	T-C	mSNP
rs150320510	chr4:g.110659015	C-T	-
rs188009540	chr4:g.148463789	C-T	-
rs200104039	chr4:g.148463742	G-A	-
rs200186870	chr4:g.148463809	C-T	-
rs200476599	chr4:g.148463779	A-T	-
rs200681001	chr4:g.148463898	A-G	-
rs200799891	chr4:g.148463908	A-G	-
rs200945454	chr4:g.148463745	G-C	-
rs201337276	chr4:g.148463901	C-T	-
rs201768930	chr4:g.148463784	C-G	-
rs201989285	chr4:g.148463791	C-G	-
rs202086451	chr4:g.148463911	C-T	-

rs202195925	chr4:g.148463760	G-C	-
rs369664098	chr4:g.110659007	C-T	-
rs375287100	chr4:g.148463786	C-T	-
rs5335	chr4:g.148463840	G-C	mSNP
rs5336	chr4:g.148463907	A-G	-
rs76335033	chr4:g.110659188	C-G	-
rs78314565	chr4:g.110659072	G-T	-
rs10038113	chr5:g.25902342	T-C	mSNP
rs112131429	chr5:g.25902355	C-T	-
rs13187934	chr5:g.25902368	C-T	-
rs139418397	chr5:g.25902245	A-G	-
rs182552453	chr5:g.25902350	A-G	-
rs184468575	chr5:g.15555426	G-T	-
rs191854777	chr5:g.25902343	G-A	-
rs373030117	chr5:g.15555457	G-A	-
rs56967160	chr5:g.25902396	G-A	-
rs73754137	chr5:g.15555509	A-G	-
rs983889	chr5:g.15555486	G-T	mSNP
rs138941900	chr6:g.13370562	G-A	-
rs139499603	chr6:g.33636880	C-A	-
rs140973588	chr6:g.33636890	C-T	-
rs146493384	chr6:g.33636828	C-T	-
rs150278166	chr6:g.13370492	T-C	-
rs182289168	chr6:g.13370449	G-A	-
rs202109196	chr6:g.33636901	C-T	-
rs2077163	chr6:g.33636907	C-T	mSNP
rs367915100	chr6:g.13370410	G-C	-
rs368168994	chr6:g.33636854	C-T	-
rs369432667	chr6:g.33636960	C-G	-
rs373777173	chr6:g.33636891	G-A	-
rs376933400	chr6:g.33636895	C-T	-
rs552655	chr6:g.13370488	T-C	mSNP
rs117449302	chr7:g.116140526	C-A	-
rs137971888	chr7:g.116140448	C-T	-
rs143606406	chr7:g.116140430	G-C	-
rs148169812	chr7:g.103489732	C-T	-
rs149320804	chr7:g.116140419	A-G	-
rs199577791	chr7:g.116140445	T-C	-
rs199812446	chr7:g.116140509	G-A	-
rs199984351	chr7:g.116140459	T-C	-
rs200183803	chr7:g.116140518	C-T	-
rs200666871	chr7:g.116140499	C-T	-
rs200803879	chr7:g.116140457	C-T	-
rs200858684	chr7:g.116140431	G-T	-
rs201434174	chr7:g.116140442	C-T	-
rs201801614	chr7:g.116140508	C-A	-

rs201827952	chr7:g.116140458	A-G	-
rs2270188	chr7:g.116140524	G-T	mSNP
rs2270189	chr7:g.116140616	A-G	-
rs370290952	chr7:g.103489715	C-A	-
rs370692837	chr7:g.116140592	G-C	-
rs372467774	chr7:g.116140528	G-T	-
rs374420882	chr7:g.103489728	C-T	-
rs375402156	chr7:g.116140542	T-G	-
rs377760396	chr7:g.103489719	C-T	-
rs39395	chr7:g.103489729	G-A	mSNP
rs10505477	chr8:g.128407443	A-G	mSNP
rs112183774	chr8:g.12957594	C-T	-
rs138054457	chr8:g.128407518	C-T	-
rs138749997	chr8:g.12957581	C-G	-
rs139243841	chr8:g.12957498	T-C	-
rs141272973	chr8:g.12957587	G-T	-
rs144923726	chr8:g.12957567	G-A	-
rs150342076	chr8:g.12957605	C-T	-
rs150424033	chr8:g.128407417	G-A	-
rs17854877	chr8:g.12957440	T-C	-
rs186925500	chr8:g.128407352	A-G	-
rs191801620	chr8:g.128407357	T-C	-
rs200152731	chr8:g.12957537	T-C	-
rs201189821	chr8:g.12957582	G-A	-
rs34027003	chr8:g.128407337	T-C	-
rs372929532	chr8:g.12957522	T-G	-
rs373770104	chr8:g.12957468	T-C	-
rs376046492	chr8:g.12957545	C-T	-
rs376194381	chr8:g.12957497	A-C	-
rs377708345	chr8:g.12957566	C-T	-
rs532841	chr8:g.12957475	C-T	mSNP
rs10124636	chr9:g.14442529	G-A	-
rs115648561	chr9:g.14442572	G-A	-
rs115944264	chr9:g.91669307	T-G	-
rs141285483	chr9:g.91669446	A-G	-
rs150542537	chr9:g.14442617	C-T	-
rs2297313	chr9:g.91669362	A-G	mSNP
rs2297314	chr9:g.91669370	A-T	-
rs424539	chr9:g.14442595	C-G	mSNP
rs141538707	chr10:g.13155707	A-G	-
rs1561570	chr10:g.13155726	T-C	mSNP
rs17723910	chr10:g.104581242	G-A	-
rs184429534	chr10:g.13155601	T-C	-
rs186980432	chr10:g.104581261	A-G	-
rs619824	chr10:g.104581288	A-C	mSNP
rs117001766	chr11:g.44082957	C-T	-

rs148667283	chr11:g.44082930	C-A	-
rs178503	chr11:g.44082931	G-A	mSNP
rs185380402	chr11:g.44082843	C-T	-
rs198464	chr11:g.61521621	G-A	mSNP
rs1126758	chr12:g.103248924	C-T	mSNP
rs150187239	chr12:g.53294408	G-A	-
rs182135145	chr12:g.103248939	C-T	-
rs199475576	chr12:g.103248948	A-C	-
rs199475656	chr12:g.103248916	T-G	-
rs199475673	chr12:g.103248931	A-C	-
rs199935897	chr12:g.53294330	C-A	-
rs201245932	chr12:g.103248952	T-C	-
rs281865443	chr12:g.103248988	G-A	-
rs281865444	chr12:g.103248938	C-T	-
rs368267213	chr12:g.53294357	C-T	-
rs373782868	chr12:g.103248978	T-C	-
rs375402364	chr12:g.53294324	C-A	-
rs376057081	chr12:g.103248877	A-G	-
rs376161788	chr12:g.103248874	C-T	-
rs376422291	chr12:g.103248838	G-C	-
rs5030845	chr12:g.103248929	A-G	-
rs62507323	chr12:g.103248949	A-G	-
rs62507348	chr12:g.103248926	G-A	-
rs62508572	chr12:g.103249001	T-C	-
rs62508577	chr12:g.103248928	G-A	-
rs62508592	chr12:g.103248897	C-A	-
rs62508593	chr12:g.103249006	T-G	-
rs62508615	chr12:g.103248942	C-G	-
rs62508617	chr12:g.103248970	C-T	-
rs62508696	chr12:g.103248944	G-A	-
rs62508721	chr12:g.103249000	T-C	-
rs62508728	chr12:g.103249003	T-C	-
rs62509013	chr12:g.103248972	G-C	-
rs62514931	chr12:g.103248989	G-T	-
rs62514933	chr12:g.103248967	C-A	-
rs62514934	chr12:g.103248958	T-C	-
rs62516109	chr12:g.103248982	A-G	-
rs62516152	chr12:g.103248932	C-T	-
rs62517170	chr12:g.103249004	A-C	-
rs62517198	chr12:g.103248985	A-G	-
rs62517201	chr12:g.103249002	G-C	-
rs62517204	chr12:g.103248946	G-C	-
rs62517208	chr12:g.103248921	G-T	-
rs76887545	chr12:g.103248858	C-T	-
rs8608	chr12:g.53294381	T-C	mSNP
rs1061472	chr13:g.52524488	T-C	mSNP

rs146250029	chr13:g.19735985	G-A	-
rs181388674	chr13:g.52524498	C-T	-
rs183519418	chr13:g.52524474	C-A	-
rs188903683	chr13:g.52524481	G-C	-
rs200219625	chr13:g.52524584	G-A	-
rs200450017	chr13:g.52524546	C-T	-
rs200996053	chr13:g.52524439	G-A	-
rs201647718	chr13:g.52524576	C-T	-
rs34402231	chr13:g.19735925	C-T	-
rs368589213	chr13:g.52524503	C-T	-
rs373528664	chr13:g.52524508	C-T	-
rs373641722	chr13:g.52524410	G-A	-
rs374592960	chr13:g.52524492	C-T	-
rs504544	chr13:g.19735891	A-T	mSNP
rs61945038	chr13:g.19735973	A-G	-
rs10143250	chr14:g.104723433	C-T	mSNP
rs111654129	chr14:g.63669675	C-A	-
rs143879923	chr14:g.104723366	G-A	-
rs147234702	chr14:g.104723418	C-T	-
rs148538694	chr14:g.63669614	C-T	-
rs186908217	chr14:g.104723384	A-G	-
rs189858143	chr14:g.104723490	C-T	-
rs191940884	chr14:g.63669599	G-T	-
rs1957779	chr14:g.63669647	A-G	mSNP
rs368700793	chr14:g.104723511	G-A	-
rs373404833	chr14:g.63669583	G-A	-
rs55806265	chr14:g.63669587	G-A	-
rs74058898	chr14:g.63669668	G-A	-
rs146072700	chr15:g.94879769	G-A	-
rs181869522	chr15:g.35006036	G-A	-
rs191816259	chr15:g.94879694	C-T	-
rs2117215	chr15:g.94879684	A-G	mSNP
rs374531656	chr15:g.94879713	G-T	-
rs634990	chr15:g.35006073	T-C	mSNP
rs76157757	chr15:g.94879690	C-A	-
rs78720274	chr15:g.94879615	T-A	-
rs113022744	chr16:g.7720934	G-A	-
rs113134069	chr16:g.66184058	C-T	-
rs116104532	chr16:g.7720871	C-T	-
rs121893	chr16:g.66183995	T-C	mSNP
rs138190186	chr16:g.7721008	A-C	-
rs139179216	chr16:g.66183953	C-T	-
rs141520676	chr16:g.7720919	A-C	-
rs150796285	chr16:g.66183929	T-C	-
rs187098577	chr16:g.7720959	C-T	-
rs2191125	chr16:g.7720923	T-C	mSNP

rs235092	chr16:g.66184097	G-A	-
rs370937247	chr16:g.7720949	T-C	-
rs75909127	chr16:g.7720907	C-T	-
rs9938697	chr16:g.7720861	C-A	-
rs11540685	chr17:g.62400489	T-A	-
rs140262430	chr17:g.40514163	C-A	-
rs143501773	chr17:g.40514283	G-C	-
rs14753	chr17:g.62400513	G-T	-
rs2278680	chr17:g.62400602	C-A	-
rs56350444	chr17:g.62400661	G-C	-
rs6808	chr17:g.62400575	G-C	mSNP
rs744166	chr17:g.40514201	A-G	mSNP
rs183907786	chr18:g.48509129	G-A	-
rs191649987	chr18:g.57831441	G-C	-
rs367918858	chr18:g.48509236	C-T	-
rs368784324	chr18:g.57831432	G-A	-
rs620898	chr18:g.48509148	T-A	mSNP
rs633265	chr18:g.57831468	G-T	mSNP
rs72913222	chr18:g.48509170	C-T	-
rs75038383	chr18:g.48509237	G-A	-
rs108295	chr19:g.34224816	G-A	mSNP
rs112897571	chr19:g.10837667	G-T	-
rs147257650	chr19:g.10837735	G-T	-
rs150702249	chr19:g.34224617	C-A	-
rs182323372	chr19:g.10837711	G-A	-
rs369964019	chr19:g.34224762	C-T	-
rs372742481	chr19:g.10837745	C-T	-
rs376218830	chr19:g.10837616	G-A	-
rs376404909	chr19:g.34224667	C-T	-
rs376995977	chr19:g.10837706	G-C	-
rs80293523	chr19:g.10837708	C-T	-
rs892086	chr19:g.10837677	G-A	mSNP
rs137975099	chr20:g.39797439	C-T	-
rs146544228	chr20:g.50055341	G-C	-
rs202246756	chr20:g.39797473	G-C	-
rs369687296	chr20:g.39797424	C-G	-
rs56132252	chr20:g.39797469	G-A	-
rs6063653	chr20:g.50055330	C-T	-
rs715147	chr20:g.50055350	G-A	mSNP
rs753381	chr20:g.39797465	T-C	mSNP
rs182030596	chr21:g.43728964	C-T	-
rs182180151	chr21:g.35721653	C-A	-
rs187852386	chr21:g.35721477	G-A	-
rs191555026	chr21:g.35721605	C-A	-
rs225436	chr21:g.43729034	C-T	mSNP
rs7277338	chr21:g.43728926	G-A	-

rs7282516	chr21:g.43728908	C-T	-
rs78593835	chr21:g.35721609	A-T	-
rs78681357	chr21:g.35721583	A-G	-
rs79757900	chr21:g.43728990	T-C	-
rs8128316	chr21:g.35721560	C-T	mSNP
rs113648111	chr22:g.38563521	G-C	-
rs132985	chr22:g.38563471	C-T	mSNP
rs132986	chr22:g.38563548	A-G	-
rs132987	chr22:g.38563572	G-A	-
rs138613303	chr22:g.38563512	G-A	-
rs140129123	chr22:g.31205425	G-A	-
rs182361106	chr22:g.31205298	C-G	-
rs187017139	chr22:g.31205440	G-A	-
rs2284067	chr22:g.38563511	C-T	-
rs370404784	chr22:g.31205337	G-A	-
rs4444	chr22:g.31205334	T-C	mSNP
rs4497	chr22:g.38563541	A-G	-

Supplementary table 1. HP286SNPs Bed file. The main SNP are marked with the abbreviation mSNP.

8.2. Non disclosure agreement DKMS

REGIONE AUTONOMA FRIULI VENEZIA GIULIA
ISTITUTO DI RICOVERO E CURA a CARATTERE SCIENTIFICO
Burlo Garofolo di Trieste
BURLO
DIREZIONE GENERALE
34137 Trieste, via dell'Istria 65/1 cf. 00124430323
tel.+39 040 3785.225 email:direzione.generale@burlo.trieste.it
fax +39 040 666019 pec:OIBurloTS.protgen@certsanita.fvg.it

Trieste,

Oggetto: "Non-disclosure agreement" between DKMS Life Science Lab GmbH ('the Discloser') and Institute for Maternal and Child Health IRCCS "Burlo Garofolo" ('the Recipient')

Parties:

Dott. Gianluigi Scannapieco,
Institute for Maternal and Child Health IRCCS "Burlo Garofolo"
via dell'Istria 65/1 34137 Trieste (Italy)
('**the Recipient**')

and

DKMS Life Science Lab GmbH, Blasewitzer Str. 43 01307 Dresden, Germany
(**the Discloser**)

1. The Discloser intends to disclose information on the primers used for next generation sequencing or NGS (the Confidential Information) to the Recipient for the sole purpose of non- commercial scientific research (the Purpose).

2. The Recipient undertakes not to use the Confidential Information for any purpose except the Purpose, without first obtaining the written agreement of the Discloser.

3. The Recipient undertakes to keep the Confidential Information secure and not to disclose it to any third party, except as specifically set forth herein. Recipient may order primers to be synthesized at a third-party vendor, which requires disclosure of the primer sequences to such a vendor. To the extent Recipient needs to disclose Confidential Information to order such primers, it may disclose such information, provided that these primers once received will be used by Recipient solely for non-commercial research purposes and provided that such a vendor is contractually obliged by the Recipient to keep the information they receive strictly confidential.

4. Where the Recipient receives any personal data, he or she must ensure to fully comply with the provisions of applicable law and only use the data to fulfill the Purpose. In particular, the Recipient may not disclose any personal data to third parties and must take all necessary steps in order to avoid unintentional disclosure of such data.

5. The undertakings in clauses 2 and 3 above apply to all of the information disclosed by the Discloser to the Recipient, regardless of the way or form in which it is disclosed or recorded but they do not apply to:

a) any information which is or in future comes into the public domain (unless as a result of the breach of this Agreement);

b) any information which is already known to the Recipient and which was not subject to any obligationofconfidencebeforeitwasdisclosedtotheRecipientbytheDiscloser;

c) is lawfully received by Recipient from a third party without restrictions; or

d) information that Recipient can demonstrate is developed independently by Recipient's employees who have no knowledge of or access to Discloser's Confidential Information.

6. Nothing in this Agreement will prevent the Recipient from making any disclosure of the Confidential Information required by law or by any competent authority.

7. The Recipient will, on request from the Discloser, return all copies and records of the Confidential Information to the Discloser and will not retain any copies or records of the Confidential Information;

8. Neither this Agreement nor the supply of any information grants the Recipient any license, interest or right in respect of any intellectual property rights of the Discloser except the right to u s e the Confidential Information solely for the Purpose.

9. In the event that any of the obligations under this Agreement are breached, or any provisions are not complied with or violated in any way, in whole or in part, by the Recipient, DKMS shall have the right to seek any remedies available at law or equity resulting from the breach, non-compliance and/or violation.

10. The undertakings in clauses 2, 3 and 5 will continue in force indefinitely from the date of this Agreement.

11. This Agreement is governed by, and is to be construed in accordance with German law. The German Cõurts will have non-exclusive jurisdiction to deal with any dispute which has arise nor may arise out of, or in connection with, this Agreement.

Date

Recipient

Date 06.06.18

DKMS Life Science Lab GmbH
Blasewitzer Str. 43
01307 Dresden
T 0351 450 45-30
F 0351 450 45-45
dkms-lab.de

Discloser

IRCCS BURLO GAROFOLO - TRIESTE
Il Direttore Scientifico
prof. Fabio Barbone

Documento informatico redatto e sottoscritto digitalmente ai sensi degli artt. 20 e 21 del d.lgs. 82/2005 (codice dell'amministrazione digitale)

MOD 0103V°5

*pag.*2 di 2

Supplementary Figure 1. Non disclosure agreement document.

8.3. HP763SNPs Bed BLOCK file

SNP ID	Genome position	Variation	Block TAG
rs374707893	chr4:g.114018992	A-G	PCR-3
rs531885350	chr4:g.114019018	G-A	PCR-3
rs113967111	chr4:g.114019022	G-A	PCR-3
rs113886905	chr4:g.114019047	C-T	PCR-3
rs547947903	chr4:g.114019056	A-G	PCR-3
rs566046345	chr4:g.114019068	A-G	PCR-3
rs6844654	chr4:g.114019075	C-T	PCR-3
rs28532593	chr4:g.114019101	T-A	PCR-3
rs567458411	chr4:g.114019146	T-C	PCR-3
rs374095763	chr4:g.114019150	G-A	PCR-3
rs542162067	chr7:g.7902080	G-T	PCR-5
rs369382758	chr7:g.7902083	G-T	PCR-5
rs572444920	chr7:g.7902118	A-G	PCR-5
rs6463729	chr7:g.7902123	T-C	PCR-5
rs564791175	chr7:g.7902147	C-T	PCR-5
rs6463730	chr7:g.7902171	A-G	PCR-5
rs6972172	chr7:g.7902214	C-T	PCR-5
rs561773612	chr7:g.7902236	T-C	PCR-5
rs73061204	chr7:g.7902242	G-A	PCR-5
rs547507851	chr7:g.7902244	C-T	PCR-5
rs78580561	chr7:g.7902254	G-A	PCR-5
rs73061208	chr7:g.7902258	G-A	PCR-5
rs115943357	chr7:g.7902282	A-G	PCR-5
rs181686777	chr9:g.31000592	G-A	PCR-6
rs139572309	chr9:g.31000609	A-G	PCR-6
rs543034571	chr9:g.31000614	G-A	PCR-6
rs554636962	chr9:g.31000647	T-A	PCR-6
rs10813408	chr9:g.31000648	T-C	PCR-6
rs187112566	chr9:g.31000654	C-A	PCR-6
rs564741178	chr9:g.31000656	T-A	PCR-6
rs10970034	chr9:g.31000664	C-T	PCR-6
rs11534129	chr9:g.31000689	T-G	PCR-6
rs11532869	chr9:g.31000698	T-C	PCR-6
rs190020177	chr9:g.31000710	G-A	PCR-6
rs11521418	chr9:g.31000712	T-A	PCR-6
rs11548860	chr9:g.31000727	T-C	PCR-6
rs528783371	chr9:g.31000741	G-A	PCR-6
rs547362017	chr9:g.31000790	C-A	PCR-6
rs10970035	chr9:g.31000803	C-T	PCR-6
rs542925425	chr11:g.118558299	T-A	PCR-8
rs693306	chr11:g.118558333	A-G	PCR-8
rs139097373	chr11:g.118558366	C-T	PCR-8
rs540588612	chr11:g.118558367	G-A	PCR-8
rs565147672	chr11:g.118558413	G-C	PCR-8
rs532425127	chr11:g.118558421	C-T	PCR-8
rs149509653	chr11:g.118558425	G-A	PCR-8
rs562733795	chr11:g.118558428	G-A	PCR-8
rs530266982	chr11:g.118558429	G-A	PCR-8
rs514143	chr11:g.118558449	G-A	PCR-8
rs567126926	chr11:g.118558451	C-T	PCR-8
rs558907	chr11:g.118558455	A-G	PCR-8

rs146008906	chr11:g.118558488	G-A	PCR-8
rs533428917	chr11:g.118558497	C-T	PCR-8
rs139963710	chr11:g.118558510	G-T	PCR-8
rs367725604	chr11:g.5906018	C-A	PCR-9
rs4758168	chr11:g.5906048	G-A	PCR-9
rs4757986	chr11:g.5906073	G-T	PCR-9
rs199681194	chr11:g.5906083	C-A	PCR-9
rs61876191	chr11:g.5906117	A-G	PCR-9
rs11823809	chr11:g.5906143	T-C	PCR-9
rs146818333	chr11:g.5906162	T-C	PCR-9
rs1453435	chr11:g.5906192	A-C	PCR-9
rs137865578	chr11:g.5906193	G-A	PCR-9
rs201376951	chr11:g.5906196	C-G	PCR-9
rs11823828	chr11:g.5906203	T-G	PCR-9
rs4757987	chr11:g.5906205	G-A	PCR-9
rs141226100	chr11:g.5906226	G-A	PCR-9
rs576234801	chr12:g.3194482	G-A	PCR-10
rs545149759	chr12:g.3194497	G-T	PCR-10
rs535885095	chr12:g.3194544	A-C	PCR-10
rs377233734	chr12:g.3194558	C-T	PCR-10
rs6489432	chr12:g.3194564	G-A	PCR-10
rs11614782	chr12:g.3194570	A-G	PCR-10
rs77514274	chr12:g.3194573	T-C	PCR-10
rs74815384	chr12:g.3194576	A-G	PCR-10
rs11614783	chr12:g.3194579	A-G	PCR-10
rs76640570	chr12:g.3194603	G-A	PCR-10
rs190215240	chr12:g.3194611	C-G	PCR-10
rs6489433	chr12:g.3194625	T-C	PCR-10
rs6489434	chr12:g.3194631	T-C	PCR-10
rs181423644	chr12:g.3194632	G-A	PCR-10
rs11062497	chr12:g.3194668	T-C	PCR-10
rs536549330	chr12:g.3194683	G-A	PCR-10
rs143632458	chr12:g.3194686	G-A	PCR-10
rs114794539	chr12:g.3194705	C-T	PCR-10
rs557024654	chr12:g.29829618	C-T	PCR-11
rs759888	chr12:g.29829639	C-T	PCR-11
rs537667501	chr12:g.29829649	G-T	PCR-11
rs1079365	chr12:g.29829663	A-C	PCR-11
rs759889	chr12:g.29829680	T-G	PCR-11
rs115382545	chr12:g.29829687	C-G	PCR-11
rs370216194	chr12:g.29829690	T-C	PCR-11
rs574778047	chr12:g.29829722	G-A	PCR-11
rs759890	chr12:g.29829723	T-G	PCR-11
rs188302116	chr12:g.29829728	C-T	PCR-11
rs145304851	chr12:g.29829734	G-A	PCR-11
rs145062394	chr12:g.29829735	A-T	PCR-11
rs564323142	chr12:g.29829740	T-C	PCR-11
rs528388827	chr12:g.29829741	T-C	PCR-11
rs10843476	chr12:g.29829743	T-C	PCR-11
rs181832412	chr12:g.29829777	C-T	PCR-11
rs2033041	chr12:g.29829794	T-C	PCR-11
rs550371563	chr12:g.29829831	T-C	PCR-11
rs186480127	chr12:g.29829834	C-T	PCR-11
rs563986983	chr14:g.63141835	T-C	PCR-12
rs533066013	chr14:g.63141940	A-C	PCR-12
rs148341614	chr14:g.63141941	C-T	PCR-12
rs4899091	chr14:g.63141952	A-G	PCR-12
rs4899092	chr14:g.63141981	A-C	PCR-12

rs8009247	chr14:g.63141989	C-T	PCR-12
rs8009249	chr14:g.63142000	C-T	PCR-12
rs8009109	chr14:g.63142001	A-G	PCR-12
rs116108019	chr14:g.63142004	G-A	PCR-12
rs571615369	chr14:g.63142021	G-A	PCR-12
rs4899093	chr14:g.63142026	A-G	PCR-12
rs554332368	chr14:g.63142061	C-T	PCR-12
rs74546560	chr14:g.63142079	T-G	PCR-12
rs142931548	chr16:g.81204011	G-A	PCR-13
rs569469885	chr16:g.81204026	G-A	PCR-13
rs540176270	chr16:g.81204043	G-A	PCR-13
rs191611226	chr16:g.81204055	C-G	PCR-13
rs567119725	chr16:g.81204060	G-T	PCR-13
rs1901816	chr16:g.81204066	A-T	PCR-13
rs1901817	chr16:g.81204069	A-C	PCR-13
rs573691630	chr16:g.81204070	G-A	PCR-13
rs537925777	chr16:g.81204077	T-C	PCR-13
rs12934430	chr16:g.81204081	T-C	PCR-13
rs78515382	chr16:g.81204113	T-A	PCR-13
rs1901818	chr16:g.81204125	C-G	PCR-13
rs560019649	chr16:g.81204131	G-T	PCR-13
rs11648241	chr16:g.81204172	A-C	PCR-13
rs112508375	chr16:g.81204185	G-A	PCR-13
rs1901819	chr16:g.81204188	A-G	PCR-13
rs529413241	chr16:g.81204200	G-C	PCR-13
rs551044236	chr16:g.81204221	C-T	PCR-13
rs143472886	chr16:g.81204230	T-C	PCR-13
rs185068103	chr16:g.6713342	A-G	PCR-15
rs532102385	chr16:g.6713369	A-G	PCR-15
rs12922407	chr16:g.6713387	A-G	PCR-15
rs568873515	chr16:g.6713391	G-A	PCR-15
rs535868044	chr16:g.6713405	C-T	PCR-15
rs8053333	chr16:g.6713419	T-G	PCR-15
rs566663669	chr16:g.6713422	A-G	PCR-15
rs148205521	chr16:g.6713425	A-C	PCR-15
rs78367422	chr16:g.6713429	C-T	PCR-15
rs115803164	chr16:g.6713436	C-A	PCR-15
rs8047260	chr16:g.6713468	G-C	PCR-15
rs12921979	chr16:g.6713471	C-T	PCR-15
rs574825852	chr16:g.6713473	C-G	PCR-15
rs189524883	chr16:g.6713488	C-G	PCR-15
rs8047273	chr16:g.6713497	G-A	PCR-15
rs572669967	chr16:g.6713552	T-C	PCR-15
rs546063957	chr16:g.6713561	T-G	PCR-15
rs12922020	chr16:g.6713563	C-G	PCR-15
rs531778264	chr16:g.6713579	G-A	PCR-15
rs112093495	chr16:g.6713585	A-G	PCR-15
rs562446763	chr16:g.6713591	A-C	PCR-15
rs529755603	chr16:g.6713608	A-C	PCR-15
rs147296344	chr17:g.6165637	C-T	PCR-16
rs140851163	chr17:g.6165673	G-A	PCR-16
rs118064277	chr17:g.6165687	G-A	PCR-16
rs9303200	chr17:g.6165696	A-G	PCR-16
rs530353734	chr17:g.6165724	T-C	PCR-16
rs9303201	chr17:g.6165758	T-C	PCR-16
rs560437314	chr17:g.6165776	C-A	PCR-16
rs532623327	chr17:g.6165790	G-C	PCR-16
rs186081726	chr17:g.6165803	G-C	PCR-16

rs568955375	chr17:g.6165808	G-T	PCR-16
rs9906240	chr17:g.6165815	T-C	PCR-16
rs9303202	chr17:g.6165818	G-A	PCR-16
rs9284355	chr17:g.6165834	G-A	PCR-16
rs9906271	chr17:g.6165853	T-C	PCR-16
rs73974745	chr17:g.6165856	G-A	PCR-16
rs570769401	chr17:g.6165860	G-A	PCR-16
rs539784678	chr17:g.6165861	G-A	PCR-16
rs9303203	chr17:g.6165863	G-T	PCR-16
rs576136070	chr17:g.6165912	C-G	PCR-16
rs73136476	chr21:g.25405287	C-T	PCR-19
rs34737404	chr21:g.25405328	T-C	PCR-19
rs568428054	chr21:g.25405329	G-A	PCR-19
rs2251998	chr21:g.25405350	T-A	PCR-19
rs2828739	chr21:g.25405363	A-G	PCR-19
rs572747157	chr21:g.25405377	A-G	PCR-19
rs538765385	chr21:g.25405429	A-G	PCR-19
rs34473781	chr21:g.25405451	C-A	PCR-19
rs36074329	chr21:g.25405453	A-G	PCR-19
rs35398670	chr21:g.25405455	G-C	PCR-19
rs554772186	chr21:g.25405460	A-G	PCR-19
rs574796525	chr21:g.25405472	A-G	PCR-19
rs540111012	chr21:g.25405474	A-G	PCR-19
rs143432105	chr22:g.19581889	G-A	PCR-20
rs540239264	chr22:g.19581905	C-A	PCR-20
rs560131767	chr22:g.19581939	T-C	PCR-20
rs5993715	chr22:g.19581946	G-T	PCR-20
rs5993716	chr22:g.19581999	C-T	PCR-20
rs562682281	chr22:g.19582002	T-C	PCR-20
rs530575904	chr22:g.19582016	T-C	PCR-20
rs5993717	chr22:g.19582021	G-T	PCR-20
rs5748329	chr22:g.19582029	A-C	PCR-20
rs5993718	chr22:g.19582036	A-G	PCR-20
rs147132604	chr22:g.19582042	C-T	PCR-20
rs5992433	chr22:g.19582066	C-T	PCR-20
rs538685409	chr22:g.19582135	A-G	PCR-20
rs116260586	chr1:g.188247566	C-T	PCR-21
rs187509030	chr1:g.188247585	T-C	PCR-21
rs75818240	chr1:g.188247616	T-A	PCR-21
rs3010658	chr1:g.188247654	C-T	PCR-21
rs2931598	chr1:g.188247674	G-A	PCR-21
rs2931599	chr1:g.188247734	A-G	PCR-21
rs61811374	chr1:g.188247741	C-T	PCR-21
rs3011972	chr1:g.188247742	A-G	PCR-21
rs573481618	chr1:g.188247775	G-C	PCR-21
rs1487488	chr1:g.188247807	G-C	PCR-21
rs3010657	chr1:g.188247809	T-A	PCR-21
rs114572039	chr1:g.188247818	T-C	PCR-21
rs557513078	chr3:g.11955792	G-T	PCR-22
rs56065911	chr3:g.11955797	T-C	PCR-22
rs56180123	chr3:g.11955805	A-G	PCR-22
rs55943645	chr3:g.11955809	C-T	PCR-22
rs370635296	chr3:g.11955827	C-T	PCR-22
rs28412762	chr3:g.11955851	T-C	PCR-22
rs9840548	chr3:g.11955862	C-T	PCR-22
rs9840851	chr3:g.11955866	G-C	PCR-22
rs9860599	chr3:g.11955872	T-C	PCR-22
rs9878130	chr3:g.11955875	A-C	PCR-22

rs115110074	chr3:g.11955887	T-C	PCR-22
rs303846	chr3:g.11955892	C-T	PCR-22
rs303847	chr3:g.11955899	G-A	PCR-22
rs186777109	chr3:g.11955907	G-A	PCR-22
rs547189871	chr3:g.11955921	G-A	PCR-22
rs565761821	chr3:g.11955931	G-A	PCR-22
rs2600246	chr3:g.11955938	T-C	PCR-22
rs557719662	chr3:g.11955945	C-A	PCR-22
rs62247767	chr3:g.11955947	C-G	PCR-22
rs62247768	chr3:g.11955948	A-G	PCR-22
rs2618383	chr3:g.11955958	A-C	PCR-22
rs2618382	chr3:g.11955959	A-G	PCR-22
rs55955151	chr3:g.11955972	C-T	PCR-22
rs9878282	chr3:g.11955986	A-G	PCR-22
rs58820061	chr3:g.11956001	T-A	PCR-22
rs372657560	chr3:g.11956012	C-T	PCR-22
rs564485849	chr3:g.11956030	T-C	PCR-22
rs531571830	chr3:g.11956052	C-G	PCR-22
rs191264107	chr3:g.11956057	C-G	PCR-22
rs376075344	chr3:g.11956072	C-T	PCR-22
rs369081080	chr3:g.11956073	C-T	PCR-22
rs375254223	chr10:g.127833938	T-C	PCR-23
rs180990657	chr10:g.127833944	C-T	PCR-23
rs564940057	chr10:g.127833979	T-C	PCR-23
rs532254669	chr10:g.127833989	C-T	PCR-23
rs74158340	chr10:g.127833992	A-G	PCR-23
rs10794061	chr10:g.127833998	A-G	PCR-23
rs11244838	chr10:g.127834000	G-C	PCR-23
rs11244839	chr10:g.127834046	A-G	PCR-23
rs11244840	chr10:g.127834067	T-C	PCR-23
rs12773842	chr10:g.127834080	G-C	PCR-23
rs9422812	chr10:g.127834139	T-G	PCR-23
rs12776125	chr10:g.127834142	A-G	PCR-23
rs569589517	chr15:g.31405220	C-G	PCR-25
rs148190079	chr15:g.31405221	C-G	PCR-25
rs551804565	chr15:g.31405230	G-A	PCR-25
rs71474644	chr15:g.31405278	A-C	PCR-25
rs534189958	chr15:g.31405284	G-A	PCR-25
rs7170825	chr15:g.31405285	A-G	PCR-25
rs182474293	chr15:g.31405289	C-T	PCR-25
rs536071768	chr15:g.31405305	G-A	PCR-25
rs141157823	chr15:g.31405317	G-A	PCR-25
rs564825786	chr15:g.31405339	T-C	PCR-25
rs571977208	chr15:g.31405340	G-C	PCR-25
rs373199012	chr15:g.31405371	C-G	PCR-25
rs12907809	chr15:g.31405378	C-T	PCR-25
rs529763821	chr15:g.31405379	G-A	PCR-25
rs79259282	chr15:g.31405380	G-A	PCR-25
rs187496047	chr15:g.31405393	A-C	PCR-25
rs537022557	chr15:g.31405396	T-G	PCR-25
rs552144134	chr15:g.31405397	G-A	PCR-25
rs571894622	chr15:g.31405399	T-C	PCR-25
rs533968390	chr15:g.31405403	T-C	PCR-25
rs4041971	chr15:g.31405408	A-T	PCR-25
rs567662642	chr15:g.31405413	C-A	PCR-25
rs116603240	chr15:g.31405422	A-C	PCR-25
rs556236009	chr15:g.31405426	C-A	PCR-25
rs2338861	chr15:g.31405436	T-C	PCR-25

rs2879276	chr15:g.31405439	A-G	PCR-25
rs11637099	chr15:g.31405442	A-G	PCR-25
rs572273663	chr15:g.31405448	G-A	PCR-25
rs193122402	chr15:g.31405456	C-G	PCR-25
rs184977185	chr15:g.31405479	T-C	PCR-25
rs574453612	chr15:g.31405481	C-T	PCR-25
rs543591251	chr15:g.31405502	A-C	PCR-25
rs535829775	chr16:g.83740020	G-A	PCR-26
rs6563972	chr16:g.83740033	G-T	PCR-26
rs569591921	chr16:g.83740042	T-C	PCR-26
rs116872632	chr16:g.83740057	A-G	PCR-26
rs7201416	chr16:g.83740068	G-A	PCR-26
rs1364121	chr16:g.83740107	C-T	PCR-26
rs1364120	chr16:g.83740126	T-C	PCR-26
rs1364119	chr16:g.83740136	T-C	PCR-26
rs572537872	chr16:g.83740149	A-T	PCR-26
rs181719685	chr16:g.83740150	T-A	PCR-26
rs7205784	chr16:g.83740167	C-T	PCR-26
rs1364118	chr16:g.83740179	T-A	PCR-26
rs10853817	chr19:g.30071237	A-G	PCR-28
rs12608877	chr19:g.30071243	A-T	PCR-28
rs10853818	chr19:g.30071263	C-A	PCR-28
rs10853819	chr19:g.30071274	G-A	PCR-28
rs10853820	chr19:g.30071287	T-C	PCR-28
rs184518376	chr19:g.30071288	G-A	PCR-28
rs555466901	chr19:g.30071306	C-T	PCR-28
rs572268098	chr19:g.30071341	C-T	PCR-28
rs534820676	chr19:g.30071354	G-A	PCR-28
rs10853821	chr19:g.30071359	T-C	PCR-28
rs529450508	chr19:g.30071399	G-T	PCR-28
rs12608899	chr19:g.30071417	A-G	PCR-28
rs116247894	chr19:g.30071418	C-T	PCR-28
rs574005168	chr19:g.30071439	A-G	PCR-28
rs542748909	chr19:g.30071471	C-G	PCR-28
rs559457817	chr19:g.30071472	C-T	PCR-28
rs548181231	chr20:g.59309954	G-A	PCR-29
rs2865944	chr20:g.59309997	A-G	PCR-29
rs114260772	chr20:g.59310005	C-T	PCR-29
rs114800846	chr20:g.59310006	T-C	PCR-29
rs116091174	chr20:g.59310007	C-T	PCR-29
rs10485469	chr20:g.59310123	T-G	PCR-29
rs546812515	chr20:g.59310136	T-C	PCR-29
rs1202013	chr20:g.59310148	G-A	PCR-29
rs1202014	chr20:g.59310162	A-G	PCR-29
rs556886810	chr20:g.59310173	T-G	PCR-29
rs575939325	chr20:g.59310184	G-A	PCR-29
rs9260128	chr6:g.29910492	G-T	HLA-A-E2
rs2904759	chr6:g.29910496	C-G	HLA-A-E2
rs41549214	chr6:g.29910506	G-T	HLA-A-E2
rs41554614	chr6:g.29910507	C-T	HLA-A-E2
rs9260129	chr6:g.29910508	A-G	HLA-A-E2
rs45620138	chr6:g.29910516	G-A	HLA-A-E2
rs560549696	chr6:g.29910518	C-T	HLA-A-E2
rs180796996	chr6:g.29910519	A-C	HLA-A-E2
rs9260130	chr6:g.29910522	G-C	HLA-A-E2
rs41541319	chr6:g.29910541	C-A	HLA-A-E2
rs1136659	chr6:g.29910557	T-A	HLA-A-E2
rs41545116	chr6:g.29910566	G-A	HLA-A-E2

rs1059423	chr6:g.29910581	C-A	HLA-A-E2
rs1059426	chr6:g.29910583	C-T	HLA-A-E2
rs12721672	chr6:g.29910586	G-A	HLA-A-E2
rs41549014	chr6:g.29910587	G-A	HLA-A-E2
rs12721675	chr6:g.29910604	C-A	HLA-A-E2
rs1059438	chr6:g.29910623	A-T	HLA-A-E2
rs1059441	chr6:g.29910636	G-A	HLA-A-E2
rs3173429	chr6:g.29910640	C-T	HLA-A-E2
rs41559117	chr6:g.29910660	A-G	HLA-A-E2
rs707910	chr6:g.29910663	G-A	HLA-A-E2
rs199474404	chr6:g.29910679	G-A	HLA-A-E2
rs41555915	chr6:g.29910681	C-T	HLA-A-E2
rs1059449	chr6:g.29910698	G-A	HLA-A-E2
rs41541222	chr6:g.29910703	G-T	HLA-A-E2
rs1059455	chr6:g.29910716	C-G	HLA-A-E2
rs199474424	chr6:g.29910721	G-C	HLA-A-E2
rs199474430	chr6:g.29910725	C-G	HLA-A-E2
rs199474436	chr6:g.29910730	T-A	HLA-A-E2
rs79361534	chr6:g.29910731	G-A	HLA-A-E2
rs199474443	chr6:g.29910737	G-A	HLA-A-E2
rs554470055	chr6:g.29910740	C-A	HLA-A-E2
rs78306866	chr6:g.29910742	G-C	HLA-A-E2
rs199474457	chr6:g.29910750	C-T	HLA-A-E2
rs1136683	chr6:g.29910752	G-C	HLA-A-E2
rs199474459	chr6:g.29910756	G-T	HLA-A-E2
rs1136688	chr6:g.29910761	G-A	HLA-A-E2
rs2231004	chr6:g.29910762	A-G	HLA-A-E2
rs1136689	chr6:g.29910767	G-C	HLA-A-E2
rs41552317	chr6:g.29910768	G-C	HLA-A-E2
rs1136690	chr6:g.29910771	C-T	HLA-A-E2
rs9260138	chr6:g.29910773	C-G	HLA-A-E2
rs9260139	chr6:g.29910774	T-C	HLA-A-E2
rs9260140	chr6:g.29910777	G-T	HLA-A-E2
rs1136691	chr6:g.29910779	G-C	HLA-A-E2
rs199474463	chr6:g.29910782	T-C	HLA-A-E2
rs1136692	chr6:g.29910801	C-A	HLA-A-E2
rs41562315	chr6:g.29910816	G-A	HLA-A-E2
rs41559412	chr6:g.29910817	G-C	HLA-A-E2
rs9260142	chr6:g.29910820	G-C	HLA-A-E2
rs41541123	chr6:g.29910822	G-T	HLA-A-E2
rs63225861	chr6:g.29910834	G-C	HLA-A-E2
rs377364793	chr6:g.29910849	C-T	HLA-A-E2
rs576190891	chr6:g.29910856	G-A	HLA-A-E2
rs41555717	chr6:g.29910861	A-G	HLA-A-E2
rs41540614	chr6:g.29910865	C-G	HLA-A-E2
rs41543916	chr6:g.29910866	G-A	HLA-A-E2
rs540533064	chr6:g.29910871	C-G	HLA-A-E2
rs560554218	chr6:g.29910887	G-T	HLA-A-E2
rs9260148	chr6:g.29911003	G-A	HLA-A-E3
rs41547618	chr6:g.29911005	G-A	HLA-A-E3
rs41562213	chr6:g.29911011	C-T	HLA-A-E3
rs9260149	chr6:g.29911019	T-C	HLA-A-E3
rs9260150	chr6:g.29911029	G-T	HLA-A-E3
rs9260151	chr6:g.29911030	C-T	HLA-A-E3
rs9260152	chr6:g.29911037	G-C	HLA-A-E3
rs41554914	chr6:g.29911038	G-A	HLA-A-E3
rs41556913	chr6:g.29911052	C-T	HLA-A-E3
rs199474485	chr6:g.29911063	T-G	HLA-A-E3

rs1136695	chr6:g.29911064	A-G	HLA-A-E3
rs41551416	chr6:g.29911068	T-C	HLA-A-E3
rs12721702	chr6:g.29911073	C-T	HLA-A-E3
rs199474490	chr6:g.29911077	G-C	HLA-A-E3
rs1136700	chr6:g.29911086	T-C	HLA-A-E3
rs1136702	chr6:g.29911092	G-T	HLA-A-E3
rs1059488	chr6:g.29911098	T-C	HLA-A-E3
rs41563718	chr6:g.29911103	C-G	HLA-A-E3
rs41540615	chr6:g.29911112	C-T	HLA-A-E3
rs41540315	chr6:g.29911113	C-G	HLA-A-E3
rs3173420	chr6:g.29911114	G-A	HLA-A-E3
rs3173418	chr6:g.29911124	C-T	HLA-A-E3
rs41553319	chr6:g.29911127	C-T	HLA-A-E3
rs1059506	chr6:g.29911149	C-T	HLA-A-E3
rs1059509	chr6:g.29911154	C-A	HLA-A-E3
rs1059514	chr6:g.29911190	G-A	HLA-A-E3
rs41541518	chr6:g.29911194	C-G	HLA-A-E3
rs1059517	chr6:g.29911203	A-C	HLA-A-E3
rs1059526	chr6:g.29911218	G-A	HLA-A-E3
rs1059535	chr6:g.29911222	C-T	HLA-A-E3
rs1059536	chr6:g.29911225	A-G	HLA-A-E3
rs9260155	chr6:g.29911239	T-C	HLA-A-E3
rs1059542	chr6:g.29911256	T-G	HLA-A-E3
rs3098019	chr6:g.29911271	G-C	HLA-A-E3
rs76185201	chr6:g.29911272	T-G	HLA-A-E3
rs41563116	chr6:g.29911284	T-C	HLA-A-E3
rs2517725	chr6:g.29911326	A-G	HLA-A-E3
rs539679368	chr6:g.29911332	C-A	HLA-A-E3
rs556330213	chr6:g.29911333	A-G	HLA-A-E3
rs570061466	chr6:g.29911334	C-T	HLA-A-E3
rs535706174	chr6:g.29911339	C-A	HLA-A-E3
rs201850551	chr6:g.29911348	G-C	HLA-A-E3
rs3132687	chr6:g.29911356	G-A	HLA-A-E3
rs17878860	chr6:g.29911360	A-G	HLA-A-E3
rs2507989	chr6:g.31324415	A-C	HLA-B-E2
rs17881210	chr6:g.31324448	G-A	HLA-B-E2
rs41559314	chr6:g.31324467	G-T	HLA-B-E2
rs3180379	chr6:g.31324491	C-A	HLA-B-E2
rs9266161	chr6:g.31324494	A-G	HLA-B-E2
rs9266162	chr6:g.31324495	G-C	HLA-B-E2
rs3190923	chr6:g.31324499	C-G	HLA-B-E2
rs1050388	chr6:g.31324506	C-T	HLA-B-E2
rs1131217	chr6:g.31324507	T-C	HLA-B-E2
rs41553715	chr6:g.31324509	T-A	HLA-B-E2
rs1131215	chr6:g.31324516	C-A	HLA-B-E2
rs1131214	chr6:g.31324523	T-C	HLA-B-E2
rs1131212	chr6:g.31324526	C-G	HLA-B-E2
rs1140404	chr6:g.31324527	T-G	HLA-B-E2
rs41546313	chr6:g.31324530	G-C	HLA-B-E2
rs41541616	chr6:g.31324535	G-C	HLA-B-E2
rs41543121	chr6:g.31324537	A-T	HLA-B-E2
rs707909	chr6:g.31324538	G-C	HLA-B-E2
rs1131202	chr6:g.31324539	A-T	HLA-B-E2
rs1131201	chr6:g.31324542	T-C	HLA-B-E2
rs1065386	chr6:g.31324547	G-C	HLA-B-E2
rs1050570	chr6:g.31324549	T-C	HLA-B-E2
rs141484466	chr6:g.31324552	G-C	HLA-B-E2
rs41556614	chr6:g.31324561	A-G	HLA-B-E2

rs1050564	chr6:g.31324562	C-T	HLA-B-E2
rs9266175	chr6:g.31324574	C-T	HLA-B-E2
rs143961508	chr6:g.31324580	T-C	HLA-B-E2
rs145974360	chr6:g.31324582	T-C	HLA-B-E2
rs1050556	chr6:g.31324586	C-T	HLA-B-E2
rs151341132	chr6:g.31324590	G-C	HLA-B-E2
rs1050543	chr6:g.31324595	C-G	HLA-B-E2
rs1050538	chr6:g.31324599	T-G	HLA-B-E2
rs9266178	chr6:g.31324603	C-T	HLA-B-E2
rs1050529	chr6:g.31324615	C-T	HLA-B-E2
rs41552717	chr6:g.31324622	G-A	HLA-B-E2
rs1065378	chr6:g.31324633	T-G	HLA-B-E2
rs1050518	chr6:g.31324641	T-A	HLA-B-E2
rs1050517	chr6:g.31324643	G-C	HLA-B-E2
rs9266183	chr6:g.31324647	T-C	HLA-B-E2
rs9266184	chr6:g.31324664	T-G	HLA-B-E2
rs1050502	chr6:g.31324667	G-A	HLA-B-E2
rs3177890	chr6:g.31324682	C-T	HLA-B-E2
rs3177891	chr6:g.31324691	G-A	HLA-B-E2
rs1050486	chr6:g.31324702	C-T	HLA-B-E2
rs41559921	chr6:g.31324703	G-A	HLA-B-E2
rs1131170	chr6:g.31324705	A-C	HLA-B-E2
rs41555216	chr6:g.31324725	G-A	HLA-B-E2
rs151341084	chr6:g.31324740	G-A	HLA-B-E2
rs147324178	chr6:g.31324742	T-C	HLA-B-E2
rs41558815	chr6:g.31324750	G-A	HLA-B-E2
rs41552714	chr6:g.31324756	G-A	HLA-B-E2
rs41560522	chr6:g.31324764	C-A	HLA-B-E2
rs41558918	chr6:g.31324774	C-A	HLA-B-E2
rs9266193	chr6:g.31324788	T-C	HLA-B-E2
rs9266194	chr6:g.31324790	A-G	HLA-B-E2
rs41552116	chr6:g.31324798	C-A	HLA-B-E2
rs9266195	chr6:g.31324820	G-T	HLA-B-E2
rs9266197	chr6:g.31324830	G-A	HLA-B-E2
rs41549217	chr6:g.31323885	G-T	HLA-B-E3
rs2596493	chr6:g.31323945	A-C	HLA-B-E3
rs1131285	chr6:g.31323953	C-G	HLA-B-E3
rs1131279	chr6:g.31323958	T-G	HLA-B-E3
rs1131275	chr6:g.31323960	G-C	HLA-B-E3
rs148138285	chr6:g.31323962	C-T	HLA-B-E3
rs41551517	chr6:g.31323992	A-C	HLA-B-E3
rs41551919	chr6:g.31323993	C-G	HLA-B-E3
rs9266141	chr6:g.31324023	C-G	HLA-B-E3
rs151341293	chr6:g.31324036	T-A	HLA-B-E3
rs41551018	chr6:g.31324051	C-A	HLA-B-E3
rs41541519	chr6:g.31324064	T-A	HLA-B-E3
rs41562716	chr6:g.31324074	C-T	HLA-B-E3
rs709053	chr6:g.31324077	C-G	HLA-B-E3
rs1050370	chr6:g.31324089	G-A	HLA-B-E3
rs137854729	chr6:g.31324126	T-C	HLA-B-E3
rs12697944	chr6:g.31324143	G-T	HLA-B-E3
rs1050628	chr6:g.31324152	A-G	HLA-B-E3
rs1050379	chr6:g.31324154	G-A	HLA-B-E3
rs41558819	chr6:g.31324166	G-A	HLA-B-E3
rs1131112	chr6:g.31324184	C-G	HLA-B-E3
rs3179865	chr6:g.31324194	G-A	HLA-B-E3
rs41562913	chr6:g.31324198	A-G	HLA-B-E3
rs41556417	chr6:g.31324202	T-C	HLA-B-E3

rs1131235	chr6:g.31324206	G-C	HLA-B-E3
rs41556312	chr6:g.31324207	A-C	HLA-B-E3
rs12721829	chr6:g.31324208	G-A	HLA-B-E3
rs1131233	chr6:g.31324209	G-A	HLA-B-E3
rs4999716	chr6:g.31324227	C-A	HLA-B-E3
rs5006140	chr6:g.31324245	A-C	HLA-B-E3
rs41558413	chr6:g.31239340	G-A	HLA-C-E2
rs11547346	chr6:g.31239346	T-G	HLA-C-E2
rs41546813	chr6:g.31239363	C-T	HLA-C-E2
rs527516951	chr6:g.31239367	G-T	HLA-C-E2
rs1131122	chr6:g.31239376	C-T	HLA-C-E2
rs1131123	chr6:g.31239378	T-G	HLA-C-E2
rs281860421	chr6:g.31239402	C-T	HLA-C-E2
rs17408553	chr6:g.31239407	G-T	HLA-C-E2
rs2308557	chr6:g.31239417	C-T	HLA-C-E2
rs539016289	chr6:g.31239429	G-C	HLA-C-E2
rs41543814	chr6:g.31239430	C-T	HLA-C-E2
rs28626310	chr6:g.31239449	C-G	HLA-C-E2
rs41555917	chr6:g.31239455	T-C	HLA-C-E2
rs41540214	chr6:g.31239480	C-T	HLA-C-E2
rs1050409	chr6:g.31239501	G-T	HLA-C-E2
rs1050414	chr6:g.31239506	C-G	HLA-C-E2
rs1050420	chr6:g.31239518	C-T	HLA-C-E2
rs1050428	chr6:g.31239543	C-T	HLA-C-E2
rs707911	chr6:g.31239577	A-C	HLA-C-E2
rs1050437	chr6:g.31239585	C-T	HLA-C-E2
rs41542719	chr6:g.31239593	T-C	HLA-C-E2
rs281860337	chr6:g.31239601	C-T	HLA-C-E2
rs281860336	chr6:g.31239602	G-A	HLA-C-E2
rs281860333	chr6:g.31239607	G-A	HLA-C-E2
rs1050444	chr6:g.31239614	G-A	HLA-C-E2
rs1050445	chr6:g.31239616	C-A	HLA-C-E2
rs1050446	chr6:g.31239617	G-T	HLA-C-E2
rs1131151	chr6:g.31239630	C-T	HLA-C-E2
rs41556116	chr6:g.31239652	G-A	HLA-C-E2
rs41560121	chr6:g.31239653	C-A	HLA-C-E2
rs41559317	chr6:g.31239660	G-T	HLA-C-E2
rs538276256	chr6:g.31239661	G-A	HLA-C-E2
rs9264669	chr6:g.31239681	A-T	HLA-C-E2
rs281860317	chr6:g.31239697	C-T	HLA-C-E2
rs542265088	chr6:g.31239711	C-T	HLA-C-E2
rs560644504	chr6:g.31239712	C-A	HLA-C-E2
rs9264670	chr6:g.31239722	A-C	HLA-C-E2
rs29029490	chr6:g.31239727	C-A	HLA-C-E2
rs2074494	chr6:g.31239733	C-T	HLA-C-E2
rs529374079	chr6:g.31238782	C-T	HLA-C-E3
rs181531333	chr6:g.31238783	C-T	HLA-C-E3
rs41561913	chr6:g.31238793	G-A	HLA-C-E3
rs41544614	chr6:g.31238801	C-G	HLA-C-E3
rs552098447	chr6:g.31238807	A-C	HLA-C-E3
rs9264653	chr6:g.31238814	T-C	HLA-C-E3
rs41544212	chr6:g.31238815	A-G	HLA-C-E3
rs41561812	chr6:g.31238822	G-A	HLA-C-E3
rs41552417	chr6:g.31238874	C-T	HLA-C-E3
rs2308598	chr6:g.31238889	T-C	HLA-C-E3
rs201707949	chr6:g.31238898	A-C	HLA-C-E3
rs138584390	chr6:g.31238899	C-G	HLA-C-E3
rs79636386	chr6:g.31238930	A-C	HLA-C-E3

rs697743	chr6:g.31238931	G-A	HLA-C-E3
rs41552817	chr6:g.31238943	C-T	HLA-C-E3
rs142570222	chr6:g.31238970	T-A	HLA-C-E3
rs281860496	chr6:g.31238980	C-T	HLA-C-E3
rs2308585	chr6:g.31238983	G-C	HLA-C-E3
rs281860494	chr6:g.31238989	C-G	HLA-C-E3
rs281860475	chr6:g.31239033	C-A	HLA-C-E3
rs281860471	chr6:g.31239044	T-C	HLA-C-E3
rs281860470	chr6:g.31239047	G-A	HLA-C-E3
rs1065406	chr6:g.31239049	G-T	HLA-C-E3
rs2308575	chr6:g.31239057	C-T	HLA-C-E3
rs2308574	chr6:g.31239060	A-G	HLA-C-E3
rs41563421	chr6:g.31239073	G-T	HLA-C-E3
rs41544015	chr6:g.31239081	C-G	HLA-C-E3
rs1050384	chr6:g.31239082	G-C	HLA-C-E3
rs34592426	chr6:g.31239090	G-C	HLA-C-E3
rs1131114	chr6:g.31239100	A-G	HLA-C-E3
rs1131118	chr6:g.31239108	T-A	HLA-C-E3
rs1071649	chr6:g.31239114	G-A	HLA-C-E3
rs41543218	chr6:g.31239124	C-A	HLA-C-E3
rs28367580	chr6:g.31239157	C-G	HLA-C-E3
rs72558143	chr6:g.31239172	C-T	HLA-C-E3
rs9274371	chr6:g.32632553	A-G	HLA-DQ-B1-E2
rs9274372	chr6:g.32632555	G-A	HLA-DQ-B1-E2
rs41263826	chr6:g.32632557	A-C	HLA-DQ-B1-E2
rs28746798	chr6:g.32632558	G-C	HLA-DQ-B1-E2
rs142144985	chr6:g.32632561	G-T	HLA-DQ-B1-E2
rs150769232	chr6:g.32632563	G-A	HLA-DQ-B1-E2
rs139172132	chr6:g.32632566	G-C	HLA-DQ-B1-E2
rs1140322	chr6:g.32632578	T-G	HLA-DQ-B1-E2
rs1140321	chr6:g.32632581	T-G	HLA-DQ-B1-E2
rs1130397	chr6:g.32632587	A-G	HLA-DQ-B1-E2
rs1140320	chr6:g.32632589	A-G	HLA-DQ-B1-E2
rs1140319	chr6:g.32632592	C-G	HLA-DQ-B1-E2
rs1140318	chr6:g.32632593	C-T	HLA-DQ-B1-E2
rs17412833	chr6:g.32632598	A-T	HLA-DQ-B1-E2
rs1130391	chr6:g.32632624	C-T	HLA-DQ-B1-E2
rs1130392	chr6:g.32632628	G-C	HLA-DQ-B1-E2
rs9274384	chr6:g.32632635	A-C	HLA-DQ-B1-E2
rs3204379	chr6:g.32632642	C-T	HLA-DQ-B1-E2
rs1130390	chr6:g.32632647	T-C	HLA-DQ-B1-E2
rs1130386	chr6:g.32632650	C-T	HLA-DQ-B1-E2
rs9274390	chr6:g.32632659	C-T	HLA-DQ-B1-E2
rs281874783	chr6:g.32632660	T-G	HLA-DQ-B1-E2
rs1130382	chr6:g.32632672	G-A	HLA-DQ-B1-E2
rs41552812	chr6:g.32632689	C-T	HLA-DQ-B1-E2
rs1130381	chr6:g.32632691	G-A	HLA-DQ-B1-E2
rs1049074	chr6:g.32632711	C-T	HLA-DQ-B1-E2
rs3210148	chr6:g.32632714	G-C	HLA-DQ-B1-E2
rs1049073	chr6:g.32632717	G-A	HLA-DQ-B1-E2
rs1049083	chr6:g.32632724	C-T	HLA-DQ-B1-E2
rs1049082	chr6:g.32632744	C-T	HLA-DQ-B1-E2
rs1063318	chr6:g.32632745	G-A	HLA-DQ-B1-E2
rs1049068	chr6:g.32632777	C-T	HLA-DQ-B1-E2
rs1140310	chr6:g.32632783	A-C	HLA-DQ-B1-E2
rs546609663	chr6:g.32632785	G-A	HLA-DQ-B1-E2
rs41540813	chr6:g.32632790	C-A	HLA-DQ-B1-E2
rs3204373	chr6:g.32632801	G-A	HLA-DQ-B1-E2

rs1130368	chr6:g.32632818	T-G	HLA-DQ-B1-E2
rs1130375	chr6:g.32632820	C-G	HLA-DQ-B1-E2
rs9274407	chr6:g.32632832	A-T	HLA-DQ-B1-E2
rs12722107	chr6:g.32632833	A-G	HLA-DQ-B1-E2
rs9274408	chr6:g.32632850	G-A	HLA-DQ-B1-E2
rs74186130	chr6:g.32632859	C-G	HLA-DQ-B1-E2
rs9274409	chr6:g.32632864	G-A	HLA-DQ-B1-E2
rs555329021	chr6:g.32632872	A-C	HLA-DQ-B1-E2
rs201184533	chr6:g.32632887	C-T	HLA-DQ-B1-E2
rs281862010	chr6:g.32632890	C-G	HLA-DQ-B1-E2
rs41270915	chr6:g.32629692	G-T	HLA-DQ-B1-E3
rs9273902	chr6:g.32629716	T-C	HLA-DQ-B1-E3
rs36233043	chr6:g.32629718	G-A	HLA-DQ-B1-E3
rs9273912	chr6:g.32629737	C-T	HLA-DQ-B1-E3
rs1130399	chr6:g.32629755	G-A	HLA-DQ-B1-E3
rs1130398	chr6:g.32629764	C-T	HLA-DQ-B1-E3
rs1049092	chr6:g.32629802	A-G	HLA-DQ-B1-E3
rs9273942	chr6:g.32629806	C-T	HLA-DQ-B1-E3
rs701564	chr6:g.32629809	C-T	HLA-DQ-B1-E3
rs9273951	chr6:g.32629822	T-C	HLA-DQ-B1-E3
rs1049133	chr6:g.32629847	A-G	HLA-DQ-B1-E3
rs1049130	chr6:g.32629859	A-G	HLA-DQ-B1-E3
rs1049088	chr6:g.32629868	A-G	HLA-DQ-B1-E3
rs1049087	chr6:g.32629889	G-A	HLA-DQ-B1-E3
rs1063323	chr6:g.32629891	C-T	HLA-DQ-B1-E3
rs1049086	chr6:g.32629904	A-G	HLA-DQ-B1-E3
rs2647032	chr6:g.32629905	T-C	HLA-DQ-B1-E3
rs41544112	chr6:g.32629920	C-T	HLA-DQ-B1-E3
rs41542812	chr6:g.32629931	C-G	HLA-DQ-B1-E3
rs1049107	chr6:g.32629936	C-T	HLA-DQ-B1-E3
rs1063321	chr6:g.32629955	C-T	HLA-DQ-B1-E3
rs1049100	chr6:g.32629963	C-T	HLA-DQ-B1-E3
rs68027833	chr6:g.32630028	G-A	HLA-DQ-B1-E3
rs9274031	chr6:g.32630033	C-T	HLA-DQ-B1-E3
rs9274038	chr6:g.32630047	A-G	HLA-DQ-B1-E3
rs9274045	chr6:g.32630055	T-C	HLA-DQ-B1-E3
rs184050448	chr6:g.32551844	C-T	HLA-DR-B1-E2
rs144319718	chr6:g.32551850	C-T	HLA-DR-B1-E2
rs548637317	chr6:g.32551854	G-A	HLA-DR-B1-E2
rs9269939	chr6:g.32551878	A-G	HLA-DR-B1-E2
rs9269940	chr6:g.32551879	T-C	HLA-DR-B1-E2
rs558916921	chr6:g.32551899	T-C	HLA-DR-B1-E2
rs17882028	chr6:g.32551911	C-A	HLA-DR-B1-E2
rs17885482	chr6:g.32551912	A-C	HLA-DR-B1-E2
rs17424145	chr6:g.32551915	A-G	HLA-DR-B1-E2
rs17879599	chr6:g.32551959	C-G	HLA-DR-B1-E2
rs17881965	chr6:g.32551961	G-C	HLA-DR-B1-E2
rs17885388	chr6:g.32551962	C-T	HLA-DR-B1-E2
rs1059584	chr6:g.32551996	G-T	HLA-DR-B1-E2
rs17880292	chr6:g.32552000	C-T	HLA-DR-B1-E2
rs17885908	chr6:g.32552020	A-G	HLA-DR-B1-E2
rs17884945	chr6:g.32552029	A-T	HLA-DR-B1-E2
rs16822820	chr6:g.32552060	A-T	HLA-DR-B1-E2
rs17882603	chr6:g.32552067	C-T	HLA-DR-B1-E2
rs17879995	chr6:g.32552072	T-G	HLA-DR-B1-E2
rs1064664	chr6:g.32552075	A-G	HLA-DR-B1-E2
rs3175105	chr6:g.32552080	T-C	HLA-DR-B1-E2
rs11554462	chr6:g.32552081	A-G	HLA-DR-B1-E2

rs16822516	chr6:g.32552092	A-T	HLA-DR-B1-E2
rs17879981	chr6:g.32552121	A-T	HLA-DR-B1-E2
rs17879702	chr6:g.32552123	G-A	HLA-DR-B1-E2
rs17885011	chr6:g.32552127	C-T	HLA-DR-B1-E2
rs1136756	chr6:g.32552134	T-G	HLA-DR-B1-E2
rs9269958	chr6:g.32552143	C-T	HLA-DR-B1-E2
rs9269959	chr6:g.32552144	A-C	HLA-DR-B1-E2
rs9269962	chr6:g.32552177	G-A	HLA-DR-B1-E2
rs112353158	chr6:g.32552188	G-A	HLA-DR-B1-E2
rs9269964	chr6:g.32552212	C-T	HLA-DR-B1-E2
rs9269965	chr6:g.32552219	A-G	HLA-DR-B1-E2
rs149614319	chr6:g.32552223	T-A	HLA-DR-B1-E2
rs144353907	chr6:g.32552226	G-C	HLA-DR-B1-E2
rs115462544	chr6:g.32552229	G-A	HLA-DR-B1-E2
rs199704140	chr6:g.32549362	G-A	HLA-DR-B1-E3
rs701831	chr6:g.32549401	T-C	HLA-DR-B1-E3
rs557962729	chr6:g.32549443	G-A	HLA-DR-B1-E3
rs2308759	chr6:g.32549596	T-C	HLA-DR-B1-E3
rs199927008	chr6:g.32549627	A-C	HLA-DR-B1-E3
rs72850244	chr6:g.32549638	G-A	HLA-DR-B1-E3
rs143082263	chr6:g.32549642	T-G	HLA-DR-B1-E3
rs9269813	chr6:g.32549681	G-T	HLA-DR-B1-E3
rs535956089	chr6:g.32549683	G-T	HLA-DR-B1-E3
rs374639997	chr6:g.33048379	T-C	HLA-DP-B1-E2
rs3128959	chr6:g.33048380	G-A	HLA-DP-B1-E2
rs141207777	chr6:g.33048417	C-A	HLA-DP-B1-E2
rs1126504	chr6:g.33048457	C-G	HLA-DP-B1-E2
rs1126506	chr6:g.33048459	T-G	HLA-DP-B1-E2
rs12722013	chr6:g.33048460	T-C	HLA-DP-B1-E2
rs1126509	chr6:g.33048461	T-A	HLA-DP-B1-E2
rs1126511	chr6:g.33048466	G-T	HLA-DP-B1-E2
rs1126513	chr6:g.33048467	G-T	HLA-DP-B1-E2
rs188627284	chr6:g.33048483	C-T	HLA-DP-B1-E2
rs41555313	chr6:g.33048484	G-C	HLA-DP-B1-E2
rs12722018	chr6:g.33048532	G-C	HLA-DP-B1-E2
rs9277348	chr6:g.33048538	T-C	HLA-DP-B1-E2
rs1042117	chr6:g.33048539	T-A	HLA-DP-B1-E2
rs1042121	chr6:g.33048542	C-T	HLA-DP-B1-E2
rs12722022	chr6:g.33048564	G-A	HLA-DP-B1-E2
rs553686794	chr6:g.33048576	G-A	HLA-DP-B1-E2
rs575237777	chr6:g.33048579	G-A	HLA-DP-B1-E2
rs707958	chr6:g.33048599	C-A	HLA-DP-B1-E2
rs9277351	chr6:g.33048600	T-G	HLA-DP-B1-E2
rs1042131	chr6:g.33048602	C-A	HLA-DP-B1-E2
rs1042133	chr6:g.33048606	G-C	HLA-DP-B1-E2
rs1042136	chr6:g.33048628	A-C	HLA-DP-B1-E2
rs1042140	chr6:g.33048640	A-G	HLA-DP-B1-E2
rs12722027	chr6:g.33048641	A-G	HLA-DP-B1-E2
rs41546618	chr6:g.33048649	G-T	HLA-DP-B1-E2
rs41556420	chr6:g.33048651	G-T	HLA-DP-B1-E2
rs41551920	chr6:g.33048653	C-T	HLA-DP-B1-E2
rs1042151	chr6:g.33048661	A-G	HLA-DP-B1-E2
rs1042153	chr6:g.33048663	G-A	HLA-DP-B1-E2
rs534577141	chr6:g.33048689	G-A	HLA-DP-B1-E2
rs552975041	chr6:g.33048691	C-G	HLA-DP-B1-E2
rs553665868	chr6:g.33048694	A-G	HLA-DP-B1-E2
rs536113120	chr6:g.33048703	C-G	HLA-DP-B1-E2
rs41541915	chr6:g.33048707	G-A	HLA-DP-B1-E2

rs200678831	chr6:g.33052712	C-T	HLA-DP-B1-E3
rs9277452	chr6:g.33052717	A-C	HLA-DP-B1-E3
rs561674320	chr6:g.33052719	G-A	HLA-DP-B1-E3
rs9277454	chr6:g.33052723	C-G	HLA-DP-B1-E3
rs1042187	chr6:g.33052768	T-C	HLA-DP-B1-E3
rs562166558	chr6:g.33052802	C-T	HLA-DP-B1-E3
rs1042212	chr6:g.33052803	G-A	HLA-DP-B1-E3
rs61736934	chr6:g.33052836	G-A	HLA-DP-B1-E3
rs566678973	chr6:g.33052852	G-T	HLA-DP-B1-E3
rs534162897	chr6:g.33052867	G-A	HLA-DP-B1-E3
rs1042331	chr6:g.33052950	T-C	HLA-DP-B1-E3
rs1042335	chr6:g.33052958	C-T	HLA-DP-B1-E3
rs14362	chr6:g.33052981	C-A	HLA-DP-B1-E3
rs1071597	chr6:g.33052986	T-C	HLA-DP-B1-E3
rs41559424	chr6:g.33052998	C-T	HLA-DP-B1-E3
rs539695276	chr6:g.33053025	C-T	HLA-DP-B1-E3

Supplementary table 2. HP763SNPs bed BLOCK file. The SNPs appertaining to the same block were identified with the same block identification Tag (column : Block tag).

8.4. Binomial Analysis

We used the binomial distribution to compute, for each population, the probability to find the number of informative markers, given the average informativity of the 20 markers in the ACCh_hb panel. In the pairs of related the average informativity was considered a half of the one computed for the unrelated pairs.

The probability to find n informative markers in a panel containing N markers with average informativity x is given by the formula:

$$p_{N,x}(n) = \frac{N!}{n!\,(N-n)!}x^n(1-x)^{N-n}$$

In the following table we report the result of this analysis on our ACCh_hb panel, in which the average informativity was 0,639 in EUR, 0,496 in EAS and 0,596 in ALL_p.

Number of informative markers	**Probability EUR**		**Probability EAS**		**Probability ALL_p**	
	Unrelated	Related	Unrelated	Related	Unrelated	Related
0	0,0	0,0	0,0	0,0	0,0	0,0
1	0,0	0,4	0,0	2,4	0,0	0,6
2	0,0	1,9	0,0	6,7	0,0	2,9
3	0,0	5,4	0,1	13,4	0,0	7,4
4	0,0	10,7	0,5	19,0	0,0	13,3
5	0,0	16,0	1,6	20,2	0,1	18,1
6	0,2	18,8	3,9	16,9	0,5	19,2
7	0,6	17,7	7,8	11,2	1,6	16,3
8	1,7	13,5	12,4	6,1	3,8	11,2
9	4,0	8,4	16,3	2,7	7,5	6,4
10	7,9	4,4	17,6	1,0	12,1	3,0
11	12,7	1,9	15,8	0,3	16,2	1,1
12	16,8	0,7	11,6	0,1	18,0	0,4
13	18,4	0,2	7,0	0,0	16,3	0,1
14	16,2	0,0	3,5	0,0	12,1	0,0
15	11,6	0,0	1,4	0,0	7,1	0,0
16	6,4	0,0	0,4	0,0	3,3	0,0
17	2,6	0,0	0,1	0,0	1,1	0,0
18	0,8	0,0	0,0	0,0	0,3	0,0
19	0,1	0,0	0,0	0,0	0,0	0,0
20	0,0	0,0	0,0	0,0	0,0	0,0

Supplementary table 3. Binomial Analysis. Probability to find n informative markers in a ACCh_hb panel, containing 20 markers.

9. ACKNOWLEDGMENTS

I would like to express my gratitude and appreciation to my supervisor, Doctor Giovanni Maria Severini, for his support and also for his humanity during the whole PhD period.

I would also like to thank my supervisor, Prof. Roberta Bulla, for her assistance.

A big thanks goes also to people who helped me to carry out this work:

- Dr. Emmanouil Athanasakis, Dr. Danilo Licastro and Dr. Massimiliano Cocca for their bioinformatics support;
- Prof. Fiorella Florian and Prof. Alberto Pallavicini for mentoring in the custom NGS library construction.

I also want to show my gratitude to AGMEN, "Associazione Genitori Malati Emopatici Neoplastici, Friuli Venezia Giulia" and the Institute for Maternal and Child Health "I.R.C.C.S. Burlo Garofolo" Trieste for supporting my PhD project.

I'm sincerely grateful to my family for their continuous encouragement.

Last, but not least, my appreciation goes to all my colleagues and friends met during this period.