Application of Single Nucleotide Polymorphism to Quantification of Hematopoietic Chimerism in Children with Allogeneic Hematopoietic Stem Cell Transplants

LAU, Wai Hung

A Thesis Submitted in Partial Fulfilment
of the Requirements for the Degree of
Master of Philosophy
in
Medical Sciences

The Chinese University of Hong Kong
September 2013

Abstract of thesis entitled:

Application of Single Nucleotide Polymorphism to Quantification of Hematopoietic Chimerism in Children with Allogeneic Hematopoietic Stem Cell Transplants

Submitted by LAU, Wai Hung

for the degree of Master of Philosophy in Medical Sciences

at The Chinese University of Hong Kong in September 2013

Background

Allogeneic hematopoietic stem cell transplantation (HSCT) is considered as an effective therapeutic option for patients with malignant or nonmalignant hematopoietic diseases. The degree of graft rejection and disease relapse is correlated to the proportion of donor and recipient hematopoietic cells in the post-transplant period. Therefore, early analysis of the chimerism pattern after HSCT with high sensitivity and accuracy is clinically important in monitoring the process of transplantation and follow-up immunotherapy. Currently, polymerase chain reaction of short tandem repeats (STR-PCR) is regarded as a standard tool for chimerism analysis after HSCT. STR-PCR, however, is limited by a number of factors: (I) STR marker-dependent detection limit, (II) amplification bias and (III) reduced specificity and inaccurate quantification due to presence of shutter peaks. It has been shown that single nucleotide polymorphisms (SNPs) can be used as genetic markers for forensic identification of individuals. This M.Phil. project was aimed to develop a quantitative SNP genotyping assay, which was based on a sensitive matrix-assisted laser desorption ionization time-of-flight mass spectrometry (MALDI-TOF MS) platform, for quantification of levels of hematopoietic mixed chimerism in post-HSCT paediatric patients.

Methods

MALDI-TOF MS has emerged as a reliable tool for DNA analysis. The first part of this study was to identify a panel of informative SNPs for differentiating genomic identities among individuals. A panel of informative SNPs in Han Chinese in Beijing population was selected from the HapMap database. Then, an in-house MALDI-TOF MS-based SNP genotyping assay with aid of competitive PCR and primer extension reactions was developed and optimized for qualitative (genotyping of the targeted SNPs) and quantitative (chimerism assessment in post-transplant patients) analysis of SNPs. Finally, the assay linearity, accuracy, precision and detection limit were assessed. Performance of the developed SNP genotyping assay and the routine STR-PCR on quantification of hematopoietic chimerism in 30 recruited paediatric patients following allogeneic HSCT was compared.

Results

A panel of 8 genetically linkage-independent SNPs on different human chromosomes was identified by bioinformatic analysis. The probability of having at least one targeted informative SNP between two unrelated individuals was calculated to be 0.975. Our SNP genotyping assay comprised three molecular biology reactions, namely competitive multiplex PCR, shrimp alkaline phosphatase (SAP) treatment and primer extension reactions. The entire assay was developed. Using the optimized assay, the SNP genotyping results indicated that over 90% of the examined donor-recipient pairs had informative SNPs. The assay displayed a high linearity across a predicted working range (R^2>0.99). It could accurately measure the level of mixed chimerism with detection limit down to 1.4%. Its precision increased with the level of mixed chimerism. Our developed assay and the STR-PCR assay generated highly similar results. The level of recipient DNA were measurable only by the STR-PCR assay for the cases with mixed chimerism <1.4%.

Conclusions

An assay capable of genotyping 8 targeted SNPs was successfully developed. It also proven to be useful in measuring the level of mixed chimerism in clinical samples in quantitative manner. Nevertheless, further optimizations may be required before it can be applied to routine clinical use due to its limitations, such as long assay turnaround time. This M.Phil. study provided a foundation for future development of MALDI-TOF MS-based SNP genotyping assay for routine assessment of chimerism.

撮要

背景

異基因造血幹細胞移植（HSCT）被認為是在治療惡性或良性造血系統疾病的一個有效方法。移植物排斥反應的程度和疾病的復發是與移植後供體和受體的造血幹細胞所佔的比例有關。因此，透過高靈敏度和精確的技術，分析早期造血幹細胞嵌合情況對監測移植後的過程及決定後續免疫療法有臨床上的意義。移植後造血幹細胞嵌合體分析的標準工具是短串聯重複序列聚合酶鏈反應（STR-PCR），但是此技術有多項限制：（I）依賴 STR 標記的檢測限，（II）倍大的偏差和（III）由影子峰做成的特異性下降和定量均不準確。單核苷酸多態性（SNPs）已被證明可作為遺傳標記來識別個人身份。本哲學碩士專題研究的目的是透過高靈敏的基質輔助激光脫附游離飛行時間質譜儀（MALDI-TOF-MS）技術來研發 SNP 基因分型檢測的定量技術，並且以此技術來檢測已接受了異基因 HSCT 的兒童的造血幹細胞混合嵌合體的程度。

方法

MALDI-TOF-MS 已發展成為一個可靠的 DNA 分析工具。此專題研究的第一部分是從世界人類基因組單體型圖數據庫的中國漢族(北京)中，找尋具資訊性的 SNPs，並以此作為識別個人基因型的標記。然後，為了作定性（辨別個人 SNPs 的基因型）和定量（檢測病人混合嵌合體的程度）的 SNPs 分析，我們研發了一個以 MALDI-TOF MS 為基礎，並結合競爭 PCR 和引物延伸反應的 SNP 基因分型檢測技術。最後，當此技術被優化後，我們評估了它的線性度，準確度，精密度和檢測限，另外，我們利用此技術和常用的 STR-PCR 檢測法為 30 名接受了異基因 HSCT 的兒科病人進行了定量的造血幹細胞混合嵌合體檢測，並比較其結果。

實驗結果

透過生物信息學分析，我們在不同人類染色體上選擇了八個基因聯動獨立的目標 SNPs，透過計算，在兩個沒有血緣關係的人找到至少一個具資訊性的目標 SNPs 的概率為 0.975。所研發的 SNP 基因分型檢測技術包括三個分子生物學反應: 多套式競爭性 PCR、蝦鹼性磷酸酶（SAP）處理和引物延伸反應。優化後的檢測技術在超過 90%的實驗對象中找到了具資訊性的 SNPs。實驗結果顯示此技術在特定範圍內有高線性（R^2> 0.99），其檢測限為 1.4%，並且可以精確地測量混合嵌合體的程度，其精確度會隨嵌合體的程度上升而提高。我們研發的技術和 STR-PCR 檢測法產生了非常相似的結果，但是，對於混合嵌合體的程度低於 1.4%的病人，我們只能透過 STR-PCR 檢測法作檢測。

結論

一個 SNP 基因分型檢測技術已被成功研發，它能辨別每個人的目標 SNPs 基因型，及後也被證明能檢測在臨床樣品中混合嵌合體的程度。然而，由於其局限性(如測定時間較長)，需要進一步優化這個技術才可應用到真正臨床上。本哲學碩士專題研究為未來將 MALDI-TOF MS 為基礎的 SNP 基因分型檢測技術應用在檢測混合嵌合體上提供了一個重要的基礎。

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

I wish to express my thanks to my supervisor, Prof. Terence C.W. Poon, for his academic advice, clear guidance, kindly encouragement, emotional support during the whole period of my study, and preparation of this thesis. In addition, I wish to acknowledge my colleagues. In particular, I am grateful to Ms. Eddy Ng who has taught me techniques in MALDI-TOF MS and proteomic study. Dr. Melody Lam is acknowledged for her help in bioinformatic analysis. Mr. Henry Poon is also acknowledged for his technical assistance in genomic study and his work in identification of SNP markers and STR-PCR assay.

I would like to thank Dr. C.K. Li and his team for provision of clinical specimens for this M.Phil. study.

Last but not least, I would like to thank my parents, Mr. Y.H. Lau and Mrs. S.Y. Yiu, and friends for supporting me continuously. In particular, I would like to give thanks to my friends Mr. Jacky Chan, Mr. Nicky Lam, Ms. Elas Sung and Ms. Hiedi Wong for accompanying me to go travelling before working on this thesis.

The financial support from The Chinese University of Hong Kong is highly appreciated.

LIST OF ABBREVIATIONS

Units:

Da	Dalton
g	Gram
hr	Hour
M	Molarity
mg	Milligram
μg	Microgram
min	Minute
mL	Milliliter
μL	Microliter
mM	Millimolarity
μM	Micromolarity
mol	Mole
m/z	Mass-to-charge ratio
pmol	Picomole
V	Voltage
v/v	Volume by volume
sec	Second
S/N	Signal-to-noise ratio

Diseases and complications:

ALCL	Anaplastic large-cell lymphoma
ALL	Acute lymphoblastic leukemia
AML	Acute myeloid leukemia
CLL	Chronic lymphocytic leukemia
CML	Chronic myeloid leukemia
DKC	Dyskeratosis congenita
FA	Fanconi's anemia
GVHD	Graft-versus-host disease
MDS	Myelodysplastic syndromes
MPD	Myeloproliferative disease
MRD	Minimal residual disease
SAA	Severe aplastic anemia
SCID	Severe combined immunodeficiency

Others:

ACN	Acetonitrile
CAR	Chimeric antigen receptor
CC	Complete chimerism
CV	Coefficient of variation
CHB	Han Chinese population in Beijing
CHCA	α-cyano-4-hydroxycinnamic acid

CHS	Han Chinese population in Shanghai
CNV	Copy number variation
CVC	Central venous catheter
dGTP	3’-deoxyguanosine triphosphate
dNTP	3’-deoxynucleoside triphosphate or deoxynucleotide
dTTP	3’-deoxythymidine triphosphate
ddATP	2’,3’-dideoxyadenosine triphosphate
ddCTP	2’,3’-dideoxycytidine triphosphate
ddGTP	2’,3’-dideoxyguanosine triphosphate
ddNTP	2’,3’-dideoxynucleoside triphosphate or dideoxynucleotide
ddTTP	2’,3’-dideoxythymidine triphosphate
DLI	Donor lymphocyte infusion
EDTA	Ethylenediaminetetraacetic acid
FAM	6-carboxyfluorescein
FISH	Fluorescent *in situ* hybridization
GVTE	Graft-versus-tumor effect
HLA	Human leukocyte antigen
3-HPA	3-hydroxypicolinic acid
HPLC	High performance liquid chromatography
HSCT	Hematopoietic stem cell transplantation
HWE	Hardy-Weinberg equilibrium
KRAS	Kirsten rat sarcoma viral oncogene homolog
LD	Linkage disequilibrium
MALDI	Matrix-assisted laser desorption/ionization
MC	Mixed chimerism
mHA	Minor histocompatibility antigen
MHC	Major histocompatibility complex
M_r	Molecular weight
MS	Mass spectrometry
NN	Nearest neighbor algorithm
PCR	Polymerase chain reaction
SAP	Shrimp alkaline phosphatase
SNP	Single nucleotide polymorphism
STR	Short tandem repeat
TBE	Tris/borate/EDTA
TEAA	Triethylammonium acetate
T_m	Melting temperature
TOF	Time-of-flight
VNTR	Variable number tandem repeat

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

CHAPTER 1: REVIEW OF LTERATURES

1.1 Hematopoietic Stem Cell Transplantation (HSCT)

1.1.1 Types of HSCT

Hematopoietic stem cell transplantation (HSCT) refers to the intravenous infusion of pulripotent hematopoietic stem cells derived from a range of sources, such as bone marrow, peripheral blood, or umbilical cord blood to replace the defective bone marrow eliminated by disease, chemotherapy or radiotherapy.[1] Over the last twenty years, it has been widely used worldwide to treat malignant or nonmalignant hematopoietic diseases, such as aplastic anemia and severe combined immunodeficiency.[2] Depending on different donor sources, it can be divided into three classes: autologous (sources from the patients), syngeneic (sources from identical twins) and allogeneic (sources from siblings or unrelated donors). The indications for particular type of transplantation are influenced by many factors, namely patient age, disease status, response to prior treatment, and more importantly, availability of donor sources.[3]

In autologous HSCT, stem cells harvested from patients' peripheral blood or bone marrow are first cryopreserved and later returned back to them, typically after high-dose myeloablative conditioning regimen. The main purpose is to restore their hematopoiesis. The patients seldom experience graft-versus-host disease (GVHD) since the reinfused stem cells are genetically identical to host cells. The absence of GVHD may result in lower treatment-related morbidity and mortality.[3] For those with malignant diseases, any residual tumor cells, however, may escape from immune surveillance and contribute to disease relapse. A trace amount of tumor cells may

contaminate the graft, which also leads to relapse.[4] Another disadvantage is the lack of therapeutic graft-versus-tumor effect (GVTE), which limits the clinical use of autologous HSCT.[3] Recently, a modified autologous stem cell transplant was developed. By infusion of genetically modified autologous T cells that express functional chimeric antigen receptors (CARs) *in vivo* into a patient with chronic lymphocytic leukemia (CLL), leukemia cells were rapidly eliminated from blood.[5, 6] Syngeneic HSCT, which is rarely performed due to limitation of graft sources from identical twins[7], is similar to the autologous one in which donor and recipient stem cells have the same genetic make-up. Like the recipients of autologous HSCT, patients with syngeneic transplantation seldom experience GVHD but likely develop relapse of malignant diseases as residual tumor cells may evade immune surveillance.[8, 9]

Allogeneic HSCT not only reproduces hematopoiesis in patients following myeloablative therapy, but also exerts therapeutic GVTE. It is preferentially used for treatment of certain malignant hematopoietic disorders, mainly due to its GVTE against residual tumor cells. Donor immunocompetent T cells may play a role in GVTE in the presence of alloantigens, ranging from host minor histocompatibility antigens (mHAs) to leukemia-specific epitopes.[10, 11] Since the graft comes from an individual with normal hematopoiesis, there is no risk of tumor contamination of it. On the other hand, GVHD is a common problem when performing allogeneic HSCT. Patients may experience higher morbidity of certain organs like skin, liver and gut.[12] In order to reduce the severity of GVHD, matching of human leukocyte antigen (HLA), which is also known as major histocompatibility complex (MHC), is essential to find compatible donor tissues. They are expressed on cell surface and mediate the immune response of

leukocytes. An ideal graft is derived from completely matched related donors. In most cases in which donors are unrelated, grafts with a single HLA mismatch are considered acceptably matched tissues.[3] Nevertheless, only around 30% of paediatric patients could receive the transplants from HLA-matched related donors. In spite of the establishment of worldwide volunteer bone marrow donor registries, finding an acceptably HLA-matched unrelated donor remains an obstacle to successful allogeneic transplantation, partially due to extensive polymorphism of HLA genes.[13] For many paediatric patients, umbilical cord blood from unrelated donors is the common source of hematopoietic stem cells. Having an advantage over peripheral blood or bone marrow, cord blood can be transplanted with less-stringent HLA matching as cord blood stem cells are more immunologically native and less likely to induce GVHD without losing GVTE. Cord blood transplantation is especially important when there is limited time for finding, matching and harvesting grafts from unrelated donors. The problem of low cell dose in cord blood can be overcome by second cord blood transplantation to increase the cell count and improve immune reconstitution. Similar to the bone marrow donor registries, cord blood banks have been established to find suitable cord blood transplants.[14]

1.1.2 Indications for HSCT in Children

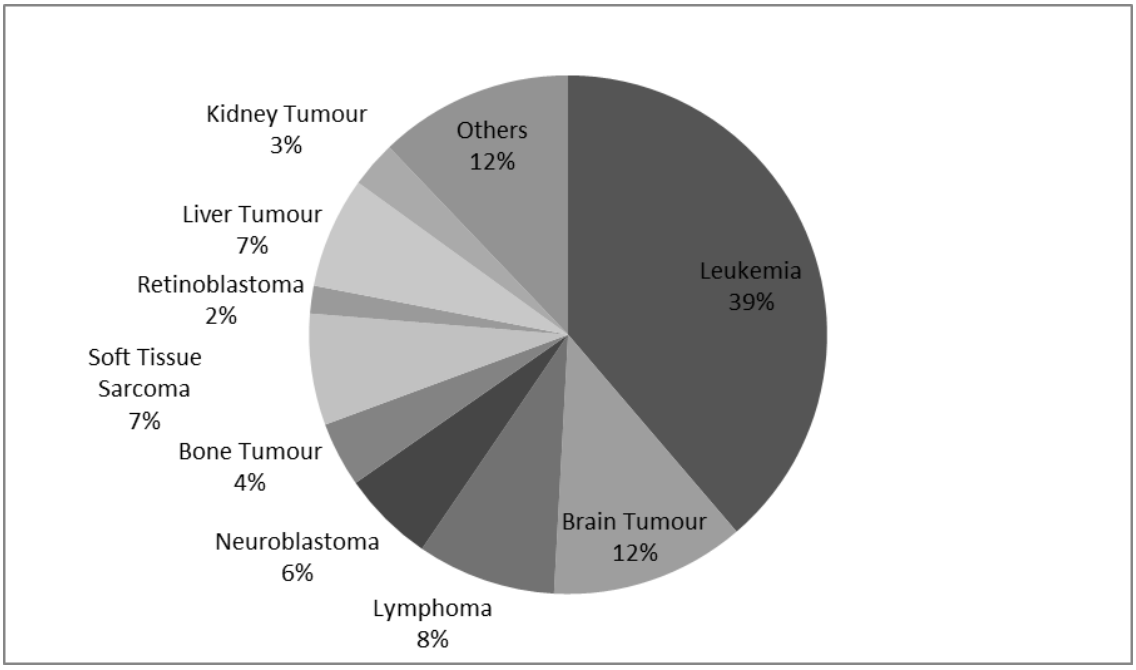
Table 1 reveals common indications for HSCT in paediatric patients. HSCT is primarily used for patients with hematopoietic malignancies, but also those with a variety of acquired and inherited non-malignant disorders. In order to improve the cure rate and reduce post-transplant complications of HSCT, the indications for HSCT in children have to be standardized. The standardization strategy is developed based on biological characteristics of their underlying diseases and their response to conventional therapeutic treatments, which may be greatly different from those in adults.[15]

Among different types of hematopoietic malignancies, leukemia is the main indication for allogeneic HSCT globally.[2] Acute lymphoblastic leukemia (ALL), the subtype of leukemia, accounts for more than half of paediatric leukemia cases.[16] Accounting for around 40% of typical malignancies diagnosed in Hong Kong children aged below 15, leukemia is the most common paediatric malignancy (**Table 2**). Again, ALL is the most common form of childhood leukemia.[17] Currently, most children with ALL are cured with combined chemotherapeutic approaches since the corresponding outcome has generally improved over the years.[18] In high-risk ones with ALL in first complete remission, allogeneic HSCT was shown to be a successful treatment strategy.[19] Besides, patients suffering from ALL and being predicted to respond poorly to current chemotherapy can be treated by allogeneic HSCT successfully.[18] On the other hand, autologous HSCT is typically applied to treat the first relapse of chemo-sensitive solid tumors like neuroblastoma and non-Hodgkin's lymphoma.[1]

Table 1. Common indications for allogeneic and autologous hematopoietic stem cell transplantation in children[(1)]

Allogeneic transplantation			**Autologous transplantation**	
Malignant diseases	Non-malignant diseases		Malignant diseases	Non-malignant diseases
	Acquired	Inherited		
Acute myeloid leukemia (AML)	Severe acquired idiopathic aplastic anemia	β-thalassemia major	Non-Hodgkin's lymphoma	Autoimmune disorders
Acute lymphoblastic leukemia (ALL)	Paroxysmal nocturnal hemoglobinuria	Sickle cell anemia	Hodgkin's disease	Amyloidosis
Chronic myeloid leukemia (CML)		Fanconi's anemia (FA)	Neuroblastoma	
Myelodysplastic syndromes (MDS)		Diamond-Blackfan anemia	Acute myeloid leukemia (AML)	
Acute myeloid leukemia (AML)		Wiskott-Aldrich syndrome		
Non-Hodgkin's lymphoma		Familial hemophagocytic lymphohistiocytosis		
Hodgkin's disease		Severe combined immunodeficiency (SCID)		
		Osteopetrosis		
		Dyskeratosis congenital (DKC)		
		Histiocytic disorders		
		Inborn errors of metabolism		

Table 2. Top 9 common cancers in paediatric patients in Hong Kong in 2011 (total number: 173) (Source: Children's Cancer Foundation, http://www.ccf.org.hk/)



Apart from treating hematopoietic malignancies, allogeneic HSCT is the potential curative therapy for some acquired disorders (e.g. severe acquired idiopathic aplastic anemia), hereditary metabolic disorders (e.g. sphingolipidoses), bone marrow failure syndromes (e.g. Fanconi's anemia), hemoglobinopathies (e.g. sickle cell anemia), congenital immunodeficiencies (e.g. severe combined immunodeficiency). A majority of children with these disorders are cured by allogeneic HSCT with satisfactory outcomes.[20] Although these disorders are rare, they would be otherwise lethal if the patients were not treated properly. The incidence of some non-malignant disorders is geographically dependent. For instance, it was reported that severe acquired idiopathic aplastic anemia occurs more frequently in Asia. In contrast, the disease is seldom found in the West.[21] On the other hand, autologous HSCT can cure patients with autoimmune disorders (e.g. systemic lupus erythematosus) by transfusing back auto-reactive lymphocytes-reduced peripheral blood.[15]

Despite advances in current HSCT that leads to improved outcomes, it often comes with risk of post-transplant complications. Therefore, the indications for HSCT have to be evaluated regularly after current therapeutic treatments, such as chemotherapy.

1.1.3 Complications of Allogeneic HSCT

Allogeneic HSCT cures many patients who suffer from a wide variety of lethal diseases. The post-transplant survival rate has improved a lot so far and is higher from transplantation in children than adults.[22] Nevertheless, there is great concern about some life-threatening adverse effects. The risk-benefit balance must be evaluated by transplant team prior to the transplantation.

Inability to establish complete donor-derived hematopoiesis in recipients often represents graft rejection or relapse of underlying malignancy.[23] Graft rejection refers to immunologic destruction of transplanted tissue by functional immune system of recipients, leading to graft failure. It is uncommon due to the pre-transplant myeloablative treatment and carful HLA matching process.[4] Disease relapse, which is defined as the return of disease after a remission, is the most common cause of treatment failure after allogeneic HSCT.[24] It may be attributable to insufficient myeloablative conditioning regimens or a deficient GVTE due to decreasing amounts and effectiveness of donor immunocompetent T cells.[25] Surviving the conditioning regimen, residual tumor cells may proliferate and re-emerge.[26] Even after development of donor hematopoietic cells, other negative outcomes may develop and impede the successful transplantation. Acute GVHD is a consequence of inflammation of recipient tissues recognized as non-self by donor immunocompetent T cells. It causes damages to many organs like liver and intestines[27] and is the main cause of transplant-related mortality and morbidity.[10] Although HLA-mismatched grafts are risk factor for acute GVHD, patients with HLA-matched sibling donors may also develop acute GVHD.[27] The incidence and severity of GVHD can be greatly reduced by administration of

immunosuppressive drugs (e.g. methotrexate) in the pre-transplant and post-transplant period.[1] However, the risk of opportunistic infections increases.[28] Alternatively, T cell depletion of the graft was shown to be an effective strategy of avoiding GVHD, but increased the incidence of disease relapse.[29] Apart from disease relapse and acute GVHD, post-transplant patients may experience other short-term complications, such as nausea and vomiting, mucositis, hepatic veno-occlusive disease, transplantation-related acute infection by bacteria, virus or fungi and pulmonary complications. For children undergoing transplantation, special management of central venous catheters (CVCs) is required as they are commonly associated with bloodstream infection and thrombosis.[22] In addition, these patients would face other long-term complications, which assume greater significance in them than in adults. Chronic GVHD can be a serious late complication in some apparently healthy post-transplant paediatric survivors. Older recipients/donors and grafts from unrelated donors increase the risk of chronic GVHD.[27] Growth and development are typically impaired following myeloablative preparation regimens.[30] Children with acute leukemia following cranial radiation therapy and transplantation are likely to have growth failure and growth hormone deficiency. Infertility is another frequently observed problem.[31] As the young long-term survivors of transplantation grow up, the incidence of secondary malignancies of skin, thyroid and bone increases. In order to avoid the post-transplant malignancies, they should get rid of any carcinogens (e.g. tobacco smoke) and regularly be examined by physicians to detect early cancer precursor lesions.[1] Nowadays, treatment-related mortality has decreased significantly. Post-transplant care and life support is important to improve their quality of life continuously following transplantation.

1.2 Chimerism Analysis after Allogeneic HSCT

1.2.1 Introduction to Chimerism

Allogeneic HSCT is an effective therapeutic regimen to reconstitute normal hematopoiesis in patients suffering from a number of hematopoietic diseases. In the follow-up of patients after allogeneic HSCT, repopulation of donor hematopoietic cells should start. Most transplanted patients have both donor and recipient hematopoietic cells in the short early post-transplant period. Chimerism is defined by the presence of cells from at least two different individuals in a body.[26] Patients undergoing autologous HSCT do not have this state since the graft is genetically the same as the host cells. Different states of chimerism may develop after engraftment. Complete chimerism (CC) refers to the condition in which all hematopoietic cells are derived from donors, whereas mixed chimerism (MC) denotes the persistent coexistence of hematopoietic cell populations originated from donors and recipients.[32] Recipient cells in MC may be normal or malignant. The presence of a trace amount of recipient hematopoietic malignant cells is considered as minimal residual disease (MRD), resulting in disease recurrence.[26]

It is important to realize that the post-transplant chimerism status is mostly not static. CC may be a transient stage at a particular time point following transplantation and the stage would ultimately develop into MC, or vice versa.[33] Also, development of MC is a dynamic process. Patients with increasing and decreasing level of recipient cells over time are referred to as having increasing/progressive and decreasing MC respectively. However, displaying stable MC, some do not have significant changes of

donor and recipient cells over time.[26, 32] MC has the potential to develop into autologous recovery, i.e., 100% chimerism (all donor cells are eliminated and only recipient cells are found in post-transplant samples). This indicates the failure in first transplantation and second transplantation is required.[34] There is a rare phenomenon called split chimerism, which occurs when one or more hematopoiesis cell lineages, for example, natural killer cells, are of recipient origin and others are of donor origin.[35]

The close post-transplant monitoring in short time intervals is required to recognize the dynamic changes of chimerism status. Peripheral blood is the most typical source for chimerism analysis. However, the level of MC is affected by the sensitivity of techniques implemented for chimerism analysis. If a method has a detection limit of 1% of recipient cells, patients with less than 1% of recipient cells will be regarded to have CC. When less than 1% of recipient cells are detected in peripheral blood with a sensitive technique, this state is defined as microchimerism.[26]

1.2.2 Importance of Hematopoietic Chimerism Analysis

Analysis of chimerism status, i.e., determination of the proportion of donor and recipient cells, after allogeneic HSCT generally assists in two aspects: (I) predict which transplanted patients have higher risk of developing negative outcomes and (II) decide on and monitor sequential strategies of therapeutic interventions to rescue the graft and achieve sustained remissions.[26, 36] Different types of chimerism may correlate with the prognosis of post-transplant complications, such as disease relapse. MC is commonly detected after allogeneic HSCT by various techniques, so the clinical significance and incidence of it for different hematopoietic diseases has been studied by many researchers.[37, 38]

Determination of chimerism status helps monitor the engraftment status. Development and maintenance of CC indicates that the engraftment process is successful and malignancy is thoroughly eliminated. Patients may suffer from less relapse for the rest of their lives, but may experience more severe and frequent GVHD.[39] Their reconstituted T cell receptor repertoire was reported to be more diverse, so their immune system may improve after transplantation.[40] In contrast, MC is usually thought to reveal that the residual malignancy is not thoroughly cleared. Patients may suffer from more frequent relapse with short-term disease-free survival, although their GVHD is less severe.[32, 39] However, the true clinical relevance of MC remains unclear as the degree of MC depends on several factors, namely the sensitivity and timing of molecular techniques employed to detect chimerism, the underlying disease for allogeneic HSCT, the stage of disease when allogeneic HSCT is used and the type of conditioning regimen.[26] Stable MC indicates that transplantation tolerance between

donors and recipients is induced.[41] Patients with stable MC have lower chances of developing GVHD but higher chances of suffering from relapse.[42] Some studies found that patients with higher levels of MC, i.e., higher proportion of recipient cells, or increasing MC have enhanced risk of relapse in acute leukemia[25, 37, 43] while others demonstrated such a correlation in patients with chronic myeloid leukemia (CML) only.[44] Nonetheless, an opposite result which suggests MC not associate with leukemia relapse was also found by several researchers.[45-47] In fact, some leukemia patients may relapse within a very short period of time after transplantation. If the chimerism analysis is not performed during this period, the association between MC and leukemia relapse will not be found. Therefore, the quantification of MC with shorter time intervals in the follow-up of patients may be more useful to predict disease relapse.[25, 37] Published data showed that regular monitoring of chimerism in lymphoid and myeloid cells from recipients after HSCT permits detection of residual leukemia cells and estimates the risk of relapse or graft rejection.[48, 49]

Once negative events, such as disease relapse, graft rejection and GVHD, following allogeneic HSCT are recognized, early administration of appropriate therapeutic interventions, such as cessation of immunosuppressive drugs and administration of donor lymphocyte infusion (DLI), is important to rescue the graft or eradicate the re-emerged malignant cells. DLI involves the infusion of donor's lymphocytes into recipients. By inducing GVTE, it is a common immunomodulatory therapeutic strategy to manage relapse of CML, but the response rate of patients with acute myeloid and lymphoid leukemia (AML and ALL) is low.[50] Moreover, overdose of DLI may induce lethal GVHD.[51] In order to improve the efficiency of DLI, applying

incremental doses of DLI before the onset of relapse may maximize the effect of GVTE and minimize the risk of GVHD[52], and hence early recognition of relapse by chimerism analysis is a prerequisite for success of DLI.[53] Furthermore, DLI process should be guided according to chimerism status of patients since extra unnecessary DLI may be toxic to patients.[54] With the invention of more advanced procedures in HSCT, analysis of chimerism is useful to monitor the effect of the procedures. For instance, for the sake of facilitating engraftment of donor cells, some patients may undergo allogeneic HSCT without myeloablative regimen for conditioning. This approach must be followed up by DLI. Prompt assessment of chimerism status permits timely administration of DLI with suitable dose.[55]

1.2.3 Techniques for Chimerism Analysis: Principles and Limitations

Chimerism status is analyzed by taking advantage of polymorphic markers between individuals. Various techniques have been developed and used in patients post allogeneic HSCT. Making use of different molecular techniques and markers, these methods of chimerism analysis offer various sensitivities and specificities with certain limitations. All approaches follow, however, the same principle in which samples from donors and pre-transplant recipients are screened initially to find informative markers, which can show differences between donors and recipients. Following the initial screening, analysis of these informative markers in post-transplant recipient samples is performed for MC quantification.

1.2.3.1 Erythrocyte Phenotyping

After allogeneic HSCT, the engrafted stem cells would take some time to produce erythrocytes. The phenotyping of erythrocytes of donor and recipient origin is performed by discriminating different antigens or markers on their surface. However, it is only restricted to analyze chimerism in transplanted patients suffering from CML. Erythrocytes are derived from myeloid lineages, which is negatively affected in CML. It is believed that relapse in CML correlates with a rising level of recipient erythrocytes and/or a reducing level of donor erythrocytes while lymphoid linkages remain of donor origin. It is a simple, accurate and sensitive technique.[26] Flow cytometric analysis using a set of labeled monoclonal antibodies against different types of blood group antigens like ABH and Rh antigens is widely used for detection of dual erythrocyte populations. It was shown that expression of donor erythrocyte antigens occurs within days 15 – 30 after transplantation and indicates successful engraftment.[56]

This protein-based technique for chimerism analysis has several limitations. Apart from the lineage-specific detection of chimerism, determination of chimerism states shortly after transplantation is not possible due to the extended lifespan of erythrocytes (around 120 days) In addition, other processes which involve infusion of erythrocytes such as blood transfusion can confuse the phenotyping results. Relying on informative protein polymorphism, erythrocyte phenotyping may not discriminate all donor-recipient pairs as disparities in protein are relatively smaller than genetic ones. In fact, polymorphic genetic markers are more commonly used for chimerism analysis.[136]

1.2.3.2 Sex Chromosomes Genotyping

In the case of sex-mismatched allogeneic HSCT, Y-chromosome is present either in donor or recipient hematopoietic cells. Such a macroscopic genetic difference can be visualized and is reliable for differentiating donor and recipient cells. Fluorescent *in situ* hybridization (FISH) using X- and/or Y-chromosome-specific probes on routine smears of bone marrow or peripheral blood is a widely used technique for chimerism analysis in these sex-mismatched cases. The fluorescent-labeled probes bind to their complementary sequences on corresponding sex chromosomes at interphase nuclei and give detectable signals. It is a simple and quantitative method, which is particular reliable when only a few cells for analysis are available.[57] Utilization of a Y-specific[58] or X-specific[59] probe in FISH assay was shown to be a useful method for MC examination in sex-mismatched transplantation. More advanced FISH assay called XY-FISH using duel-color X- and Y-specific probes seems to be more powerful for chimerism analysis, as it permits simultaneous detection of two sex chromosomes with an internal quality control parameter for successful hybridization (**Figure 1**).[60] A large number of cells can be screened within a short period of time. Analysis of over 1000 interphase cells within 4 hours was documented.[61] The sensitivity of XY-FISH was reported to be 1% when 500 cells were scored. However, it requires laborious workflow.[62]

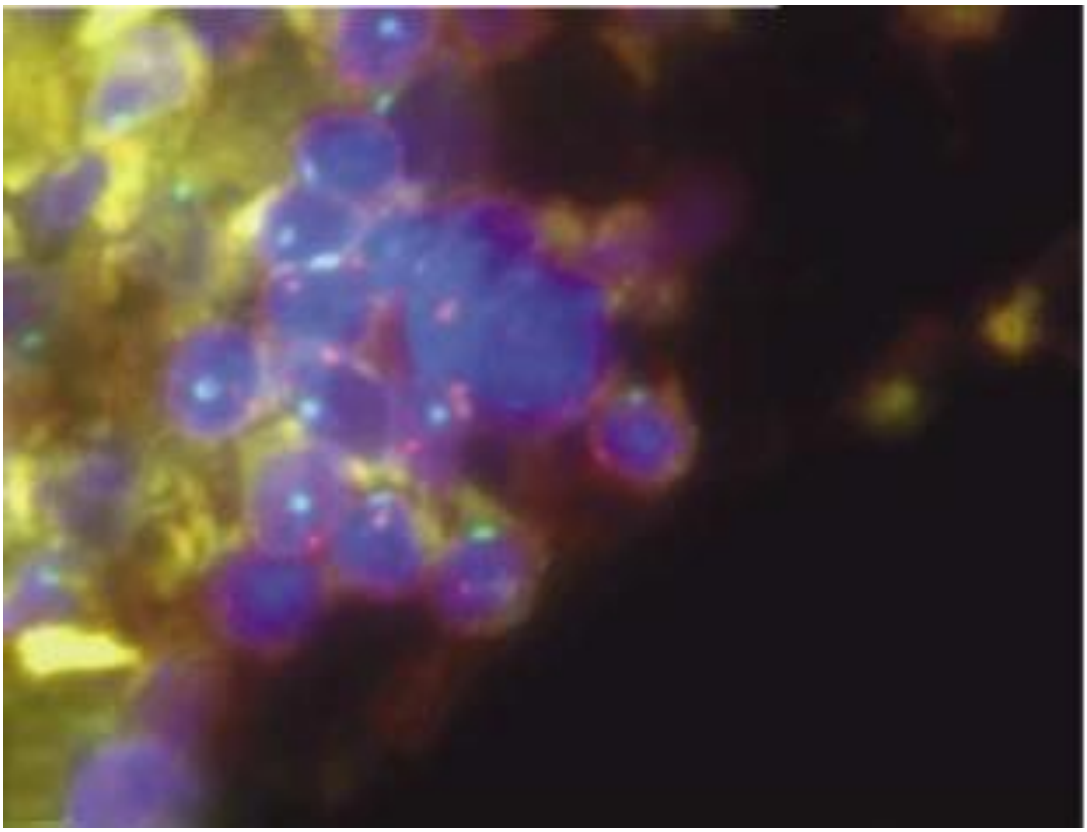


Figure 1. XY-FISH analysis of nucleated erythroid precursor cells in a female patient with CML after sex-mismatched allogeneic HSCT (red: X-chromosome; green: Y-chromosome)[63]

Another documented sex chromosome genotyping technique is real-time quantitative PCR, which was reported to be very sensitive. Lo *et al* established a real-time quantitative PCR assay for the analysis of Y-linked testis-determination gene to detect a trace amount of fetal male DNA in maternal DNA during pregnancy.[64] For the assessment of chimerism after allogeneic HSCT involving male recipients and female donors, Fehse *et al* reported a similar method by using chromosome Y-specific gene with high sensitivity of 0.001%. However, this technique only limits to male recipients transplanted from female donors, which represents only around 25% of all cases.[65]

The major drawback of sex chromosome genotyping for chimerism analysis is low applicability as it is suitable for sex-mismatched transplantation only. Sometimes, changes in sex chromosomes during post-transplant period can cause confusion to the results of chimerism analysis. Loss of Y-chromosome in the leukemia cells of the male elderly patients with AML leukemia cells was reported.[66] There is a case report concerning a male child with ALL whose leukemia cells have lost Y-chromosome and duplicated X-chromosome after allogeneic HSCT.[67] Although sex chromosome changes are rare in paediatric ALL[68], such changes can definitely complicate the chimerism analysis.

1.2.3.3 STR/VNTR Genotyping

In addition to the macroscopic genetic difference, microscopic genetic differences are also well-defined markers for chimerism analysis. Repeating sequences are widely dispersed in the human genome. Short tandem repeat (STR)/ variable number tandem repeats (VNTR) genotyping is routinely used to assess chimerism status after allogeneic HSCT. STRs, or microsatellites, and VNTRs, or minisatellites, are defined as DNA regions with 2-6 and 15-50 nucleotides long of a repetitive sequence respectively. The repeat numbers of a particular sequence varies between individuals, resulting in a number of alleles.[26] Since they exhibit the high degree of length polymorphism, they are suitable markers to discriminate donors and recipients based on their different alleles. PCR-based genotyping is frequently employed. Generally, STRs rather than VNTRs are often preferred for complete amplification by PCR, for many VNTRs are GC-rich and larger in size.[69] STRs were also reported to be more sensitive for genotyping than VNTRs.[70]

At present, most clinical laboratories utilize fluorescence-based STR/VNTR-PCR for chimerism analysis. Following PCR using fluorescent-labeled primers of informative STR/VNTR loci, the amplified products of various sizes are fractionated and detected in a DNA sequencer with aid of capillary electrophoresis (**Figure 2**).[71] The level of MC is calculated based on the peak ratio of donor and recipient signals. This method allows rapid quantification of MC. However, its sensitivity is only moderately high (1-5%)[34], which is mainly due to the preferential amplification of shorter alleles. The sensitivity will be lower if recipient reporting alleles are longer than donor ones.[72] Another drawback is the formation of stutter peaks, which is due to 'slippery' amplification of

repetitive sequences. The PCR products generated will be shorter than the expected.[73] These two drawbacks may influence the interpretation of results of STR-based chimerism analysis.

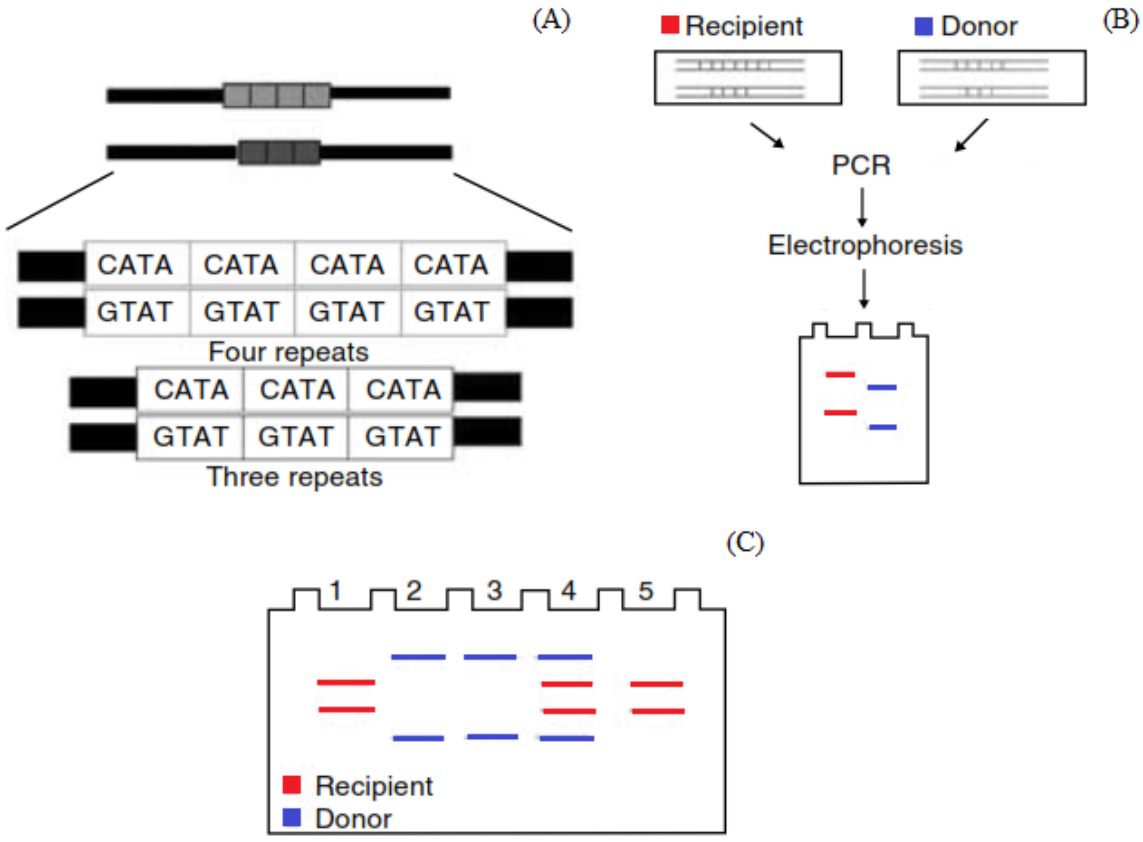


Figure 2. STR genotyping for chimerism analysis. **(A)** The number of tandem repeats of a STR locus on two chromosomes is different, resulting in different alleles. **(B)** Different alleles from donors and recipients are amplified by PCR, followed by electrophoresis. Informative markers are indicated by the presence of a difference in allele size. **(C)** Samples from pre-transplant recipient and donor are analyzed in lane 1 and 2 respectively. Different chimerism states can be indicated by STR analysis of post-transplant recipient samples: lane 3- CC, lane 4- MC, lane 5- graft rejection and/or disease relapse.[26]

1.3 Use of Single Nucleotide Polymorphisms (SNPs) for Chimerism Analysis

1.3.1 Background on SNPs

The human genome among individuals contains roughly 0.1% DNA variation, or polymorphism.[74] Single nucleotide polymorphisms (SNPs), which describe the substitution of a nucleotide for another throughout the coding and non-coding regions of genomes, are the most plentiful and simplest form of genetic variation in the human genome. On average SNPs occur every 1.9 kilobases in the human genome[75] and contribute to about 90% of human DNA polymorphism.[76] In humans, they are one of the biallelic genetic markers, though theoretically they can be multiallelic.[77] A/G polymorphism, for example, gives rise to three possible genotypes on a human autosomal chromosome: homozygous A, homozygous G and heterozygous A/G. The least frequent allele of SNPs is defined to have an abundance of at least 1% in a given population.[74]

The abundance of SNPs in the human genome makes them valuable for human genetics research. More than three million SNPs have been genotyped from the human genome so far. Each SNP can exist as either a homozygote or heterozygote and therefore more than nine million possible SNP patterns.[78] Many mapped SNPs have undefined functions, but some are thought to associate with genetic changes which predispose people to disease or affect their responses to environmental factors.[77] Apart from identification of potential disease genes, SNPs also provide a powerful tool for human identification. Genetic markers with high informativity are useful for identification of individuals. Although biallelic SNPs are less informative than typical multiallelic

genetic markers which display size polymorphism, like STRs, the great density of SNPs in the human genome outweighs the disadvantage of low informativity of a single SNP marker. Use of a large panel of SNPs with high minor allele frequencies would enhance the informativity to distinguish any two individuals (except identical twins)[79] and improve the statistical validity of MC quantification.[80] The informative SNPs should be selected according to the population being studied as allele frequencies of a SNP may vary in different populations.[75] However, relative to the gene function identification, the clinical application of discrimination of individuals by SNPs is seldom investigated.

SNPs have recently attracted the interest of many physicians who perform allogeneic HSCT. Genotyping informative SNP loci of donors and recipients should be reliable for the evaluation of post-transplant chimerism. Various SNP genotyping methods have been developed for the analysis of chimerism after allogeneic HSCT.[80-82] These methods will be reviewed in the following section. They are different in terms of assay chemistry, signal detection, cost and assay flexibility.

1.3.2 Techniques for SNP-based Chimerism Analysis and Limitations

Theoretically, since the human genome is estimated to contain more than three million SNPs, there are a practically unlimited number of SNP markers for distinguishing two individuals. However, single SNP has low informativity as each SNP locus typically has two alleles only. Therefore, it is necessary to examine a panel of SNPs to search for one or more informative markers, which enable clear differentiation of donors from recipients.[32]

1.3.2.1 Real-time Quantitative PCR using TaqMan technology

TaqMan-based real-time quantitative PCR was reported to be a very sensitive approach for SNP genotyping[83] or quantification of MC.[81] This method measures the quantity of PCR products at the onset of the exponential phase of PCR process, rather than the plateau phase in conventional PCR. This is achieved by using two sequence-specific dual-labeled fluorogenic probes, which are complementary to the two SNP allele loci respectively and hybridize between a pair of typical PCR forward and reverse primers. Each probe is labeled with either a 6-carboxyfluorescein (FAM) or 2'-chloro-7'-phenyl-1,4-dichloro-6-carboxyfluorescein (VIC) fluorescent reporter dye, the emission spectrum of which is quenched by another attached fluorescent quencher, 6-carboxy-tetramethylrhodamine (TAMRA). During the extension phase of the PCR cycle, as a consequence of the 5' to 3' exonuclease activity of Taq polymerase, the fluorogeneic probe is cleaved to release the reporter from the quencher, resulting in a rise in fluorescent signal as the reporter fluoresces. A rise in either FAM or VIC signal represents homozygous SNPs for FAM- or VIC-specific alleles while a rise in both signals indicates heterozygous SNPs. The intensity of the signal is proportional to the

quantity of PCR products as well as the starting target DNA (SNP loci) copy number. Higher the initial copy number of amplifiable DNA, lower is the number of PCR cycles at which a significant increase in the fluorescent signal is detected. This PCR cycle number is termed cycle threshold. The proportion of two allelic variants is reflected by the relative signals generated by the two probes (**Figure 3**).[64, 81]

Utilization of real-time PCR technique aiming at chimerism analysis has been evaluated by several researchers. Alizadeh *et al* enabled to discriminate more than 90% of recipient/donor pairs by using a panel of 11 SNPs. The reproducible sensitivity of 0.1% was reported.[81] Another publication demonstrated that a set of 7 SNPs was able to quantify 0.1-0.01% of recipient cells in donor cell populations, but only 67% of sibling pairs could be discriminated.[36] It was shown that real-time PCR can detect very low percentages of recipient cells in donor cells, and thereby provide early detection of MC. However, although quantification of low levels of MC is accurate, the accuracy in quantification of high levels of MC is relatively lower due to high standard errors (30%-50%).[36, 81] In addition, in order to find the relative amounts of two different alleles of a SNP in mixed DNA samples, a standard curve, which is prepared in separated reactions, is required for both alleles. This is laborious when multiple SNPs are analyzed. Real-time PCR only provides semi-quantitative information of chimerism.[36]

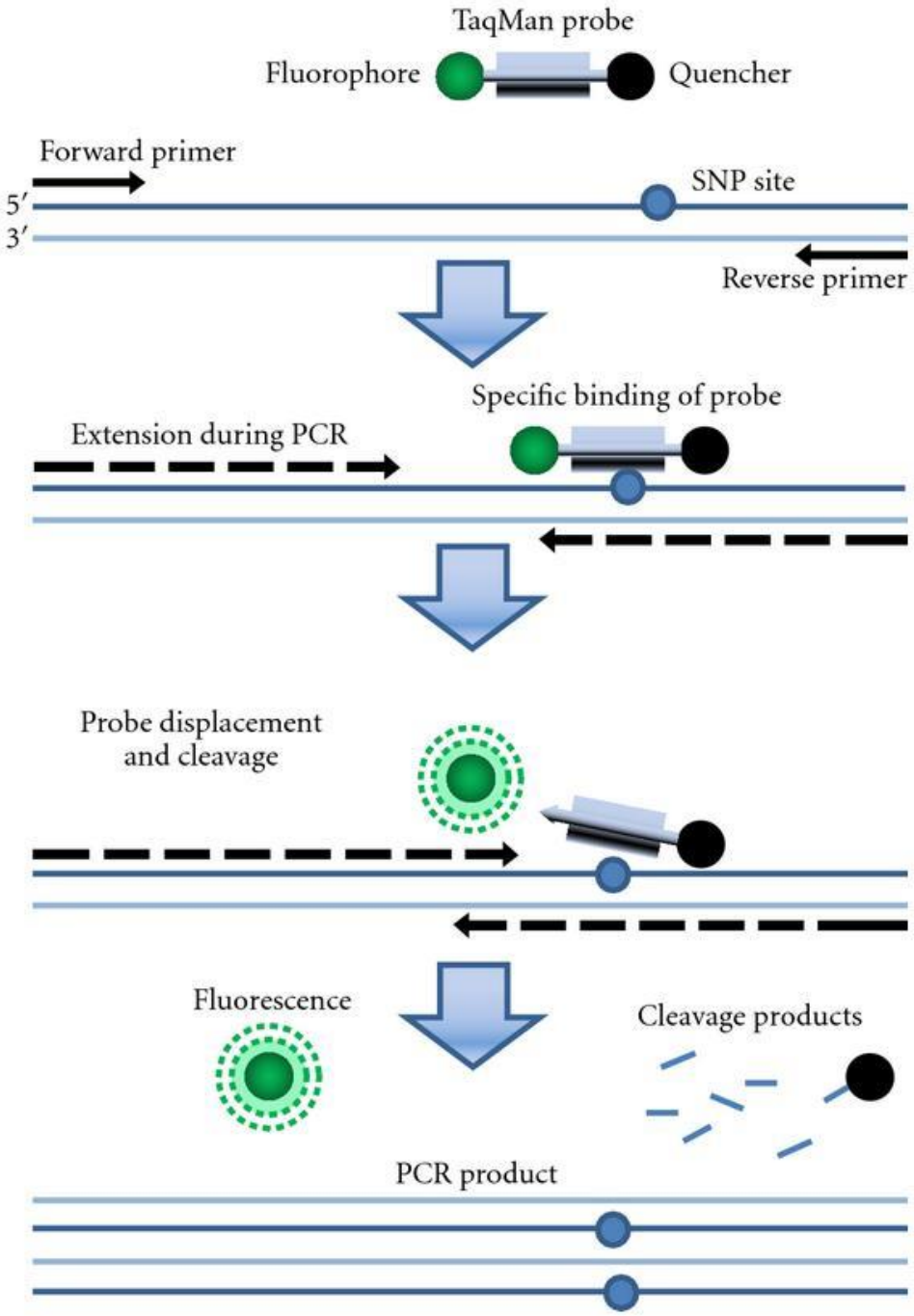


Figure 3. Simplified workflow of TaqMan SNP genotyping[72]

1.3.2.2 Pyrosequencing

Apart from using real-time PCR technique, DNA sequencing technique in a real time fashion is an alternative for SNP genotyping. Differing from real-time PCR, pyrosequencing requires not only DNA polymerase but also three other enzymes ─ ATP sulfurylase, luciferase and apyrase. After PCR amplification of regions around targeted SNP sites, incorporation of nucleotides into a sequencing primer, which is hybridized with target DNA (SNP loci), occurs and pyrophosphates are released during the extension reaction. In the presence of a substrate adenosine 5'phosphosulfate, the pyrophosphates are converted to ATP by ATP sulfurylase. ATP subsequently mediates a conversion of another substrate luciferin to oxyluciferin catalyzed by luciferase. At the same time, measureable light signals are emitted. More identical bases incorporated to the growing primer simultaneously, more is the amount of pyrophosphates released and higher is the signal. The sequencing reaction continues by stepwise addition of four deoxynucleotides (dNTPs) and simultaneous degradation of any incorporated nucleotides by apyrase (**Figure 4**).[84] The relative signals which correspond to the two SNP alleles directly indicate the proportion of them in chimeric DNA samples. Unlike real-time PCR technique, this quantitative method gives the neighboring sequences of a SNP, which helps confirm the correct genotyping of a SNP.[80]

Publications related to SNP-based pyrosequencing for assessment of chimerism after allogeneic HSCT are not numerous. Several automated pyrosequencing platforms have been developed to offer rapid and accurate quantification of DNA. This technique was previously assessed for detection of kirsten rat sarcoma viral oncogene homolog

(KRAS) mutations, which belongs to point mutations and are similar to SNPs. A mixture containing 5% of mutant alleles could be identified.[85] With regard to chimerism analysis, by using a panel of 14 SNPs, Hochberg *et al* reported that pyrosequencing technique enables detection of at least 5% of recipient cells after allogeneic HSCT.[80] The sensitivity of pyrosequencing is worse than that of real-time PCR, which may be due to the use of the enzyme cascade system for the indirect generation and detection of the chemiluminescent signal in the former while the fluorescent signal is produced and measured directly in the latter.[82] The sensitivity of pyrosequencing method has to be improved before it is applied in routine clinical use as early detection of increasing recipient cells from a very low level is predictive for impending relapse.[25, 37]

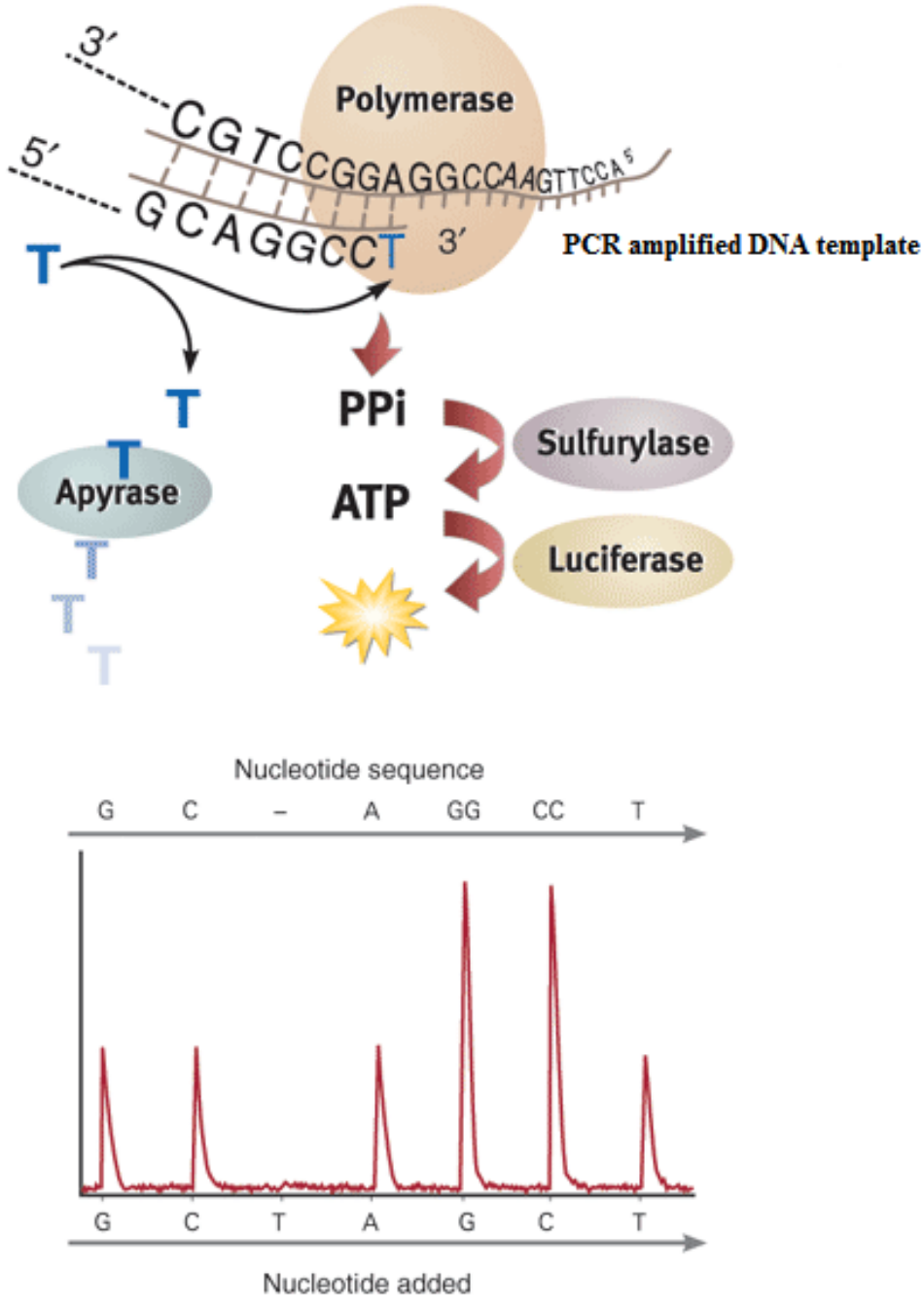


Figure 4. Enzyme cascade system in pyrosequencing for signal detection. Peak heights indicate the number of identical bases incorporated in the growing strand.

1.3.2.3 Minisequencing

Pyrosequencing indirectly measures the fluorescent signal from the incorporated nucleotides. In contrast, minisequencing directly measures the signal from incorporated modified nucleotides, fluorescently labeled dideoxynucleotides (ddNTPs). This technique requires a detection primer, which is designed to anneal immediately upstream of a SNP site. This primer carries a 5' tag sequence, complementary to capturing oligonucleotides immobilized on a microarray. Initially, PCR reaction is required to amplify SNP regions of interest. After annealing of the detection primer followed by a single-base primer extension reaction using four specific labeled ddNTPs, the tagged and extended detection primer is hybridized to the capturing oligonucleotides. The fluorescent signals corresponding to specific bases are measured from the microarray and the genotypes of the SNP are read accordingly. Relative signals corresponding to the two SNP alleles indicate their relative amounts in DNA samples.[86, 87] Compared with pyrosequencing, minisequencing offers a higher sensitivity due to the direct measurement of the fluorescent signal.[82] However, this approach only provides semi-quantitative measurement[88] of each allele of targeted SNPs. Another limitation is its inability to provide confirmatory evidence of targeted SNPs being genotyped, as the detection method is fluorescent signal-based but not DNA sequence-based.

Microarray-based minisequencing permits large-scale and highly multiplexed SNP genotyping. This is an advantage for high-throughput chimerism analysis of post-transplant patient samples. Fredriksson *et al* applied this technique in two phases. After initial screening of pre-transplant DNA samples obtained from recipients and donors to identify informative SNPs, semi-quantitative minisequencing method in a microarray

setup using identified informative SNPs was used for analysis of chimerism developed after allogeneic HSCT. Detection threshold for MC was reported to be 1%.[82]

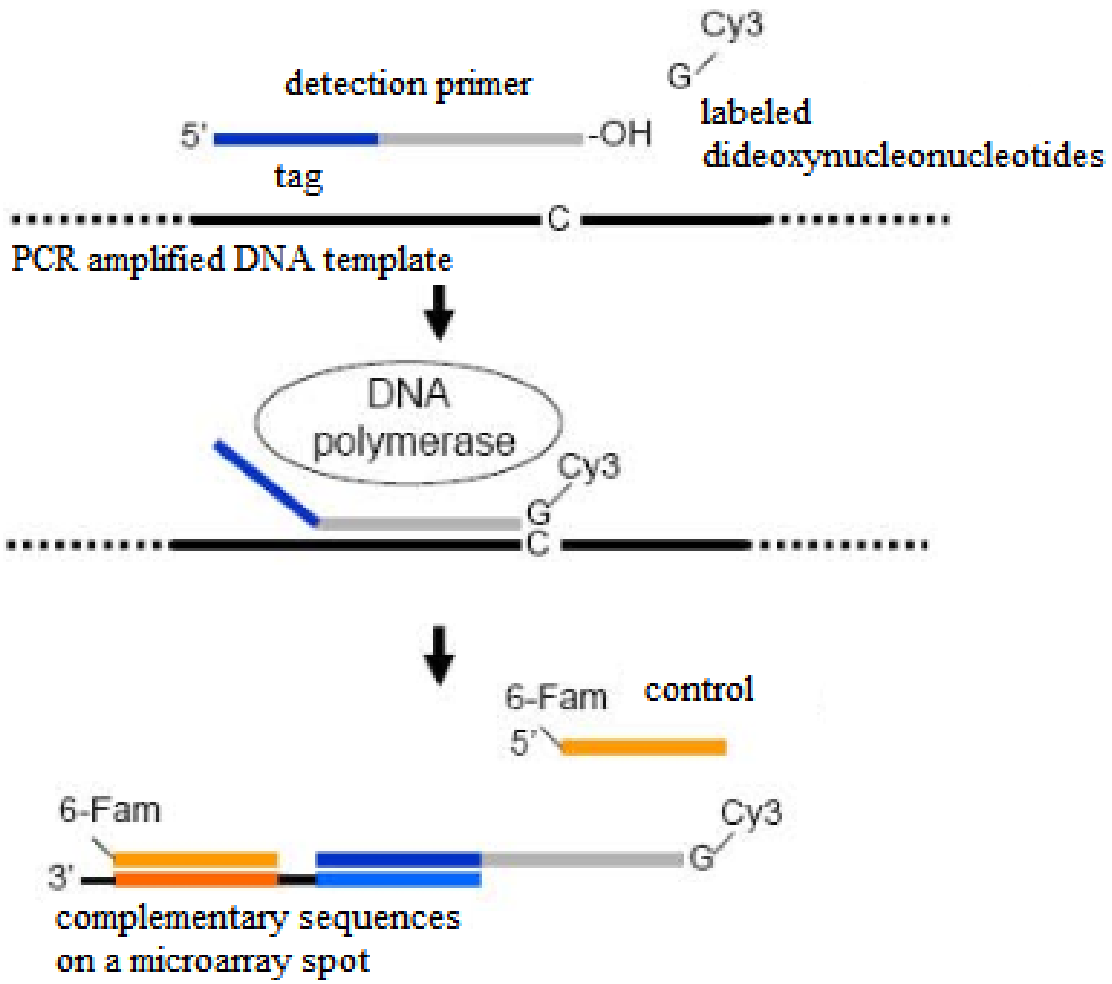


Figure 5. Principle of minisequencing[87]

1.3.3 MALDI-TOF MS for DNA SNP analysis

1.3.3.1 Principles and Challenges

Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) has evolved as a versatile tool for analysis of biomolecules. While it has been being one of the two major standard tools for proteomic study, it has been explored extensively for the analysis of DNA[89, 90] and RNA[91, 92] molecules. This ability has made it an attractive and reliable platform for genotyping genetic variations, especially SNPs.[93, 94] Different nucleotides have different specific masses, which can be well differentiated by using MALDI-TOF MS. Sequence of an oligonucleotide can be deduced from its mass. The general principal of this genotyping platform is to resolve the mass difference between allele-specific products by ionizing these oligonucleotides followed by detecting their masses. The allele-specific products are first co-crystalized with a small organic acid (matrix) on a stainless steel target plate. The matrix molecules can absorb energy from a pulsed laser source, resulting in decomposition of the irradiated allele-specific products-matrix co-crystals and ionization of the allele-specific products. The charged gaseous products are accelerated and separated in an electric field and eventually reach a detector. Their mass-to-charge ratios (m/z) are calculated based on their TOF, i.e., larger the analytes, longer is their flight times. Finally, the ionized allele-specific products will appear as separate peaks in a mass spectrum, in which the X-axis and Y-axis correspond to the m/z value and signal intensity, respectively (**Figure 6**).[96] Because charged products are predominantly singly charged, "m/z value – 1" will be equivalent to the mass value of an observed product. Furthermore, a product peak

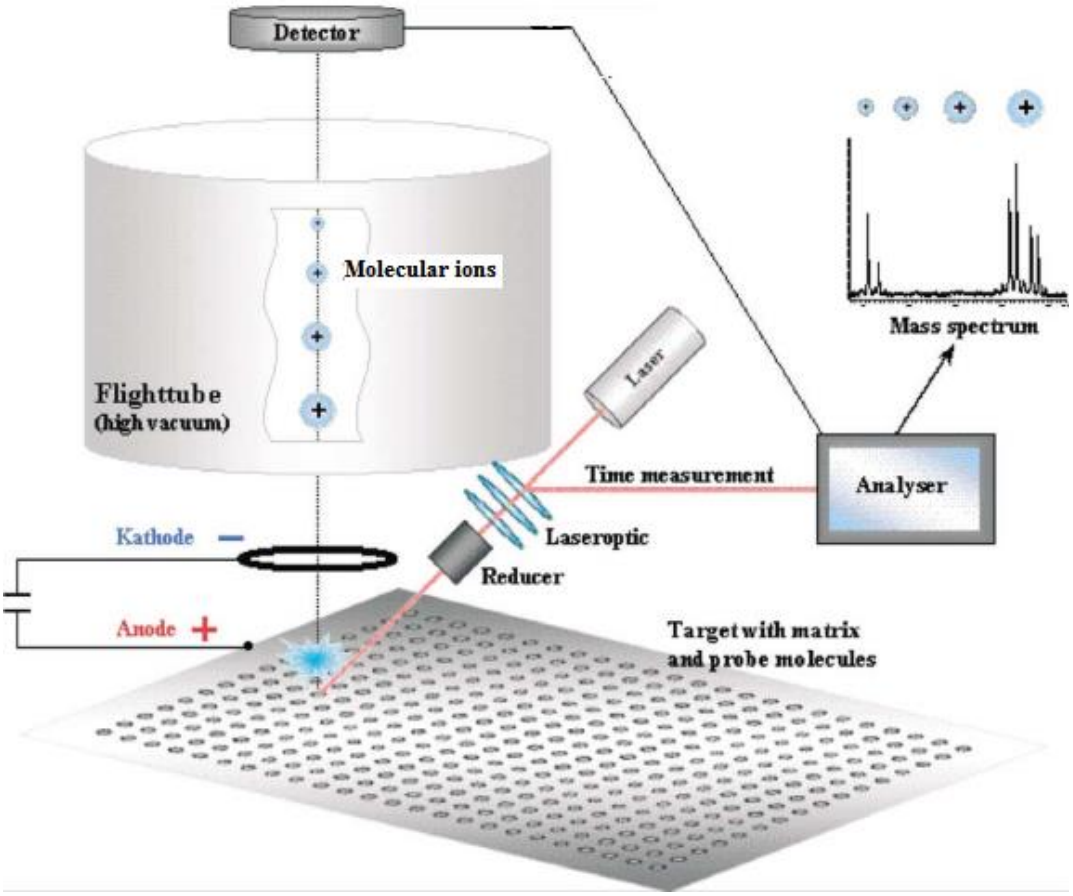


Figure 6. Schematic representation of MALDI-TOF MS[95]

area will be directly proportional to quantity of the product. There are various approaches for generation of allele-specific products for genotyping by MALDI-TOF MS, namely hybridization with allele-specific probes[97], ligation of allele-specific probes and target sequences[98], allele-specific cleavage of probes[99], and allele-specific primer extension reaction.[100] The most widely used platform is the last one, which is based on a series of molecular biology procedures comprising (I) region-specific PCR, (II) removal of remaining nucleotides with phosphatase and (III) allele-specific primer extension reaction. Because different nucleotides have different specific masses, the sequence of an extended nucleotide can be easily obtained by tracing the mass increment of an extension primer after allele-specific primer extension.[101]

In general quantitative analysis of oligonucleotides by MALDI-TOF MS is more difficult compared to those of peptides and proteins. In order to achieve multiplex SNP genotyping by MS with high-quality signal intensity, it is necessary to overcome several challenges that would lower the overall sensitivity of the technique, such as that large biomolecules DNA may lose parts of their structures due to the heating process (ionization by laser) in the matrix. Depurination is the most common fragmentation process, which denotes the loss of purine bases from nucleotides. It leads to formation of unwanted peaks having masses 134Da (adenine) or 150Da (guanine) less than the parent peak's ones, and thereby the sensitivity is lowered.[95, 101] However, it was reported that one of pyridine bases thymine is more resistant to decomposition.[102] The sensitivity and resolution decrease further with increasing mass of oligonucleotides because of the inverse relationship between analyte mass and signal intensity. Therefore, for the sake of obtaining high-quality spectra, primers and their corresponding extension products have

a restricted mass range of 1000-7000Da (around 3-25 nucleotides) for easy and successful analysis.[101] This is especially important for resolving extension primers with low mass differences.[103] Another challenge is related to the negatively charged phosphodiester backbone of DNA. Unlike protein and peptide analysis, DNA analysis by MS is much more sensitive to the presence of salts, which are normally found in buffered solutions for the preparation steps (e.g. PCR) of the extension products. The phosphate backbone is susceptible to bind metallic cations such as Na^+ (23Da) and K^+ (39Da). These ions interfere with the ionization process by inducing multiple adduct peaks, resulting in broadening of peaks and reduction in sensitivity and resolution.[94] Furthermore, the presence of adducts may interfere in the accurate discrimination for some heterozygous SNPs. The mass separations between A/C, G/T and C/G heterozygotes are 24Da, 25Da and 40Da respectively. The typical sodium and potassium adducts and the true heterozygotes may be indistinguishable when several extension products are examined.[95] Aside from the salt contaminants, some chemicals like detergents, which may be present in some molecular biology reaction buffers, are not compatible with the matrix crystallization process and therefore weaken the MS analysis.[94] As a result, stringent clean-up procedures have to be implemented prior to the MS analysis of the products. Some popular procedures include cation-exchange chromatography, reverse-phase column chromatography and ethanol precipitation. These techniques can circumvent the problem of MS signal interference by contaminants at the expense of automation, time and cost.[101] For successful analysis of oligonucleotides, suitable matrices are required. Rather than α-cyano-4-hydroxycinnamic acid (CHCA), which is a typical matrix for peptide analysis, 3-

hydroxypicolinic acid (3-HPA) is preferentially used for ionization of oligonucleotides. Sometimes the use of matrix additives like sugars enables a decrease in in-source fragmentation and an increase in signal resolution.[104]

1.3.3.2 Advantages over conventional methods for DNA analysis

A main advantage of MS is its direct method for analyte detection, which depends on an intrinsic physical property (i.e., m/z) of molecules. Conventional methods for DNA analysis often involve reporting probes for indirect detection of target oligonucleotides such as fluorescent reporting tags used in TaqMan-based real-time quantitative PCR.[81] Nevertheless, the number of commercially available reporting probes is limited, so a genotyping platform with a high level of multiplexing is rarely supported. On the other hand, MS is able to acquire multiple mass spectra from multiple reporting oligonucleotides for each sample simultaneously, as long as they have different sequences and m/z. In case of SNP genotyping, the sequence-based detection method provides highly accurate data for genotype calling. The detection of peaks with specific m/z in mass spectra provides confirmatory evidence of targeted SNPs being genotyped. Because of the ability to acquire many data points at the same time by MS, a high-level multiplex SNP genotyping platform can be designed.[101, 105] In case of chimerism analysis, a peak area ratio of two alleles of an informative SNP is an indicator for percentages of MC in chimeric DNA samples. Since nowadays mass spectrometers are able to detect analytes in a femtomole or even attomole range[106], a very low quantity of minor alleles and percentages of MC can be theoretically detected.

1.3.3.3 Application of MALDI-TOF MS-based SNP genotyping to chimerism analysis

Combined with the primer extension strategy, MALDI-TOF MS has emerged as a powerful technology for determination of allele frequencies of SNPs. In the field of clinical applications, this combination is able to discover genetic markers that are phenotype-related and detect mutant strains of pathogens.[95] Nevertheless, SNP genotyping by MALDI-TOF MS is rarely applied for chimerism analysis. The existence of small amounts fetal DNA in the background of maternal DNA in maternal plasma, which is similar to the situation of chimerism, has facilitated the development of a noninvasive detection method for fetal-specific alleles using maternal plasma rather than amniotic fluid.[107] Ding *et al* has developed an approach based on single-allele base extension reaction followed by MS analysis of fetal allele-specific extended primers.[108] The utilization of MS for the detection of fetal β-thalassemia point mutations in maternal plasma proposed that MS is a reliable tool for detection of single-nucleotide differences (e.g. SNPs) between DNA species from two individuals after transplantation. This concept has been applied in the detection of donor-specific SNPs in the urine of recipients after renal transplantation.[109] This proof-of-concept study, however, only provided preliminary result without quantification. In addition, the scale of this study was too small for evaluation. It is worth noting that MALDI-TOF MS-based genotyping has not been applied to quantitative chimerism analysis in patients after receiving allogeneic HSCT. Hence, its feasibility and detection limit for MC remain unknown.

CHAPTER 2: PROJECT RATIONALE AND OBJECTIVES

As mentioned before, MALDI-TOF MS has not been applied to quantitative chimerism analysis. Its practical values and detection limit for MC are undefined. On the basis of the advantages of MALDI-TOF MS in quantitative SNP genotyping and quantitative DNA analysis, the primary objective of this M.Phil. research project was to develop a MALDI-TOF MS-based SNP genotyping method for quantitative chimerism analysis.

SNP-based analysis of chimerism is a promising approach for analysis of outcomes after allogeneic HSCT. Selection of a panel of informative SNPs with high allele frequency that can distinguish donor-recipient pairs is a prerequisite for the analysis. In view of the fact that most of donors and recipients are Han Chinese in Hong Kong**,** the first secondary objective was to identify a panel of informative SNPs in Han Chinese from a public SNP database by informatics analysis.

After selection, the SNP markers have to be analyzed by certain molecular techniques. The number of available techniques is limited, however. MALDI-TOF MS has emerged as a very useful technique for SNP genotyping. Different strategies for generation of genotyping products for subsequent MALDI-TOF MS analysis have been developed. These products are important in allele discrimination and quantification. Combined use of competitive PCR and primer extension is the most popular strategy. Commercially, based on this strategy, several SNP genotyping platforms have been developed and optimized to achieve a high degree of automation and a high multiplex level. The Sequenom^{TM} (Chicago, USA) MassEXTEND^{TM} Assay and iPLEX^{TM} Assay are the most widely used MALDI-TOF MS platforms. Despite the commercial

availability, applications of these methods are restricted to a tailored-made MALDI-TOF MS platform, of which the detection limit was at femtomole level. Because a more advanced MALDI-TOF/TOF MS platform, of which the detection limit was at attomole level, was available in our university, the second objective of this project was to develop an in-house primer extension-based multiplex SNP genotyping assay for quantitative chimerism assessment by using our more sensitive MALDI-TOF/TOF MS platform. A number of molecular biological procedures were optimized for achieving highest detection limit.

After the in-house assay was developed, its linearity, accuracy, precision and limit of detection were evaluated. Currently, STR-PCR is the mainstay of chimerism analysis after allogeneic HSCT. It has been being used to measure the percentage of MC of the post-transplant paediatric patients in the bone marrow transplant clinics in the Prince of Wales Hospital. The last secondary objective was to evaluate practical values of the developed assay for quantitative chimerism assessment and compare it with the current STR-PCR method in 30 post-transplant paediatric patients in the Prince of Wales Hospital after allogeneic HSCT in terms of detection limit and accuracy.

CHAPTER 3: SECTION 1 – SELECTION OF REPORTING SNPs

3.1 Introduction

Prior to identification among individuals by using a panel of SNPs, reporting SNPs was selected according to several criteria in order to identify at least one informative SNP in donor-recipient pairs. First, all SNPs should be selected from different chromosomes so that linkage disequilibrium (LD) cannot occur among them. Allele frequencies of each SNP rather than haplotype frequencies can be used in calculation of informativity of the selected SNP panel. Second, population being studied should be well defined. Allele frequencies of SNPs across ethnic populations vary significantly. Such variations in allele frequency may be related to differences in the prevalence of genetically inherited diseases among ethnic groups.[110] Serving as genetic markers, SNPs are chosen based their reported allele frequencies in a particular population from a public database. Third, all SNPs should display high level of heterozygosity in the study population. The more similar the major and minor frequencies, the more likely the SNP can be heterozygous in the population. These SNPs have higher discrimination rates. Global efforts have been made to find SNPs in the human genome and create public SNP databases[111], the SNP contents of which are highly redundant but complementary also.[112] The public SNP databases are valuable resources for searching reporting SNPs that meet the mentioned criteria.

To control the assay cost, minimum number of SNPs in a panel is determined to find at least a single informative SNP between donors and recipients. Due to biallelic nature of SNPs, there are usually only three possible genotypes (e.g. AA, GG, AG/GA

for A/G polymorphism). Since biallelic SNPs have fewer variations than typical multiallelic genetic markers, the discrimination rate of a SNP marker generally is lower than that of a multiallelic marker like STRs. Therefore, a sufficient and reasonable number of SNPs are needed for distinguishing two individuals.[26] Hochberg *et al* calculated that theoretically seven SNPs of 50% allele frequency would be sufficient to find at least one informative SNP in around 99% of HLA-identical siblings or unrelated pairs. In case of a panel of SNPs of allele frequency below 50%, the number of SNPs in the panel has to be raised to maintain the high possibility to identify an informative SNP. Experimentally, they showed that the uses of six and eleven SNPs were able to identify at least one informative SNP in 99.6% of unrelated and related pairs respectively by pyrosequencing.[80] Another group showed that the use of 10 SNPs was able to discriminate 92.6% of donor-recipient pairs (125 out of 135 pairs) by TaqMan real-time PCR.[32] Expanding the SNP panel increases the probability of distinguishing each donor-recipient pair and, more importantly, the chance of identifying more than one informative SNP in all pairs. This also increases the statistical validity of MC quantification.[80] However, the assay cost will be increased. In addition, the size of a SNP panel also depends on a definition of an informative SNP, which varies according to the nature of clinical applications. The possible informative combinations of donor and recipient genotypes at a SNP locus have a great impact on calculating the probability of identifying at least one informative SNP in all donor-recipient pairs.

In the first part of our study, we aimed at searching a number of SNPs from a public SNP database, HapMap. Their reported allele frequencies were obtained from different public databases for comparison. The theoretical informativity of the selected

panel of SNPs was determined by calculating the probability of finding at least one informative SNP allele between two unrelated individuals whom were randomly selected form the population. After that, the informativity of the selected SNPs was evaluated in recruited donor-recipient pairs by SNP genotyping their pre-transplant DNA samples.

3.2 Materials and Methods

3.2.1 Patient and Donor Materials

All human specimens used in this study were clinical samples left behind after routine clinical laboratory analysis. This study was part of a clinical program to improve the quantitative chimerism analysis of patients after receiving allogeneic HSCT. During 1998-2012 in the Prince of Wales Hospital, 30 paediatric patients who received allogeneic HSCT were recruited. In order to evaluate if there is a difference in the number of informative SNPs between cases with related and unrelated donors, the recruited patients either received allogeneic HSCT from related donors (n=15) or unrelated individuals (n=15). Among them 13 patients received double cord blood transplantation (i.e., receiving grafts from two donors in a single HSCT). All donors for double cord blood transplantation were unrelated to their corresponding recipients. Pre-transplant blood or bone marrow samples were collected from the patients and donors and post-transplant samples were collected from the transplanted patients at regular intervals (1-2 months). These patients suffered from a variety of hematopoietic diseases, including acute lymphoblastic leukemia (ALL, n=3), myelodysplastic syndromes (MDS, n=2), acute myeloid leukemia (AML, n=5), T-leukemia (n=1), T-lymphoma (n=2), anaplastic large-cell lymphoma (ALCL, n=1), myeloproliferative disease (MPD, n=1), NK cell lymphoma (n=1), severe aplastic anemia (SAA, n=4), Fanconi's anemia (FA, n=2), thalassemia (n=6), severe combined immunodeficiency (SCID, n=1) and dyskeratosis congenita (DKC, n=1). Characteristics of these patients are summarized in **Table 3**.

Table 3. Demographic data of the 30 paediatric patients

	Cases with unrelated donors	Cases with related donors
Number of patients	15	15
Gender of patients		
Male	11	11
Female	4	4
Mean age of patients (range)	7.6 (1-18)	9.6 (1-18)
Diagnosis-Number		
ALL	2	1
MDS	1	1
AML	3	2
T-leukemia	1	0
T-lymphoma	0	2
ALCL	0	1
MPD	0	1
NK cell lymphoma	1	0
SAA	0	4
FA	2	0
Thalassemia	4	2
SCID	0	1
DKC	1	0
Number of cases with double cord blood transplantation	13	0

3.2.2 Selection of Reporting SNPs from HapMap Database

Over 90% of the HSCT donors and paediatric recipients are Chinese in Hong Kong. Using the HapMap database (http://www.hapmap.org/), reporting SNPs were selected on the basis of three criteria. Firstly, they should be located in different chromosomes. Secondly, their major and minor allele frequencies should be near 0.5 in Han Chinese population in Beijing (CHB). Thirdly, the neighboring sequences of the SNPs were analyzed by a public database SNP and CNV Annotation Database (www.scandb.org/). This database search ensured that none of the SNPs are co-localized in one or more regions with copy-number-variations (CNVs). To examine the robustness of the selected SNPs, their major and minor allele frequencies in CHB and in Han Chinese population in Shanghai (CHS) as well as their global major and minor allele frequencies were also obtained from the 1000 Genomes Project database (http://browser.1000genomes.org/) and the NCBI database (http://www.ncbi.nlm.nih.gov/SNP). The heterozygosities of each selected SNP were calculated by multiplying the product of its reported major and minor allele frequency by two.

3.2.3 Statistical Evaluation of the SNP Panel

Hardy-Weinberg Equilibrium (HWE) (expected genotype frequencies are calculated under the assumptions of independent assortment and no population admixture)[101] was assessed for each SNP. According to Hardy-Weinberg relationship ($p^2 + 2pq + q^2=1$), p^2 or q^2 denotes the possibility of having respective homozygotes (or genotype frequencies of respective homozygotes) while 2pq represents the possibility of having heterozygotes (or genotype frequencies of heterozygotes).[113] Using an in-house developed SNP genotyping assay (see Chapter 4 for detailed method development), observed genotype

frequencies for each selected SNP in all recruited samples from the patients (n=30) and donors (n=43) were obtained and counted. Based on the observed genotype frequencies, the expected allele frequencies for each selected SNP were calculated. For example, for an A/G polymorphism, if the frequencies for AA, GG and AG genotypes are 10/73, 15/73 and 48/73 respectively, the expected allele frequencies for A and G will be 68/146 [i.e., (10×2+48)÷146] and 78/146 [i.e., (15×2+48)÷146], respectively. After that, under the assumption of HWE, the expected genotype frequencies were calculated. In order to assess if HWE was established for all SNPs selected, the expected and observed genotype frequencies were compared by carrying Chi-square test with two degree of freedom using a genetic analysis tool POPGENE (version 1.32). Another test multi-locus linkage disequilibrium tests were also performed by this tool. Then, in order to calculate the panel informativity, a SNP was initially defined as informative if a recipient has a SNP allele that the donor does not have. This included donor-recipient pairs comprised a homozygous donor-a heterozygous recipient or disparate homozygotes either. The informativity ***i*** of a single SNP is the sum of possibilities that given donor-recipient pairs yield possible informative combinations. Assuming all SNPs were in HWE and not genetically linked to each other, the probability ***P*** (i.e., the theoretical panel informativity) of finding at least one informative SNP between two unrelated individuals when ***n*** SNP loci were counted is:

$$P = 1-(1-i_1)(1-i_2)\ldots(1-i_{n-1})(1-i_n)$$

According to the definition of an informative SNP, the numbers of informative SNPs in the related and unrelated donor-recipient pairs were counted and compared to the theoretical calculations by Mann Whitney test (2-tailed, SPSS).

3.3 Results

3.3.1 Characteristics of the SNP Panel

Using the HapMap database, we found a total of eight SNPs at different chromosomal locations. The eight selected SNPs have allele frequencies of 0.33-0.67 and possess a high level of calculated heterozygosities ($2pq > 0.44$) in CHB population. The reported allele frequencies of most selected SNPs were similar in each group of populations. However, the 1000 Genomes Project and NCBI databases do not provide allele frequencies for one of the SNPs rs12944065, and thereby the global allele frequencies of the SNP were estimated from the HapMap database by combining the allele frequencies in different ethnic groups. The detailed information about the SNP panel is shown in **Table 4**. The CNV database search found none of the selected SNPs are co-localized in CNV regions. **Appendix 1** showed an example of the CNV database search results. The SNPs were validated to be biallelic in the recruited pre-transplant samples from patients and donors (n=73 in total) by genotyping them using the developed assay. As shown in **Table 5**, the observed genotype frequencies are very close to the expected ones. The chi-square test revealed that the difference between the observed and expected genotype frequencies for any of the SNPs is not statistically significant ($P > 0.05$). Therefore, all SNPs should be in HWE. The multi-locus LD analysis (**Appendix 2**) revealed that LD was not observed for all SNPs. Their respective alleles were not associated with each other. According to the definition of an informative SNP, there are six possible informative combinations of donor and recipient genotypes (**Table 6 and 7**). The theoretical panel informativity ***P*** was calculated to be 0.975 for both CHB population (**Table 6**) and global population (**Table 7**). Therefore, our panel should be suitable for

Table 4. Reported allele frequencies of the selected SNPs from different public databases

Chromosome	dbSNP ID	Variant alleles	Reported allele frequencies *(Calculated heterozygosities)*			
			CHB population (HapMap)	CHB population (1000 Genomes Project)	CHS population (1000 Genomes Project)	Global population (NCBI)
1	rs891700	A G	0.48 0.52 *(0.50)*	0.46 0.54 *(0.50)*	0.47 0.53 *(0.50)*	0.47 0.53 *(0.50)*
4	rs714825	A G	0.51 0.49 *(0.50)*	0.50 0.50 *(0.50)*	0.58 0.42 *(0.49)*	0.57 0.43 *(0.49)*
9	rs10757713	A G	0.47 0.53 *(0.50)*	0.45 0.55 *(0.50)*	0.46 0.54 *(0.50)*	0.51 0.49 *(0.50)*
13	rs1335873	A T	0.67 0.33 *(0.44)*	0.66 0.34 *(0.49)*	0.67 0.33 *(0.44)*	0.55 0.45 *(0.50)*
14	rs715463	A G	0.45 0.55 *(0.50)*	0.44 0.56 *(0.49)*	0.54 0.46 *(0.50)*	0.59 0.41 *(0.48)*
17	rs12944065	A G	0.53 0.47 *(0.50)*	N/A	N/A	0.39 (HapMap) 0.61 (HapMap) *(0.48)*
20	rs163781	A G	0.39 0.61 *(0.48)*	0.41 0.59 *(0.48)*	0.34 0.66 *(0.45)*	0.54 0.46 *(0.50)*
21	rs2831700	A G	0.45 0.55 *(0.50)*	0.46 0.54 *(0.50)*	0.50 0.50 *(0.50)*	0.58 0.42 *(0.49)*

Table 5. Comparison of expected and observed genotype frequencies for the 8 SNPs

Chromosome	dbSNP ID	Genotype	Observed genotype frequencies, 95% CI	Expected allele frequencies	Expected genotype frequencies	Chi-square statistic	P-value
1	rs891700	AA GG AG	0.178, 0.090 – 0.266 (13/73) 0.247, 0.148 – 0.345 (18/73) 0.575, 0.462 – 0.689 (42/73)	A: 68/146 (p) G: 78/146 (q)	0.217 (p^2) 0.285(q^2) 0.498(2pq)	1.765	0.414
4	rs714825	AA GG AG	0.315, 0.209 – 0.442 (23/73) 0.205, 0.113 – 0.298 (15/73) 0.479, 0.365 – 0.594 (35/73)	A: 81/146 (p) G: 65/146 (q)	0.308 (p^2) 0.198 (q^2) 0.494 (2pq)	0.064	0.969
9	rs10757713	AA GG AG	0.219, 0.124 – 0.314 (16/73) 0.301, 0.197 – 0.407 (22/73) 0.479, 0.365 – 0.594 (35/73)	A: 67/146 (\p) G: 79/146 (q)	0.210 (p^2) 0.293 (q^2) 0.497(2pq)	0.092	0.955
13	rs1335873	AA TT AT	0.389, 0.276 – 0.501 (28/72*) 0.125, 0.049 – 0.201 (9/72*) 0.486, 0.371 – 0.602 (35/72*)	A: 91/144 (p) T: 53/144 (q)	0.399 (p^2) 0.135 (q^2) 0.465 (2pq)	0.141	0.932
14	rs715463	AA GG AG	0.247, 0.148 – 0.345 (18/73) 0.192, 0.101 – 0.282 (14/73) 0.562 0.448 – 0.675 (41/73)	A: 77/146 (p) G: 69/146 (q)	0.278 (p^2) 0.224 (q^2) 0.499 (2pq)	1.173	0.556
17	rs12944065	AA GG AG	0.219, 0.124 – 0.314 (16/73) 0.370, 0.259 – 0.481 (27/73) 0.411, 0.298 – 0.524 (30/73)	A: 62/146 (p) G: 84/146 (q)	0.181 (p^2) 0.331 (q^2) 0.489 (2pq)	1.832	0.400
20	rs163781	AA GG AG	0.096, 0.028 - 0.163 (7/73) 0.438, 0.325 – 0.552 (32/73) 0.466, 0.351 – 0.580 (34/73)	A: 48/146 (p) G: 98/146 (q)	0.108 (p^2) 0.450 (q^2) 0.441 (2pq)	0.222	0.895
21	rs2831700	AA GG AG	0.288, 0.184 – 0.392 (21/73) 0.233, 0.136 – 0.330 (17/73) 0.479, 0.365 – 0.594 (35/73)	A: 77/146 (p) G: 69/146 (q)	0.278 (p^2) 0.223 (q^2) 0.499 (2pq)	0.112	0.945

*The genotype of one recipient could not be identified as no allele-specific extension products were generated

Table 6. Use of informative genotype frequencies to determine the panel informativity in CHB

Chromosome	dbSNP ID	Variant alleles	Reported allele frequencies (CHB population#)	Informative combinations		Probability of finding informative combinations	Cumulative probability of finding informative combinations (***i****)	Probability of finding non-informative combinations (***1-i***)
				Donor	Recipient			
1	rs891700	A G	0.48 0.52	AA AA AA GG GG GG	GA AG GG GA AG AA	0.0575 0.0575 0.0623 0.0675 0.0675 0.0623	0.3746	0.6254
4	rs714825	A G	0.51 0.49	AA AA AA GG GG GG	GA AG GG GA AG AA	0.0650 0.0650 0.0625 0.0600 0.0600 0.0625	0.3749	0.6251
9	rs10757713	A G	0.47 0.53	AA AA AA GG GG GG	GA AG GG GA AG AA	0.0550 0.0550 0.0621 0.0700 0.0700 0.0621	0.3741	0.6259
13	rs1335873	A T	0.67 0.33	AA AA AA TT TT TT	TA AT TT TA AT AA	0.0997 0.0997 0.0484 0.0235 0.0235 0.0484	0.3444	0.6556

****i***: Informativity of a single SNP, # allele frequencies were obtained from the HapMap database

Table 6 (con't). Use of informative genotype frequencies to determine the panel informativity in CHB

Chromosome	dbSNP ID	Variant alleles	Reported allele frequencies (CHB population#)	Informative combinations		Probability of finding informative combinations	Cumulative probability of finding informative combinations (***i****)	Probability of finding non-informative combinations (***1-i***)
				Donor	Recipient			
14	rs715463	A G	0.45 0.55	AA AA AA GG GG GG	GA AG GG GA AG AA	0.0501 0.0501 0.0613 0.0749 0.0749 0.0613	0.3725	0.6275
17	rs12944065	A G	0.53 0.47	AA AA AA GG GG GG	GA AG GG GA AG AA	0.0700 0.0700 0.0621 0.0550 0.0550 0.0621	0.3741	0.6259
20	rs163781	A G	0.39 0.61	AA AA AA GG GG GG	GA AG GG GA AG AA	0.0362 0.0362 0.0566 0.0885 0.0885 0.0566	0.3626	0.6374
21	rs2831700	A G	0.45 0.55	AA AA AA GG GG GG	GA AG GG GA AG AA	0.0501 0.0501 0.0613 0.0749 0.0749 0.0613	0.3725	0.6275
P = 1-(1-0.3746)*(1-0.3749)*(1-0.3741)*(1-0.3444)*(1-0.3725)*(1-0.3741)*(1-0.3626)*(1-0.3725) ≈ 0.975								

****i***: Informativity of a single SNP, # allele frequencies were obtained from the HapMap database

Table 7. Use of informative genotype frequencies to determine the panel informativity in global population

Chromosome	dbSNP ID	Variant alleles	Reported allele frequencies (global population#)	Informative combinations		Probability of finding informative combinations	Cumulative probability of finding informative combinations (***i****)	Probability of finding non-informative combinations (***1-i***)
				Donor	Recipient			
1	rs891700	A G	0.47 0.53	AA AA AA GG GG GG	GA AG GG GA AG AA	0.0550 0.0550 0.0621 0.0700 0.0700 0.0621	0.3741	0.6259
4	rs714825	A G	0.57 0.43	AA AA AA GG GG GG	GA AG GG GA AG AA	0.0796 0.0796 0.0601 0.0453 0.0453 0.0601	0.3701	0.6299
9	rs10757713	A G	0.51 0.49	AA AA AA GG GG GG	GA AG GG GA AG AA	0.0650 0.0650 0.0625 0.0600 0.0600 0.0625	0.3749	0.6251
13	rs1335873	A T	0.55 0.45	AA AA AA TT TT TT	TA AT TT TA AT AA	0.0749 0.0749 0.0613 0.0501 0.0501 0.0613	0.3725	0.6275

****i***: Informativity of a single SNP, # allele frequencies were obtained from the NCBI database

Table 7 (con't). Use of informative genotype frequencies to determine the panel informativity in global population

Chromosome	dbSNP ID	Variant alleles	Reported allele frequencies (global population#)	Informative combinations		Probability of finding informative combinations	Cumulative probability of finding informative combinations (***i****)	Probability of finding non-informative combinations (***1-i***)
				Donor	Recipient			
14	rs715463	A G	0.59 0.41	AA AA AA GG GG GG	GA AG GG GA AG AA	0.0842 0.0842 0.0585 0.0407 0.0407 0.0585	0.3668	0.6332
17	rs12944065	A G	0.39 (from HapMap) 0.61 (from HapMap)	AA AA AA GG GG GG	GA AG GG GA AG AA	0.0362 0.0362 0.0566 0.0885 0.0885 0.0566	0.3626	0.6374
20	rs163781	A G	0.54 0.46	AA AA AA GG GG GG	GA AG GG GA AG AA	0.0724 0.0724 0.0617 0.0526 0.0526 0.0617	0.3734	0.6266
21	rs2831700	A G	0.58 0.42	AA AA AA GG GG GG	GA AG GG GA AG AA	0.0819 0.0819 0.0593 0.0430 0.0430 0.0593	0.3685	0.6315
P = 1-(1-0.3741)*(1-0.3701)*(1-0.3749)*(1-0.3425)*(1-0.3668)*(1-0.3626)*(1-0.3734)*(1-0.3685) ≈ 0.975								

****i***: Informativity of a single SNP, # allele frequencies were obtained from the NCBI database

Han Chinese, non-Han Chinese and non-Chinese.

3.3.2 Number of Informative SNPs in the Related and Unrelated Donor-Recipient Pairs

In order to find the number of informative SNPs in the examined donor-recipient pairs, the pre-transplant samples from the patients and donors were genotyped by the developed assay. For the cases with double cord blood transplantation, there were two donor-recipient pairs. In total there were 28 unrelated donor-recipient pairs [26 from the cases with double cord blood transplantation (i.e., 2 times of 13 donors) and 2 from the cases with single graft transplantation] and 15 related donor-recipient pairs. Of the unrelated pairs, the genotyping results identified 0-6 informative SNPs (mean 2.8). Of the related pairs, the genotyping results identified 0-7 informative SNPs (mean 2.4). Unfortunately, no informative SNPs could be found in two unrelated and one related pairs by using the selected SNP panel. Based on our genotyping results, the panel informativity was found to 93% (40 out of 43; 95% CI = 85-100%), which was not significantly different from the expected value of 97.5% (P = 0.060, Chi-square test). Almost identical informativity vales were observed for the related (93%, 26 out of 28; 95% CI = 83-100%) and related (93%, 14 out of 15; 95% CI = 81-100%) pairs. **Figure 7** summarizes these results. It should be noted that the mean number of informative SNPs in the unrelated pairs was higher than that in the related pairs, but not statistically significant (P = 0.178, Mann Whitney test, 2-tailed). Informative SNPs were rarely found among the cases with double cord blood transplantation. Only one informative SNP was identified in the four cases.

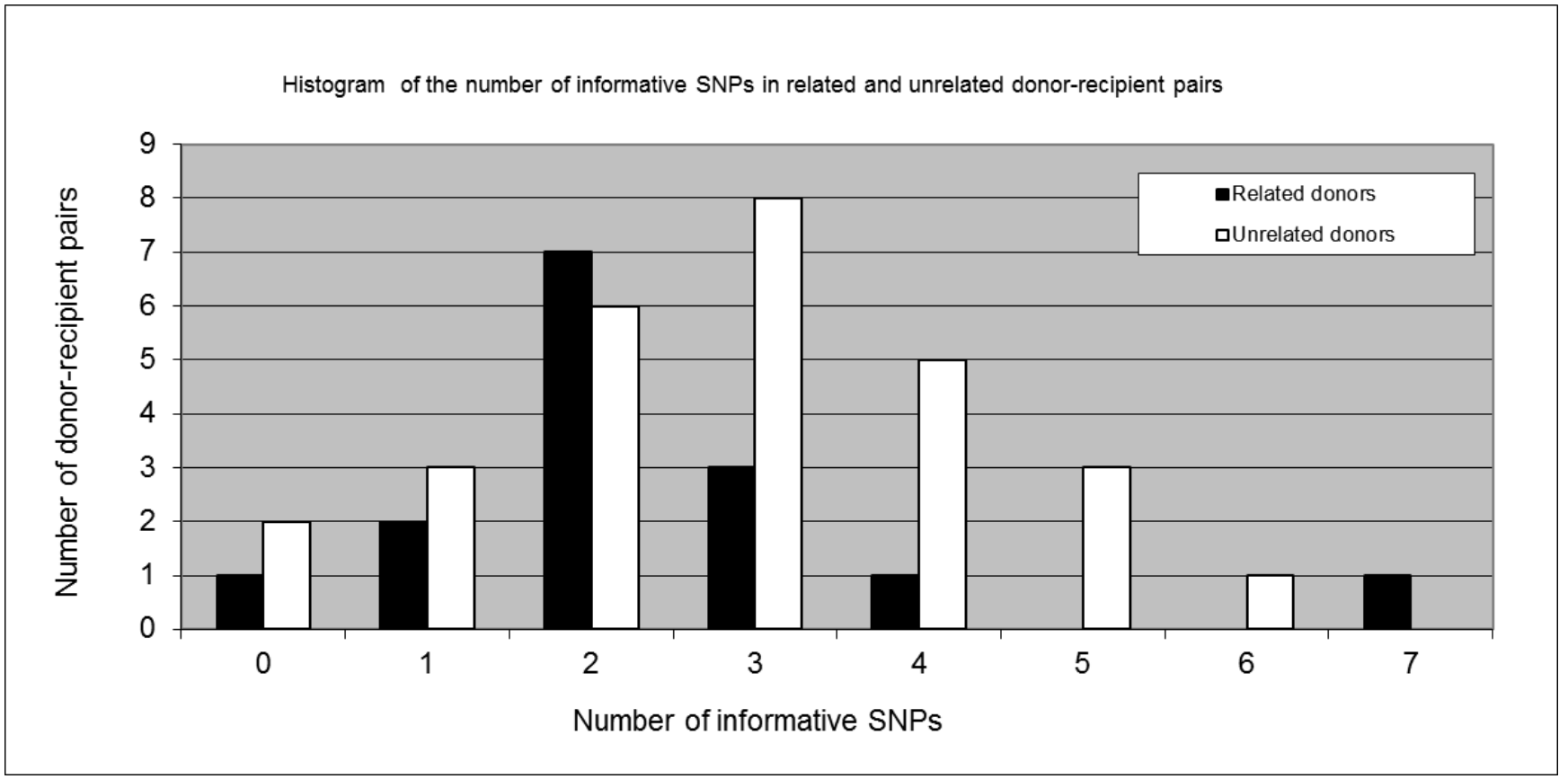


Figure 7. Histogram showing the number of informative SNPs identified in related (n=15) and unrelated (n=28) donor-recipient pairs

3.4 Discussion

In this part of our study, we identified a panel of eight SNPs with high allele frequencies on different chromosomes from the HapMap database. On the basis of the expected allele frequencies, the theoretical panel informativity was calculated and evaluated by genotyping the recruited pre-transplant patient samples and corresponding donor samples using the in-house developed SNP genotyping assay. Detailed method development would be discussed in Chapter 4.

The SNPs were selected from various chromosomes to ensure that they are independently inherited. Statistical analysis confirmed the LD did not significantly exist among the selected SNPs. Allele frequencies of a single SNP rather than haplotype frequencies were preferentially used for calculating the panel informativity since haplotypes are relatively less stable than a single SNP.[83] In addition, the selected SNPs are not co-localized in CNV regions. Otherwise, they can exist in more than one copy number in the human genome. Then, the allelic ratio of an informative SNP in a post-transplant patient sample will be different from the expected value. This greatly affects the calculation of MC percentages since allele-specific genotyping products are not calculated on the basis of the expected allelic ratio under the assumption that there is only one allele copy in each chromosome. (See Chapter 5 for detailed calculation of MC percentages). All SNPs were shown to be biallelic without any mutation detected and informative with high value of heterozygosities in the examined samples. The statistical analysis showed that they were not deviated significantly from HWE. Therefore, all these results confirmed that our strategy for selection of reporting SNP markers was

valid, and the eight selected SNP markers were suitable markers for distinguishing recipient genomic materials from the donor genomic materials in CHB population.[83]

For the SNP panel, theoretically there is a probability of 97.5% of finding at least one informative SNP among unrelated individuals. Experimentally, informative alleles were observed in 93% of the donor-recipient pairs regardless of the fact whether individual donor-recipient pairs were related or not. There was a trend (P = 0.060) that the observed informativity (93%) was lower than the expected value of 97.5%. In future, more donor-receipt pairs should be examined in order to confirm this trend. It is worth noting that the observed informativity values for both related and unrelated pairs were both about 93%. Theoretically, the informativity value should be lower in related donor-recipient pairs. The observation of no significant difference could have been caused by insufficient sample sizes of the both groups in this study. Based on our definition of an informative SNP, informativity of a single SNP of 50% allele frequency is 37.5% [(0.5×0.5×0.5×0.5)×6]. When four more of such SNPs are included in our SNP panel, the panel informativity can increase from 93% to 99% [$1-(1-0.93)(1-0.375)^4$].

Although it appeared that more informative SNPs could be identified in unrelated pairs, the difference was not statistically significant. Theoretically, more informative peaks should be observed in the unrelated pairs. The insignificant finding should have caused by insufficient sample sizes in both related and unrelated groups. The experimental informativity may match the theoretical one after increasing sample size. Despite not being as polymorphic as multiallelic markers, biallelic SNPs seemed to enable distinguishing unrelated and related individuals from one another, as long as the

number of SNP markers was high enough. Previously it was estimated that seven SNPs with 0.5 allele frequency would be sufficient to identify 99% of related or unrelated donor-recipient pairs.[80] We showed that eight SNPs were able to identify 93% of examined pairs. Although it was lower than the expected value of 97.5%, the number of informative pairs found was still acceptable.

The successful rates of identification of the eight SNP markers among the studied subjects were 100%, except 98% for 1 SNP marker (dbSNP ID: rs1335873). In the failure case (a recipient in one of the unrelated donor-recipient pairs), no primer extension products were detected. It might be due to the failure in competitive PCR, detection of the region containing the SNP or point mutation at 1 base upstream of the SNP in the recipient. However, our genotyping results could still identify three informative SNPs from the other seven SNPs in the pair. These helped provide complementary information for subsequent chimerism analysis.

Our SNP panel had two limitations. Firstly, it was not suitable for cases with double cord blood transplantation. Among the 13 cases of our studies, when we tried to identify informative SNPs that distinguished the receipt genomic material from both donor genomic materials in individual HSCT cases, informative SNPs were found to be present in only 31% of the cases (4 out of 13). Such a low percentage was expected because the reporting SNPs were selected for differentiating two individuals, but not differentiating one from two individuals. One of the donor hematopoiesis will emerge as a dominant unit within 1 month after transplantation in more than 80% of patients receiving double cord blood transplantation.[114] Hence, our SNP panel should remain

useful after 1 month of double cord blood transplantation. More SNP markers should be included if one wants to quantify donors' genomic materials within 1 month after double cord blood transplantation. On the basis of the definition of an informative SNP, the theoretical informativity of our selected panel for differentiating one from two individuals is 57% (calculation not shown). Theoretically inclusion of 40 more SNPs with 50% allele frequency can increase the probability of finding at least one informative SNP in one recipient from two donors to over 99% (calculation not shown). Another major limitation of this SNP panel was that it was mainly designed for the Han Chinese population. If this panel is applied to other non-Chinese populations, modifications will be likely required, such as addition of other SNPs with high allele frequencies in other populations.

In future studies, more donor-receipt pairs should be examined in order to confirm the informativity values observed in this study, and to answer whether related and unrelated donor-recipient pairs have similar informativity values. Furthermore, inclusion of more SNP markers in the SNP panel may increase the likelihood of identifying more informative SNPs in all donor-recipient pairs, especially in double cord blood transplantation cases. If one wants to examine non-Chinese donor-recipient pairs, modifications of the SNP panel may be required.

3.5 Conclusion

A panel of 8 SNPs that are in linkage equilibrium and HWE was identified by using the HapMap database. The panel informativity was calculated to be 0.975 for both CHB and global populations, i.e., this panel was likely to provide 97.5% probability of finding at least one informative SNP between two unrelated Han Chinese, non-Han Chinese and non-Chinese individuals.

In order to evaluate whether this panel could distinguish our recruited patients and donors reliably, the developed SNP genotyping assay was performed by using custom PCR and extension primers. In total 93% (40 out of 43) of the donor-recipient pairs was distinguishable. This percentage was insignificantly smaller than the theoretical percentage, 97.5%. More informative SNPs could be identified in the unrelated pairs, but this finding was insignificant also.

CHAPTER 4: SECTION 2 - DEVELOPMENT OF QUANTITATIVE MALDI-TOF MS-BASED SNP GENOTYPING ASSAY

4.1 Introduction

Although different strategies coupled with MALDI-TOF MS for SNP genotyping were invented, primer extension reactions are the only approach used practically.[95] Its simplicity, flexibility and robustness are attractive features for researchers desiring a tailored-made multiplex SNP genotyping assay. The principle of the entire assay can be separated into three fundamental steps with two levels of specificity (**Figure 8**). To begin with, competitive multiplex PCRs are used to amplify several SNP loci of interest simultaneously. PCR products rather than original genomic DNA act as templates for subsequent primer extension reactions. The human genome is highly complex. It contains approximately 3 billion of base pairs. Reduction of genome complexity by amplifying regions containing targeted SNPs aids in reduction of non-specific annealing and extension of genotyping primers (or extension primers).[115] In addition, amplification of sequence regions containing the targeted SNPs also helps to increase the detection limit. On the basis of the flanking sequences of each selected SNP, respective forward and reverse PCR primers and extension primers could be designed.

Following the first level of specificity, the reaction mixture containing amplified products (targeted SNP loci) are subjected to shrimp alkaline phosphatase (SAP) treatment to remove unincorporated dNTPs. The enzyme degrades monophosphate linkages from of remaining dNTPs and converts them into non-functional 3'-deoxynucleoside diphosphates (dNDPs). Otherwise, the left-behind functional dNTPs can inference the following primer extension reactions, which is the third step, resulting

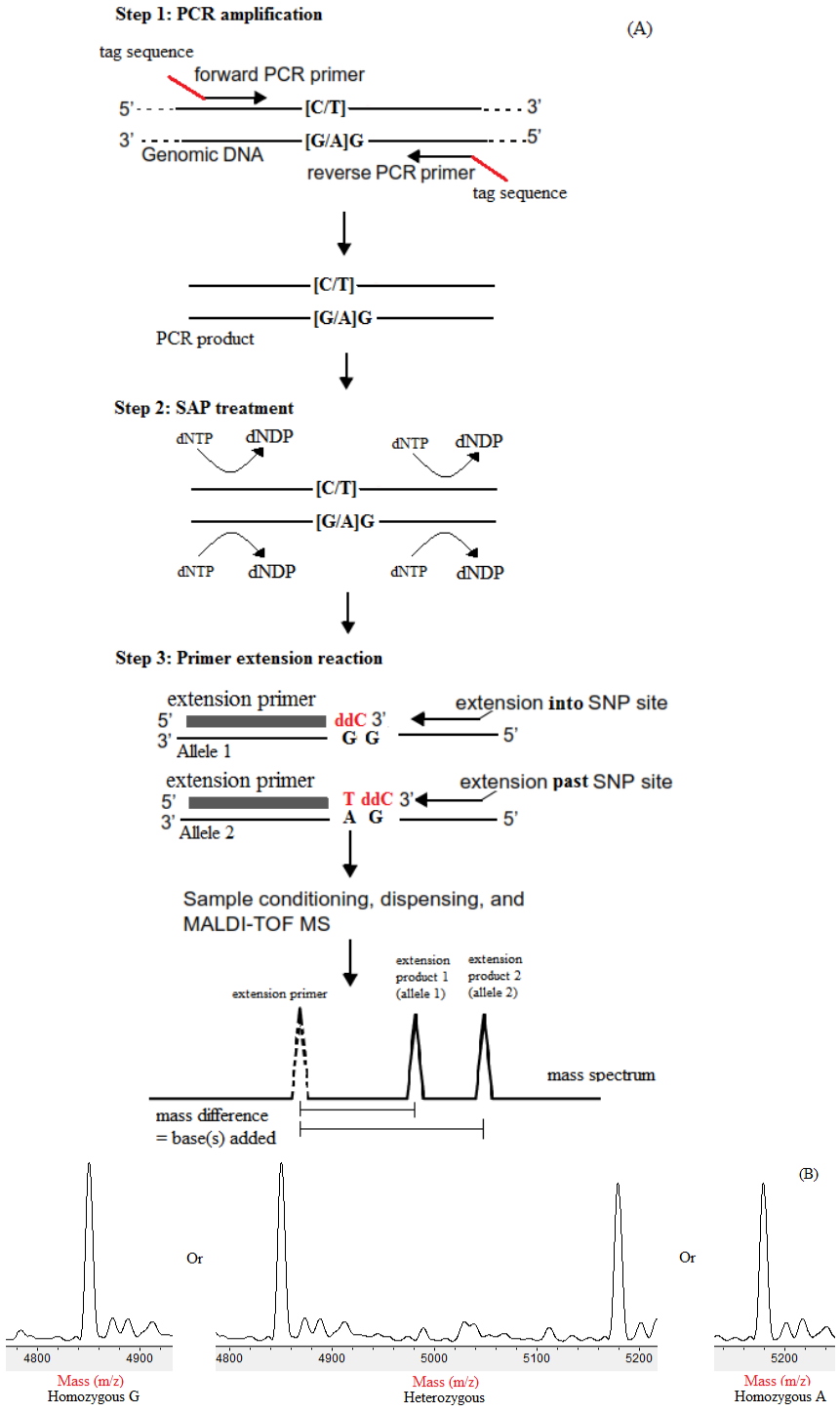


Figure 8. (A) Overview of our primer-extension based SNP genotyping assay. **(B)** Presence of two peaks represents heterozygosity while one peak represents homozygosity. Their masses (m/z) are indicators for correct genotyping of a target SNP.

in formation of unexpected peaks that complicate data analysis in mass spectra.[116] After the clean-up of dNTPs, a base extension reaction cocktail, comprising extension primers, sequencing-grade DNA polymerase (sequenase), and a proper dNTPs/ddNTPs mixture (a terminator mix), is added to the treated PCR products. After the extension primer specifically anneal immediately one-base adjacent to the targeted SNP sites, the sequenase extends the primers either by adding one or two nucleotides (this is the reason why the reaction is called multiple base primer extension reaction), depending on the allele of the SNPs. The extension reaction terminates at the nucleotide that is complementary to one of the ddNTPs in the terminator mix. In case of a heterozygous SNP, two allele-specific extension products are produced. By design, they differ by one nucleotide or the most two nucleotides, and appear as two well-separated signal peaks with defined masses in MALDI-TOF MS analysis. In case of a homozygous SNP, only one signal peak with an allele-dependent mass can be observed. Accurate measurement of the masses and signal peak areas of extension products by MALDI-TOF MS is important for qualitative analysis, such as SNP genotyping, and quantitative analysis, such as chimerism analysis. **Figure 8** shows an overview of the whole assay design.

Prior to setting up the SNP genotyping assay, MALDI TOF-MS can be used to assess quality and integrity of the extension primers supplied by the commercial companies. Mass spectrum of a newly purchased primer can be assessed with respect to the theoretical mass of the primer. Presence of unexpected peaks or absence of expected peaks indicates poor primer quality.[117] Adjustments of concentrations of the extension primers with different masses in the extension reaction cocktail is the next important step to develop successful multiplex assay. There is an inverse relationship between

analyte mass and signal-to-noise ratio (S/N) in MALDI-TOF MS.[115] Oligonucleotides with higher masses generally give lower S/N. The detection sensitivity of MALDI-TOF MS decreases when the mass of a molecule increases. Besides, mass-independent variations in peak heights can occur, as oligonucleotides may behave differently in ionization/ desorption process in MALDI and have inconsistent quality.[117] Hence, even at the same concentration, oligonucleotides with different masses usually do not produce comparable peak signals. It is necessary to adjust the amount of the extension primers so that sufficient peak intensity can be provided for reliable genotype calling. After the adjustments, components for the competitive multiplex PCR and primer extension reaction and thermal cycling conditions to these reactions have to been optimized. A concentration ratio of dNTPs to $MgCl_2$, for example, is a critical parameter for a successful multiplex PCR assay. Studies in the area of optimization are abundant and can be used as references when tailored-made competitive multiplex PCRs and primer extension reactions are designed.[118, 119]

Following the generation of the allele-specific extension products, it is critical to optimize the analyte preparation and clean-up steps in order to generate high-quality mass spectra of the products. It is well known that it is much more difficult to analyze nucleic acids than peptides by MALDI-TOF MS. It is mainly attributable to impurity contamination and analyte fragmentation, which cause reduction of sensitivity, resolution and accuracy in the analysis.[96] Carrying a negatively charged phosphate backbone, nucleic acids are highly susceptible to form salt adducts with metallic cations that are commonly present in molecular biology reaction buffers. Other common components in the reaction buffers are ionic detergents, some of which are matrix-

analyte cocrystallization-disturbing agents, such as Tween-20.[120] They severely affected the MALDI detection of DNA products. Several clean-up processes have been reported, such as cation-exchange chromatography[121], ethanol precipitation[122], reverse-phase chromatography by ZipTips[120], etc. On the other hand, DNA fragmentation during MALDI ionization is a common problem, which results in formation of fragmented peaks. Nucleobase elimination and backbone cleavage of oligonucleotides are two predominant fragmentation effects. In fact, excess laser energy applied on an analyte during MALDI ionization may cause high degree of fragmentation. The nature of a matrix chemical used in MALDI-TOF MS analysis of DNA also affects the degree of fragmentation.[104]

After clean-up, a DNA sample is spotted on a stainless steel target plate in the presence of a suitable matrix chemical solution for subsequent MALDI-TOF MS analysis. The most commonly used matrix chemical for DNA analysis is 3-hydroxypicolinic acid (3-HPA).[123] Besides, other matrix additives can be included to facilitate DNA analysis. For example, dibasic ammonium citrate can suppress formation of the salt adducts.[124] Fructose is able to reduce the DNA fragmentation and raise the yield of parent peaks.[104] Ascorbic acid can reduce the formation of adducts in MALDI-TOF MS.[125]

A well-established SNP genotyping platform, called MassEXTEND™, is commercially available. It is based on a low-end MALDI-TOF MS platform with detection limit of femtomole. The aim of this part of the M.Phil. research project was to design and optimize a specific and sensitive quantitative SNP genotyping assay by using

an advanced MALDI-TOF/TOF MS platform with attomole detection sensitivity. Hopefully, using a more sensitive MS platform, an in-house assay could be developed to allow quantification of MC with detection limit down to 0.1%. An eight-plex competitive PCR reaction and two quadruplex primer extension reactions were optimized. This SNP genotyping assays was then utilized in a qualitative (i.e., SNP genotyping of examined pre-transplant patients and donors) and quantitative (i.e., chimerism assessment in post-transplant patients) analyses of donor/recipient genomic materials.

4.2 Materials and Methods

4.2.1 Isolation of Genomic DNA from the Human Blood Samples

The blood samples were initially collected into EDTA tubes and transferred to the laboratory at room temperature for DNA extraction by a standard phenol-chloroform method.[126] Eventually the extracted DNA, which was dissolved in autoclaved 100μL of 10mM tris buffer (pH = 8.1) (AMRESCO™, Solon, USA), was stored at -80°C until further processing. DNA concentrations were determined by ultraviolet spectrophotometry at 260nm using a NanoDrop™ 2000 Spectrophotometer (Thermo Fisher Scientific™, Wilmington, USA).

4.2.2 Design for PCR and Extension Primers

All primers were designed manually. Their masses were calculated based on the absence of phosphate group on the 5' ends and their melting temperatures (T_ms) were calculated based on nearest neighbor algorithm (NN) by using the Oligonucleotide Properties Calculator (http://www.basic.northwestern.edu/biotools/OligoCalc.html). Prior to primer extension reactions, targeted SNP regions in the genomic DNA must be amplified by competitive PCRs. Each pair of PCR primers was designed according to the template sequence upstream and downstream of from a targeted SNP site. On the 5' ends of all PCR primers, a tag or non-binding sequence was included to increase their masses. After the competitive PCR, any residual PCR primers should fall outside the mass range of the extension primers, which would be our analytical targets. All PCR primers were screened for possible secondary structures like hairpin and primer-dimer interactions. With regard to extension primers annealing immediately adjacent to a SNP site being

genotyped on templates (PCR products), they either extend up to the SNP site or through the site by 1 nucleotide by addition of a mixture of ddNTPs and dNTPs, which is so-called a terminator mix. They were designed with similar effective binding length on the templates and T_m. A nucleotide T was included at their 5' ends to minimize fragmentation during ionization process in MALDI-TOF MS.[127] Absence of unwanted intra- and inter-primer interactions that could possibly result in non-specific extension was confirmed manually. The minimum separation of the extension primers and extension products was restricted to 160Da so as to minimize chances for peaks overlapping in mass spectra. The upper limit of mass was set to 8500Da as signal intensity and resolution decreased with primer mass.[115] All sequencing-grade custom primers were synthesized by and purchased from Applied Biosystems™ (Foster City, USA). They were delivered in a lyophilized and desalted form. After the delivery, stock primer solutions of 100μM were prepared by dissolving the DNA pellet in UltraPure™ DNase/RNase-Free distilled water (Invitrogen™, Grand Island, USA) and stored at -20°C.

4.2.3 Adjustments of Amounts of the Extension Primers

To begin with, a 100μL mixture solution comprising 5pmol of each extension primer (eight primers in total) was prepared. After the primer mixture was cleaned up by using cation exchange resin and purified with ZipTip microextraction, it was dispensed on a stainless steel target plate for MALDI-TOF MS analysis (the detailed procedures will be discussed in section 4.2.7 to 4.2.9). The resulting primer peaks were adjusted based on the highest peak. For any peaks with low peak heights, more primers were added until

all peaks were at least around half of the height of the highest peak. For example, if a given peak is about 40% intensity of the highest peak, instead of 5pmol, 7pmol [5pmol×(1+40%)] of the primer was added to the primer mixture for the second analysis. The m/z values were externally calibrated with Protein Calibration Standard I (Bruker Daltonics^{TM}, Leipzig, Germany), observed masses of the extension primers were obtained by a formula "m/z value – 1" (as they are predominantly singly charged by a proton) and verified if they matched the expected masses. Components in the calibration standard were shown in **Table 8**.

Table 8. List of the proteins in the Protein Calibration Standard I

Proteins	Molecular ion form	Average mass (m/z)
Insulin	$[M+H]^{+}$	5734.51
Cytochrome C	$[M+2H]^{2+}$	6180.99
Myoglobin	$[M+2H]^{2+}$	8476.65
Ubiquitin I	$[M+H]^{+}$	8565.76
Cytochrome C	$[M+H]^{+}$	12360.97
Myoglobin	$[M+H]^{+}$	16952.30

4.2.4 Competitive Multiplex PCR Conditions and Reaction Clean-up

For the development of the SNP genotyping assay, some DNA samples were obtained from several healthy subjects without known hematopoietic diseases with their prior consent The competitive multiplex PCR was performed in a final volume of 10μL using the PCR primers listed in **Table 10** (See section 4.3.1) and 80ng of the template DNA. Two types of DNA polymerase, HotStarTaq™ *Plus* DNA Polymerase (QIAGEN™, Hilden, Germany), and FastStart™ Taq DNA Polymerase (Roche Molecular Biochemicals™, Mannheim, Germany), were used for comparison in terms of PCR efficiency and suitability for MS analysis. Final concentrations or amounts of other PCR reagents in a PCR Master Mix were: 0.35units of DNA polymerase, 1X Taq PCR buffer, 0.1mM of dNTPs (GE Healthcare™, Freiburg, Germany), 1.5mM $MgCl_2$, and 0.13μM of each forward and reverse PCR primer. Thermal cycling conditions to the competitive multiplex PCR were 94°C for 7min (5min for the HotStar polymerase) for initial denaturation and enzyme activation, followed by repeating thermocycles 45 times (denaturation at 94°C for 30sec, annealing at 60°C for 90sec and extension at 72°C for 90sec), and final incubation at 72°C for 10min for final extension. The cycling programmes for the competitive PCR and subsequent SAP treatment and primer extension reactions were carried out in DNA Engine™ Peltier Thermal Cycler (MJ Research™, South San Francisco, USA). For initial evaluation of performance of the PCR amplification, parts of PCR products were subjected to polyacrylamide gel electrophoresis (See section 4.2.6). After the PCR products, which were stored at 4°C until further processing, were obtained, they were subjected to SAP (GE Healthcare™, Freiburg, Germany) treatment to remove remaining dNTPs. 2μL of a SAP working

solution with 0.5 units of SAP and 1X Sequenase reaction buffer (GE Healthcare™, Freiburg, Germany) was added to each 10μL post-PCR solution. The mixture was incubated at 37°C for 50min followed by 20min at 85°C to inactivate SAP. The SAP-treated PCR products were stored at 4°C until setting up primer extension reaction.

4.2.5 Primer Extension Reaction Conditions

Thermo Sequenase™ (GE Healthcare™, Freiburg, Germany) was used in multiple base primer extension reactions. For each sample to be analyzed, two sets of base extension reaction cocktails were prepared and utilized. Each set used a particular terminator mix (dT or dG mix) and a set of the extension primers (4 primers for each set) (An extension primer list will be shown in **Table 11** in section 4.3.1). 2μL of the reaction cocktail, consisting of 1.28units of the Thermo Sequenase, 1X Sequenase reaction buffer, a dG mix (10mM of dGTP, ddCTP, ddATP, and ddTTP) or dT mix (10mM of dTTP, ddCTP, ddATP, and ddGTP) (GE Healthcare™, Freiburg, Germany), and 4 specific extension primers (referred as a PrimerMix) with adjusted concentrations, was added to the 12μL of SAP-treated PCR products. Thermal cycling conditions to the primer extension reaction were 94°C for 30sec for initial denaturation and enzyme activation, followed by repeating thermocycles 120 times (denaturation at 94°C for 5sec, 5 cycles of annealing at 52°C for 5sec and extension at 72°C for 5sec), and final incubation at 72°C for 3min for final extension. Extension products in a final volume of 14μL were stored at 4°C until they were ready for clean-up procedures.

4.2.6 Gel Electrophoresis Analysis of the PCR products

Unless otherwise specified, all components for electrophoresis, including buffers and gels, were purchased from Bio-Rad Laboratories™ (Hercules, USA) A mixture of 5μL of the PCR products, 2μL of 5X sample loading buffer (50mM Tris-HCl, 5mM EDTA, 25% glycerol, 0.2% bromophenol blue, 0.2% xylene cyanole FF, pH 8.0) and 3μL of autoclaved water was fractionated on a 15% polyacrylamide gel (Criterion™ Precast TBE Gel) run in 1X TBE buffer (89mM Tris, 89mM boric acid, 2mM EDTA, pH 8.4) for 3hr at 150V. A 100bp DNA ladder (Invitrogen™, Grand Island, USA) served as a size marker in the gel. The PCR products were visualized by staining the gel with 3X GelRed™ staining solution (Biotium, Hayward, USA) followed by exposure to ultraviolet light for visualization.

4.2.7 Sample Clean-up prior to MALDI-TOF MS Analysis

Three sample clean-up approaches were proposed. They are summarized as follows. All working solutions, including aqueous solutions and organic solvents with various percentages, were prepare by UltraPure™ DNase/RNase-Free distilled water (Invitrogen™, Grand Island, USA). In order to find an optimized and time-efficient methodology for sample clean-up, four combinations among the three clean-up approaches were tested: (I) ZipTip microextraction alone; (II) removal of alkali metal cations with cation exchange resin followed by ZipTip microextraction; (III) ethanol precipitation followed by ZipTip microextraction; (IV) ethanol precipitation and removal of alkali metal cations with cation exchange resin followed by ZipTip microextraction.

A. Ethanol precipitation:

1.4μL of 10M ammonium acetate (Sigma-Aldrich™, St. Louis, USA) solution was added to the extension products in a final volume of 14μL. After 5min for incubation, 100μL of absolute ethanol (Merck™, Darmstadt, Germany; HPLC gradient grade) was added and the mixture was placed at -20°C for 30min. The mixture was then centrifuged at 16000g at 4°C for 10min, followed by removal of supernatant without disturbing a white or translucent pellet. It was washed by 100μL of 85% ethanol, which was pre-chilled at -20°C. After the washing step, again, the solution was centrifuged at 16000g at 4°C for 10min, followed by removal of supernatant. The extension products were re-suspended by adding 200μL of the UltraPure™ distilled water. For further clean-up by cation exchange resin and ZipTips, half of the re-suspended sample (100μL) was used each time.

B. Removal of alkali metal cations with cation exchange resin:

DOWEX™ 50WX8-400 ion exchange resin was purchased from Sigma-Aldrich™ (St. Louis, USA). For initial preparation of the resin, it (roughly 5g) was washed three times with 10mL of 1M ammonium acetate solution. Following that, the resin was immersed in 15mL of 2M ammonium acetate solution and mixed for 2 days. The resulting resin in ammonium form was washed three times with 10mL of 1M ammonium acetate solution. Then, it was gently washed two times with the UltraPure™ distilled water, acetone (Lab-Scan™, Dublin, Ireland; analytical grade), and n-hexane (Duksan™, Korea; HPLC grade) in order. The air-dry resin was stored at room temperature until use. To use the prepared resin for sample desalting, 3mg of it was first transferred to a 0.6-mL microcentrifuge tube. 100μL of the sample, which was previously treated by ethanol

precipitation, was vortexed with the resin for 1hr. The supernatant (~100μL) was transferred to a 96-well PCR plate for further clean-up by ZipTips.

C. ZipTip microextraction:

μ-C18 ZipTip™ (Millipore™, Billerica, Germany) pipette tips were ordered for this microextraction step. The entire working process (except elution process) was performed on a Thermo-Fast™ 96 Skirted PCR plate (Thermo Fisher Scientific™, Waltham, USA) by a pipetting robot, JANUS™ Automated Workstation (PerkinElmer™, Waltham, USA), with 4-tip dispense arms equipped with the ZipTips. 100μL of the ethanol-precipitated and cation resin-treated sample (for samples without treatment of the ethanol precipitation and ZipTip microextraction, 100μL of the sample was prepared by adding 86μL of the UltraPure™ water to the 14μL extension products) was transferred to 100μL of 0.2M triethylammonium acetate (TEAA) (Merck™, Darmstadt, Germany) solution on the 96-well PCR plate. The programme shown in **Table 9** was automatically run by the JANUS™ machine. For elution of the sample from the ZipTips, 0.7μL of 50% acetonitrile (ACN) (v/v) (Merck™, Darmstadt, Germany; HPLC gradient grade) solution was aspirated and dispensed manually through the sample-bound ZipTips at least 15 times without introducing air. The eluant was ready to be deposited on a stainless steel target plate for MALDI-TOF MS analysis.

Table 9. Programme run by JANUS™ Automated Workstation

Steps	Solution, volume in a well	Number of aspirating and dispensing cycle	Pipetting volume (μL)
1. ZipTip pre-wetting and equilibration	Absolute ACN, 100μL	3	15
	50% ACN (v/v) solution, 100μL	3	15
	30% ACN (v/v) solution, 100μL	3	15
	0.1M TEAA solution, 100μL	3	15
2. Sample binding	Sample solution, 200μL	20	170
3. ZipTip washing	0.1M TEAA solution, 100μL	4	30
	0.1M TEAA solution, 100μL	4	30
	0.1M TEAA solution, 100μL	4	30
	UltraPure™ distilled water, 100μL	1	30

4.2.8 Matrix Preparation

UltraPure™ distilled water was used during the matrix preparation. Initially a saturated 3-HPA (Fluka™, Neu-Ulm, Germany; mass spectrometry grade) solution was prepared freshly by dissolving 2mg of 3-HPA in 50μL of 60% ACN (v/v) solution. A 100μL of working 3-HPA matrix solution was made by loading 4μL of 100mg/mL dibasic ammonium citrate (Sigma™, St. Louis, USA) solution, 10μL of 50mg/mL β-D-fructose (≥97% purity) (Sigma-Aldrich™, St. Louis, USA) solution, and 8μL of 0.2M L-ascorbic acid (Sigma™, St. Louis, USA; reagent grade) into the saturated 3-HPA solution. Prior to use the working 3-HPA matrix solution (100μL) was briefly desalted by mixing with 3mg of the prepared cation exchange resin in ammonium form for 20min. To deposit the sample on a target plate, a dried droplet spotting method was employed. 0.15μL of the working matrix solution was mixed with the eluant (0.7μL) and then spotted onto a 384/400-μm AnchorChip™ MALDI sample plate (Bruker Daltonics™, Leipzig, Germany) where it dried under vacuum at room temperature.

4.2.9 MALDI-TOF MS Analysis

The analysis was performed on a MALDI-TOF-TOF-mass spectrometer Ultraflex III (Bruker Daltonics™, Leipzig, Germany) equipped with a 355nm Nd:YAG laser. The MS was operated in linear positive ion mode and the laser power was optimized before use. All mass spectra were acquired automatically using the MS's AutoXecute function as a result of 5040 laser shots at different spot positions. Two software were used with the MS. Acquisition of mass spectra was performed in flexControl (Version 3.3, Bruker Daltonics™) while processing and annotation of spectra in flexAnalysis (Version 3.3,

Bruker Daltonics™). Results were acquired in terms of several peak characteristics, m/z, intensity, and area.

4.3 Results

4.3.1 PCR and Extension Primers for Our SNP Genotyping Assay

Table 10 and 11 list the PCR and extension primers respectively. All T_ms of the PCR and extension primers were calculated under the assumption of each primer being 130 and 50nM respectively (the tag sequences were excluded during the calculation). All PCR primers were designed to have the same size 32bp, including a tag sequence of 7-13 bases long on their 5'-end, their masses were therefore over 9700Da. The tag sequence assisted in uniform amplification of competitive multiplex PCR and increased the masses of the PCR primers. The range of their T_ms lied within 47.0^{o}C-53.2^{o}C (mean 50.8^{o}C). The amplicon sizes calculated included the sizes of amplified regions and the tag sequences. Since the masses of all extension primers and extension products must be separated by at least 160Da with upper mass limit of 8500Da, two sets of extension primers (PrimerMix1 and PrimerMix2) were designed. Each set contained 4 extension primers and used a specific terminator mix (dT and dG mix) for the primer extension reaction. The extension primers were designed with effective binding length of 15-23bp and T_m of 38.2-50.9^{o}C (mean 46.3^{o}C). For both PCR and extension primers, a nucleotide T was preferentially introduced to their 5'-ends. The extension primers and corresponding extension products were ranked according to their expected molecular weights (M_r).

Table 10. PCR primers for our SNP genotyping assay

Chromosome	dbSNP ID	Direction	Sequence* (5'---3'), 32bp	M_r (Da)	T_m, 130nM (^oC)	Amplicon size
1	rs891700	Forward	**TTCAGACTAAC**CAGAGGTGGTATTTCTAGCTG	9854.5	53.2	117bp+21b
		Reverse	**TTGACGTACC**CACTATGTGTGAGCTATGACAC	9799.4	53.2	
4	rs714825	Forward	**TTTCGAGCA**GCTTCCTGATTAGTTAGTTGAAT	9835.5	51.8	81bp+16b
		Reverse	**TTCTTGC**GTCCTATACTTTCAGTGTTTAGAAC	9746.4	52.2	
9	rs10757713	Forward	**TTACTAGACAGAC**TTCATTTGAGTGGCATATG	9853.5	47.0	121bp+23b
		Reverse	**TTTCGTCAAC**ATCACATACACATCTAAGCTTG	9702.4	51.3	
13	rs1335873	Forward	**TTATTGTAAGTACT**CAAACGTGTGATGCTCTC	9804.5	49.9	81bp+27b
		Reverse	**TTGTAAGACGTAC**CAGGGTGCAGGTATGTATT	9934.5	49.6	
14	rs715463	Forward	**TTAGTGGATATACA**TGCCAGTCTTCTTGGATT	9835.5	49.0	82bp+27b
		Reverse	**TTTAGTTCAGACC**TTCTATCCATTGGCAGAAC	9749.4	48.3	
17	rs12944065	Forward	**TTCGTAGCTTGTAT**TGCAACATGTGGGTGTTA	9891.5	48.0	118bp+27b
		Reverse	**TTTGACGTACGAC**TGTGGTTAGAAGAAGCCAT	9903.5	49.4	
20	rs163781	Forward	**TTGGATTGGAAG**CCTAGGAAGCAACTTTCAGT	9903.5	51.8	96bp+26b
		Reverse	**TTGCTCATACTTAT**CACAGTCGGCTGAGTGAC	9790.4	53.2	
21	rs2831700	Forward	**TTAATTTGAACA**GACCTATGCTGTGCAGTATC	9813.5	52.0	88bp+21b
		Reverse	**TTCCTGAGA**CCTAGAACCACAATTATCTGTCT	9718.4	52.5	

*Bold and underlined: Tag sequences

Table 11. Extension primers for our SNP genotyping assay

Primer Mix#	Primer Code	Chromosome	dbSNP ID	Sequence* (5'---3')	Effective binding length (bp)	T_m, 50nM (^oC)	Expected M_r of extension primers (Da)	Variant alleles	Expected M_r of extension products (Da)
Primer Mix1	Ex4	4	rs714825	TGAGGACTGAGAAAC	15	38.2	4650.1	G	4923.3
								A	5251.5
	Ex9	9	rs10757713	**T**GCTCTAAGCCAACTTCTT	18	47.1	5713.8	G	5987.0
								A	6331.2
	Ex1	1	rs891700	**TT**ACTCCTTAGAACTATGCAAG	20	48.1	6693.4	G	6966.6
								A	7294.9
	Ex13	13	rs1335873	**TT**GTACCTAGCTATGTACTCAGTAT	23	50.2	7622.0	T	7919.2
								A	8199.4
Primer Mix2	Ex20	20	rs163781	TGAGTCCGATTTTCT	15	38.5	4549.0	A	4846.2
								G	5175.4
	Ex17	17	rs12944065	TTGCACCCCACTCTACAC	18	48.8	5339.5	A	5636.7
								G	5941.9
	Ex14	14	rs715463	**TTA**TTCTTGGATTCCAGGCCT	18	49.0	6378.2	A	6675.4
								G	6995.6
	Ex21	21	rs2831700	**TTCT**GCTAAACTATTGCCGGAGAT	20	50.9	7342.9	A	7640.1
								G	7969.3

*Bold and underlined: Tag sequences for adjusting the masses of the extension primers and their products
#The terminator mix used in PrimerMix1was dG mix (a mixture of dGTP, ddATP, ddTTP and ddCTP) while that used in PimerMix2 was dT mix (a mixture of dTTP, ddATP, ddCTP and ddGTP)

4.3.2 Adjusted Concentrations of Each Extension Primer for the Primer Extension Reaction

In case of the analysis of the equimolar primer mixture, obviously the S/N decreased with the length of the extension primers. Signals for the extension primers with masses over 6000Da (Ex1, Ex13, Ex14, and Ex21) were severely suppressed. Therefore, more primers with low-peak intensity were spiked in to even out the peak intensities based on the relative peak heights between them and the primer with highest peak. **Figure 9** shows the obvious changes in peak intensities of the high-mass extension primers after the adjustments of their amounts. The primer quality was acceptable as the observed masses matched expected the ones (The small deviations from the expected masses were due to systematic mass shift during MALDI-TOF MS analysis). The adjusted amounts of each extension primer are listed in **Table 12**. The changes in their amounts led to changes in their concentrations and T_ms. in a 14μL solution (the final volume for undergoing the primer extension reaction). The new average T_m (51^oC) was higher than the previous one ($46.3\ ^oC$). Using the two sets of adjusted PrimerMix, we found the genotypes of 30 patients and 43 donors (See section 4.3.6).

4.3.3 Visualization of the PCR Products by Gel Electrophoresis

An optimal competitive multiplex PCR was designed to amplify multiple targeted SNP loci of genomic DNA evenly and completely with minimum amount of PCR primers left and no non-specific amplified byproducts after the reaction. In our assay an eight-plex competitive PCR was designed to amplify eight targeted SNP loci on different chromosomes. **Figure 10** is a polyacrylamide gel image showing bands corresponding to

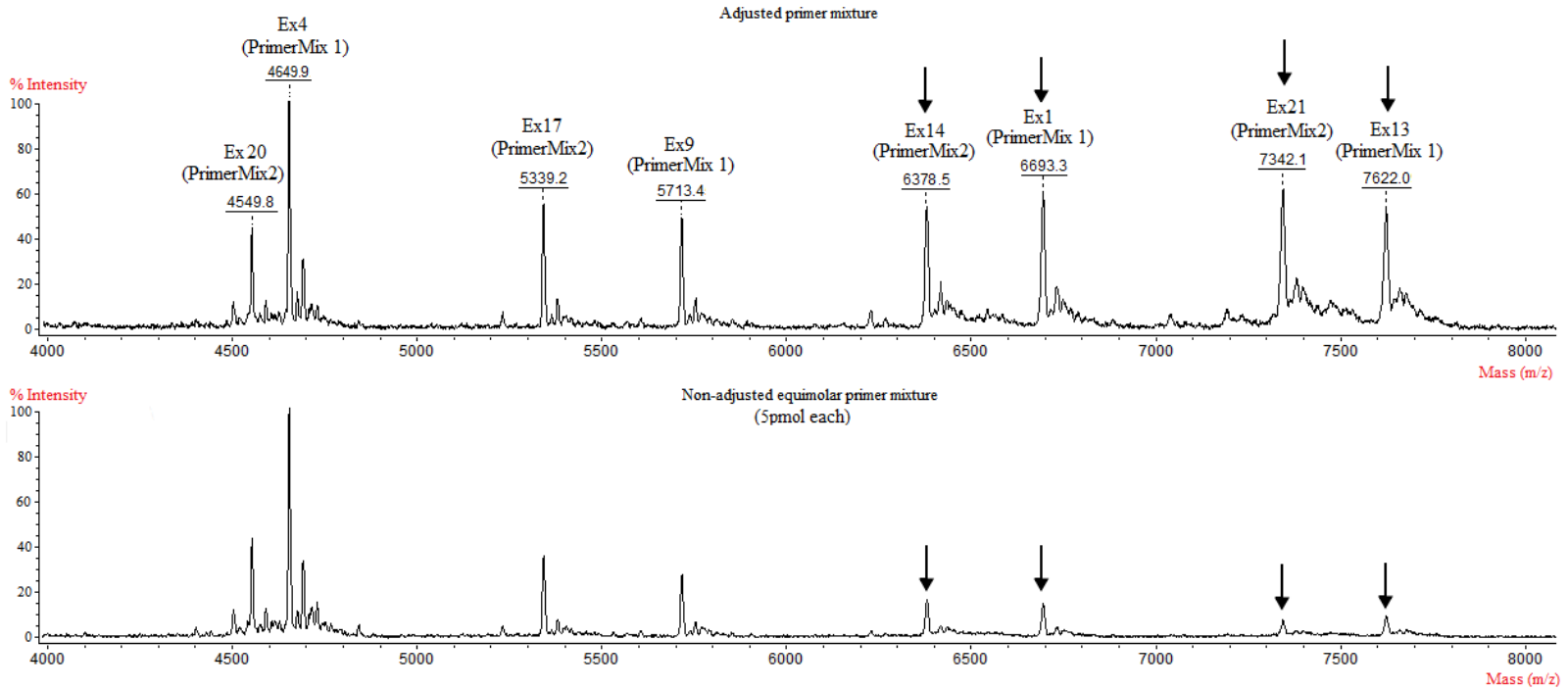


Figure 9. Mass spectra of extension primers with and without concentration adjustment. The observed m/z values and corresponding primer code of each extension primer in PrimerMix1 and PrimerMix2 are indicated. The arrows show the apparent peak intensity differences before and after the amount adjustments. The masses of the extension primers were calculated based on their m/z values:

Ex4: 4649.9–1 = 4648.9Da (expected: 4650.1Da)
Ex9: 5713.4–1 = 5712.4Da (expected: 5713.8Da)
Ex1: 6693.3–1 = 6692.3Da (expected: 6693.4Da)
Ex13: 7622.0–1 = 7621.0Da (expected: 7622.0Da)

Ex20: 4549.8–1 = 4548.8Da (expected: 4549.0Da)
Ex17: 5339.2–1 = 5338.2Da (expected: 5339.5Da)
Ex14: 6378.5–1 = 6377.5Da (expected: 6378.2Da)
Ex21: 7342.1–1 = 7341.1Da (expected: 7342.9Da)

Table 12. Adjusted concentrations and corresponding T_ms of each extension primer in the two PrimerMix setups for the primer extension reaction

PrimerMix	Primer Code	Original amount (pmol)	Adjusted amount (pmol)	Final concentration in a 14μL solution (nM)	Original T_m (^oC) (based on 50nM)	New T_m (^oC) (based on the adjusted concentration)	Average T_m of the new T_ms (^oC)
PrimerMix1	Ex4	5	5	357	38.2	42.0	50.5
	Ex9	5	7.5	536	47.1	51.2	
	Ex1	5	15	1071	48.1	53.1	
	Ex13	5	30	2143	50.2	55.6	
PrimerMix2	Ex20	5	5	357	38.5	42.4	51.5
	Ex17	5	7.5	536	48.8	52.6	
	Ex14	5	15	1071	49.0	54.1	
	Ex21	5	30	2143	50.9	56.8	

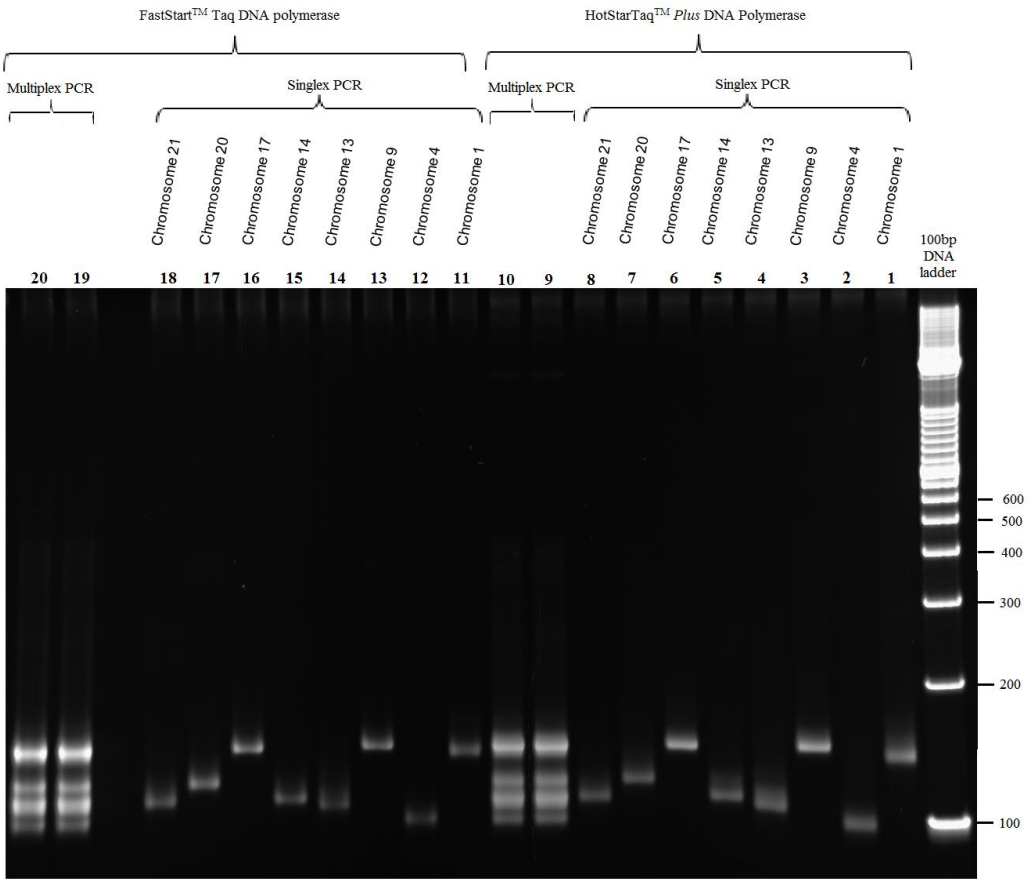


Figure 10. Polyacrylamide gel analysis of the PCR products from the setups using the HotStarTaq™ *Plus* DNA Polymerase and FastStart™ Taq DNA Polymerase. The expected sizes for each PCR product were listed as followings:
Lane 1 and 11 (Ex1): 138bp
Lane 2 and 12 (Ex4): 97bp
Lane 3 and 13 (Ex9): 144bp
Lane 4 and 14 (Ex13): 108bp
Lane 5 and 15 (Ex14): 109bp
Lane 6 and 16 (Ex17): 145bp
Lane 7 and 17 (Ex20): 122bp
Lane 8 and 18 (Ex21): 109bp

the PCR products generated in multiplex and singleplex setups using the HotStarTaq™ *Plus* DNA Polymerase and FastStart™ Taq DNA Polymerase. Single bands were observed in each singleplex PCR, indicating that specific amplification was obtained by both polymerases and all PCR primers were exhausted. The annealing temperature (60^{o}C) should have been optimal. The PCR products appeared as correctly sized fragments (smaller the size of PCR products, faster is the migration speed and lower is the band position). Because of the limited resolving power of the electrophoresis setup, products in the eight-plex competitive PCR could not be separated into 8 distinct bands, and only 4 bands were visualized. However, the intensities of the 4 bands (lanes 9, 10, 19 and 20 in **Figure 10**) were directly proportional to the number of PCR products with size differences < 7bp. All the first (top) bands had intensity values close to the summation of band intensities of the PCR products for Ex1 (138bp), Ex9 (144bp) and Ex17 (145bp), while the third bands had intensity values close to summation of band intensities for Ex13 (108bp), Ex14 (109bp) and Ex21 (109bp). The second and forth (bottom) bands had intensity values similar to the band intensities of the distinct PCR products for Ex20 (122bp) and Ex4 (97bp), respectively. No obvious differences were observed between the competitive PCR setups using the FastStart DNA polymerase and those using HotStar DNA polymerase The FastStart DNA polymerase was eventually used for our assay, for it aided in generating good-quality mass spectra (See section 4.3.4).

4.3.4 Comparison of the two DNA polymerases with Different PCR Buffer Systems in terms of MS Compatibility

One set of the extension primers (PimerMix1) was spiked into SAP-treated PCR products generated by using either FastStart or HotStar DNA polymerase and their corresponding buffer systems, and then analyzed by MALDI-TOF MS following an optimized procedure for sample clean-up (See section 4.3.5). The use of HotStar DNA polymerase PCR system generated a mass spectrum with higher background noise (**Figure 11**). For the sake of generating high-quality mass spectra, FastStart DNA polymerase PCR system was chosen for the subsequent development of a method for SNP genotyping of the pre-transplant samples from allogeneic HSCT donors and recipients and chimerism assessment of the post-transplant patient samples.

4.3.5 Optimization of the Sample Clean-up

After the primer extension reaction, sample clean-up is a must prior to the MS analysis, due to the presence of non-volatile salts and detergents in the reaction buffers. Three clean-up strategies were proposed and they were combined in different ways for investigation. **Figure 12** reveals two poor-quality mass spectra obtained by analysis of the extension products after ZipTip microextraction only and those after removal of alkali metal cations with cation exchange resin followed by ZipTip microextraction respectively. It seemed that some chemicals, which were most likely to be detergents, interfere the detection of the extension primers. To overcome this problem, ethanol precipitation was employed. **Figure 13A** reveals a mass spectrum obtained after ethanol precipitation followed by ZipTip microextraction. Due to the high salt concentration, a single extension primer appeared in form of multiple salt adduct peaks, instead of a single predominant peak. Moreover, the overall signal intensity was highly suppressed, compared to a high-quality mass spectrum obtained after processing with all three clean-

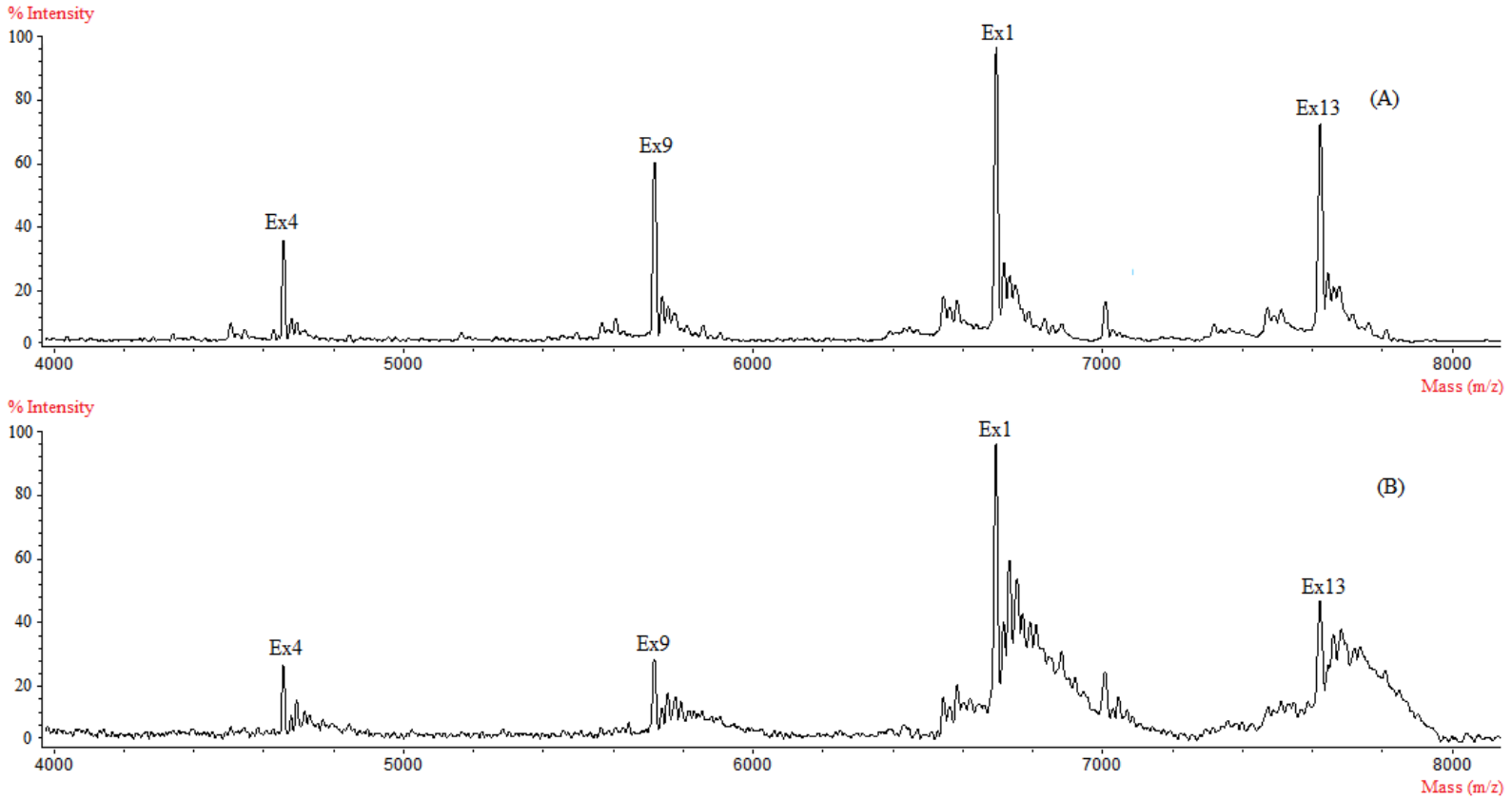


Figure 11. MALDI-TOF MS analysis of the extension primers (PrimerMix1) in the solutions containing **(A)** FastStart™ Taq DNA Polymerase and **(B)** HotStarTaq™ *Plus* DNA Polymerase

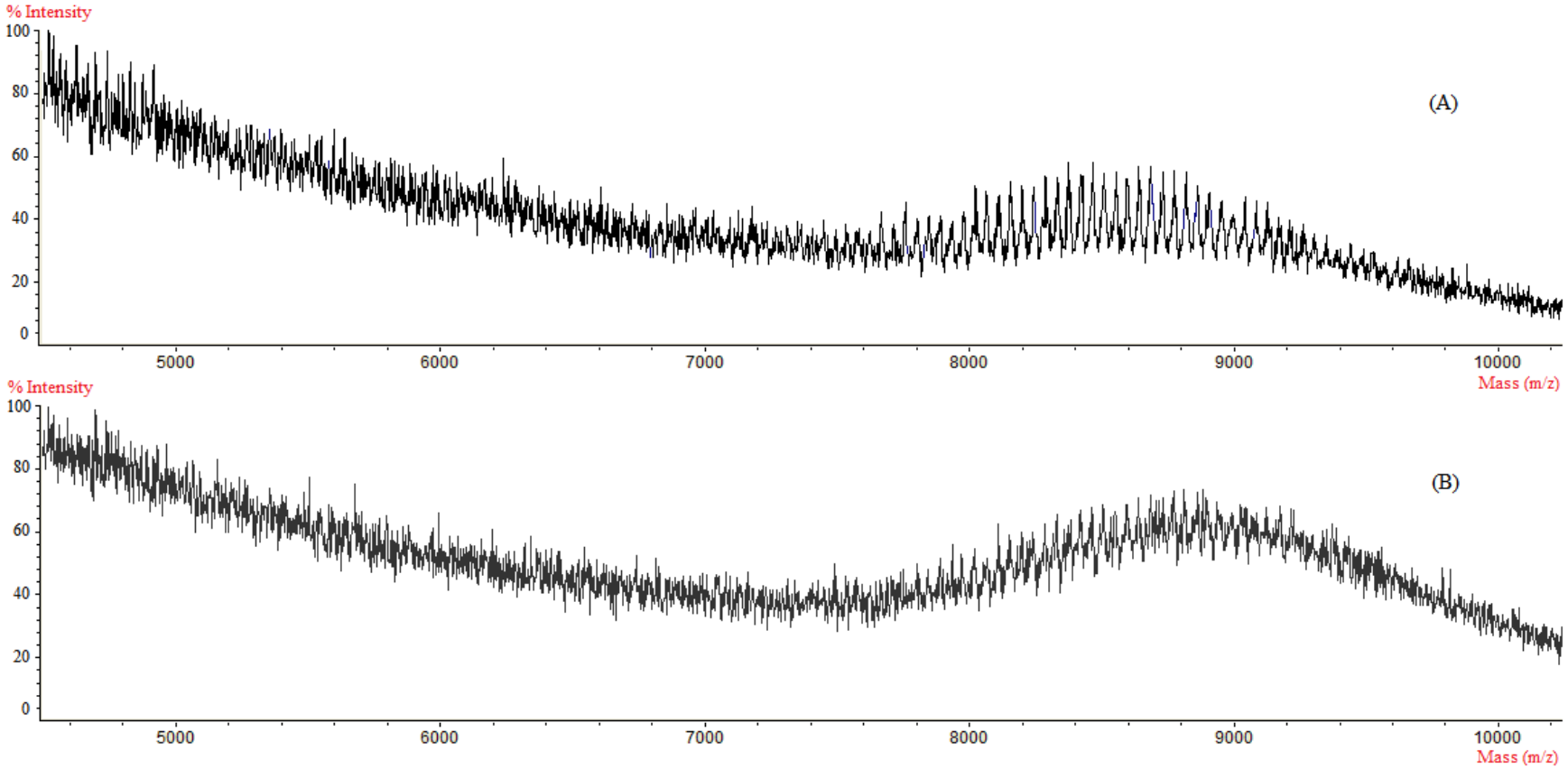


Figure 12. MALDI-TOF MS analysis of the extension products generated by PrimerMix1 after **(A)** ZipTip microextraction and **(B)** removal of alkali metal cations with cation exchange resin followed by ZipTip microextraction

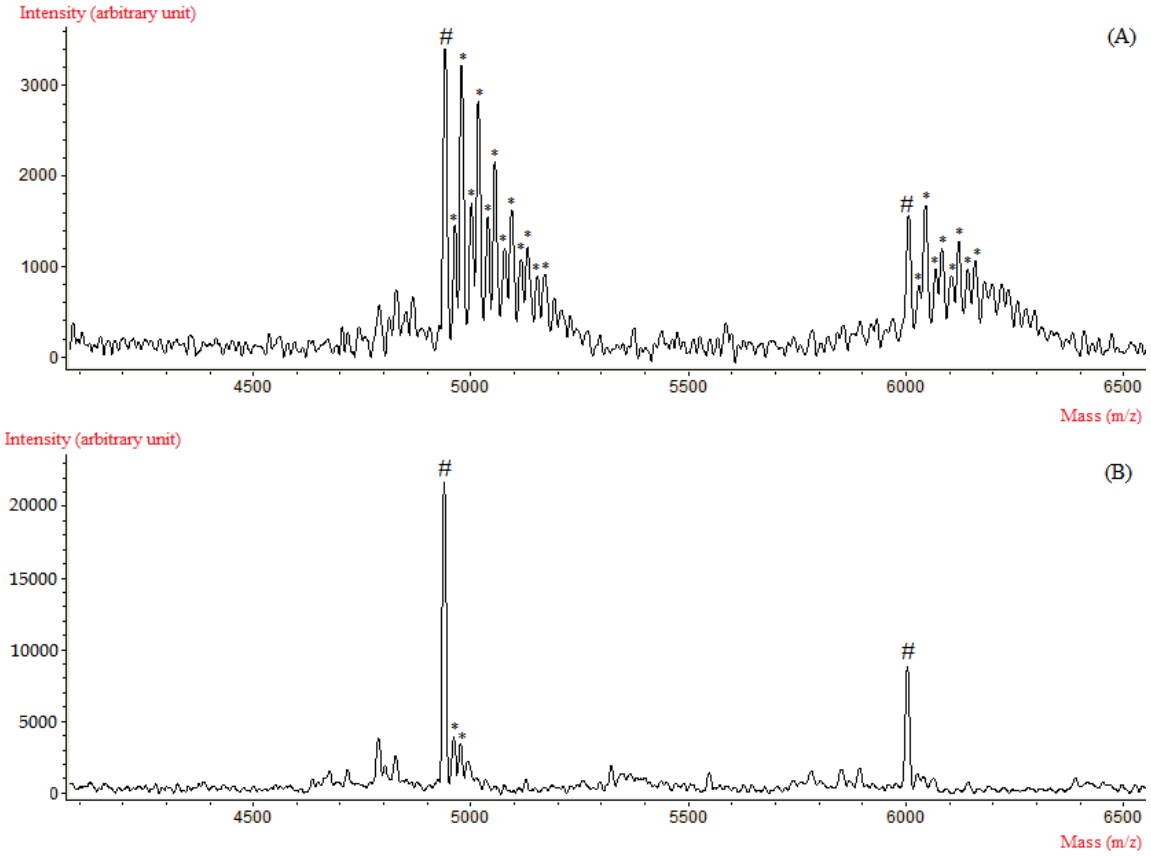


Figure 13. MALDI-TOF MS analysis of the extension products generated by two extension pirmers in PrimerMix1 (Ex4, Ex9) after **(A)** ethanol precipitation followed by ZipTip microextraction **(B)** ethanol precipitation and removal of alkali metal cations with cation exchange resins followed by ZipTip microextraction. The parent peaks and salt adduct peaks were marked as '#' and '*' respectively.

up procedures (**Figure13B**). In the latter case, the salt adduct peaks became much less significant. As a result, all three clean-up procedures were used in the subsequent qualitative (SNP genotyping) and quantitative (chimerism assessment) analyses.

4.3.6 SNP Genotyping HSCT Donors and Recipients Using the Developed In-house Assay

After the SNP genotyping assay was developed and optimized, it was used for genotyping 30 patients and 43 donors to find informative SNPs in individual donor-recipient pairs using the two adjusted PrimerMix. **Figure 14 and 15** show two typical SNP genotyping results of a pair of allogeneic HSCT donor and recipient, respectively. Genotypes of the eight targeted SNPs were marked in the figures. The observed masses of the allele-specific extension products were matched with their expected masses before the genotypes were confirmed (See **Table 11** in section 4.3.1). According to the genotyping results, two informative SNPs were identified for this donor-recipient pair, and they were rs891700 and rs2831700. Both were defined as informative because the donor was shown as a homozygote while the recipient was a heterozygote. The number of informative SNPs in the donor-recipient pairs was summarized previously (See **Figure 7** in section 3.3.2).

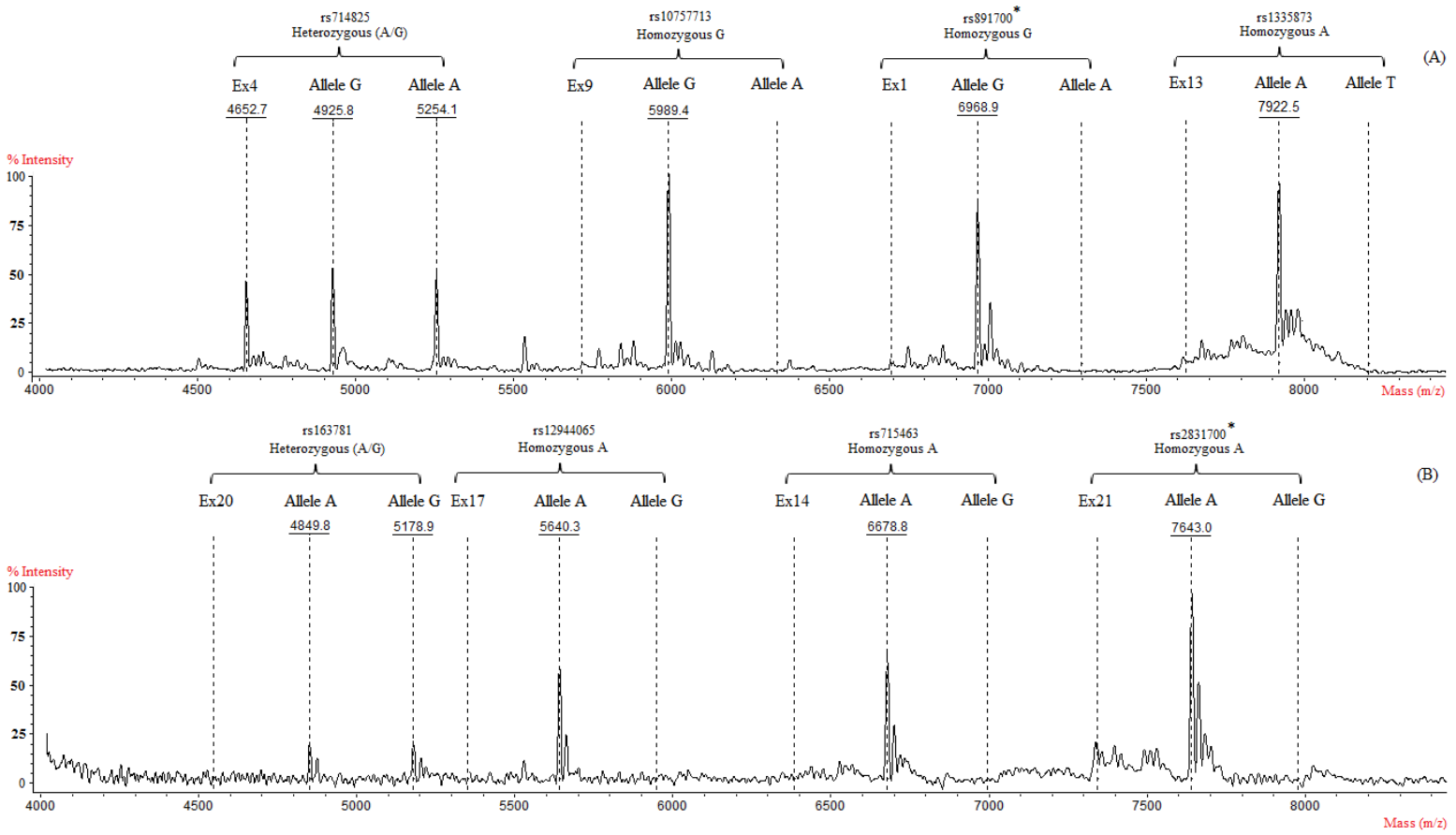


Figure 14. Typical genotyping result of a HSCT donor generated by the in-house primer extension assay. SNP genotypes of the donor were obtained by identifying the mass-specific extension primer products of **(A)** PrimerMix1 and **(B)** PrimerMix2 in the MALDI-TOF mass spectra. Genotypes of the eight target SNPs and the corresponding SNP IDs were indicated. The asterisk marked the informative SNPs, based on the genotype profile from the corresponding recipient (**Figure 15**).

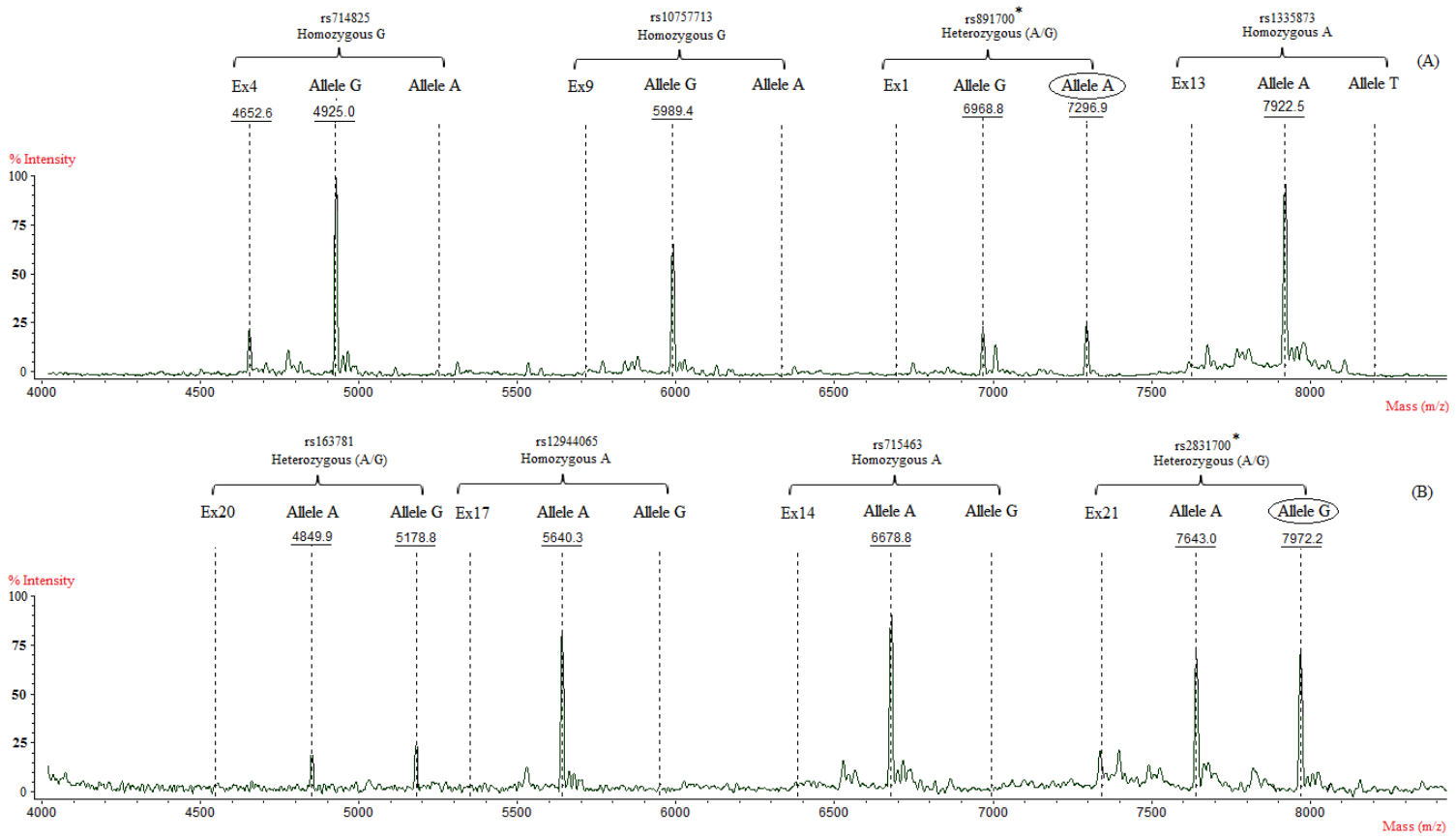


Figure 15. Typical genotyping result of a HSCT recipient generated by the in-house primer extension assay. SNP genotypes of the recipient were obtained by identifying the mass-specific extension primer products of **(A)** PrimerMix1 and **(B)** PrimerMix2 in the MALDI-TOF mass spectra. Genotypes of the eight targeted SNPs and the corresponding SNP IDs were indicated. The asterisk marked the informative SNPs, based on the genotype profile from the corresponding donor (**Figure 14)**.The reporting alleles, which were important indicators for chimerism analysis , were circled.

4.4 Discussion

The principle of our SNP genotyping assay is that using competitive PCR-amplified targeted SNP loci as a template, a genetically engineered DNA polymerase extends a primer that anneals to 1 base upstream of a targeted SNP site with a specific set of dNTPs/ddNTPs mixture, leading to formation of allele-specific extension products for MALDI-TOF MS analysis. After competitive PCR, any residual dNTPs are removed by SAP to prevent them from interfering the extension reaction. In the PCR amplicon when the DNA polymerase encounters a nucleobase which is complementary to one of the ddNTPs in the reaction solution, the extension reaction ceases. When there are two different SNP alleles, two extension primer products with different distinct masses will form. These two products are then resolved by MALDI-TOF MS, and appear as two separated signal peaks in a mass spectrum. In a multiplex assay, precise mass measurements by MALDI-TOF MS aid in identifying correct allele-specific extension products in a mass spectrum corresponding to a panel of targeted SNPs for multiple genotype assignment.

The PCR and extension primers were designed according to several criteria. A sequence tag was added to the 5'-ends of the PCR primers to improve PCR efficiency and, more importantly, increase their masses. It was to ensure that the signal peaks of the PCR primers would not interfere with those of extension primers in mass spectra. The mass separation between the extension primers and extension products was an important consideration. The peaks of salt adducts and fragmented primers might overlap with other analytical peaks. In general, oligonucleotide primers are susceptible to salt adduct formation due to their negatively charged backbones. A parent peak may be

accompanied with multiple adduct peaks. For instance, an additional peak with an additional mass of 39Da will appear when one K^+ ion is added to a primer. Other possible unwanted peaks come from fragmented primers. During MALDI-TOF MS analysis, depurination is a common fragmentation process[96] causing formation of an additional peak with a mass reduction of 134Da (adenine) or 150Da (guanine). In addition to elimination of purine bases, elimination of cytosine (111Da) was reported as well.[128] As a result, the extension primers were designed to have a minimum separation by 160Da. Hopefully, the signal peaks of the extension primer products would not overlap with unwanted peaks, and hence more quantitative results could be obtained from the signal intensities of extension primer products. Besides, a nucleotide T was introduced at the 5' ends of the custom primers as it would be more resistant to fragmentation during MALDI-TOF MS.[102] The theoretical masses of the extension primers and extension products are also important to provide confirmatory information that a specific target SNP allele has been identified, which is one of the advantages of the MADLI-TOF MS-based SNP genotyping assay.

In our assay the conditions for competitive multiplex PCR reactions with a moderate level of multiplexing and primer extension reactions were optimized. Two types of Taq DNA polymerases, that allowed "hotstart" amplification, were tested. FastStart™ Taq DNA Polymerase was chosen instead of HotStarTaq™ *Plus* DNA Polymerase, although the latter one has been is highly recommended by Sequenom™, Inc., a company that have developed several optimized SNP genotyping platforms.[115-117] The HotStarTaq™ *Plus* DNA Polymerase resulted in higher background noise in mass spectra, probably due to high concentrations of some MS-incompatible chemicals

(e.g. glycerol) in the enzyme stock solution and the corresponding Taq PCR buffer. Although recommended by the Sequenom^{TM} company, this polymerase seemed to be not compatible with our MS instrument. As a result, another tested DNA polymerase, FastStart^{TM} Taq DNA Polymerase, was chosen for the rest of experiments in this M.Phil. project. DNA polymerases function properly in the presence of free Mg^{2+}. Since dNTPs tend to bind Mg^{2+}, their concentrations in the PCR reaction must be balanced. It was found that the free Mg^{2+} concentration (total Mg^{2+} concentration minus total dNTPs concentration) should be kept within 1-2mM[118]. The free Mg^{2+} concentration used in this study was 1.1mM (total Mg^{2+} concentration = 1.5mM; total dNTPs concentration = 4×0.1 = 0.4mM). Another important consideration regarding dNTPs is that they are sensitive to frequent freeze-thaw cycles as this process may cause physical fragmentation to them, so small aliquots of dNTPs were prepared.[118] Any adjuvants or PCR enhancers (e.g. glycerol, BSA, Q-solution^{TM}) were excluded as they might adversely affect the subsequent MALDI-TOF MS analysis through signal suppression effects.[118, 129] After the competitive PCR, the only treatment was removal of unincorporated dNTPs by SAP. Since the PCR primers were not within the mass range of the extension primers, it was not necessary to remove the PCR primers prior to the extension reaction. Using different sets of the extension primers with the adjusted concentrations, the two primer extension reactions with a specific terminator mix were employed for each sample to be analyzed. The relatively long extension primers with poor ionization efficiency in MALDI-TOF MS[130] were spiked more into the extension reaction cocktail so that the equilibrated peak intensities could be produced. The T_ms of the adjusted extension primers were higher. In general T_ms are higher than the chosen

primer annealing temperature.[131] In our assay, although the pre-set annealing temperature (52^{o}C) was only similar to the average T_ms (50.5^{o}C and 51.5^{o}C in PrimerMix1 and 2 respectively), the performance of the primer extension reaction was acceptable, as shown in the examples of SNP genotyping outputs without formation of non-specific extension products (**Figure 14 and 15**). A proporion of one of the extension primer, Ex4, remained unextended. It might be due to its relatively lower T_m (38.2^{o}C). The high annealing temperature did not favor efficient binding. The primer extension reactions used three dNTPs and one ddNTP and terminated by the ddNTP. An engineered enzyme, Thermo Sequenase^{TM}, was preferentially used so that natural substrates dNTPs and synthesized substrates ddNTPs could be used with the same efficiency.[132]

Presence of non-volatile contaminants after the competitive multiplex PCR, SAP treatment and primer extension reaction definitely obstructed the MALDI-TOF MS analysis of the extension products. Sample clean-up became a bottleneck of high-throughput MS analysis of the extension products after the typical molecular biology reactions. The first issue to be considered was the sample volume, which should be small enough to be compatible for spotting on a MS target plate.[130] Therefore, ZipTip microextraction had to be used. It not only purified but also concentrated samples in a small volume. Actually, if ZipTips had not been used, the samples would have been concentrated by drying in air. However, it might also result in enrichment of the contaminants and dust contamination.[120] It was reported that dust particles are likely to be a common source of metal alkali ions.[133] The ZipTip microextraction was performed by initial wetting of a new tip by ACN with decreasing percentages followed by washing

with TEAA solution, which serves as a volatile ion-pairing reagent. This step could improve the purification yield.[120] The number of cycles of aspirating and dispensing during the sample binding step should be high enough to allow sufficient contact time of the sample solution with the stationary phase (C-18 beds).[120] Bound oligonucleotides were eluted by a small volume of concentrated organic solvent (50% ACN solution). A small volume of the eluant sitting on a target plate was dried rapidly so that the dust contamination could be minimized. The second issue to be considered was the presence of non-volatile metal cations. Salt adduct peaks would complicate the interpretation of mass spectra. They caused signal broadening with decreased sensitivity and difficulties in mass determination of the extension primers.[96] A well-established approach to solve this problem is removal of alkali metal cations with cation-exchange resin. The prepared resin was divided into small aliquots as frequent transfer of the resin might cause contamination with dust particles, which contain metal alkali cations.[120] The third issue to be considered was the presence of non-volatile MS-incompatible chemicals. Ethanol precipitation is a standard technique for DNA extraction and removal of residual salts and small organic molecules in solutions.[126] In the presence of monovalent cations such as NH_4^+, ethanol induces conformational transitions to the structure of DNA and precipitates it efficiently.[134] The precipitation process took place at low temperature (-20^{o}C) to increase recovery rates, especially when DNA concentration was low after addition of the ethanol (<0.25 mg/mL, in our assay it is around 7×10^{-4} mg/mL).[126] The DNA pellet was washed by 85% ethanol, which dissolved most salts and small organic molecules such as glycerol but not the precipitated DNA. Ammonium acetate was used rather than sodium acetate since the latter has low solubility in 85% ethanol.[126]

Stringent clean-up procedures are essential for many SNP genotyping platforms using MALDI-TOF MS. These platforms, including our SNP genotyping assay, benefit from obtaining high-quality mass spectra but lose automation and time efficiency.[101] All liquid handling steps, including liquid removal and transfer, during the ethanol precipitation were done manually. The manual removal of supernatants must be carefully performed to avoid disturbing the DNA pellet. With regard to the use of cation exchange resin, it was hard to achieve a reproducible preparation and result. For example, it was difficult to weigh a small amount of the prepared resin accurately and reproducibly. Sometimes samples were not thoroughly desalted even after mixing with resin for 1hr (data not shown). Although ZipTip microextraction was performed by an automated liquid handling robot, back pressure, caused by the reverse-phase material (C-18 beds), was a common issue when the tips were used in an automated mode. The liquid delivery accuracy might deteriorate and the tips might not be washed completely before and after the sample binding step.[120] For processing 30 samples by a single person, the time requirement was estimated to be 8-9 hours. This sample clean-up time was much longer than the pre-treatment time of the analysis of PCR products in an STR-PCR assay (This will be discussed in Chapter 5).

Originally an eight-plex primer extension reaction had been designed to genotype the eight targeted SNPs simultaneously. For the sake of accurate quantification, it was split into two quadruplex extension reactions using PrimerMix1 and 2 respectively due to prevention of peak overlapping between the parent peaks and salt adduct peaks/fragmented peaks. After the sample clean-up it seemed that those unwanted peaks were insignificant and did not mask the parent peaks during SNP genotyping, as shown

in the **Figure 14 and 15.** However, for convenient genotype calling of the eight SNPs, one still could combine the 8 assays into a single assay. In case of analysis of much more SNPs, the splitting of the primer extension reaction is not recommended since it also increases the laboriousness of the entire assay. Careful primer design with help of commercially available software (e.g. Assay Design[117]) can raise the multiplex level of an assay. The resulting spectra complexity will be higher and the extension products for a particular SNP must be picked carefully. However, in case of quantification of MC, which will be discussed in Chapter 5, lower plex-levels are preferred as higher analyte concentrations may cause signal suppression during MALDI-TOF MS analysis.[135] Sufficient peak signals must be provided for the accurate quantification.

The major limitation of our developed assay was the lack of automation in the sample clean-up procedures. The multiplex level of the assay we used was also smaller than the documented ones, such as 40-plex by iPLEX™[115]. In future studies, for the sake of automation, the clean-up procedures should be further optimized. When more SNP markers are included in the panel, qualitative SNP genotyping in higher multiplex level should be evaluated to see if our assay can achieve the multiplex level similar to the commercial ones.

4.5 Conclusion

Coupled with primer extension reactions, an in-house MALDI-TOF MS-based SNP genotyping assay was designed and optimized. FastStart™ Taq DNA Polymerase was used in competitive multiplex PCR. Three clean-up strategies, ethanol precipitation, cation-exchange chromatography and reverse-phase chromatography, were employed for sample clean-up prior to MALDI-TOF MS analysis of extension products. This assay was successfully utilized for SNP genotyping of pre-transplant patients and donors. For chimerism analysis, informative SNP markers were observed in over 90% of donor-recipient pairs. This assay would be further applied in quantification of MC in transplanted patients.

CHAPTER 5: SECTION 3 – COMPARISON BETWEEN CONVENTIONAL STR-PCR ASSAY AND THE DEVELOPED QUANTITATIVE MALDI-TOF MS-BASED SNP GENOTYPING ASSAY IN DETERMINATION OF MIXED CHIMERISM (MC)

5.1 Introduction

Currently, analysis of STR polymorphism is a routine method for chimerism analysis after allogeneic HSCT. STRs are repeating DNA sequences that are widely dispersed and highly polymorphic in the human genome. At least 8000 STR loci (Center of Medical Genetics, www.geneticstesting.com) have been identified. Since each STR locus has multiple alleles that are different in the repeated number of core sequences, there exist over 100,000 informative STR patterns between two individuals by estimation.[136] Thanks to the multiallelic nature of STRs, they have high informativity. Even a single STR marker, such as D1S80 gene locus, is informative enough to discriminate 60%-90% of donor-recipient pairs.[26, 75, 137] Use of several STR markers can theoretically discriminate more than 99% of donor-recipient pairs. Koldehoff *et al* reported that all recipient-donor pairs could be discriminated by using 11 STR gene loci.[32] Being used in our laboratories, PCR-based analysis of STR with use of fluorescent primers is a fast and robust method for chimerism analysis. The principle of the technique is to generate PCR amplicons with different sizes by amplification of an informative STR locus followed by separation of them by capillary electrophoresis. When two individuals contain different numbers of repeated core sequences of a STR marker, this variation leads to formation of PCR products with different sizes and allows distinction of the two individuals by visualizing two separated products.[71]

Commercially several optimized kits for STR-PCR assay are available.[138] Nevertheless, the number of targeted STR loci is limited.[136]

The STR-PCR, however, faces several limitations when monitoring allogeneic HSCT. Two of which were frequently reported. First, PCR amplification bias between short and long alleles is known to occur in the STR-PCR assay. Because amplification efficiency of shorter DNA templates is generally higher than longer ones[139], shorter alleles of an informative STR seem to be amplified preferentially. This property affects the sensitivity of the STR-PCR assay, especially when the size differences between donor- and recipient-specific alleles are large. If reporting recipient-specific alleles are longer than donor ones, the signal (peak area) for the recipient-specific alleles will be lower than the expected and the resulting sensitivity will be lower as well. This phenomenon was demonstrated by Mehdi *et al*, showing underestimation of recipient cell proportions in artificial chimeric DNA mixtures with known chimerism percentages.[81] Therefore, selection of informative STRs in which recipient-specific alleles are shorter than donor ones is a critical step to increase the assay sensitivity due to the preferential amplification of short templates.[72] Second, impairing the accuracy of the assay, stutter peaks may be produced due to 'slippery' amplification of STR regions. During PCR amplification of repetitive sequences like STRs, DNA polymerase may 'slip' off the template by looping out one repeat unit, generating stutter fragments or peaks that are one repeat unit smaller than the parent peaks.[140] When a heterozygous STR with alleles differing by one repeat unit is genotyped, the smaller allele may have higher signal intensity, for its allelic signal is added by stutter signals from the larger allele. There exists a risk of SNP mistyping that a homozygous STR is regarded as a

heterozygote.[96] This problem is more important during chimerism analysis. When recipient-specific alleles are smaller than donor ones by one repeat unit of informative STRs, appearance of peaks that are one-unit smaller than donor peaks may be attributable to presence of recipient cells in the background of donor cells (existence of MC) or formation of stutter peaks only (existence of CC). Misinterpretation of the result may occur (**Figure 16**). Therefore, the difference of length between the donor and recipient alleles should be greater than one repeat unit.[73]

SNP genotyping has emerged as a reliable tool for human identification. It is advantageous over STR genotyping in several aspects. First, a huge number of SNPs in human genome potentially permits the development of a large panel of genetic markers for identification of almost all related or unrelated donor-recipient pairs.[141] Although SNPs are a much simpler polymorphism than STRs, a SNP marker is less informative than a STR marker. The discrimination rates of one SNP locus (20-25%) generally are lower than those of one STR locus (60-90%).[26, 75, 137] However, this problem can be compensated by genotyping multiple SNPs to increase the panel informativity. Second, SNPs has lower mutation rates than STRs. Due to the smaller size of SNPs, they are less likely to mutate between generations. The result interpretation of SNP genotyping is more robust due to the stability of inheritance. This is an advantage for population genetic studies.[142, 143] In contrast, longer STR loci may vary in size from generation to generation as they are more prone to mutation.[144] Third, STRs typically occur in non-coding intron regions of the human genome[145] while SNPs are not limited to the non-coding regions.[75] Inclusion of SNPs encoding minor histocompatibility antigens (mHAs) can provide additional information that predicts transplant outcomes like

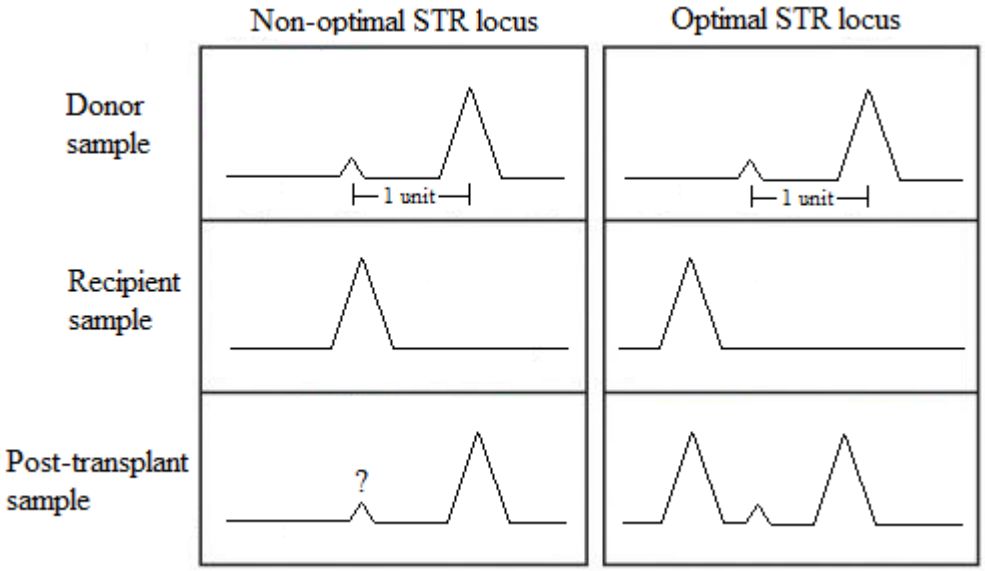


Figure 16. Schematic representation showing how stutter peaks affect the data interpretation of the STR-PCR assay

GVHD.[80] When donor immunocompetent cells encounter foreign mHAs, immune responses may be triggered. The presence of disparities of informative mHA-encoding SNPs between donors and recipients may lead to immune responses and even GVHD.[146] Therefore, genotyping clinically informative SNPs can not only assesses chimerism status but also predict the clinical outcomes of allogeneic HSCT.

Various platforms for SNP genotyping have been developed.[80-82] As mentioned in Chapter 4, we have developed an in-house MALDI-TOF MS-based SNP genotyping assay with aid of primer extension reactions. Its accuracy in terms of the precise mass measurement of the extension products had been evaluated in Chapter 4. It was also essential to evaluate the detection limit of our developed assay since it gave an idea about the minimum detectable quantity of recipient cells in patients after allogeneic HSCT. Accumulating results of primer extension- and MALDI-TOF MS-based discrimination of mutant-type DNA in a background of wild-type DNA were reported. Non-invasive prenatal diagnosis was able to detect as little as 5% of fetal point-mutated DNA in a background of maternal wild-type DNA.[147] Another study revealed that down to 5% of mitochondrial DNA with rare SNPs in a background of normal mitochondrial DNA could be detected reproducibly. An occasional detection of 2.5% of the mutated mitochondrial DNA, shown as a mini-peak in mass spectra, was reported also.[148] Besides detection limit, the assay linearity was evaluated by finding the relationship between the expected percentages of MC and the observed percentages in a set of artificial chimeric DNA samples.

There are two objectives in this last part of my M.Phil. research project. It was first aimed to evaluate the linearity, detection limit, precision and accuracy of the developed SNP genotyping assay. Secondly, practical values of the developed assay in quantification of MC were evaluated by using post-transplant DNA samples from paediatric patients and by comparing to the data obtained by a well-established in-house fluorescent-based STR-PCR assay.

5.2 Materials and Methods

5.2.1 Workflow of the In-House Fluorescent-Based STR-PCR Assay

The fluorescent-based STR-PCR was performed routinely for chimerism analysis in our laboratory. Using the purified DNA samples from the studied subjects (the purification method was mentioned in section 4.2.1), the PCR was performed in a final volume of 30μL with the use of the fluorescently-labeled PCR primers specific for amplification of informative STR systems, which depends on particular donor-recipient pairs. 100ng of the template DNA was used for each assay. Final concentrations or amounts of other PCR reagents in a PCR Master Mix were: 1units of FastStart^{TM} Taq DNA Polymerase (Roche Molecular Biochemicals^{TM}, Mannheim, Germany), 1X Taq PCR buffer, 0.2mM of dNTPs (GE Healthcare^{TM}, Freiburg, Germany), 1.5mM $MgCl_2$, and 0.13μM of each forward and reverse primer. Carrying out in a DNA Engine^{TM} Peltier Thermal Cycler (MJ Research^{TM}, South San Francisco, USA), thermal cycling conditions to the PCR were 94^{o}C for 7min for initial denaturation, followed by repeating thermocycles 28 times (denaturation at 94^{o}C for 45sec, annealing at 58^{o}C for 45sec and extension at 72^{o}C for 45sec), and final incubation at 72^{o}C for 20min for final extension. After the cycling programme, PCR products were stored at 4^{o}C until they were analyzed by electrophoresis. For electrophoresis analysis of the PCR products, a portion of the PCR products (2μL) was mixed with 8μL of formamide (Applied Biosystems^{TM}, Foster City, USA) and 1μL of GS500 size standard (Applied Biosystems^{TM}, Foster City, USA). The resulting specimen (11μL) was denatured at 95^{o}C for 5min followed by chilling for 5min in ice. After the specimen was added to a sample cassette, it was placed in an ABI

PRISM 3130 DNA sequencer (Applied Biosystems™, Foster City, USA) for separation of the PCR products by capillary electrophoresis and fluorescent detection. They were run on a POP-7™ polymer (Applied Biosystems™, Foster City, USA) for 2hr. The gel was analyzed by software GeneMapper™ Analysis (version 4.0; Applied Biosystems™).

5.2.2 Quantification of MC by the STR-PCR Assay

Percentages of MC were determined using peak areas obtained from electropherograms. For each examined donor-recipient pair, the pre-transplant patients and donors were genotyped for a particular STR marker. Different sized peaks shown in the electropherograms represented different alleles of the STR loci belonging to recipients or donors. Informative STR markers were identified when unique sized peaks are present in recipients only (reporting alleles). Using these informative markers, the post-transplant patient samples were assessed for chimerism percentages. For an informative STR marker, there are several informative allelic constellations, three of which were commonly found and each required a specific calculation approach for quantification of MC. They were shown as follows:

A. Recipients and donors were either homozygous or heterozygous and they had no shared alleles: (**Figure 17A**)

This was the simplest constellation. In this case, the sum of peak areas for recipients' alleles was divided by the total sum of peak areas for recipients' and donors' alleles:

$$\%\ \text{of recipient cells} = \frac{R_T\ (\text{or } R_A + R_B)}{R_T + D_T\ (\text{or } R_A + R_B + D_A + D_B \text{ or } R_A + R_B + D_T \text{ or } R_T + D_A + D_B)} \times 100\%$$

B. Recipients were heterozygous and donors were homozygous and they shared one common allele: (**Figure 17B)**

The reporting allele was counted twice in the calculation. In this case, the twofold peak area of the reporting allele was divided by the sum of peak areas for donors' alleles and another recipients' allele.

$$\% \text{ of recipient cells} = \frac{2R_B}{R_A + R_B + D_T} \times 100\%$$

C. Recipients and donors were heterozygous and they shared one common allele: (**Figure 17C)**

The common allele was ignored. In this case, the peak area of the reporting allele was divided by the total sum of peak areas for the reporting allele and donors' non-common alleles.

$$\% \text{ of recipient cells} = \frac{R_B}{R_B + D_A} \times 100\%$$

5.2.3 Construction of an Equation for Correction of Electronic Background Noise in MALDI-TOF MS Analysis

After the establishment of the in-house MALDI-TOF MS-based SNP genotyping assay, an equation was constructed to correct electronic noise that might be present in following chimerism analysis of real patient samples. Peripheral blood from two unrelated volunteers, one of which was named 'donor' while another was 'recipient', was collected. Their genomic DNA was purified by the previously mentioned method (See section 4.2.1). Informative SNP markers were then identified between the two individuals, and one of which was selected. 15pmol of the corresponding extension primer was used for the developed SNP genotyping assay. A set of artificial chimeric

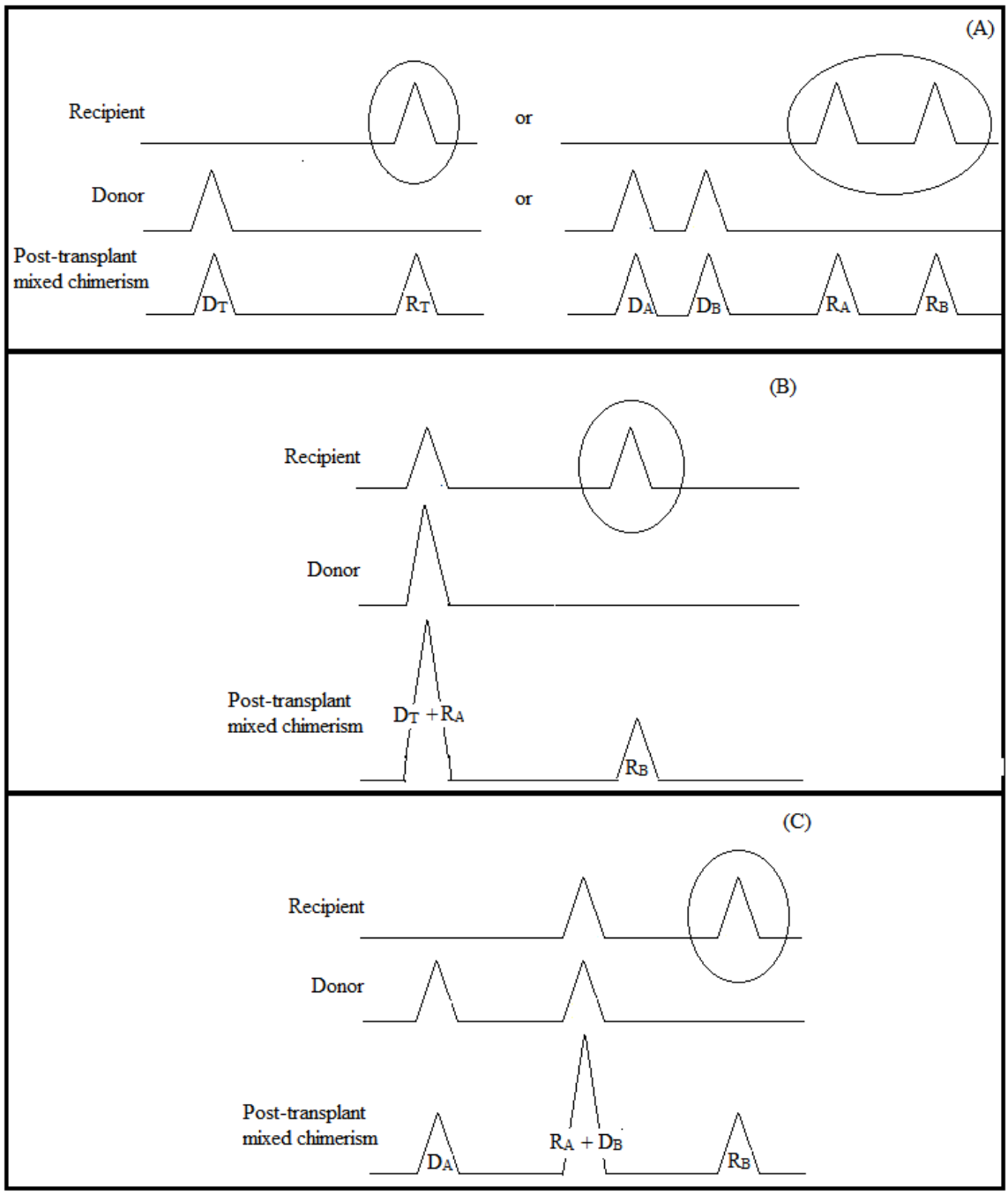


Figure 17. Schematic representation of three different informative allelic constellations for an STR-PCR assay after allogeneic HSCT. Peaks corresponding to informative alleles were circled.

R_A: peak area for allele A in recipients
R_B: peak area for allele B in recipients
R_T: total peak area in homozygous recipients
D_A: peak area for allele A in donors
D_B: peak area for allele B in donors
D_T: total peak area in homozygous donors

DNA samples with known percentages of 'recipient' cells in a background of 'donor' cells was prepared. DNA samples from the 'recipient' were added to the 'donor' ones, resulting in seven chimeric DNA mixtures covering a predicted sensitivity range of the developed SNP genotyping assay: 90%, 45%, 22.5%, 4.5%, 2.25%, 0.9% and 0%. Each standard sample was analyzed in duplicate. The observed relative peak area for a reporting allele (Relative Peak $Area_{obs}$, %) was calculated by dividing its absolute peak area by peak area for another allele (equation 1).

$$\text{Relative Peak Area}_{obs}(\%) = \frac{\text{Peak area for a reporting allele}}{\text{Peak area for a non-reporting allele}} \times 100\% \text{ — equation 1}$$

The observed relative peak area was plotted against the expected relative peak area for the reporting allele, which was calculated based on the expected percentages of MC and genotype profiles of the 'recipient' and the 'donor' (This will be calculated in section 5.3.1). An equation for correction of electronic noise was constructed by fitting the data points with the quadratic function.

5.2.4 Evaluation of Accuracy and Precision of the Developed Assay

It is important to determine the minimum detectable amount of recipient cells in donor circulations. It was reported that 5% of chimerism was commonly detected by using MALDI-TOF MS[147, 148] and 2.5% was rarely detected.[148] In order to figure out the detection limit of the assay at low levels of MC, 2.25% and 4.5% of artificial chimeric DNA samples were prepared by the same method mentioned above. The developed SNP genotyping assay was performed 10 times for each chimeric sample. In addition, it was also performed 10 times for artificial chimeric DNA samples with 67.5% recipient cells to test the assay could accurately measure high levels of MC.

5.2.5 Quantification Formulae for the Developed Assay

An SNP marker was defined as informative when a recipient had a SNP allele (reporting allele) that the donor did not have. According to the definition, there are six possible informative combinations of donor and recipient genotypes (See **Table 6 or 7** in Section 3.3.1). For each informative biallelic SNP marker, post-transplant patient samples might present two different informative allelic constellations. For both constellations, the percentage of recipient cells in the post-transplant patient samples was calculated based on the peak area fraction of a reporting allele that distinguished the donor and the recipient in each recruited donor-recipient pair. Therefore, Relative Peak $Area_{obs}$ for each examined donor-recipient pair was initially calculated (equation 1). The Relative Peak $Area_{obs}$ was corrected by the equation of the correction curve to obtain an expected relative peak area (Relative Peak $Area_{exp}$, %) (equation 2).

f(Relative Peak $Area_{obs}$, %) = Relative Peak $Area_{exp}$, % — equation 2

Then, each constellation required a specific calculation approach for quantification of MC. They were shown as follows:

A. Recipients are heterozygous and donors are homozygous: (**Figure 18A)**

$$\text{Calculated \% of recipient cells} = \frac{2 \times \text{Relative Peak Area}_{exp}(\%)}{\text{Relative Peak Area}_{exp}(\%) + 100} \times 100\% \text{ — equation 3}$$

B. Recipients and donors are disparate homozygous: (**Figure 18B)**

$$\text{Calculated \% of recipient cells} = \frac{\text{Relative Peak Area}_{exp}(\%)}{\text{Relative Peak Area}_{exp}(\%) + 100} \times 100\% \text{ — equation 4}$$

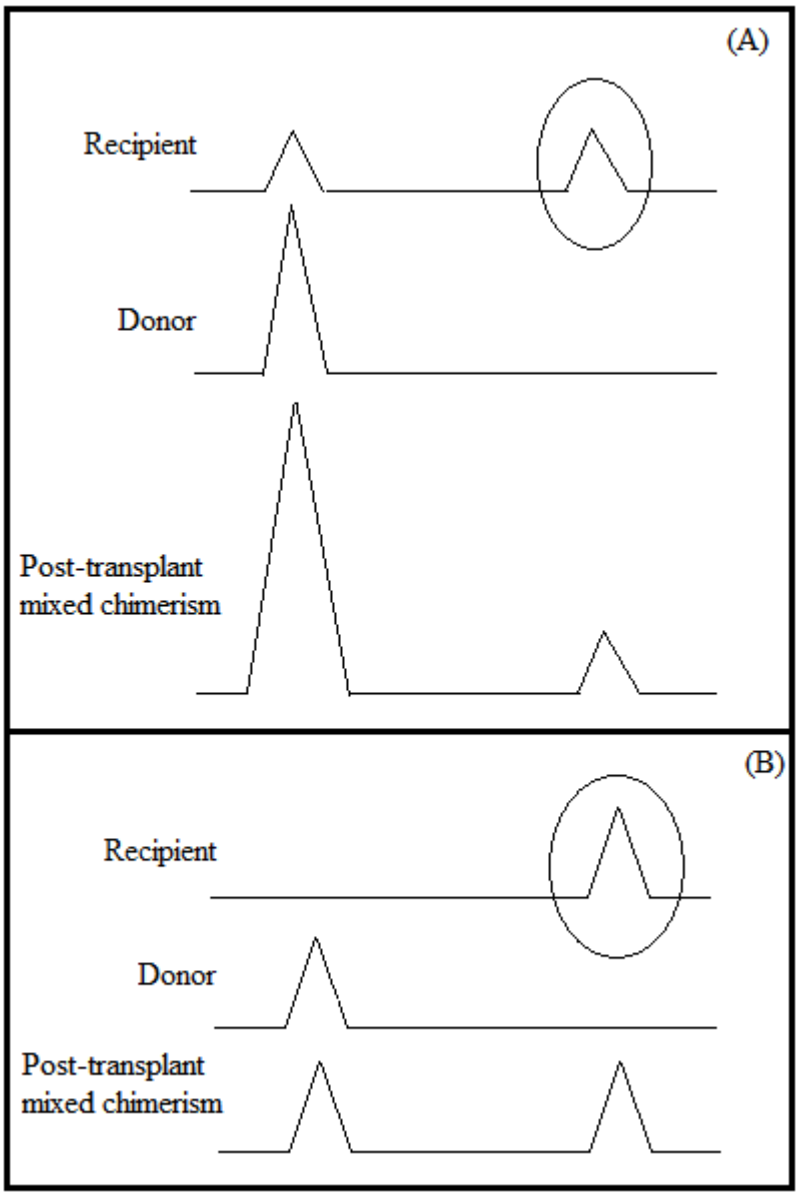


Figure 18. Schematic representation of two different informative allelic constellations for the developed assay after allogeneic HSCT. Peaks corresponding to informative alleles were circled.

5.2.6 Quantification of MC in the Post-Transplant Patients

In order to test the feasibility of the developed assay to quantify MC in clinical specimens, we determined the level of recipient cells in post-transplant patient samples which had been analyzed by the routine STR-PCR assay. After the genotype profiles of all studied subjects were obtained and informative SNPs were identified in each case (n=30), one of the informative SNPs that yielded the smallest primer extension products and were mostly common among the cases were selected for the following chimerism for each case. 15pmol of the corresponding extension primer was used for the developed SNP genotyping assay. For the cases (n=13) with double cord blood transplantation, the post-transplant samples were genotyped first to determine which donor was dominant and which SNPs were informative. Among the cases with more than one informative SNP, two cases were selected and examined by two informative SNPs to verify the consistency of the assay in MC quantification. Results obtained from the two types of assays were compared by paired T-test. Pearson correlation analysis was performed to compare the percentages of recipient cells obtained by the developed SNP genotyping assay with that found by the STR-PCR assay to see if there was a correlation between them. Among the 30 cases, three cases with longitudinal post-transplant samples from different time points after allogeneic HSCT were analyzed to check if there was a trend of changing percentages of recipient cells similar to that found by the STR-PCR assay. All statistical analyses were performed by using the statistical software SPSS (version 21.0, IBM, Chicago, USA).

5.3 Results

5.3.1 Linearity of Quantitative Analysis of MC

DNA materials were extracted from EDTA-whole blood samples collected from two healthy volunteers, and mixed at different percentages to simulate the situation of MC. In this "artificial" donor-recipient pair, SNP rs714825 was found to an informative SNP, in which the recipient was heterozygous and the donor was homozygous. The expected percentages of MC were converted into the expected peak areas for the reporting SNP allele relative to peak area of the corresponding non-informative allele (Relative Peak $Area_{exp}$, %) (equation 5).

$$\text{Relative Peak Area}_{exp}(\%) = \frac{\text{Expected \%}}{\text{Expected \%} + (100 - \text{Expected \%}) \times 2} \quad \text{— equation 5}$$

The observed area of the informative peak for "recipient" DNA (relative to the peak area for the corresponding non-informative peak) was directly and linearly proportional to the expected percentage with a slope of close to 1.00 ($R^2 > 0.99$, **Figure 19)**. However, the Y-axis interception did not pass through the origin, indicating the presence of measurement bias at low relative peak area of the reporting SNP allele. Detailed examination of signal intensities of the reporting SNP allele revealed the presence of positive signal bias, which was due to the presence of high background electronic noise. This was overcome by deriving a quadratic equation from the observed and theoretical relative peak areas to correct for the electronic noise (**Figure 20**). The correction equation was applied to quantitative analysis in the subsequent experiments.

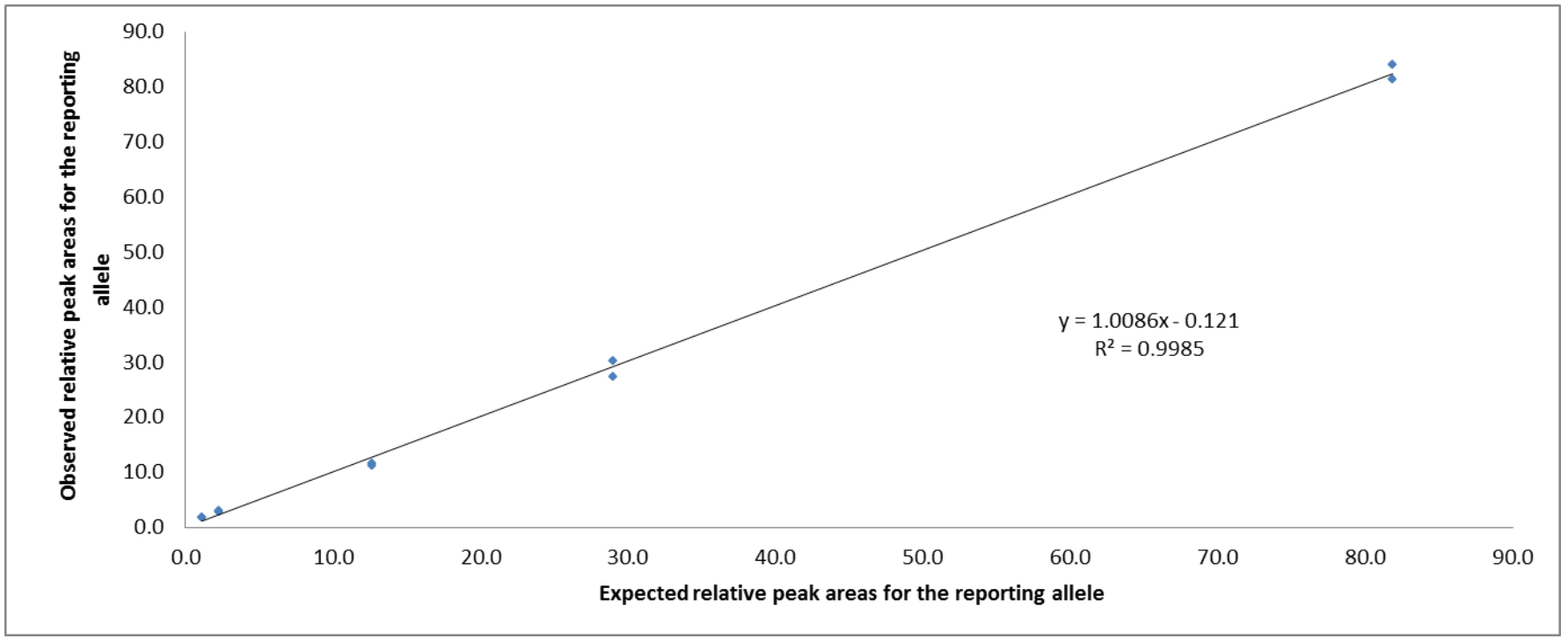


Figure 19. Relationship between the observed relative peak area and the expected relative peak area of a reporting informative SNP allele (Points for 0% and 0.9% chimerism are excluded)

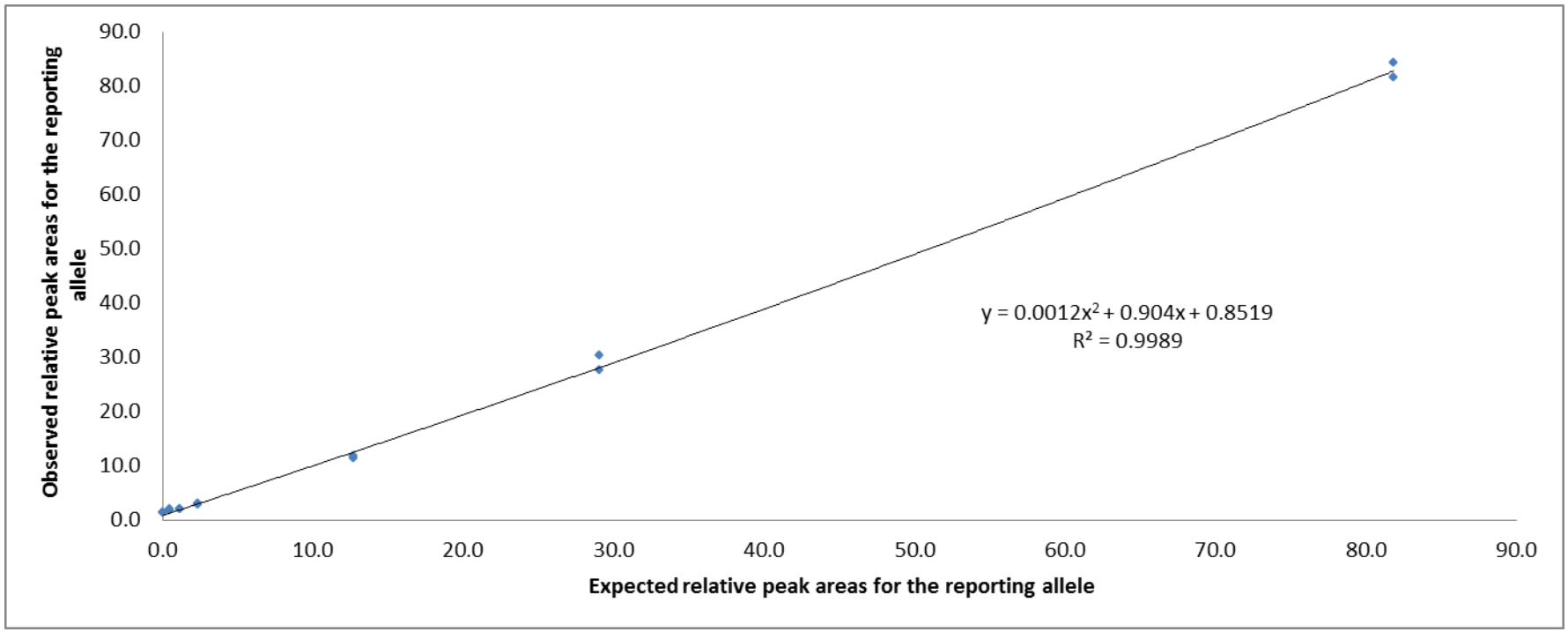


Figure 20. Relationship between the observed relative informative peak area and the expected relative informative peak area

5.3.2 Accuracy, Precision and Detection Limit of the Developed Quantitative SNP Genotyping Assay

The developed SNP genotyping assay reliably measured the percentage of recipient DNA in a donor-recipient DNA mixture in the assessed range of 2.25 – 67.5%. The relative error was 11 to 13%, whereas the intra-assay CV was 4 to 16%. The degree of accuracy seems to be fairly consistent in the assessed range. Precision improved when the percentage of recipient DNA increased. The results of the quality control samples are summarized in the **Table 13**. The limit of detection and limit of quantification were 1.4% and 4.6%, respectively.

5.3.3 Application of the Developed Assay to Measurement of MC in the Post-Transplant Patients

Excluding the cases without informative SNPs identified, we examined 25 post-transplant patient samples by one of the informative SNPs in each case, except for two cases examined by two informative SNPs. The percentages of MC measured were compared with the percentages previously obtained by the STR-PCR assay. Typical examples of the quantitative outputs by the developed SNP genotyping assay with different MC percentages were shown in **Figure 21**. **Table 14** summarizes the results of MC quantification and the corresponding STR and SNP markers for each case were indicated. Paired t-test showed that measurements from our quantitative SNP genotyping assay were not significantly different from the results of the STR assays. Pearson correlation analysis further indicated that measurements from these two types of markers were significantly and highly correlated with each other (correlation coefficient = 0.997,

Table 13. Precision and accuracy of the developed SNP genotyping assay in quantitative analysis of quality control samples. For each quality control sample, ten independent measurements were performed within one assay.

Percentage of Recipient DNA		Intra-assay CV (%)	Relative error (%)
Expected value (%)	Measured value, Mean ± SEM (%)		
0	0.1 ± 0.13	N.A.	N.A.
2.25	2.6 ± 0.13	16	13
4.5	5.0 ± 0.22	13	11
67.5	75.5 ± 0.95	4	12

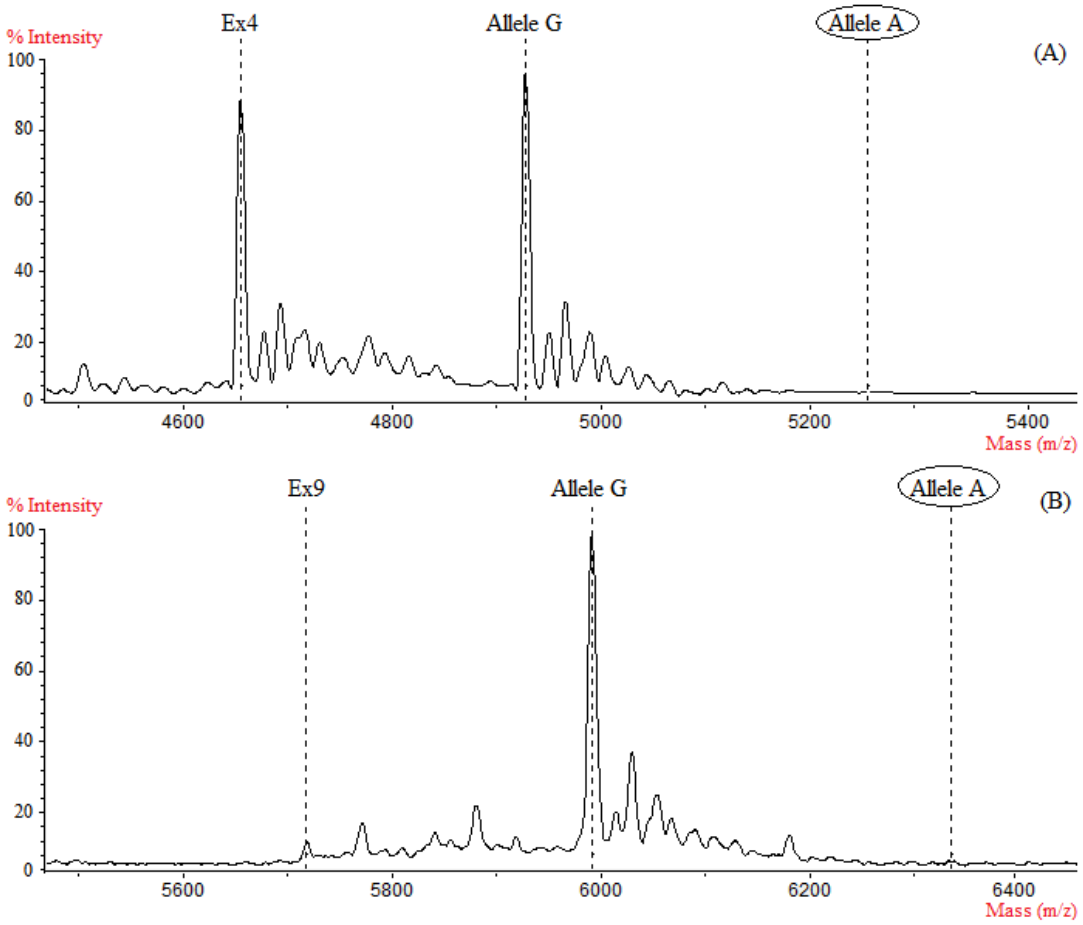


Figure 21. Examples of the quantitative outputs by the developed SNP genotyping assay The codes for particular extension primers were indicated and reporting alleles were circled. The percentages of MC in **(A)** and **(B)** were 0% since no reporting peaks could be detected.

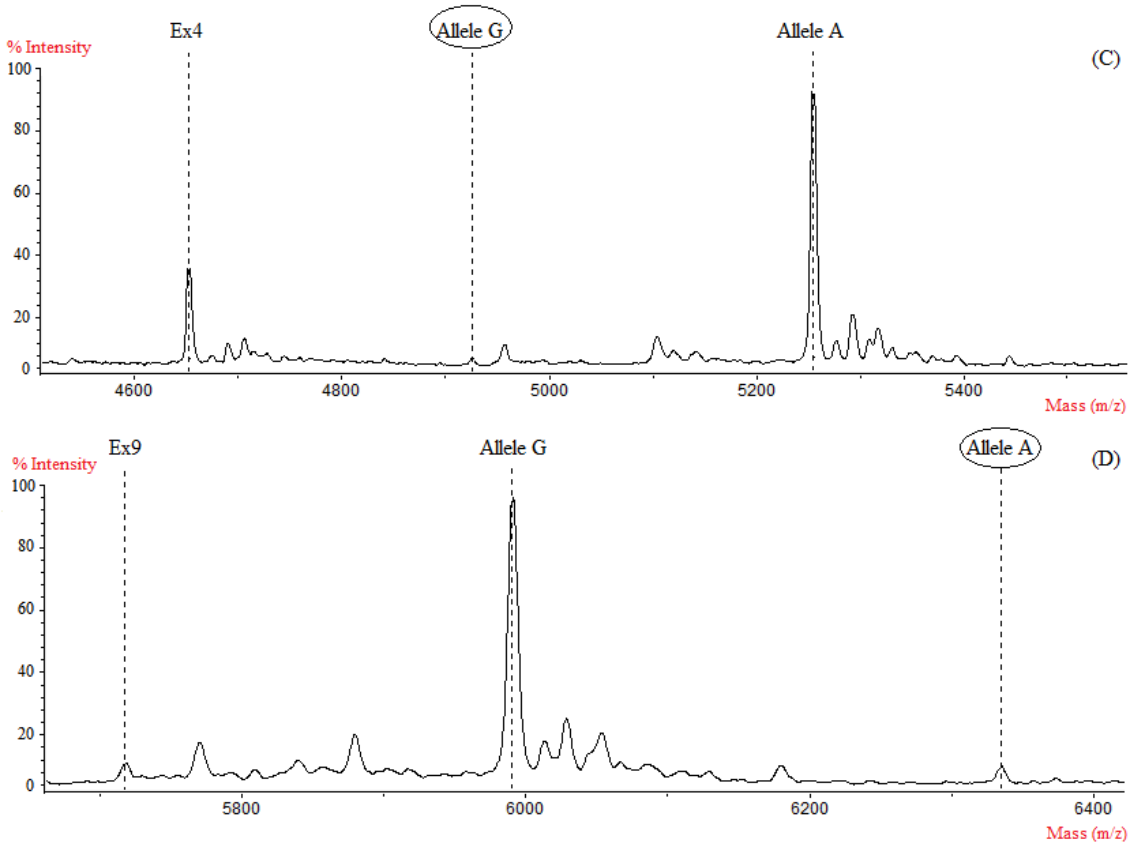


Figure 21 (con't). Examples of the quantitative outputs by the developed SNP genotyping assay. The codes for particular extension primers were indicated and reporting alleles were circled. The percentages of MC in **(C**; Recipients are heterozygous and donors are homozygous**)** and **(D**; Recipients are heterozygous and donors are homozygous**)** were 4.0% and 5.6% respectively.

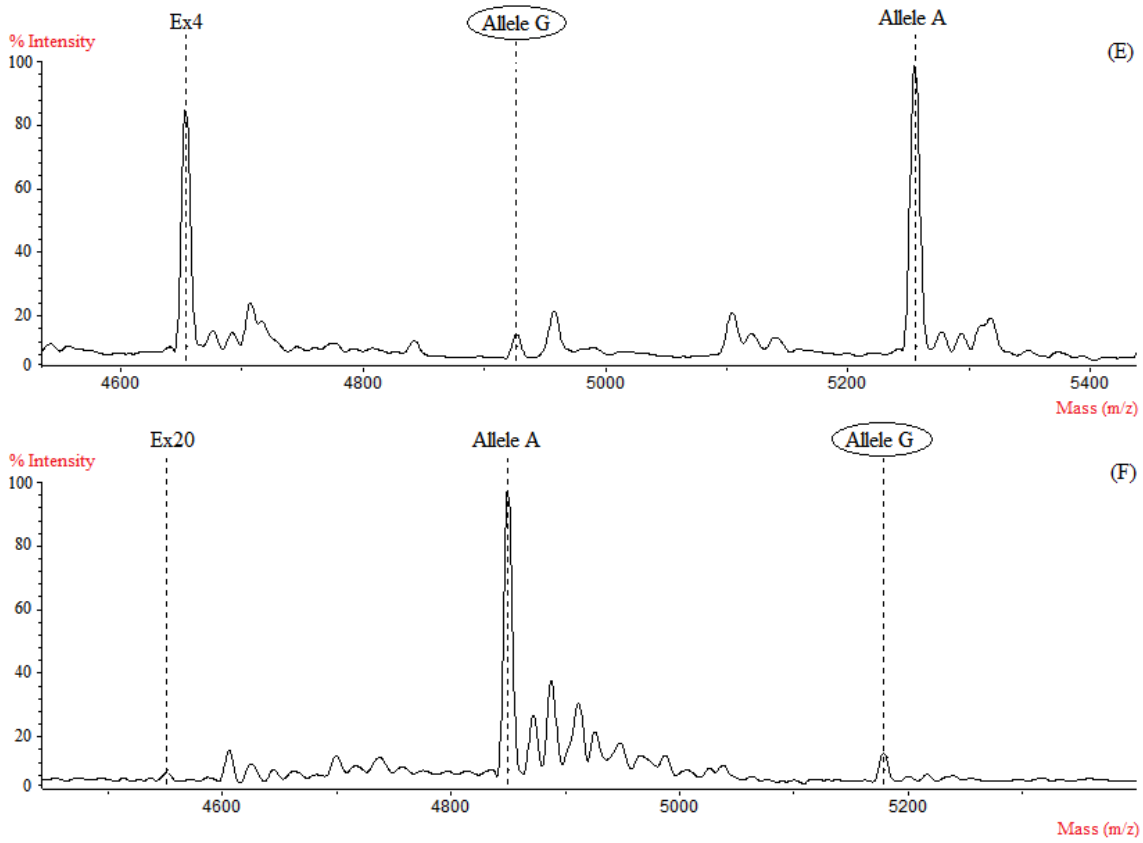


Figure 21 (con't). Examples of the quantitative outputs by the developed SNP genotyping assay. The codes for particular extension primers were indicated and reporting alleles were circled. The percentages of MC in (**E**; Recipients are heterozygous and donors are homozygous) and **(F;** Recipients are heterozygous and donors are homozygous) were 11.3% and 16.0% respectively.

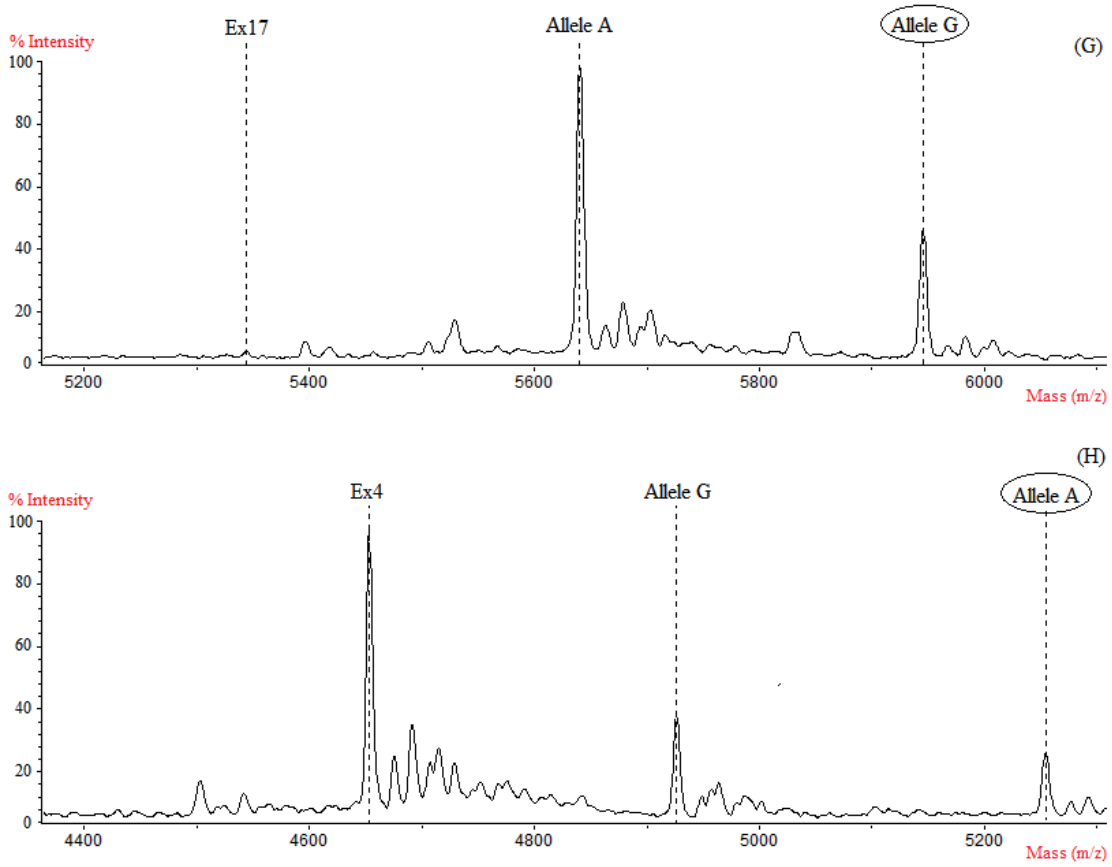


Figure 21 (con't). Examples of the quantitative outputs by the developed SNP genotyping assay. The codes for particular extension primers were indicated and reporting alleles were circled. The percentages of MC in (**G**; Recipients and donors are disparate homozygous) and (**H**; Recipients are heterozygous and donors are homozygous**)** were 29.3% and 62.8% respectively.

Table 14. Comparison of the percentages of MC measured by the STR-PCR assay and the developed SNP genotyping assay. **(A)** Cases with double cord blood transplantation (n=13). **(B)** Cases with single donor (n=17).

(A)	% of MC measured by the STR-PCR assay (STR marker)	% of MC measured by the developed SNP genotyping assay (dbSNP ID)
Cases with unrelated donors	3.9% (DXR)	No informative SNPs[a]
	16.0% (vWFA)	Mean value = 20.5% 16.0% (rs163781) 25.0% (rs1335873)
	0.3% (388II)	0.0% (rs12944065)
	0.0% (388II)	0.0% (rs714825)
	24.0% (Humara)	No informative SNPs[a]
	2.9% (388II)	No informative SNPs[c]
	80.0% (D5S818)	62.8% (rs714825)
	34.0% (D19S253)	Mean value = 30.9% 32.4% (rs10757713) 29.3% (rs12944065)
	1.0% (388II)	0.0% (rs1335873)
	0.6% (388II)	0.0% (rs12944065)
	2.0% (388II)	0.8% (rs163781)
	1.7% (388II)	No informative SNPs[c]
	1.7% (vWFA)	2.8% (rs163781)

(B)	% of MC measured by the STR-PCR assay (STR marker)	% of MC measured by the developed SNP genotyping assay (dbSNP ID)
Cases with unrelated donors	0.0% (Humara)	0.0% (rs1335873)
	0.0% (388II)	0.0% (rs1335873)
Cases with related donors	92.0% (388II)	No informative SNPs[b]
	10.6% (DXS9902)	18.3% (rs2831700)
	0.7% (vWFA)	0.0% (rs163781)
	0.7% (388II)	0.0% (rs2831700)
	1.9% (388II)	1.9% (rs2831700)
	53.0% (D5S818)	59.6% (rs2831700)
	1.3% (388II)	0.0% (rs163781)
	0.2% (DXS9902)	0.0% (rs2831700)
	16.0% (Humara)	15.7% (rs714825)
	4.1% (DXS9902)	3.2% (rs714825)
	0.2% (388II)	0.0% (rs2831700)
	0.0% (Humara)	0.0% (rs2831700)
	0.0% (388II)	0.0% (rs714825)
	0.2% (Humara)	0.0% (rs12944065)
	0.3% (Humara)	0.0% (rs2831700)

[a]No informative SNPs could be identified between the recipient and the dominant donor
[b]No informative SNPs could be identified between the recipient and the donor
[c]Both donors were dominant and no informative SNPs could be identified among the three individuals

P <0.001). For the two cases assessed by two informative SNPs, the MC percentages were found to be similar to each other. In consistence to the determined limit of detection (i.e., 1.4%), among the cases having informative SNPs, our developed assay generated readings for all 10 cases (100%) of MC >1.4% (determined by the STR-PCR assays), whereas "zero" readings were obtained for all 15 cases (100%) of MC <1.4%.

5.3.4 Application of the Developed Assay to Longitudinal Monitoring of Chimerism in the Post-Transplant Patients

Longitudinal blood samples from three typical post-transplant cases were examined in this experiment. The STR-PCR assay results showed that the first case displayed a change of levels of MC in a wavy pattern; the second case showed an increasing trend of MC; and third case showed a decreasing trend of MC. **Figure 22** shows the results of chimerism analysis for longitudinal post-transplant patient samples by the quantitative SNP genotyping assay. In all three cases, the trends obtained by the quantitative SNP genotyping assay were similar to those found by the STR-PCR assay. Consistently, MC <1.4% were undetectable. In one case (**Figure 22A**), at one time point there was a large deviation between the two MC percentages.

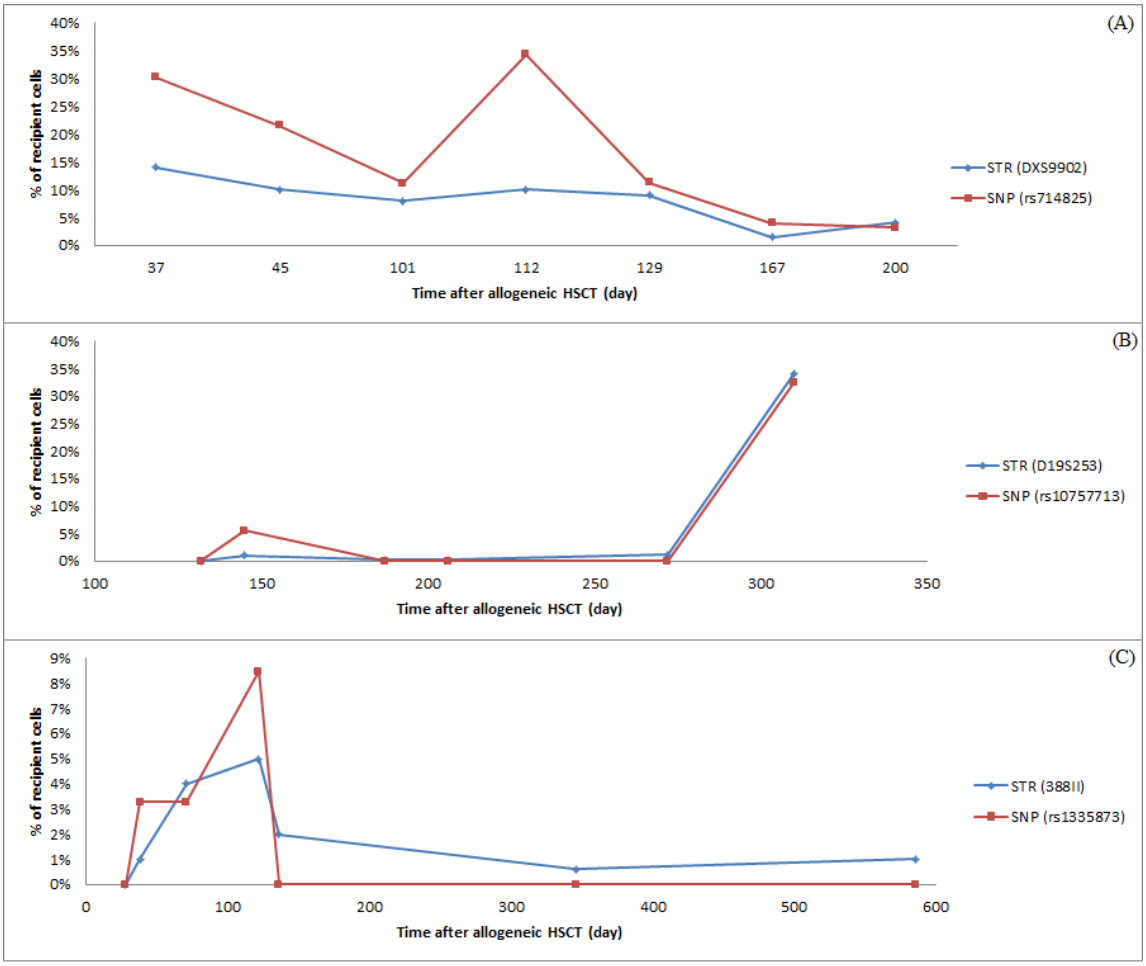


Figure 22. The percentages of recipient cells as a function of time after allogeneic HSCT in three cases

5.4 Discussion

As shown in Chapter 4, the developed SNP genotyping assay, which was based on MALDI-TOF MS, could reliably provide qualitative information of oligonucleotides (genotyping of SNPs). The use of MALDI-TOF MS in quantitative analysis of oligonucleotides would be a greater challenge. Instead of absolute quantification, relative quantification was investigated, giving rise to an important clinical application ─ quantification of MC in patients after allogeneic HSCT. For typical biallelic SNPs, a peak area ratio of a particular reporting allele to another non-reporting allele would give an estimation of MC percentages in post-transplant patients. However, it was essential to evaluate its linearity, accuracy, precision and limit of detection in assessment of chimerism before it can be used in routine clinical laboratories. Therefore, its performance was compared with the routine well-established STR-PCR assay in terms of quantification of MC using clinical human specimens.

In order to determine whether the quantitative output of the developed SNP genotyping assay was linear across the predicted working range, the relationship between the observed relative peak area and the expected relative peak area of a reporting informative SNP allele in an 'artificial' pair of donor and recipient was investigated using a set of chimeric DNA mixtures. The linear relationship was shown, but the line did not pass through the origin. Actually, MALDI-TOF MS analysis of oligonucleotides has several weaknesses, such as reduced S/N with masses and tendency to form salt adducts. Another weakness is the presence of electronic background noise, which makes it more difficult to identify peaks in mass spectra.[149] Especially when a reporting allele peak is very small, it may be indistinguishable from the electronic noise.

Therefore, the effect of the electronic noise should be more significant when samples with lower levels of MC were analyzed. Finally, a correction equation was constructed for the reporting allele peak. The peak area of the reporting SNP allele should be corrected before it can be used to calculate the level of MC.

MC percentage down to 5% can be readily detected by MALDI-TOF MS.[150, 151] Our MALDI-TOF MS-based SNP genotyping assay was optimized to increase the detection sensitivity. MC percentage down to 2.25% could be reproducibly detected with high accuracy. The calculated limit of detection (1.4%) was confirmed by the quantitative measurement of the clinical specimens in which all cases with >1.4% of MC (determined by the STR-PCR assays) were measurable by the developed assay while the others were not. Coefficient of variation (CV) monitors the deviation of an assay. The intra-assay CV decreased with increasing percentages of MC within the same assay, i.e., degree of precision increased with the levels of MC. It seemed that this assay was able to precisely detect increasing MC, which was reported to be an indicator for an enhanced risk of relapse in acute leukemia.[25, 37, 43] However, due to time limitation, the inter-assay CVs were not evaluated in this research project. However, because calibration curve is not needed in the developed assay, the intra-assay and inter-assay CVs are expected to be similar. In future studies, the inter-assay CV should be examined to confirm the precision of our assay. The value of detection limit was investigated by a set of ‘artificial’ chimeric DNA mixtures from a homozygous ‘donor’ and a heterozygous ‘recipient’. Theoretically, in the case of a pair of individuals with disparate homozygosity the peak area of the reporting allele will be doubled, as compared to a pair of a homozygous ‘donor’ and a heterozygous ‘recipient. Hence, the detection limit of

our in-house SNP genotyping assay will then become 0.7%. However, we were unable to recruit two volunteers with disparate homozygosity at any one of our targeted SNPs for the investigation. In future studies, the effect of zygosity of SNP genotypes on the limit of detection should be investigated.

With regard to the comparison between the STR-PCR assay and developed SNP genotyping assay in quantification of MC in the clinical samples, in general the two assays generated highly similar results. The major discrepancy might have been contributed by their different detection limits. The detection limit for a STR-PCR assay depends on particular STR markers. It ranges from 0.1-0.5%.[152, 153] For the cases with MC >1.4% (determined by the STR-PCR assay), the two assays gave consistent results. For the cases with MC <1.4%, levels of recipient DNA were measurable only by the STR-PCR assay, owing to its lower detection limit. For the two cases assessed by two informative SNPs, the measured percentages of MC were found to have some degree of difference. To measure the MC level more accurately, one could make use of several informative SNPs simultaneously to generate an average value for reporting. For the longitudinal assessment of chimerism in three cases, the trends obtained by the two assays were similar. It should be noted that at one time point there was a large deviation between the results obtained by the assays. This DNA sample at this time point should be re-analyzed by using another informative SNP marker. However, in this case only one informative SNP marker was available among the 8 targeted SNPs. In future studies, more SNP markers should be added to the panel of targeted SNPs. Then this discrepancy can be validated.

Table 15 shows a comparison between different characteristics of the developed SNP genotyping assay and the STR-PCR assay. The major advantage of the developed assay is that it can provide an evidence of targeted SNPs being genotyped and used for quantification. Although the biallelic SNPs have less variation and informativity, a theoretically unlimited number of SNPs in the human genome can compensate the low informativity of a single SNP. Thanks to the biallelic nature (i.e., two alleles of a SNP have the same size), PCR amplification biases, which are commonly observed during amplification of different sized alleles in the STR-PCR assay, are theoretically absent in the developed assay since SNPs do not display length polymorphism. This is the second advantage of the developed assay. So far the apparent bias has not been observed (data not shown). The third advantage is related to possible reduction of unwanted peaks. Different types of unwanted peaks can be generated in both assays. For the developed assay, those peaks (salt adduct and fragmented peaks) can be reduced by some sample preparations. However, for the STR-PCR assay, the unwanted peaks (stutter peaks) seem to be unavoidable as they form due to the nature of STR markers (repetitive sequences). Compared with the STR-PCR assay, the major drawback of the developed SNP-based assay is the long turnaround time. By estimation, for processing 30 samples, it may take 1.5-2day for the developed assay while 0.5-1day for the STR-PCR assay. Another drawback is the relatively lower detection limit (1.4% for the developed assay V.S. 0.1%-0.5% for the STR-PCR assay). The latter assay can provide earlier detection of MC following allogeneic HSCT. If the detection limit of the developed assay can be increased by further optimization of the assay, it may be suitable for clinical use, as long as manpower is enough in a clinical laboratory.

Table 15. Comparison between different characteristics of the developed SNP genotyping assay and the STR-PCR assay

	STR genotyping (STR-PCR)	SNP genotyping (MALDI-TOF MS based-primer extension reaction)
1. Discrimination rate per marker	Higher	Lower
2. Number of markers available	Less	More
3. Provision of sequence information to confirm correct genotyping	No	Yes
4. PCR amplification bias	Presence	Theoretically absence
5. Formation of stutter peaks	Yes	No
6. Formation of salt adduct peaks	No	Yes, but heavily reduced after the sample clean-up
7. Formation of fragmented peaks	No	Yes, but can be reduced by careful primer design and components in the matrix
8. Assay duration for 30 samples (rough estimation)	- 3-4hr for PCR; - 2-3hr for capillary electrophoresis	- 12-13hr for competitive PCR, SAP treatment and primer extension reaction; - 8-9hr for sample clean-up prior to MALDI-TOF MS analysis
9. Detection limit	0.1-0.5%[152, 153]	1.4%

5.5 Conclusion

The developed SNP genotyping assay has a potential for quantitative assessment of chimerism in patients following allogeneic HSCT. This assay was able to measure the level of MC accurately and precisely. Its detection limit was found to be 1.4%, which was independently confirmed through the analyses of the clinical specimens. For the cases with MC >1.4%, its performance was similar to that of the STR-PCR assay. One of the major advantages of the developed assay is that it provides confirmatory evidence of targeted SNPs being genotyped. Use of SNPs rather than STRs as genetic markers is preferred as the problem of amplification bias can be avoided. The major problem of the developed assay is that the assay time requirement is rather long, because it takes time for sample clean-up to minimize any unwanted signal peaks.

CHAPTER 6: GENERAL DISCUSSIONS AND CONCLUSIONS

SNPs are advantageous compared to other polymorphic markers for human identification. They are the most common type of genetic polymorphism. The unlimited number of SNPs in the human genome can potentially distinguish any two individuals, regardless of their sex and other phenotypic differences (e.g. blood groups). SNP-genotyping assays have much higher applicability than other techniques making use of other polymorphic markers, such as erythrocyte phenotyping and sex chromosome genotyping (e.g. XY-FISH).

MC is a common outcome in patients after allogeneic HSCT. Levels of MC (i.e., proportion of recipient cells in donor hematopoiesis) are associated with several post-transplant complications, like disease relapse. It is clinically important to measure the level of MC accurately by a sensitive technique using a panel of markers (e.g. SNPs) with high informativity. We have developed an in-house MALDI-TOF MS-based SNP genotyping assay with aid of primer extension reactions for qualitative and quantitative analysis of SNPs. The assay was demonstrated to be capable in genotyping target SNPs accurately and to measure the level of MC in a quantitative manner with limit of detection of 1.4% in the examined clinical samples. In terms of detection limit, our assay is inferior compared to the STR assay.

There are three other SNP genotyping techniques applied to quantification of MC, including real-time quantitative PCR, minisequencing in a microarray setup and pyrosequencing. The detection limits for real-time quantitative PCR and minisequencing

in a microarray setup were reported to 0.1%[81] and 1%[82], respectively. Our developed assay could not achieve a detection limit as low as those of the STR assay, real-time quantitative PCR and minisequencing in a microarray setup However, our developed assay could provide additional confirmatory evidence for the observed genotype because it made use of MALDI-TOF MS for direct mass (m/z) measurement of allele-specific genotyping products. When observed m/z values of the products match the expected ones, it was confirmed targeted SNPs have been genotyped. Although pyrosequencing could also provide confirmatory evidence for the observed genotype, its detection limit was found to be 5%[80]. Therefore, in terms of detection limit, our developed SNP genotyping assay was better than the pyrosequencing approach. Most commercial MALDI-TOF MS instruments, including the one used in this study, use a microchannel plate detector to capture and amplify ion signals. The microchannel plate detector is well known to have insufficient sensitivity and high susceptibility to signal saturation.[154] This probably was one of the major causes for the relatively low detection limit in our SNP genotyping assay. Hopefully, upon the development of a better signal detector for MALDI-TOF MS, the detection limit of our SNP genotyping assay could be greatly improved.

It is worth noting that all assays, including our developed assay, have limitations. First, the size of the SNP panel would affect the probability of finding at least one informative SNP by the developed SNP genotyping assay. Among our examined cases no informative SNPs was found in five cases. In order to distinguish all donor-recipient cases, expansion of the panel size may be useful. Another limitation of the developed

assay is its inability to have automation due to the sample clean-up procedures. It may roughly took half a day to process 30 samples in a semi-manual manner (one part was performed by a pipetting robot). The procedures cost much time and labor when sample sizes are large. The entire assay duration is much longer than that of the routine STR-PCR assay. Another limitation is related to the MS instrument. Usually this machine has high maintenance cost. It must be maintained and operated by specialized trained staff. Nowadays, a one-push-bottom MS system is not commercially available. It is difficult to apply mass spectrometry to clinical diagnostics (e.g. chimerism assessment) in routine clinical laboratories. **Table 16** shows a summary table for comparison of different techniques for quantification of MC after allogeneic HSCT.

To conclude, combined with primer extension reactions, an in-house quantitative MALDI-TOF MS-based SNP genotyping assay was successfully developed. It was able to genotype all selected SNPs and provide quantitative measurement of MC in clinical samples by using the targeted informative SNPs. However, it is still not suitable for clinical use as it has several limitations, particularly high detection limit and such as the long assay turnaround time.

Table 16. Different techniques for quantification of MC after allogeneic HSCT with use of SNP markers

Techniques	Merits	Problems	Reported limit of detection for MC
Real-time Quantitative PCR using TaqMan technology	High detection sensitivity	- Inability to confirm correct genotyping of targeted SNPs; - Inaccurate quantification of high levels of MC; - Requirement of a calibration curve for each allele of a SNP	0.1%[81]
Pyrosequencing	- Provision of evidence of correct genotyping of targeted SNPs; - Quantitative;	Low detection sensitivity	5%[80]
Minisequencing in a microarray setup	High-throughput	- Inability to confirm correct genotyping of targeted SNPs; - Semi-quantitative	1%[82]
MALDI-TOF MS (primer extension-based)	- Direct measurement of allele-specific products; - Provision of confirmatory evidence of correct genotyping of targeted SNPs; - Simultaneous analysis of multiple markers	- Being very sensitive to contaminants; - Long turnaround time; - Laboriousness	1.4%

REFERENCES:

1. Copelan EA. Hematopoietic stem-cell transplantation. N Engl J Med. 2006;354(17):1813-26.

2. Gratwohl A, Baldomero H, Aljurf M, Pasquini MC, Bouzas LF, Yoshimi A, et al. Hematopoietic stem cell transplantation. JAMA: the journal of the American Medical Association. 2010;303(16):1617-24.

3. Forman SJ, Blume KG, Thomas ED. Hematopoietic cell transplantation. Blackwell Science. 1999.

4. Blood C. Stem cell transplant (peripheral blood, bone marrow, and cord blood transplants). American Cancer Society. 2012

5. Kalos M, Levine BL, Porter DL, Katz S, Grupp SA, Bagg A, et al. T cells with chimeric antigen receptors have potent antitumor effects and can establish memory in patients with advanced leukemia. Science Translational Medicine. 2011;3(95):95ra73.

6. Porter DL, Levine BL, Kalos M, Bagg A, June CH. Chimeric antigen receptor–modified T cells in chronic lymphoid leukemia. N Engl J Med. 2011;365(8):725-33.

7. Gratwohl A, Baldomero H, Frauendorfer K, Urbano-Ispizua A, Niederwieser D. Results of the EBMT activity survey 2005 on haematopoietic stem cell transplantation: Focus on increasing use of unrelated donors. Bone Marrow Transplant. 2007;39(2):71-87.

8. Gale RP, Horowitz MM, Ash RC, Champlin RE, Goldman JM, Rimm AA, et al. Identical-twin bone marrow transplants for leukemia. Ann Intern Med. 1994;120(8):646.

9. Weiden PL, Flournoy N, Thomas ED, Prentice R, Fefer A, Buckner CD, et al. Antileukemic effect of graft-versus-host disease in human recipients of allogeneic-marrow grafts. N Engl J Med. 1979;300(19):1068-73.

10. Perreault C, Roy DC, Fortin C. Immunodominant minor histocompatibility antigens: The major ones. Immunol Today. 1998;19(2):69-74.

11. Mutis T, Schrama E, van Luxemburg-Heijs SA, Falkenburg J, Melief CJ, Goulmy E. HLA class II restricted T-cell reactivity to a developmentally regulated antigen shared by leukemic cells and CD34 early progenitor cells. Blood. 1997;90(3):1083-90.

12. Mielcarek M, Martin PJ, Leisenring W, Flowers ME, Maloney DG, Sandmaier BM, et al. Graft-versus-host disease after nonmyeloablative versus conventional hematopoietic stem cell transplantation. Blood. 2003;102(2):756-62.

13. Rocha V, Cornish J, Sievers EL, Filipovich A, Locatelli F, Peters C, et al. Comparison of outcomes of unrelated bone marrow and umbilical cord blood transplants in children with acute leukemia. Blood. 2001;97(10):2962-71.

14. Agarwal M. Umbilical cord blood transplantation: Newer trends. JAPI. 2006;54:143-7.

15. Wang Y. Indications for hematopoietic stem cell transplantation in the treatment of pediatric disorders. World J Pediatr. 2006;2(1).

16. Esparza SD, Sakamoto KM. Topics in pediatric leukemia–acute lymphoblastic leukemia. Medscape General Medicine. 2005;7(1):23.

17. Chan G, Pui C. Current status of acute lymphoblastic leukaemia in children. HK J Paediatr (new series). 2003;(8):170-83.

18. Pulsipher MA, Peters C, Pui C. High-risk pediatric acute lymphoblastic leukemia: To transplant or not to transplant? Biology of Blood and Marrow Transplantation. 2011;17(1):S137-48.

19. Tallen G, Ratei R, Mann G, Kaspers G, Niggli F, Karachunsky A, et al. Long-term outcome in children with relapsed acute lymphoblastic leukemia after time-point and site-of-relapse stratification and intensified short-course multidrug chemotherapy: Results of trial ALL-REZ BFM 90. Journal of Clinical Oncology. 2010;28(14):2339-47.

20. Soni S. Allogeneic stem cell transplantation for genetic disorders. J Ky Med Assoc. 2007;105(1):12-6.

21. Issaragrisil S, Kaufman DW, Anderson T, Chansung K, Leaverton PE, Shapiro S, et al. The epidemiology of aplastic anemia in thailand. Blood. 2006;107(4):1299-307.

22. Farzana T, Shamsi TS, Ansari SH. Pediatric bone marrow transplantation. Haematology Updates 2010.

23. Donohue J, Homge M, Kernan NA. Characterization of cells emerging at the time of graft failure after bone marrow transplantation from an unrelated marrow donor. Blood. 1993;82(3):1023-9.

24. Dazzi F, Fozza C. Disease relapse after haematopoietic stem cell transplantation: Risk factors and treatment. Best Practice & Research Clinical Haematology. 2007;20(2):311-27.

25. Bader P, Beck J, Frey A, Schlegel P, Hebarth H, Handgretinger R, et al. Serial and quantitative analysis of mixed hematopoietic chimerism by PCR in patients with acute leukemias allows the prediction of relapse after allogeneic BMT. Bone Marrow Transplant. 1998;21(5):487.

26. 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.

27. Shamsi TS, Irfan M, Farzana T, Ansari SH, Panjwani VK, Taj M, et al. Survival and graft versus host disease in first 100 patients undergoing allogeneic peripheral blood stem cell transplantation: A single centre experience. J Pak Med Assoc. 2005;55(11):469-74.

28. Dykewicz CA. Guidelines for preventing opportunistic infections among hematopoietic stem cell transplant recipients: Focus on community respiratory virus infections. Biology of Blood and Marrow Transplantation. 2001;7(12):19S-22S.

29. Butturini A, Gale R. T cell depletion in bone marrow transplantation for leukemia: Current results and future directions. Bone Marrow Transplant. 1988;3(3):185.

30. Dell'Orto M, Rovelli A, Barzaghi A, Valsecchi M, Silvestri D, Giltri G, et al. Febrile complications in the first 100 days after bone marrow transplantation in children: A single center's experience. Pediatric Hematology-Oncology. 1997;14(4):335-47.

31. Huma Z, Boulad F, Black P, Heller G, Sklar C. Growth in children after bone marrow transplantation for acute leukemia. Blood. 1995;86(2):819-24.

32. 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):735-46.

33. Bader P, Hölle W, Klingebiel T, Handgretinger R, Niethammer D, Beck J. Quantitative assessment of mixed hematopoietic chimerism by polymerase chain reaction after allogeneic BMT. Anticancer Res. 1996;16(4A):1759.

34. Bader P, Niethammer D, Willasch A, Kreyenberg H, Klingebiel T. How and when should we monitor chimerism after allogeneic stem cell transplantation? Bone Marrow Transplant. 2004;35(2):107-19.

35. Niethammer D, Bienzle U, Dieterle U, Flad H, Goldman S, Haas J, et al. Split chimerism in severe combined immunodeficiency. Pediatr Res. 1978;12(1):69.

36. Maas F, Schaap N, Kolen S, Zoetbrood A, Buno I, Dolstra H, et al. Quantification of donor and recipient hemopoietic cells by real-time PCR of single nucleotide polymorphisms. Leukemia. 2003;17(3):621-9.

37. Bader P, Hölle W, Klingebiel T, Handgretinger R, Benda N, Schlegel P, et al. Mixed hematopoietic chimerism after allogeneic bone marrow transplantation: The impact of quantitative PCR analysis for prediction of relapse and graft rejection in children. Bone Marrow Transplant. 1997;19(7):697.

38. Bader P, Hancock J, Kreyenberg H, Goulden N, Niethammer D, Oakhill A, et al. Minimal residual disease (MRD) status prior to allogeneic stem cell transplantation is a powerful predictor for post-transplant outcome in children with ALL. Leukemia: official journal of the Leukemia Society of America, Leukemia Research Fund, UK. 2002;16(9):1668.

39. Öğret Y, Kalayoğlu S, Oğuz F, Gürtekin M, Sargın D, Çarin M. Chimerism analysis hematopoietic stem cell transplantation. Adv Mol Med. 2006;2(4):171-3.

40. Wu CJ, Chillemi A, Alyea EP, Orsini E, Neuberg D, Soiffer RJ, et al. Reconstitution of T-cell receptor repertoire diversity following T-cell depleted allogeneic bone marrow transplantation is related to hematopoietic chimerism. Blood. 2000;95(1):352-9.

41. Wekerle T, Sykes M. Mixed chimerism as an approach for the induction of transplantation Tolerance1. Transplantation. 1999;68(4):459-67.

42. Mattsson J, Uzunel M, Remberger M, Ringdén O. T cell mixed chimerism is significantly correlated to A decreased risk of acute graft-versus-host disease after allogeneic stem cell transplantation 1. Transplantation. 2001;71(3):433-9.

43. Barrios M, Jiménez-Velasco A, Román-Gómez J, Madrigal ME, Castillejo JA, Torres A, et al. Chimerism status is a useful predictor of relapse after allogeneic stem cell transplantation for acute leukemia. Haematologica. 2003;88(7):801-10.

44. Serrano J, Roman J, Herrera C, Castillejo J, Navarro J, Reina M, et al. Increasing mixed haematopoietic chimaerism after BMT with total depletion of CD4 and partial depletion of CD8 lymphocytes is associated with a higher incidence of relapse. Bone Marrow Transplant. 1999;23(5):475.

45. Roy DC, Tantravahi R, Murray C, Dear K, Gorgone B, Anderson KC, et al. Natural history of mixed chimerism after bone marrow transplantation with CD6-depleted allogeneic marrow: A stable equilibrium. Blood. 1990;75(1):296-304.

46. Van Leeuwen J, Van Tol M, Joosten A, Wijnen J, Khan P, Vossen J. Mixed T-lymphoid chimerism after allogeneic bone marrow transplantation for hematologic malignancies of children is not correlated with relapse. Blood. 1993;82(6):1921-8.

47. Schattenberg A, De Witte T, Salden M, Vet J, Van Dijk B, Smeets D, et al. Mixed hematopoietic chimerism after allogeneic transplantation with lymphocyte-depleted bone marrow is not associated with a higher incidence of relapse. Blood. 1989;73(5):1367-72.

48. Bader P, Stoll K, Huber S, Geiselhart A, Handgretinger R, Niemeyer C, et al. Characterization of lineage‐specific chimaerism in patients with acute leukaemia and myelodysplastic syndrome after allogeneic stem cell transplantation before and after relapse. Br J Haematol. 2000;108(4):761-8.

49. Lion T, Daxberger H, Dubovsky J, Filipcik P, Fritsch G, Printz D, et al. Analysis of chimerism within specific leukocyte subsets for detection of residual or recurrent leukemia in pediatric patients after allogeneic stem cell transplantation. Leukemia. 2001;15(2):307-10.

50. Kolb H, Schattenberg A, Goldman JM, Hertenstein B, Jacobsen N, Árcese W, et al. Graft-versus-leukemia effect of donor lymphocyte transfusions in marrow grafted patients. european group for blood and marrow transplantation working party chronic Leukemia. Blood. 1995;86(5):2041-50.

51. Dazzi F, Goldman J. Donor lymphocyte infusions. Curr Opin Hematol. 1999;6(6):394.

52. Morecki S, Slavin S. Toward amplification of a graft-versus-leukemia effect while minimizing graft-versus-host disease. J Hematother Stem Cell Res. 2000;9(3):355-66.

53. Carlens S, Remberger M, Aschan J, Ringdén O. The role of disease stage in the response to donor lymphocyte infusions as treatment for leukemic relapse. Biology of Blood and Marrow Transplantation. 2001;7(1):31-8.

54. De Weger RA, Tilanus MG, Scheidel KC, Van Den Tweel, Jan G, Verdonck LF. Monitoring of residual disease and guided donor leucocyte infusion after allogeneic bone marrow transplantation by chimaerism analysis with short tandem repeats. Br J Haematol. 2000;110(3):647-53.

55. Slavin S, Nagler A, Naparstek E, Kapelushnik Y, Aker M, Cividalli G, et al. Nonmyeloablative stem cell transplantation and cell therapy as an alternative to conventional bone marrow transplantation with lethal cytoreduction for the treatment of malignant and nonmalignant hematologic diseases. Blood. 1998;91(3):756-63.

56. Aghaiipour M, Iravani M, David G, Bernard D. Flow cytometric evaluation of red blood cell chimerism after bone marrow transplantation in iranian patients: A preliminary study. Archives of Iranian Medicine. 2006;9(4):406-9.

57. Hibi S, Tsunamoto K, Todo S, Sawada T, Ueda Y, Taniwaki M, et al. Chimerism analysis on mononuclear cells in the CSF after allogeneic bone marrow transplantation. Bone Marrow Transplant. 1997;20(6):503-6.

58. Durnam DM, Anders KR, Fisher L, O'Quigley J, Bryant EM, Thomas ED. Analysis of the origin of marrow cells in bone marrow transplant recipients using a Y-chromosome-specific in situ hybridization assay. Blood. 1989;74(6):2220-6.

59. Anastasi J, Thangavelu M, Vardiman JW, Hooberman AL, Bian ML, Larson RA, et al. Interphase cytogenetic analysis detects minimal residual disease in a case of acute lymphoblastic leukemia and resolves the question of origin of relapse after allogeneic bone marrow transplantation. Blood. 1991;77(5):1087-91.

60. Dewald G, Schad C, Christensen E, Law M, Zinsmeister A, Stalboerger P, et al. Fluorescence in situ hybridization with X and Y chromosome probes for cytogenetic studies on bone marrow cells after opposite sex transplantation. Bone Marrow Transplant. 1993;12(2):149.

61. Najfeld V, Burnett W, Vlachos A, Scigliano E, Isola L, Fruchtman S. Interphase FISH analysis of sex-mismatched BMT utilizing dual color XY probes. Bone Marrow Transplant. 1997;19(8):829-34.

62. Buño I, Anta B, Moreno-López E, Balsalobre P, Balas A, García-Sánchez F, et al. Lineage-specific chimaerism quantification after T-cell depleted peripheral blood stem cell transplantation. Leuk Lymphoma. 2003;44(4):659-67.

63. Thiele J, Wickenhauser C, Kvasnicka H, Varus E, Schneider C, Müller H, et al. Mixed chimerism of erythro-and megakaryopoiesis following allogeneic bone marrow transplantation. Acta Haematol. 2003;109(4):176-83.

64. Lo Y, Tein MS, Lau TK, Haines CJ, Leung TN, Poon PM, et al. Quantitative analysis of fetal DNA in maternal plasma and serum: Implications for noninvasive prenatal diagnosis. The American Journal of Human Genetics. 1998;62(4):768-75.

65. Fehse B, Chukhlovin A, Kühlcke K, Marinetz O, Vorwig O, Renges H, et al. Real-time quantitative Y chromosome-specific PCR (QYCS-PCR) for monitoring hematopoietic chimerism after sex-mismatched allogeneic stem cell transplantation. J Hematother Stem Cell Res. 2001;10(3):419-25.

66. Wong AK, Fang B, Zhang L, Guo X, Lee S, Schreck R. Loss of the Y chromosome: An age-related or clonal phenomenon in acute myelogenous leukemia/myelodysplastic syndrome? Arch Pathol Lab Med. 2008;132(8):1329-32.

67. Alpár D, Nagy G, Hohoff C, Kajtár B, Bartyik K, Hermesz J, et al. Sex chromosome changes after sex‐mismatched allogeneic bone marrow transplantation can mislead the chimerism analysis. Pediatric Blood & Cancer. 2010;55(6):1239-42.

68. Huh J, Moon H, Chung WS. Incidence and clinical significance of sex chromosome losses in bone marrow of patients with hematologic diseases. The Korean Journal of Laboratory Medicine. 2007;27(1):56-61.

69. Antin JH, Childs R, Filipovich AH, Giralt S, Mackinnon S, Spitzer T, et al. Establishment of complete and mixed donor chimerism after allogeneic lymphohematopoietic transplantation: Recommendations from a workshop at the 2001 tandem meetings of the international bone marrow transplant registry and the american society of blood and marrow transplantation. Biology of Blood and Marrow Transplantation. 2001;7(9):473-85.

70. Millson AS, Spangler FL, Wittwer CT, Lyon E. Comparison of automated short tandem repeat and manual variable number of tandem repeat analysis of chimerism in bone marrow transplant patients. Diagnostic Molecular Pathology. 2000;9(2):91-7.

71. Sun JF, Han XP, Zhao DD, Wang FF, Jin HJ, Gao CJ, et al. Application of chimerism analysis to allogeneic hematopoietic stem cell transplantation by STR-PCR. Zhongguo Shi Yan Xue Ye Xue Za Zhi. 2007;15(2):337-41.

72. Acquaviva C, Duval M, Mirebeau D, Bertin R, Cave H. Quantitative analysis of chimerism after allogeneic stem cell transplantation by PCR amplification of microsatellite markers and capillary electrophoresis with fluorescence detection: The Paris–Robert debre experience. Leukemia. 2003;17(1):241-6.

73. Kreyenberg H, Hölle W, Möhrle S, Niethammer D, Bader P. Quantitative analysis of chimerism after allogeneic stem cell transplantation by PCR amplification of microsatellite markers and capillary electrophoresis with fluorescence detection: The tuebingen experience. Leukemia. 2003;17(1):237-40.

74. Kwok P. Single nucleotide polymorphisms: Methods and protocols. Springer. 2003.

75. Sachidanandam R, Weissman D, Schmidt SC, Kakol JM, Stein LD, Marth G, et al. A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. Nature. 2001;409(6822):928-33.

76. Collins FS, Brooks LD, Chakravarti A. A DNA polymorphism discovery resource for research on human genetic variation. Genome Res. 1998;8(12):1229-31.

77. Brookes AJ. The essence of SNPs. Gene. 1999;234(2):177-86.

78. Frazer KA, Ballinger DG, Cox DR, Hinds DA, Stuve LL, Gibbs RA, et al. A second generation human haplotype map of over 3.1 million SNPs. Nature. 2007;449(7164):851-61.

79. Wang DG, Fan J, Siao C, Berno A, Young P, Sapolsky R, et al. Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. Science. 1998;280(5366):1077-82.

80. Hochberg EP, Miklos DB, Neuberg D, Eichner DA, McLaughlin SF, Mattes-Ritz A, et al. A novel rapid single nucleotide polymorphism (SNP)–based method for assessment of hematopoietic chimerism after allogeneic stem cell transplantation. Blood. 2003;101(1):363-9.

81. Alizadeh M, Bernard M, Danic B, Dauriac C, Birebent B, Lapart C, et al. Quantitative assessment of hematopoietic chimerism after bone marrow transplantation by real-time quantitative polymerase chain reaction. Blood. 2002;99(12):4618-25.

82. Fredriksson M, Barbany G, Liljedahl U, Hermanson M, Kataja M, Syvänen AC. Assessing hematopoietic chimerism after allogeneic stem cell transplantation by multiplexed SNP genotyping using microarrays and quantitative analysis of SNP alleles. Leukemia. 2003;18(2):255-66.

83. Li L, Li C, Zhang S, Zhao S, Liu Y, Lin Y. Analysis of 14 highly informative SNP markers on X chromosome by TaqMan® SNP genotyping assay. Forensic Science International: Genetics. 2010;4(5):e145-8.

84. Nyrén P. The History of Pyrosequencing®. Pyrosequencing® Protocols. Humana Press. 2007:1-13.

85. Tsiatis AC, Norris-Kirby A, Rich RG, Hafez MJ, Gocke CD, Eshleman JR, et al. Comparison of sanger sequencing, pyrosequencing, and melting curve analysis for the detection of *KRAS* mutations: Diagnostic and clinical implications. The Journal of Molecular Diagnostics. 2010;12(4):425-32.

86. Syvänen A. From gels to chips:"minisequencing" primer extension for analysis of point mutations and single nucleotide polymorphisms. Hum Mutat. 1999;13(1):1-10.

87. Ritari J, Paulin L, Hultman J, Auvinen P. Application of hybridization control probe to increase accuracy on ligation detection or minisequencing diagnostic microarrays. BMC research notes. 2009;2(1):249.

88. Uebelhoer M, Nätynki M, Kangas J, Mendola A, Nguyen H, Soblet J, et al. Venous malformation-causative *TIE2* mutations mediate an AKT-dependent decrease in PDGFB. Hum Mol Genet. 2013.

89. Sarracino D, Richert C. Quantitative MALDI-TOF MS of oligonucleotides and a nuclease assay. Bioorg Med Chem Lett. 1996;6(21):2543-8.

90. Monforte JA, Becker CH. High-throughput DNA analysis by time-of-flight mass spectrometry. Nat Med. 1997;3(3):360-2.

91. Nordhoff E, Cramer R, Karas M, Hillenkamp F, Kirpekar F, Kristiansen K, et al. Ion stability of nucleic acids in infrared matrix-assisted laser desorption/ionization mass spectrometry. Nucleic Acids Res. 1993;21(15):3347-57.

92. Zhu Y, Chung C, Taranenko N, Allman S, Martin S, Haff L, et al. The study of 2, 3, 4 - Trihydroxyacetophenone and 2, 4, 6 - Trihydroxyacetophenone as matrices for DNA detection in Matrix - assisted laser Desorption/Ionization Time - of - flight mass spectrometry. Rapid Communications in Mass Spectrometry. 1996;10(3):383-8.

93. Griffin TJ, Smith LM. Single-nucleotide polymorphism analysis by MALDI–TOF mass spectrometry. Trends Biotechnol. 2000;18(2):77-84.

94. Sauer S, Reinhardt R, Lehrach H, Gut IG. Single-nucleotide polymorphisms: Analysis by mass spectrometry. Nature protocols. 2006;1(4):1761-71.

95. Tost J, GLYNNE GUT I. Genotyping single nucleotide polymorphisms by MALDI mass spectrometry in clinical applications. Clin Biochem. 2005;38(4):335-50.

96. Jurinke C, Oeth P, van den Boom D. MALDI-TOF mass spectrometry. Mol Biotechnol. 2004;26(2):147-63.

97. Ross PL, Lee K, Belgrader P. Discrimination of single-nucleotide polymorphisms in human DNA using peptide nucleic acid probes detected by MALDI-TOF mass spectrometry. Anal Chem. 1997;69(20):4197-202.

98. Jurinke C, van den Boom D, Jacob A, Tang K, Wörl R, Köster H. Analysis of ligase chain reaction products via matrix-assisted laser Desorption/Ionization time-of-Flight–Mass spectrometry. Anal Biochem. 1996;237(2):174-81.

99. Lyamichev V, Mast AL, Hall JG, Prudent JR, Kaiser MW, Takova T, et al. Polymorphism identification and quantitative detection of genomic DNA by invasive cleavage of oligonucleotide probes. Nat Biotechnol. 1999;17(3):292-6.

100. Ross P, Hall L, Smirnov I, Haff L. High level multiplex genotyping by MALDI-TOF mass spectrometry. Nat Biotechnol. 1998;16(13):1347-51.

101. Sauer S, Gut IG. Genotyping single-nucleotide polymorphisms by matrix-assisted laser-desorption/ionization time-of-flight mass spectrometry. Journal of Chromatography B. 2002;782(1):73-87.

102. Paracchini S, Arredi B, Chalk R, Tyler-Smith C. Hierarchical high-throughput SNP genotyping of the human Y chromosome using MALDI-TOF mass spectrometry. Nucleic Acids Res. 2002;30(6):e27.

103. Fei Z, Ono T, Smith LM. MALDI-TOF mass spectrometric typing of single nucleotide polymorphisms with mass-tagged ddNTPs. Nucleic Acids Res. 1998;26(11):2827-8.

104. Shahgholi M, Garcia BA, Chiu NH, Heaney PJ, Tang K. Sugar additives for MALDI matrices improve signal allowing the smallest nucleotide change (A:T) in a DNA sequence to be resolved. Nucleic Acids Res. 2001;29(19):e91.

105. Meyer K, Magne Ueland P. Use of matrix-assisted laser desorption/ionization time-of-flight mass spectrometry for multiplex genotyping. Adv Clin Chem. 2011;53:1.

106. Duncan MW, Roder H, Hunsucker SW. Quantitative matrix-assisted laser desorption/ionization mass spectrometry. Briefings in functional genomics & proteomics. 2008;7(5):355-70.

107. Chiu RW, Lo YD. Application of fetal DNA in maternal plasma for noninvasive prenatal diagnosis. Expert review of molecular diagnostics. 2002;2(1):32-40.

108. Ding C, Chiu RW, Lau TK, Leung TN, Chan LC, Chan AY, et al. MS analysis of single-nucleotide differences in circulating nucleic acids: Application to noninvasive prenatal diagnosis. Proc Natl Acad Sci U S A. 2004;101(29):10762-7.

109. Li Y, Hahn D, Holzgreve W, Hahn S. Ready detection of donor-specific single-nucleotide polymorphisms in the urine of renal transplant recipients by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. Clin Chem. 2005;51(10):1903a-1904.

110. Ioannidis JP, Ntzani EE, Trikalinos TA. 'Racial'differences in genetic effects for complex diseases. Nat Genet. 2004;36(12):1312-8.

111. Jaing R, Duan J, Windemuth A, Stephens JC, Judson R, Xu C. Genome-wide evaluation of the public SNP databases. Pharmacogenomics. 2003;4(6):779-89.

112. Aerts J, Wetzels Y, Cohen N, Aerssens J. Data mining of public SNP databases for the selection of intragenic SNPs. Hum Mutat. 2002;20(3):162-73.

113. Stern C. The hardy-weinberg law. Science. 1943;97(2510):137-8.

114. Sideri A, Neokleous N, De La Grange, Philippe Brunet, Guerton B, Kerdilles MLB, Uzan G, et al. An overview of the progress on double umbilical cord blood transplantation. Haematologica. 2011;96(8):1213-20.

115. Gabriel S, Ziaugra L, Tabbaa D. SNP genotyping using the sequenom MassARRAY iPLEX platform. Current Protocols in Human Genetics. 2009:2.12.1-2.12.18.

116. Oeth P, Beaulieu M, Park C, Kosman D, del Mistro G, van den Boom D, et al. iPLEX™ assay: Increased plexing efficiency and flexibility for MassARRAY® system through single base primer extension with mass-modified terminators. Sequenom Application Note. 2005:8876-006.

117. Beaulieu M, Hong P. Multiplexing the homogeneous MassEXTEND assay. Sequenom Application Note. 2003.

118. Markoulatos P, Siafakas N, Moncany M. Multiplex polymerase chain reaction: A practical approach. J Clin Lab Anal. 2002;16(1):47-51.

119. Henegariu O, Heerema NA, Dlouhy SR, Vance GH, Vogt PH. Multiplex PCR: Critical parameters and step-by-step protocol. BioTechniques. 1997;23(3):504-11.

120. Nordhoff E, Schürenberg M, Thiele G, Lübbert C, Kloeppel K, Theiss D, et al. Sample preparation protocols for MALDI-MS of peptides and oligonucleotides using

prestructured sample supports. International Journal of Mass Spectrometry. 2003;226(1):163-80.

121. Vollmer DL, Gross ML. Cation‐exchange resins for removal of alkali metal cations from oligonucleotide samples for fast atom bombardment mass spectrometry. Journal of Mass Spectrometry. 1995;30(1):113-8.

122. Srinivasan JR, Liu Y, Venta PJ, Siemieniak D, Killeen AA, Zhu Y, et al. Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry as a rapid screening method to detect mutations causing Tay–Sachs disease. Rapid Communications in Mass Spectrometry. 1997;11(10):1144-50.

123. Wu KJ, Steding A, Becker CH. Matrix‐assisted laser desorption time‐of‐flight mass spectrometry of oligonucleotides using 3‐hydroxypicolinic acid as an ultraviolet‐sensitive matrix. Rapid Communications in Mass Spectrometry. 1993;7(2):142-6.

124. Pieles U, Zürcher W, Schär M, Moser H. Matrix-assisted laser desorption ionization time-of-flight mass spectrometry: A powerful tool for the mass and sequence analysis of natural and modified oligonucleotides. Nucleic Acids Res. 1993;21(14):3191-6.

125. Becker T. Compositions and processes for improved mass spectrometry analysis. WIPO Patent No. 2009091841. 2009.

126. Moore D, Dowhan D. Purification and concentration of DNA from aqueous solutions. Current Protocols in Molecular Biology. 2002: 2.1.1-2.1.10.

127. Haff LA, Smirnov IP. Multiplex genotyping of PCR products with MassTag-labeled primers. Nucleic Acids Res. 1997;25(18):3749-50.

128. Gross J, Leisner A, Hillenkamp F, Hahner S, Karas M, Schafer J, et al. Investigations of the metastable decay of DNA under ultraviolet matrix-assisted laser desorption/ionization conditions with post-source-decay analysis and hydrogen/deuterium exchange. J Am Soc Mass Spectrom. 1998;9(9):866-78.

129. Annesley TM. Ion suppression in mass spectrometry. Clin Chem. 2003;49(7):1041-4.

130. Gilar M, Belenky A, Wang BH. High-throughput biopolymer desalting by solid-phase extraction prior to mass spectrometric analysis. Journal of Chromatography A. 2001;921(1):3-13.

131. Innis MA, Gelfand DH, Sninsky JJ, White TJ. PCR protocols: A guide to methods and applications. 1990.

132. Vander Horn P, Davis M, Cunniff J, Ruan C, McArdle B, Samols S, et al. Thermo sequenase [TM] DNA polymerase and T. acidophilum pyrophosphatase: New thermostable enzymes for DNA sequencing. BioTechniques. 1997;22(4):758-65.

133. Daltonik B. Anchorchip technology: Preparation for ultrasensitive automated MALDI-TOF MS. 2009.

134. Eickbush TH, Moudrianakis EN. The compaction of DNA helices into either continuous supercoils or folded-fiber rods and toroids. Cell. 1978;13(2):295-306.

135. Zenobi R, Knochenmuss R. Ion formation in MALDI mass spectrometry. Mass Spectrom Rev. 1998;17(5):337-66.

136. Atkinson K. Clinical bone marrow and blood stem cell transplantation. Cambridge University Press. 2004.

137. Elmaagacli AH, Runkel K, Steckel N, Opalka B, Trenschel R, Seeber S, et al. A comparison of chimerism and minimal residual disease between four different allogeneic transplantation methods in patients with chronic myelogenous leukemia in first chronic phase. Bone Marrow Transplant. 2001;27(8):809-15.

138. Watson S, Kelsey Z, Webb R, Evans J, Gill P. The development of a third generation STR multiplex system (TGM). Progress in Forensic Genetics 7: Proceedings of the 17th International ISFH Congress, Oslo. 1997.

139. Hori M, Fukano H, Suzuki Y. Uniform amplification of multiple DNAs by emulsion PCR. Biochem Biophys Res Commun. 2007;352(2):323-8.

140. Nollet F, Billiet J, Selleslag D, Criel A. Standardisation of multiplex fluorescent short tandem repeat analysis for chimerism testing. Bone Marrow Transplant. 2001;28(5):511-8.

141. Krawczak M. Informativity assessment for biallelic single nucleotide polymorphisms. Electrophoresis. 1999;20(8):1676-81.

142. Weber JL, Wong C. Mutation of human short tandem repeats. Hum Mol Genet. 1993;2(8):1123-8.

143. Kruglyak L. The use of a genetic map of biallelic markers in linkage studies. Nat Genet. 1997;17(1):21-4.

144. Werner F, Durstewitz Gv, Habermann F, Thaller G, Krämer W, Kollers S, et al. Detection and characterization of SNPs useful for identity control and parentage testing in major european dairy breeds. Anim Genet. 2004;35(1):44-9.

145. Subramanian S, Mishra RK, Singh L. Genome-wide analysis of microsatellite repeats in humans: Their abundance and density in specific genomic regions. Genome Biol. 2003;4(2):R13.

146. Mutis T, Gillespie G, Schrama E, Falkenburg JF, Moss P, Goulmy E. Tetrameric HLA class I–minor histocompatibility antigen peptide complexes demonstrate minor histocompatibility antigen-specific cytotoxic T lymphocytes in patients with graft-versus-host disease. Nat Med. 1999;5(7):839-42.

147. Zhong XY, Holzgreve W. MALDI-TOF MS in prenatal genomics. Transfusion Medicine and Hemotherapy. 2009;36(4):263-72.

148. Xiu-Cheng Fan A, Garritsen HS, Tarhouny SE, Morris M, Hahn S, Holzgreve W, et al. A rapid and accurate approach to identify single nucleotide polymorphisms of mitochondrial DNA using MALDI-TOF mass spectrometry. Clinical Chemical Laboratory Medicine. 2008;46(3):299-305.

149. Malyarenko DI, Cooke WE, Adam B, Malik G, Chen H, Tracy ER, et al. Enhancement of sensitivity and resolution of surface-enhanced laser desorption/ionization time-of-flight mass spectrometric records for serum peptides using time-series analysis techniques. Clin Chem. 2005;51(1):65-74.

150. Thiede C, Lion T. Quantitative analysis of chimerism after allogeneic stem cell transplantation using multiplex PCR amplification of short tandem repeat markers and fluorescence detection. Leukemia. 2001;15:303-6.

151. Schraml E, Daxberger H, Watzinger F, Lion T. Quantitative analysis of chimerism after allogeneic stem cell transplantation by PCR amplification of microsatellite markers and capillary electrophoresis with fluorescence detection: The vienna experience. Leukemia. 2003;17(1):224-7.

152. Buño I, Nava P, Simón A, González-Rivera M, Jiménez JL, Balsalobre P, et al. A comparison of fluorescent in situ hybridization and multiplex short tandem repeat polymerase chain reaction for quantifying chimerism after stem cell transplantation. Haematologica. 2005;90(10):1373-9.

153. Matsuda K, Yamauchi K, Tozuka M, Suzuki T, Sugano M, Hidaka E, et al. Monitoring of hematopoietic chimerism by short tandem repeats, and the effect of CD selection on its sensitivity. Clin Chem. 2004;50(12):2411-4.

154. Wenzel RJ, Zenobi R, Nazabal A. Comparison of sensitivity and saturation of MALDI-TOF detectors for high-mass ions. In Proceedings of the 54th ASMS Conference on Mass Spectrometry. 2006.

APPENDICES:

(A)

rsnum	chromosome	position	alleles
rs714825	4	74427063	C/T

(B)

region	CNVs
chr4:73427063-75427063	CNVR1962_full CNVR1963.1 CNVR1964.1 CNVR1965.1 CNVR1966_full

(C)

CNV	CNV	CNV	CNV	CNV
CNVR1966_full	CNVR1962_full	CNVR1963.1	CNVR1964.1	CNVR1965.1
Chromosome	Chromosome	Chromosome	Chromosome	Chromosome
4	4	4	4	4
Start	Start	Start	Start	Start
75297425	73637646	73770982	74774226	75161168
End	End	End	End	End
75302511	73668082	73777311	74775626	75164305
Copy Number	Copy Number	Copy Number	Copy Number	Copy Number
1,2	1,2	0,1,2	1,2	1,2

Appendix 1. An example of the CNV database search results for one of the targeted SNPs, rs714825. **(A)** The SNP is located in chromosome 4. **(B)** The range in a size of 2 mega bases covering the position of the SNP was searched. Five CNV regions were found within this range. **(C)** The ranges of the CNV regions is shown and none of them covers the the position of the SNP.

[Locus	Allele]	-	[Locus	Allele]	Burrows	Correlation	Chisq*	Prob.
[rs714825	A]	-	[rs1075771	A]	0.0057	0.0222	0.04	0.8495
[rs714825	A]	-	[rs1075771	B]	-0.0057	-0.0222	0.04	0.8495
[rs714825	B]	-	[rs1075771	A]	-0.0057	-0.0222	0.04	0.8495
[rs714825	B]	-	[rs1075771	B]	0.0057	0.0222	0.04	0.8495
[rs714825	A]	-	[rs891700	A]	-0.0084	-0.0363	0.10	0.7562
[rs714825	A]	-	[rs891700	B]	0.0084	0.0363	0.10	0.7562
[rs714825	B]	-	[rs891700	A]	0.0084	0.0363	0.10	0.7562
[rs714825	B]	-	[rs891700	B]	-0.0084	-0.0363	0.10	0.7562
[rs714825	A]	-	[rs1335873	A]	-0.0108	-0.0455	0.15	0.6992
[rs714825	A]	-	[rs1335873	C]	0.0108	0.0455	0.15	0.6992
[rs714825	B]	-	[rs1335873	A]	0.0108	0.0454	0.15	0.7001
[rs714825	B]	-	[rs1335873	C]	-0.0108	-0.0454	0.15	0.7001
[rs714825	A]	-	[rs163781	A]	0.0128	0.0556	0.23	0.6346
[rs714825	A]	-	[rs163781	B]	-0.0128	-0.0556	0.23	0.6346
[rs714825	B]	-	[rs163781	A]	-0.0128	-0.0556	0.23	0.6346
[rs714825	B]	-	[rs163781	B]	0.0128	0.0556	0.23	0.6346
[rs714825	A]	-	[rs1294406	A]	-0.0061	-0.0229	0.04	0.8449
[rs714825	A]	-	[rs1294406	B]	0.0061	0.0229	0.04	0.8449
[rs714825	B]	-	[rs1294406	A]	0.0061	0.0229	0.04	0.8449
[rs714825	B]	-	[rs1294406	B]	-0.0061	-0.0229	0.04	0.8449
[rs714825	A]	-	[rs715463	A]	0.0053	0.0227	0.04	0.8460
[rs714825	A]	-	[rs715463	B]	-0.0053	-0.0227	0.04	0.8460
[rs714825	B]	-	[rs715463	A]	-0.0053	-0.0227	0.04	0.8460
[rs714825	B]	-	[rs715463	B]	0.0053	0.0227	0.04	0.8460
[rs714825	A]	-	[rs2831700	A]	-0.0118	-0.0459	0.15	0.6949
[rs714825	A]	-	[rs2831700	B]	0.0118	0.0459	0.15	0.6949
[rs714825	B]	-	[rs2831700	A]	0.0118	0.0459	0.15	0.6949
[rs714825	B]	-	[rs2831700	B]	-0.0118	-0.0459	0.15	0.6949
[rs1075771	A]	-	[rs891700	A]	0.0123	0.0529	0.20	0.6512
[rs1075771	A]	-	[rs891700	B]	-0.0123	-0.0529	0.20	0.6512
[rs1075771	B]	-	[rs891700	A]	-0.0123	-0.0529	0.20	0.6512
[rs1075771	B]	-	[rs891700	B]	0.0123	0.0529	0.20	0.6512
[rs1075771	A]	-	[rs1335873	A]	0.0194	0.0811	0.47	0.4915
[rs1075771	A]	-	[rs1335873	C]	-0.0194	-0.0811	0.47	0.4915
[rs1075771	B]	-	[rs1335873	A]	-0.0194	-0.0813	0.48	0.4905
[rs1075771	B]	-	[rs1335873	C]	0.0194	0.0813	0.48	0.4905
[rs1075771	A]	-	[rs163781	A]	-0.0139	-0.0600	0.26	0.6082
[rs1075771	A]	-	[rs163781	B]	0.0139	0.0600	0.26	0.6082
[rs1075771	B]	-	[rs163781	A]	0.0139	0.0600	0.26	0.6082
[rs1075771	B]	-	[rs163781	B]	-0.0139	-0.0600	0.26	0.6082

Appendix 2. The result of multi-locus LD test. Allele B and C stand for allele G and T respectively. All Chi-square tests have one degree of freedom. The number of significant ($P < 0.05$) LD is zero

[Locus	Allele]	-	[Locus	Allele]	Burrows	Correlation	Chisq*	Prob.
[rs1075771	A]	-	[rs1294406	A]	0.0038	0.0139	0.01	0.9054
[rs1075771	A]	-	[rs1294406	B]	-0.0038	-0.0139	0.01	0.9054
[rs1075771	B]	-	[rs1294406	A]	-0.0038	-0.0139	0.01	0.9054
[rs1075771	B]	-	[rs1294406	B]	0.0038	0.0139	0.01	0.9054
[rs1075771	A]	-	[rs715463	A]	-0.0331	-0.1401	1.43	0.2314
[rs1075771	A]	-	[rs715463	B]	0.0331	0.1401	1.43	0.2314
[rs1075771	B]	-	[rs715463	A]	0.0331	0.1401	1.43	0.2314
[rs1075771	B]	-	[rs715463	B]	-0.0331	-0.1401	1.43	0.2314
[rs1075771	A]	-	[rs2831700	A]	0.0011	0.0044	0.00	0.9702
[rs1075771	A]	-	[rs2831700	B]	-0.0011	-0.0044	0.00	0.9702
[rs1075771	B]	-	[rs2831700	A]	-0.0011	-0.0044	0.00	0.9702
[rs1075771	B]	-	[rs2831700	B]	0.0011	0.0044	0.00	0.9702
[rs891700	A]	-	[rs1335873	A]	0.0020	0.0092	0.01	0.9375
[rs891700	A]	-	[rs1335873	C]	-0.0020	-0.0092	0.01	0.9375
[rs891700	B]	-	[rs1335873	A]	-0.0020	-0.0096	0.01	0.9352
[rs891700	B]	-	[rs1335873	C]	0.0020	0.0096	0.01	0.9352
[rs891700	A]	-	[rs163781	A]	-0.0093	-0.0444	0.14	0.7045
[rs891700	A]	-	[rs163781	B]	0.0093	0.0444	0.14	0.7045
[rs891700	B]	-	[rs163781	A]	0.0093	0.0444	0.14	0.7045
[rs891700	B]	-	[rs163781	B]	-0.0093	-0.0444	0.14	0.7045
[rs891700	A]	-	[rs1294406	A]	-0.0129	-0.0527	0.20	0.6524
[rs891700	A]	-	[rs1294406	B]	0.0129	0.0527	0.20	0.6524
[rs891700	B]	-	[rs1294406	A]	0.0129	0.0527	0.20	0.6524
[rs891700	B]	-	[rs1294406	B]	-0.0129	-0.0527	0.20	0.6524
[rs891700	A]	-	[rs715463	A]	-0.0230	-0.1077	0.85	0.3573
[rs891700	A]	-	[rs715463	B]	0.0230	0.1077	0.85	0.3573
[rs891700	B]	-	[rs715463	A]	0.0230	0.1077	0.85	0.3573
[rs891700	B]	-	[rs715463	B]	-0.0230	-0.1077	0.85	0.3573
[rs891700	A]	-	[rs2831700	A]	0.0078	0.0334	0.08	0.7753
[rs891700	A]	-	[rs2831700	B]	-0.0078	-0.0334	0.08	0.7753
[rs891700	B]	-	[rs2831700	A]	-0.0078	-0.0334	0.08	0.7753
[rs891700	B]	-	[rs2831700	B]	0.0078	0.0334	0.08	0.7753
[rs1335873	A]	-	[rs163781	A]	0.0220	0.1026	0.76	0.3842
[rs1335873	A]	-	[rs163781	B]	-0.0220	-0.1004	0.73	0.3944
[rs1335873	C]	-	[rs163781	A]	-0.0220	-0.1026	0.76	0.3842
[rs1335873	C]	-	[rs163781	B]	0.0220	0.1004	0.73	0.3944
[rs1335873	A]	-	[rs1294406	A]	0.0196	0.0787	0.45	0.5045
[rs1335873	A]	-	[rs1294406	B]	-0.0196	-0.0770	0.43	0.5133
[rs1335873	C]	-	[rs1294406	A]	-0.0196	-0.0787	0.45	0.5045
[rs1335873	C]	-	[rs1294406	B]	0.0196	0.0770	0.43	0.5133
[rs1335873	A]	-	[rs715463	A]	0.0172	0.0782	0.44	0.5072
[rs1335873	A]	-	[rs715463	B]	-0.0172	-0.0780	0.44	0.5080
[rs1335873	C]	-	[rs715463	A]	-0.0172	-0.0782	0.44	0.5072
[rs1335873	C]	-	[rs715463	B]	0.0172	0.0780	0.44	0.5080

Appendix 2 (con't). The result of multi-locus LD test.

[Locus	Allele]	-	[Locus	Allele]	Burrows	Correlation	Chisq*	Prob.
[rs163781	A]	-	[rs1294406	A]	0.0111	0.0456	0.15	0.6970
[rs163781	A]	-	[rs1294406	B]	-0.0111	-0.0456	0.15	0.6970
[rs163781	B]	-	[rs1294406	A]	-0.0111	-0.0456	0.15	0.6970
[rs163781	B]	-	[rs1294406	B]	0.0111	0.0456	0.15	0.6970
[rs163781	A]	-	[rs715463	A]	0.0252	0.1185	1.02	0.3114
[rs163781	A]	-	[rs715463	B]	-0.0252	-0.1185	1.02	0.3114
[rs163781	B]	-	[rs715463	A]	-0.0252	-0.1185	1.02	0.3114
[rs163781	B]	-	[rs715463	B]	0.0252	0.1185	1.02	0.3114
[rs163781	A]	-	[rs2831700	A]	-0.0159	-0.0683	0.34	0.5597
[rs163781	A]	-	[rs2831700	B]	0.0159	0.0683	0.34	0.5597
[rs163781	B]	-	[rs2831700	A]	0.0159	0.0683	0.34	0.5597
[rs163781	B]	-	[rs2831700	B]	-0.0159	-0.0683	0.34	0.5597
[rs1294406	A]	-	[rs715463	A]	0.0089	0.0359	0.09	0.7590
[rs1294406	A]	-	[rs715463	B]	-0.0089	-0.0359	0.09	0.7590
[rs1294406	B]	-	[rs715463	A]	-0.0089	-0.0359	0.09	0.7590
[rs1294406	B]	-	[rs715463	B]	0.0089	0.0359	0.09	0.7590
[rs1294406	A]	-	[rs2831700	A]	-0.0048	-0.0177	0.02	0.8799
[rs1294406	A]	-	[rs2831700	B]	0.0048	0.0177	0.02	0.8799
[rs1294406	B]	-	[rs2831700	A]	0.0048	0.0177	0.02	0.8799
[rs1294406	B]	-	[rs2831700	B]	-0.0048	-0.0177	0.02	0.8799
[rs715463	A]	-	[rs2831700	A]	-0.0144	-0.0609	0.27	0.6029
[rs715463	A]	-	[rs2831700	B]	0.0144	0.0609	0.27	0.6029
[rs715463	B]	-	[rs2831700	A]	0.0144	0.0609	0.27	0.6029
[rs715463	B]	-	[rs2831700	B]	-0.0144	-0.0609	0.27	0.6029

Appendix 2 (con't). The result of multi-locus LD test.