MULTIPLEXED GENOTYPING OF SINGLE NUCLEOTIDE POLYMORPHISMS USING MICROARRAY TECHNOLOGY

LI JINGGUANG

A THESIS SUBMITTED

FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

DEPARTMENT OF PAEDIATRICS

NATIONAL UNIVERSITY OF SINGAPORE

2005

FOR MY PARENTS,

MY WIFE, SON AND DAUGHTER,

AND ALL OTHER DEARESTS,

TO WHOM MY HEART BELONGS

ACKNOWLEDGEMENTS

SPECIAL THANKS GOES TO MY SUPERVISOR

DR. HENG CHEW KIAT

&

MY APPRECIATIONS ALSO GO TO OTHER MEMBERS OF OUR GROUP

Dr. Suman Lal Chirammal Sugunan

Miss Chen Zhen Cheng, Clarice

Miss Yi Hong

Miss Gan Chew Ping, Amy

Madam Hong, Sally

Miss Lee Siang Ling, Karen

Madam Fatimah Bte Mustafa

Miss Lye Hui Jen

Miss Zhou Mingyi

Miss Chan Mei Yen

Madam He Xuelian

Madam Ong Ee Tze, Debora

Miss Siti Rafeah Binte Mohamed Rafe

Miss Tan Hwee Ching

Miss Zhao Yulan

Miss Zhou Shuli

TABLE OF CONTENTS

SUMMARY

The first study was initiated to develop a multiplexed genotyping platform to study many candidate genes of coronary artery disease (CAD) in the hope that disease-susceptible SNPs could be rapidly identified. For this purpose, 20 representative SNPs were chosen to develop a microarray-based genotyping platform using minisequencing. Sequencing and RFLP were performed to serve as the gold standard and bench mark, respectively. The array platform could call 90.5% of all genotypes, demonstrating great potential in interrogating multiple SNPs. However, this platform can be improved if a better strategy is in use to replace the one-color approach, and if target amplification can be ensured and the probes are approximately designed.

In the second study, we devised a novel two-color strategy and integrated it with universal tag/anti-tag system (TAT) and liquid-phase primer extension. In brief, TAT allowed fabrication of universal arrays for any combination of SNPs, and liquid-phase primer extension facilitated efficient incorporation of ddNTPs, and the two-color scheme permitted all SNPs and some short insertions/deletions (INDELs) to be genotyped simultaneously using only two fluorescent ddNTPs. This platform (TATLIPEXA) showed a remarkable concordance rate of 100% with both sequencing and RFLP. Significantly, TATLIPEXA was potentially able to provide quantitative information on allele frequencies in pooled DNA samples.

To utilize the unique feature of terminal deoxyribonucleotidyl transferase (TdT) in catalyzing polymerization in a template-independent manner, we developed a novel tool (TdT-assisted probe elongation, TAPE) in the third study. By TAPE, oligonucleotide arrays could be fabricated with regular oligonucleotides. Also by TAPE, introduction of fluorescence was enabled. TAPE permitted detection of

10fmols of oligonucleotides. Significantly, TAPE could proceed very rapidly without using any sophisticated equipments. The application of TAPE in genotyping SNPs was demonstrated by running allele-specific minisequencing with regular ddNTPs, followed by TAPE to introduce Cyanine 5-ddCTP. All 5 representative SNPs were correctly genotyped using synthetic targets. RNA fragments could be directly elongated by TAPE, indicating that it could also be used for gene expression studies.

The fourth study was conducted because we observed that TdT could add fluorescent nucleotides into probes immobilized with 3'-NH_2, despite the absence of 3'-OH group. This activity was subsequently observed on oligonucleotides with 3'-biotin and 3'-C3 linker. These modifiers have in common, an additional -OH group on the linkers. In contrast, 3'-phosphorylated oligonucleotide was effective in blocking elongation by TdT because no extra -OH group was present. Thus, a common belief that 3' chemically modified oligonucleotides could not be elongated was challenged. This implied that -OH group which was not from DNA could also be recognized by TdT, and elongation by TdT could also proceed in a primer-independent manner. This study also found that classical phosphodiester bonds were formed during such elongation.

In summary, minisequencing was successfully adapted to solid support to genotype multiple SNPs, and a novel two-color strategy demonstrated great potential for genotyping all SNPs and some INDELs, and a versatile tool for both genotyping and gene expression was demonstrated. Moreover, another unusual property of TdT was discovered during these efforts.

Key words: single nucleotide polymorphisms (SNPs), microarray, minisequencing, tag/anti-tag system (TAT), liquid-phase primer extension, tag-anti-tag liquid phase primer extension array (TATLIPEXA), TdT-assisted probe elongation (TAPE)

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ASMS: allele-specific minisequencing

ASO: allele-specific oligonucleotide

CAD: coronary artery disease

CNPs: copy-number polymorphisms

ddNTPs: dideoxynucleotides

dNTPs: deoxynucleotides

Exo I: Exonuclease I

Exo III: Exonuclease III

FRET: fluorescence resonance energy transfer

gDNA: genomic DNA

HGP: human genome project

hME: homogenous MassEXTEND assay

INDELs: insertions or deletions

LCR: ligase chain reaction

LSO: locus-specific oligonucleotide

MALDI-TOF: matrix-assisted laser desorption/ionization time-of-flight

MBE: multiple-base extension

MIP: molecular inversion probe

OLA: oligonucleotide ligation assay

PCR: polymerase chain reaction

RFLP: restriction fragment length polymorphism

SAP: shrimp alkaline phosphatase

SBE: single base extension

SNPs: single nucleotide polymorphisms

TAPE: TdT-assisted probe elongation

TAT: tag/anti-tag system

TATLIPEXA: tag-anti-tag liquid phase primer extension array

TdT: terminal deoxyribonucleotidyl transferase

VNTRs: variable number tandem repeats

WGA: whole genome amplification

WGG: whole-genome genotyping

PUBLICATIONS AND PATENTS

1. **Jing-Guang Li**, Ulrika Liljedahl and Chew-Kiat Heng, Tag-anti-tag liquid-phase primer extension Array (TATLIPEXA): A flexible and versatile genotyping platform (*Genomics*, 87:151-157).

2. **Jing-Guang Li**, Chew-Kiat Heng, TAPE: a versatile tool for genetic analysis on oligonucleotide array (manuscript)

3. **Jing-Guang Li**, Chew-Kiat Heng, An unusual property of terminal deoxyribonucleotidyl transferase (TdT): Elongation of 3' chemically modified oligonucleotides (manuscript)

4. **Jing-Guang Li**, Chew-Kiat Heng, An innovative signal amplification strategy using TdT (terminal deoxyribonucleotidyl transferase)-assisted probe elongation (TAPE) (Patent pending)

CHAPTER 1 LITERATURE REVIEW

1.1 Genetic variations in human genome

The completion of the whole genome sequencing through Human Genome Project (HGP) is undoubtedly one of the most marvelous achievements in human history (Lander *et al.,* 2001; Venter *et al.,* 2001). Consequently, a reference genome sequence is now available. Logically, the next big inquiry of the post-genome era would be the interpretation of this so-called genetic blueprint. The secrets lie within the simple double-helix polymers composed of only four basic nucleotides, but it can be a very long journey for inspired scientists with innovative technologies to decipher the meanings behind such polymers.

In brief, the "central dogma of life" can be summarized as production of a particular protein is instructed by DNA via an intermediate called RNA. Thus, there are three potential candidate target molecules which may facilitate the elucidation of the lifebook, and they are DNA, RNA and protein. Only DNA will be discussed here.

It is estimated that 99.9% of the genome sequences between any two individuals are identical, and the remaining 0.1% are represented by many genetic variations (Wang *et al.,* 1998; Cargill *et al.,* 1999). Even though such variations are only a small fraction of the total genome, they are widely believed to be the reason why individuals are different from each other and why some are susceptible to certain diseases and resistant to certain drugs or even viruses like HIV while the others are not. For these reasons, genetic variations are greatly valued for their great potential applications in biomedical studies, drug development and medical diagnostics. Unfortunately, there is a huge gap between genetic variations and complex phenotypes. In order to bridge such gap, enormous funding from the public and

private sectors have been invested and a large number of scientists, converging from many fields, are intensely studying genetic variations, from which disease-genes are expected to be identified and personalized medicines developed. Such efforts are more explicitly expressed through the International HapMap Project, which is "a partnership of scientists and funding agencies from Canada, China, Japan, Nigeria, the United Kingdom and the United States to develop a public resource that will help researchers find genes associated with human disease and response to pharmaceuticals" (http://www.hapmap.org).

Many forms of genetic variations are randomly scattered throughout the whole genome. For example, some nucleotide sequences are repeated many times in tandem, and they are named variable number tandem repeats (VNTRs). Depending on the length, VNTRs can be further subdivided into microsatellites (usually 2-4 base pairs) and minisatellites (more than 12 base pairs). There are also many insertions or deletions (INDELs) of various sizes that are distributed widely across the genome, but they are less frequent than VNTRs. While many DNA sequences may vary in the number of copies (copy-number polymorphisms, CNPs) (Sebat *et al.,* 2004; Iafrate *et al.,* 2004), other chromosome segments may be clipped out, turned upside down and reinserted back into the chromosome (inversions) (Tuzun *et al.,* 2005). Nevertheless, the most abundant form of genetic variation identified thus far is single nucleotide polymorphisms, which accounts for about 90% of all variations in the genome (Sachidanandam *et al.,* 2001).

i) Single nucleotide polymorphisms (SNPs)

SNPs (pronounced "snips") are arbitrarily defined as single-base substitutions, insertions or deletions at a frequency of 1% or higher in the population, thus

distinguishing themselves from rare mutations (<1%) (Collins *et al.,* 1998; Brookes 1999). It was reported that more than five million SNPs have been validated thus far (Rebbeck *et al.,* 2004; Riva *et al.,* 2004).

A common belief is that SNPs result mainly from errors in DNA replication, despite the fact that such process is catalyzed and monitored by high-fidelity DNA polymerases. SNPs can be either transitions (from a purine to another purine, or from a pyrimidine to another pyrimidine) or transversions (from a purine to a pyrimidine, or vice versa). The former (G↔A and C↔T) are estimated to account for over two-thirds of all SNPs in the genome, and the remaining third are represented by the latter (C↔A, G↔T, C↔G and A↔T) (Wang *et al.,* 1998). SNPs are randomly distributed throughout the entire genome, including introns, exons and regulatory regions. Depending on whether the resultant amino acids are changed or not, SNPs located inside exons can be further subdivided into synonymous SNPs (no change) and non-synonymous SNPs (occurrence of amino acid change).

ii) The need for SNP genotyping

Even before the genome sequencing was completed, many researchers have focused their studies on genetic variations between individuals. These variations can serve as good physical markers for other analysis. More importantly, they are also believed to be involved in rendering individuals susceptibility to a particular common disease. This is known as the "common variants-common disease" hypothesis (Collins *et al.,* 1997). SNPs also confer important influence on responses to medicines (Tsuchihashi *et al.,* 2002). Thus, it is quite attractive and

also imperative to develop molecular technologies to identify SNPs associated with diseases and drug responses.

The challenges posed by associating SNPs with complex phenotypes can be seen in a few aspects. Five million SNPs have been validated thus far, but it is expected that a few million more will be identified. However, only a small fraction of these SNPs have potential influence on predisposition to diseases or resistance to drugs. SNPs that are potentially associated with common diseases are themselves common in both healthy population and sick people. From the point of diseases, the situation is also very complicated. Many diseases (e.g., acquired immunodeficiency syndrome, AIDS) can be largely attributed to environmental factors such as social culture, economic, education, ethic and etc. In contrast, other diseases (such as cystic fibrosis, thalessemia and haemophilia, etc) are essentially caused by genetic variations of specific genes. These are called Mendelian diseases. However, the pathogenesis of most diseases, such as cardiovascular disorders and cancers, is attributed to both genetic and environmental factors to different extent. Even if only genetic factors are considered, the association between these diseases and SNPs is still very difficult because they are more likely to be caused by multiple SNPs from a number of genes instead of a point mutation from a single gene. Thus, it usually requires genotyping of a large number of SNPs among many individuals by association studies. To meet these challenges, therefore, it is imperative to develop robust, accurate, high-throughput and affordable SNP genotyping technologies.

1.2 Main events in the analysis of SNPs

The analysis of SNPs embraces two different aspects, discovery and genotyping. Today, the complete SNP map of our genome is not available. In order to identify the candidate SNPs, many gene- or disease-focused studies are still ongoing. This constitutes the discovery aspect. For such purpose, sequencing is often employed, but a few alternative approaches are also widely in use. In principle, they are dependent on either differential electrophoretic mobility or chemical/enzymatic cleavage reactions (Fakhrai-Rad *et al.,* 2002; Lilleberg 2003; Clifford *et al.,* 2004; Twyman 2004; Suh *et al.,* 2005).

However, more attention in the last decade has been given to the identification of SNPs potentially associated with diseases or drug responses, motivated by both scientific and industrial purposes. This constitutes the aspect of genotyping. Thus far, dozens of robust genotyping technologies are commercially available in the analysis of SNPs (Syvanen 2001; Dearlove 2002; Chen *et al.,* 2003; Kwok *et al.,* 2003; Olivier 2005; Shen *et al.,* 2005; Tost *et al.,* 2005), and new approaches are continuously emerging because no method can satisfy the requirements of any linkage studies. On the other hand, many current technologies need to be further improved in order to simplify the analysis, increase the throughput or lower the cost, etc.

An ideal genotyping technology should have high sensitivity and specificity. In other words, it should permit accurate allele discrimination and detection through an appropriate format on an appropriate instrument. To meet the demands of large-scale association studies, multiplexing and automation should also be considered so that tens of thousands of SNPs can be interrogated simultaneously at an affordable price.

This section discusses the three main events during SNPs analysis, i.e., target amplification, allele discrimination and detection. It should be kept in mind that these

events may not be performed in such order. For example, allele discrimination by the gap-filling reaction of the molecular inversion probe (MIP) technology (ParAllele) is carried out before the amplification of the target by universal PCR (Hardenbol *et al.,* 2003). Since most technologies employ a target amplification strategy, Invader technology which amplifies the signal molecules will be discussed separately as an exception. As the focus of this section is on the general principles of each event, which is often directly represented by traditional genotyping methodologies, array-based technologies will not be discussed in detail.

i) Target amplification

As genomic DNA (gDNA) is normally a substance of very low abundance, amplification of DNA is usually considered first before any subsequent analysis. Fortunately, DNA regions of interest can be amplified at will by conventional PCR technology due to the complementary nature of double-stranded DNA (dsDNA) and our good understanding of DNA replication. For some unknown reasons, however, DNA amplification is rarely addressed when SNP genotyping approaches including high-throughput ones are reviewed. Frequently, only various allele discrimination biochemistries and detection approaches are meticulously discussed and compared. It is true that targets are almost unanimously prepared by locus-specific PCR amplification for most traditional low-throughput technologies. For high-throughput approaches, however, target amplification is very crucial for subsequent analysis. For these reasons, a few alternative amplification methods have been developed. Without an efficient target amplification procedure, many current platforms cannot be as robust and highly multiplexible as they are, which explains why such platforms endeavor to

introduce universal PCR primer sequences into the allele-discrimination oligonucleotides.

Currently, target amplification is indispensible for most genotyping approaches. Other than locus-specific PCR, universal PCR amplification and whole genome amplification have also been successfully developed and widely applied to large-scale or genome-wide association applications.

a. Locus-specific PCR

Usually, gDNA regions containing SNPs of interest are amplified by many PCR reactions with multiple locus-specific primers. This amplification strategy is very simple and suitable for most gene- or diseases-focused association studies with a limited number of custom SNPs. Indeed, it is still in use by some commercial high-throughput platforms such as homogenous MassEXTEND assay (hME, Sequenom). Obviously, this approach is very time-consuming and labour-intensive, and it will consume a large amount of gDNA which is often precious and limiting. To overcome this problem, multiplex PCR has been developed, by which a few loci of gDNA can be amplified simultaneously in a single reaction. However, the capacity of multiplex PCR to amplify multiple targets simultaneously is limited. As such, it is not well-suited for comprehensive association studies to analyze hundreds or thousands of SNPs.

b. Universal PCR

The application of universal PCR to amplify multiple loci of interest in a single reaction has significantly promoted the development of several

commercial high-throughput genotyping platforms such as SNPlex technology (Applied Biosystems) (De la Vega *et al.,* 2005), GoldenGate Array (Illumina) (Shen *et al.,* 2005) and MIP technology (Hardenbol *et al.,* 2005). The allele discrimination of these technologies is achieved by either oligonucleotide ligation assay (SNPlex), or allele-specific primer extension (GoldenGate), or single-base primer extension and ligation (MIP). Regardless of the different allele discrimination biochemistry employed, one common effect of these approaches is to join two universal PCR primer sequences which are originally located on separate oligonucleotides. Only if proper allele discrimination reaction takes place should both PCR primers be joined at the 5' and 3' ends of the full-length product, which is usually composed of locus-specific sequences, SNP site and a universal sequence. Following this, the full-length product can be amplified by a universal PCR primer pair. The main advantage of universal PCR amplification is that hundreds to thousands of loci of interest can be prepared simultaneously by one reaction. For example, GoldenGate permits 1,536 SNPs to be interrogated in a single reaction, and MIP has an even higher limit of 12,000 SNPs. In average, less gDNA is required by universal PCR to produce sufficient target for each genotyping. For instance, only 250-500ng of gDNA is required by GoldenGate and MIP technologies to interrogate 1,536 and 12,000 SNPs, respectively. Compared to typical locus-specific PCR, universal PCR amplifies much smaller regions of DNA flanking the SNP sites. Consequently, genomic sequence complexity is significantly reduced to facilitate accurate and specific genotyping. Similar to locus-specific PCR, however, amplification by universal PCR still requires a large number of

tagged locus-specific oligonucleotides. For each SNP, for example, SNPlex requires three locus-specific oligonucleotides and five universal probes.

c. Whole genome amplification

Genome-wide genotyping of hundreds of thousands of SNPs have been made possible by the development of whole genome amplification (WGA) technologies, which can be either PCR-dependent or entirely PCR-free (Hawkins *et al.,* 2002; Hughes *et al.,* 2005; Tzvetkov *et al.,* 2005). For example, GeneChip Mapping Array (Affymetrix) for whole genome analysis is based on PCR-WGA (Kennedy *et al.,* 2003; Matsuzaki *et al.,* 2004). Briefly, gDNA is randomly cut by restriction enzymes into numerous fragments, which are subsequently flanked with the common adaptor sequence through ligation. Following this, the whole genome can be amplified with only one generic primer to generate sufficient DNA fragments for genome-wide SNP mapping. Recently, scientists from Illumina also reported a bead array-based whole-genome genotyping assay (WGG), by which no PCR is required at all (Gunderson *et al.,* 2005). The most obvious advantage of WGA over universal PCR is the complete elimination of locus-specific primers, which can number up to tens of thousand for MIP assay. Today, only WGA-based genotyping technologies are potentially able to map the whole genome, it is thus expected that more WGA-dependent SNP genotyping technologies will emerge in the near future.

ii) Allele discrimination

Allele discrimination is a procedure by which the allele of a given sample is identified. Undoubtedly, this is the heart of genotyping and has considerable impact on the whole assay. For example, the specificity of any approach is largely determined by allele-discrimination biochemistry employed at this stage. Allele discrimination can be achieved by many strategies, but only a few of them are discussed here, which include hybridization, primer extension, ligation and nuclease cleavage. Indeed, many commercial genotyping platforms are based on one or more of these four mechanisms (Table 1.1).

Table 1.1 Allele-discrimination chemistries of some commercial genotyping systems

Chemistries	Commercial platforms*						
	GeneChip	TaqMan	MIP	Invader	SNPlex	GoldenGate	hME
hybridization	√	√					
single-base extension (SBE)			√				√
multiple-base extension (MBE)						√	√
ligation			√		√	√	
nuclease cleavage		√		√			

* MIP: molecular inversion probe; hME: Homogenous MassEXTEND assay

a. Hybridization

The application of hybridization to analyze SNPs is based on two properties of DNA. First, two complementary oligonucleotides will hybridize to form DNA duplex. Second, perfectly matched duplex can be separated from single-mismatched duplex under appropriate conditions due to their different thermodynamics (Hall *et al.,* 2001; Urakawa *et al.,* 2003). This can be explained by the speculation that DNA duplex formation proceeds through a

transient nucleation complex from the interaction of very few base pairs and not from all base pairs simultaneously (Southern *et al.,* 1999). There are quite a few genotyping approaches that are solely dependent on hybridization, including GeneChip Mapping Array (Khrapko *et al.,* 1991; Southern 1992; Guo *et al.,* 1994; Kennedy *et al.,* 2003). Indeed, TaqMan (Applied Biosystems) also largely depends on allele-specific hybridization (Livak 1999), even though it is often classified into other category. However, the discrimination power of allele-specific hybridization alone is very limited. Affymetrix addressed this problem with the use of redundant probes, which means that multiple different oligonucleotide probes are designed to interrogate the same SNP (Lipshutz *et al.,* 1999). For example, each SNP is represented by about 40 probes on GeneChip Mapping Array.

b. Primer extension

Genotyping SNPs through primer extension is a simulation of the *in vivo* process of polymerase-mediated DNA synthesis during cell division. The greatest benefit of this strategy can be seen in that accurate and specific result can be obtained due to the high-fidelity feature of DNA polymerases. Many DNA polymerases also have 3' exonuclease activities, which further ensure incorporation of proper nucleotides when the primers perfectly anneal just 5' or exactly at the SNP site on target DNA. The allele discrimination power of primer extension was found at least nine times higher than that of hybridization (Pastinen *et al.,* 1997). Another frequently overlooked advantage of primer extension is that the design of primers is very straightforward.

Numerous forms of primer extensions are currently in employment in the analysis of SNPs. For simplicity, they can be divided into two categories according to the nucleotides in use, i.e., single-base extension (SBE) by dideoxynucleotides (ddNTPs) and multiple-base extension (MBE) by deoxynucleotides (dNTPs). SBE is often referred to minisequencing, but it is not appropriate if single-base extension is achieved with dNTPs. For example, the gap-fill reaction of MIP assay is essentially single-base extension with dNTPs followed by ligation (Hardenbol *et al.,* 2003). Minisequencing is usually performed with fluorescently labeled ddNTPs, which facilitates direct allele identification of the target DNA. If unlabeled ddNTPs are used, the extended products can be resolved alternatively by mass spectrometry (Griffin *et al.,* 2000).

If two primers differing only in the last base at the 3' hybridize to proper template, the primer with 3' perfectly matched base will be extended by DNA polymerases at a much higher efficiency (10^2-10^4) than the primer with 3' mismatched base. To take advantage of this feature, allele-specific primer extension methods have been developed to analyze SNPs. If an additional primer is introduced, the locus with SNP can be amplified exponentially, which is termed allele-specific PCR (Myakishev *et al.,* 2001). Compared to minisequencing, allele-specific primer extension has lower discrimination power, but it allows both single-base and multiple-base extensions. One obvious benefit of MBE over SBE is that it permits incorporation of multiple fluorescently labeled dNTPs (Pastinen *et al.,* 2000; Gunderson *et al.,* 2005). Indeed, SBE and MBE can even be combined into a single reaction, for which an in-depth review is given later through the hME assay (Storm *et al.,* 2003).

c. Ligation

Ligase is mainly responsible for repairing DNA nicks by joining two adjacent oligonucleotides at the junction if they hybridize perfectly to the DNA template. This function of DNA ligase has been fully utilized to genotype SNPs through a few schemes. A typical method to interrogate SNPs by ligase is achieved through two allele-specific oligonucleotides and one common oligonucleotide, which is termed oligonucleotide ligation assay (OLA) (Landegren *et al.,* 1988). Some commercial genotyping platforms like SNPlex are entirely based on this strategy. Genotyping by OLA is very specific and accurate. gDNA can be potentially genotyped by other ligase-mediated technologies such as ligase chain reaction (LCR) (Barany 1991) and padlock probes (Nilsson *et al.,* 2002; Baner *et al.,* 2003; Landegren *et al.,* 2004), by which prior PCR amplification of target can be completely eliminated. An ingenious modification of the typical ligation reaction is the introduction of SBE prior to ligation (Hardenbol *et al.,* 2003). By such way, only two oligonucleotides are required to interrogate each SNP. More importantly, the incoming nucleotide (SNP) will be double-checked by both polymerase and ligase.

d. Nuclease cleavage

SNP analysis through allele-specific cleavage by DNA nucleases has been in use for quite a long time. The first generation of genetic marker map was indeed obtained by PCR-restriction fragment length polymorphism (PCR-RFLP) strategy, which utilizes restriction endonucleases to recognize and cleave allele-specific sequences with high specificity (Weber 1990). However,

both cleavage and separation by gel electrophoresis are very time consuming and labor intensive, and PCR-RFLP is not amenable to multiplexing. Moreover, certain sequences with SNPs cannot be cleaved by appropriate restriction endonucleases.

The intrinsic 5' nuclease activity of Taq DNA polymerase has also been successfully utilized for SNPs analysis, as can be seen by TaqMan assay (Livak 1999). In principle, both allele-specific probes carrying one reporter and one quencher at either end can hybridize to DNA template under PCR condition. During chain elongation, the perfectly matched probe will be degraded by Taq DNA polymerase, thus releasing the reporter from quencher to emit fluorescence. In contrast, the mismatched probe will remain intact and the reporter is still in the quencher's proximity. This approach is very robust because SNPs can be genotyped in a closed-tube, but its allele discrimination power is limited because it depends largely on the different hybridization efficiencies between perfectly matched probe and mismatched probe. Moreover, probes with dual modifications are costly. This limitation was recently addressed by the introduction of universal reporter with dual labels (Rickert *et al.,* 2004). TaqMan is also limited in its low compatibility with highly multiplexed applications.

Invader assay is the most powerful nuclease-mediated genotyping method, for which a comprehensive discussion is given later.

iii) Allele detection

Numerous assay formats are currently available to seperate and detect alleles because many properties of the oligonucleotides and some by-products after

allele-discrimination reactions are detectable. Allele discrimination products may differ in size and mass. For example, cleavage products by RFLP are separated by gel because of the different mobilities conferred by the size. Similary, detection of SNPlex is accomplished through capillary electrophoresis technology, which also depends on the mobility of allele discrimination product. The application of matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) mass spectrometry to detect alleles is based on the fact that different alleles differ in the mass, as can be seen by hME assay. Pyrosequencing technology takes advantage of pyrophosphate produced from primer extension reaction to analyze SNPs (Ronaghi *et al.,* 1996). The most straightforward approach for allele identification is achieved by capturing the signals from the fluorescent lables, which is widely used by many commercial platforms such as GeneChip Mapping Array and GoldenGate. Other effects such as fluorescence resonance energy transfer (FRET) (Livak 1999; Olivier 2005) and fluorescence polarization (Kwok, 2002) have also been utilized in the detection of allele-specific products.

It is true that the choice of allele detection format is largely determined by the biochemistry of allele discrimination, but detection device is often the bottleneck of the whole assay for some technologies. For example, minisequencing products can be accurately identified by MALDI-TOF, but MBE and ligation products of GoldenGate assay are not suitable to be detected similarly. On the other hand, minisequencing can be carried out in a highly multiplexed format, but detection by MALDI-TOF can only guarantee accurate analysis of oligonucleotides less than 40 bases in length (Kaetzke *et al.,* 2002). For this reason, hME assay permits only 24-plex at the moment. The multiplexing level of SNPlex (48-plex) is also largely limited by capillary electrophoresis.

iv) Examples of genotyping technologies

a. Homogenous MassEXTEND assay (Sequenom)

Homogenous MassEXTEND assay (hME) is a typical primer extension-based genotyping technology. The uniqueness of hME assay can be seen in two aspects. First, allele discrimination is achieved simultaneously by minisequening and MBE. Second, the detection of different alleles is based on the masses of the extended primers by MALDI-TOF mass spectromety.

Typically, three procedures are carried out to analyze SNPs by MALDI-TOF (Griffin *et al.,* 2000; Pusch *et al.,* 2002; Blondal *et al.,* 2003; Storm *et al.,* 2003). First, locus-specific PCR is performed to amplify DNA regions containing SNPs of interest. Following this, redundant dNTPs will be digested by shrimp alkaline phosphatase (SAP) to prevent their interference with subsequent primer extension reaction. Secondly, a common MassEXTEND primer and a cocktail (typically containing one ddNTP and three dNTPs) are introduced to initiate primer extension reaction. Depending on the sequence downstream the SNP site and which ddNTP is present, the primer will be extended by 1–4 bases. As a result, the extention products of different alleles and the intact primer differ from each other in length. This way, allele discrimination is accomplished. Thirdly, following the removal of extraneous salts by resin, extension products are analyzed by MALDI-TOF to determine the genotypes by mass.

The hME assay is very easy to operate and it allows multiple SNPs to be analyzed in a single reaction. As allele discrimination is mediated by high-fidelity DNA polymerase, and genotype calling is fulfilled in real time by

MassArray software, highly reproducible results can be obtained. Significantly, detection of alleles is dependent on their masses, thus eliminating the requirement of fluorescently labeled ddNTP or dNTPs for primer extension reaction. However, mass-dependent detection by MALDI-TOF also determines that the multiplexing capacity of hME assay is limited (Ross *et al.,* 1998; Nordhoff *et al.,* 2000; Kaetzke *et al.,* 2002). Even though an advanced version of hME assay called iPLEX has been launched recently, it still permits maximum 24 SNPs to be interrogated simultaneously (http://www.sequenom.com). The typical hME assay is also limited by the requirement of locus-specific PCR amplifications with multiple primers. If the noise signals from unextended primers are addressed, hME assay can be further improved. Moreover, its cost per genotype is relatively higher than other platforms. Nevertheless, hME assay is a wise choice if moderate SNPs are studied in a large number of samples.

b. SNPlex (Applied Biosystems)

SNPlex genotyping system depends exclusively on oligonucleotide ligation assay (OLA) to discriminate alleles, and the detection of different alleles is achieved by capillary electrophoresis technology. An all-inclusive review on both SNPlex and TaqMan technologies was published recently (De la Vega *et al.,* 2005).

Like conventional OLA, SNPlex requires two allele-specific oligonucleotides (ASOs) and one locus-specific oligonucleotide (LSO) for allele discrimination reaction. However, both ASOs of SNPlex are tagged with a unique ZipCode sequence, and the LSO is tagged with a universal PCR primer sequence. In

order to introduce another PCR primer sequence, two universal allele-specific oligonucleotide linkers (ASO-linker) are designed. If a particular ASO-linker and its matching ASO are successfully ligated to proper LSO and its matching LSO-linker, both universal PCR primer sequences and the unique ZipCode sequence will be joined to make up a full-length oligonucleotide which also contains the gene-specific sequences from both ASO and LSO. Thus, allele discrimination is accomplished. If a large number of SNPs are genotyped in this way, all full-length oligonucleotides can be subsequently amplified by a common PCR primer pair. One of the universal primer is biotinylated so that only one strand of the amplified full-length amplicons can be captured by streptavidin-coated plates. The main function of turning double-stranded amplicons into single-stranded is to make subsequent entrapment of ZipChute probes more efficient. ZipChute probes are fluorescently labeled, and they are composed of a universal sequence which is complementary to the allele-specific ZipCode sequence of the captured amplicons and a unique modifier which determines the mobility rate during electrophoresis. After being released from the plates, ZipChute probes can be detected by Applied Biosystems 3730 DNA Analyzer.

This ligation-based technology does not require locus-specific amplification by PCR, and it permits 48 SNPs to be genotyped simultaneously. As the detection of different alleles is based on a pair of unique and universal probes with fluorescent label, it can be easily adapted to any fine-mapping studies with custom SNPs. Significantly, seperation and detection carried out by capillary electrophoresis instruments is also very advantageous because they are widely available. Among all high-throughput technologies, however,

SNPlex might be the most complicated one because it involves eight steps to discriminate and detect alleles (De la Vega *et al.,* 2005). For each SNP, moreover, eight oligonucleotides are required, including three gene-specific probes, three universal hairpin-like linkers and two fluorescently labeled probes. Other than ligase, many other enzymes are also involved, such as kinase, exonuclease I, λ-exonuclease and DNA polymerase. Hence, SNPlex can be improved in many ways, including its multiplexing level. As the workflow of SNPlex is largely automated, however, it is suitable for moderate association studies of candidate genes or fine-mapping of custom SNPs on a larger sample size.

c. Invader assay (ThirdWave)

Invader assay requires three oligonucleotides for each SNP, i.e., one invasive probe and two allele-specific signal probes. The invasive probe is complementary to the upstream region of gDNA except the last base (SNP site), which can be any nucleotide. The signal probes are composed of a gene-specific sequence and a universal flap, and they differ from each other by the SNP site. The invasive and signal probes overlap by one base at the SNP site. To minimize background signal, the 3' ends of the signal probes are usually blocked with an amino group (Kwiatkowski *et al.,* 1999; Lyamichev *et al.,* 1999; Hall *et al.,* 2000).

A typical Invader assay contains primary and secondary reactions (de Arruda *et al.,* 2002; Lyamichev *et al.,* 2003). During primary reaction, when locus-specific sequence of the signal probe perfectly matches gDNA, the probe will be cleaved by structure-specific 5' nuclease like Cleavase (ThirdWave) to

release its 5' universal flap. As this reaction is carried out at a temperature similar to the melting temperature of the signal probes, cleaved signal probes will be replaced with intact ones to initiate many rounds of cleavage, thus producing multiple flaps. If the signal probe does not match gDNA, its 5' flap cannot be released. Thus, the primary reaction has two functions, allele discrimination and flap generation. The flaps will serve as invasive probes in the secondary reaction, during which the flap will hybridize to a hairpin-like FRET cassette. Consequently, the FRET cassette will be cleaved. As the cassette contains a reporter and a quencher in close proximity, cleavage will release the reporter from the quencher so that fluorescent signals can be produced. Because this reaction also permits cleaved cassettes to be replaced by uncleaved ones, numerous FRET cassettes will be cleaved by one copy of 5' flap to release many reporters into the solution.

In essence, Invader technology is a signal amplification approach which presents a few advantages over other genotyping methods. For example, gDNA can be potentially genotyped without prior PCR amplification. As both reactions are carried out under isothermal conditions, no dedicated and expensive instruments are required. Moreover, it can be easily automated because it is a homogeneous assay. To some extent, Invader assay is a high-throughput approach because it allows analysis of a large number of samples per day. However, one main limitation of this technology is that it is not amenable to be highly multiplexed. Invader assay was recently adapted to solid supports to analyze multiple SNPs in a multiplexed manner, but it has yet to be promising for wide use in biomedical studies (Wilkins Stevens *et al.,* 2001; Stevens *et al.,* 2003; Olivier, 2005).

1.3 Genotyping SNPs by microarray technology

Despite the fact that numerous genotyping technologies have been successfully developed and widely applied in many association studies during the last decade, none of them have satisfactorily fulfilled the call for high-throughput genotyping until the introduction of microarray technology, which promises to analyze hundreds of thousands of SNPs simultaneously on a tiny solid support. In 1992, an allele-specific oligonucleotide hybridization array was successfully developed for SNP discovery and detection (Southern 1992).

i) Features of microarray

In terms of the mechanisms of the three main events involved in SNP genotyping (target amplification, allele discrimination and detection), microarray technology differs little from traditional methodologies. In essence, microarray is just a very powerful platform which aligns millions of probes to facilitate accurate sorting of hundreds of thousands of SNPs through hybridization. In this sense, it is appropriate to describe the replacement of traditional approaches with microarray-based ones as a simple shift of "from gel to chip". However, it is this shift that makes many previously impossible tasks possible, such as genome-wide mapping. For example, DNA target is usually prepared through multiple locus-specific PCR by most traditional methods. For one reason, only a limited number of SNPs are genotyped in large populations by most case-control studies. In this situation, locus-specific PCR is well-suited for target preparation. On the other hand, allele detection by traditional methods such as gel electrophoresis usually require a large amount of amplicons. Instead, microarray is potentially able to analyze a large number SNPs as mentioned above. In order to provide corresponding number of

targets more efficiently for downstream, highly multiplexed detection by microarray, universal PCR is frequently employed. Although locus-specific PCR is still possible, it can be very time-consuming and labor-intensive since many locus-specific PCR have to be carried out seperately. Undoubtedly, the last decade has also witnessed the emergence of a few new approaches to prepare targets. For instance, some whole genome amplification methods can produce sufficient targets from very little gDNA for subsequent analysis, thus completely eliminating the barriers of PCR.

Like traditional methods, allele discrimination by microarray-based systems is also fundamentally based on allele-specific hybridization, high-fidelity polymerase-mediated primer extension or ligase-mediated reaction. The main differences are often seen in how these reactions are carried out. For instance, allele discrimination is traditionally performed after the DNA target is amplified. Instead, many commercial microarry platforms run these reactions in the reverse order, as in the cases of GoldenGate and MIP technologies. Due to this simple shift of the reaction order, preparation of multiple targets is greatly simplified in the form of a universal PCR instead of multiple locus-specific ones. Compared to traditional methods, moreover, allele discrimination by microarray-based ones is often carried out in a highly multiplexed manner and during this event, universal PCR primer sequences are introduced to facilitate subsequent target preparation. Frequently, another universal sequence is also introduced at this stage. The purpose of doing this will be discussed shortly.

The greatest advantages of microarry technology over conventional ones are seen in its robust and efficient separation and detection of different alleles. With the use of a light-directed chemical process termed photolithography, oligonucleotide

probes (typically 25 bases in length) can be chemically synthesized at specific locations at a density of over 100 million. In this way, millions of different probes can be manufactured on a small glass slide called GeneChip. Significantly, each probe on the array is addressable, which is crucial for accurate genotyping (Fodor *et al.,* 1991). GeneChip provides the most powerful platform obtainable thus far for any nucleic acids-based studies, and its contribution to biomedical studies can never be underscored. With a given number of GeneChip arrays, the whole genome can be accessed to the largest degree via hybridization. As mentioned above, another important feature of many microarray methods is the introduction of universal sequence which is inserted into allele-/SNP-specific oligonucleotide(s). This unique sequence serves as a surrogate of a particular allele or SNP, and detection of such sequence directly signifies the presence of the corresponding allele or SNP. It has different names under different platforms, such as "Address" (GoldenGate), "ZipCode and ZipChute" (SNPlex) and "Tag" (MIP). The greatest benefits of introducing surrogate sequence can be seen in two aspects. First, it allows multiple SNPs to be interrogated through a universal array in a highly flexible manner. This means that if a generic array is fabricated, it can be applied to any association studies with different SNPs. Second, genotyping through surrogate sequence can be more accurate and specific because the genomic sequence complexity is largely reduced. Unfortunately, this advantage of surrogate sequence has not been well recognized. Microarray technology can never be so successful without the introduction of fluorescence for detection. Traditional gel assay requires several picomoles of amplicons for allele detection. Through fluorescence, however, microarray enables detection of attomoles to femtomoles of DNA targets. Hence, the detection power is increased by at least

three orders, which greatly facilitates genotyping of enormous SNPs simultaneously.

ii) Workflow of microarray

Typically, genotyping by microarray technology consists of three consecutive events: array manufacture, arrayed hybridization and signal reading/data analysis.

a. Array manufacture

Microarray is usually fabricated by either *in situ* synthesis or immobilization of pre-synthesized oligonucleotides. *In situ* synthesis is well known as the Affymetrix approach, which directly synthesizes oligonucleotides on a small glass slide through photolithography (Fodor *et al.,* 1991). This process is highly automated and permits production of many arrays with consistent and reproducible features. However, this kind of array is still the most expensive one in the market, even though its price has been reduced substantially. Despite the fact that custom array can be easily manufactured due to its flexibility, it is still not cost-effective if an array with much fewer features is needed.

Alternatively, pre-synthesized probes can be immobilized on glass slides by either contact printing or inkjet dispensing (Hacia *et al.,* 1998; Hacia 1999). Compared to *in situ* synthesis, the advantages of probe immobilization can be seen in its ease of prototyping, rapid implementation, versatility and low cost. Prior to being immobilized, however, each probe has to be chemically synthesized and modified with anchors such as amino group. If many probes are required, the cost can also be very high. As many factors (e.g., incubation,

humidity and washing) are difficult to control, arrays prepared by immobilization are not very consistent.

Depending on the nature and size of the probes positioned on the glass slide, microarray can be simply divided into cDNA array and oligonucleotide array. The cDNA array is often used to monitor gene expression, but oligonucleotide array can find wide applications in both gene expression and genotyping studies.

b. Arrayed hybridization

In essence, the concept of microarray technology is the same as that of Southern blotting, which investigates genetic information through hybridization between complementary oligonucleotides.

It is true that the power of allele discrimination by hybridization alone is limited. Without hybridization, however, high-throughput genotyping by microarray-based platforms is not possible. Microarray technologies depend solely on hybridization to sort a huge number of DNA targets which bear specific sequences complementary to their corresponding probes on the chip. Usually, the analyte may contain millions of different DNA targets of different sizes. Through hybridization alone, these DNA targets can be efficiently and accurately separated from each other with probes that are only 25 nucleotides in length (GeneChip). It is no doubt that microarray provides the most powerful biological sifter that cannot be surpassed, but the successful application of microarray technology to analyze genetic information is made possible by its underlying principle, hybridization.

It should be highlighted that sorting by surrogate sequence has also greatly improved the performance of microarray-based genotyping platforms. By such a method, genomic sequence complexity of the analyte is significantly reduced, thus facilitating efficient and accurate hybridization.

Besides sorting, hybridization is also utilized to discriminate alleles by GeneChip Mapping Array. To overcome its limited power in the identification of single base mismatches, Affymetrix designs redundant probes to interrogate each SNP so that the specificity is greatly improved.

c. Signal reading and data analysis

Allele detection by microarray technologies is usually achieved through the capturing of the fluorescence signals by confocal scanners. These approaches are mainly based on either one-color or four-color detection strategy. Usually, more probes or arrays are required if one-color strategy is in use. By the four-color approach, one probe and one array are sufficient to interrogate all genotypes of each SNP. For example, one-color is employed by GeneChip Mapping Array, which requires allele-specific probes for each SNP (Kennedy *et al.,* 2003). MIP technology was first reported to call all genotypes by one-color detection via four separate arrays (Hardenbol *et al.,* 2003). To overcome the variations between arrays, a four-color version was recently reported (Hardenbol *et al.,* 2005).

Analysis of the signals from microarray-based genotyping technologies is always accomplished by some bioinformatic tools, which ensures accurate and efficient genotype calling.

iii) Array-based genotyping platforms

At present, several commercial oligonucleotide array-based platforms are widely in use and they are mainly based on allele-specific hybridization, primer extension, ligation or the latter two.

a. GeneChip Mapping Array (Affymetrix)

GeneChip Mapping Array is the most successful hybridization-based genotyping platform which claims to be able to generate more than 100,000 genotypes by a set of two arrays (Kennedy *et al.,* 2003; Matsuzaki *et al.,* 2004; Matsuzaki *et al.,* 2004). One distinctive feature of this innovative technology can be seen in that the whole genome is amplified by a universal PCR with only one generic primer. The assay begins with random digestion of gDNA by restriction enzymes (XbaI or HindIII) to generate fragments of different sizes. Following this, adaptors are ligated to all DNA fragments, regardless of their sizes. Subsequently, PCR is performed with only one generic primer which recognizes the adaptor sequence. Although all fragments with adaptor sequences are potentially able to be amplified, those falling between 250bp and 2,000bp are preferred under controlled PCR conditions. The amplicons are then fragmented and end-labeled with biotinylated ddATP before they are injected into the microarray cartridge for hybridization to the probes. Finally, the array is stained and scanned to acquire fluorescence signals.

The main procedures and reactions of this genotyping technology are represented schematically in Figure 1.1 (adapted from Kennedy *et al.,* 2003). There are a few advantages of this assay. First, only 250 ng of gDNA is required from a given sample for almost genome-wide genotyping. Secondly,

only a single primer is required to amplify all fragments of interest by a universal PCR. Thirdly, all procedures are automated and 100,000 genotypes can be called with an accuracy of more than 95%.

Since this technology is designed for genome-wide association studies, all probes on the chip are gene-specific sequences. Thus, it is not a wise choice for linkage studies with small number of SNPs or with SNPs that are not covered by the chip. Currently, each SNP is interrogated by about 40 different probes on the array. If this problem is addressed, the throughput per array can be significantly increased. Following hybridization, staining is required to label DNA fragments because they are previously end-labeled with biotinylated ddATP. If these fragments are directly end-labeled with fluorescently labeled ddNTP or dNTP, the assay can be simplified and its sensitivity may be improved. Moreover, sophisticated algorithm is indispensable for the analysis of the enormous amount of data obtained (Liu *et al.,* 2003).

To expand its applications, Affymetrix has recently acquired ParAllele Bioscience and thus its proprietary molecular inversion probe (MIP) technology. The current version of MIP allows 12,000 SNPs to be genotyped simultaneously (Hardenbol *et al.,* 2005). Once integrated with Affymetrix's universal tag array, MIP technology will offer scientists with unprecedented access to any area of the genome.

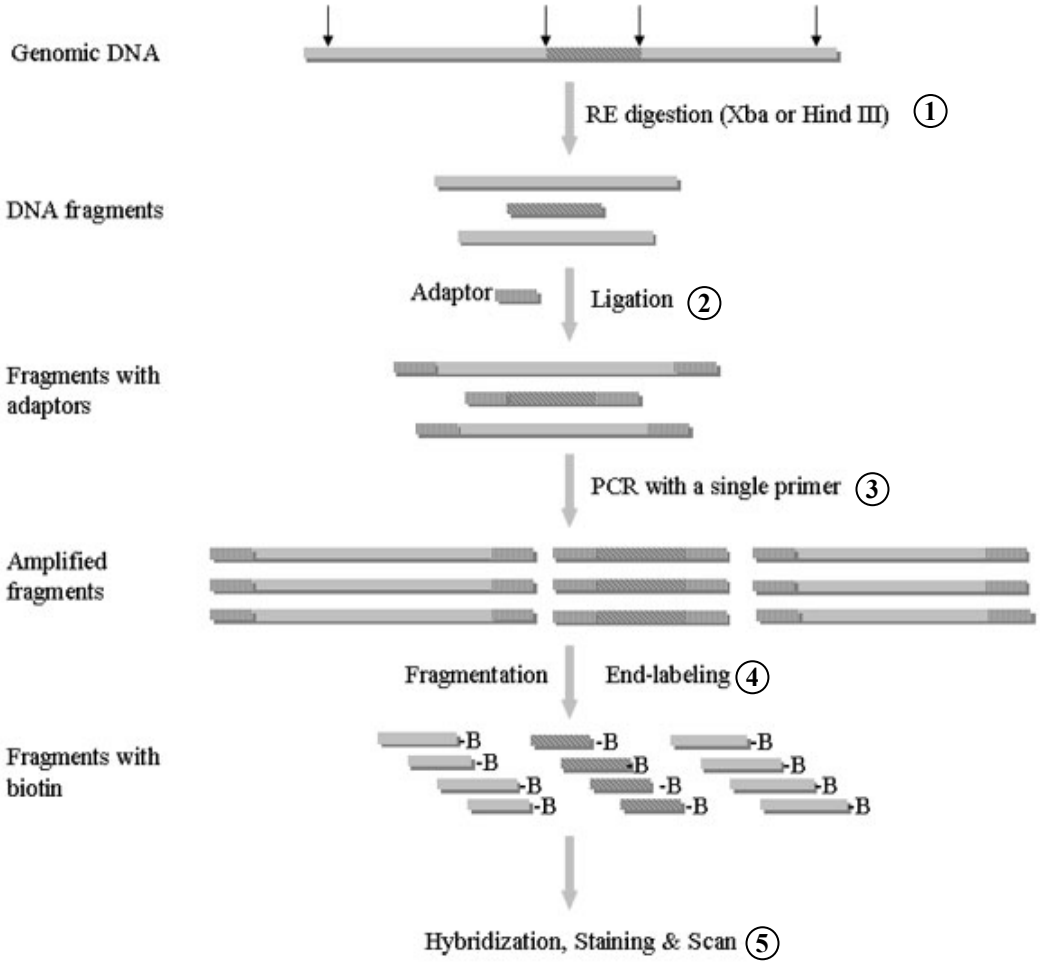

Figure 1.1 Schematic representation of Affymetrix's GeneChip Mapping Array. Genomic DNA is fragmented by RE (1) so that a common adaptor sequence can be introduced by ligation (2). Thus, all fragments with such adaptor can be amplified by PCR with a common primer (3). After another fragmentation procedure, DNA fragments are end-labelled with biotin-dATP (4). Finally, hybridization, staining and scanning are performed (5).

b. GoldenGate assay (Illumina)

The GoldenGate assay comprises three main events (Figure 1.2) (adapted from Shen *et al.,* 2005). The first event includes two reactions, primer extension and ligation. Allele discrimination is achieved through polymerase-mediated primer extension. For each SNP, two allele-specific oligonucleotides (ASOs) and a common locus-specific oligonucleotide (LSO) are required. Besides the gene-specific sequence, all oligonucleotides also contain a universal primer

sequence for subsequent PCR amplification. Insertion of an additional unique address sequence into LSO is to enable ligated and amplified oligonucleotides to hybridize to probes on Sentrix Array Matrix or BeadChip. As LSO hybridizes a few bases downstream from the SNP site, multiple nucleotides will be incorporated into appropriate ASO perfectly matching gDNA. If the last base is not complementary to gDNA, such ASO cannot be extended. Following this, extended ASO will be joined with downstream LSO by DNA ligase. Consequently, full-length ASO-LSO oligonucleotides (containing two universal primer sequences, two locus-specific sequences, one unique address sequence and a few gene-specific bases) is generated. The second event is to amplify full-length ASO-LSO oligonucleotides with three universal PCR primers, Cyanine 3-P1, Cyanine 5-P2 and unlabeled P3. Thirdly, Cy3- and/or Cy5- labeled amplicons hybridize to corresponding probes on Sentrix Array Matrix or BeadChip via the unique address sequences so that fluorescence can be detected by BeadArray Reader for subsequent genotype calling.

Like GeneChip Mapping Array, only 250ng of gDNA is required by GoldenGate to interrogate up to 1,536 SNPs in a single reaction. In terms of allele discrimination, GoldenGate is essentially based on polymerase-mediated primer extension, regardless of the use of ligase. The checkpoint of this assay is polymerase-mediated allele-specific primer extension. Only if the last incorporated nucleotide matches the gDNA, extended ASO will be ligated with LSO. In this sense, allele discrimination by ligase is limited. Thus, the main function of ligase is to ligate extended ASO and LSO so that universal PCR primer sequences and unique address sequence are joined together.

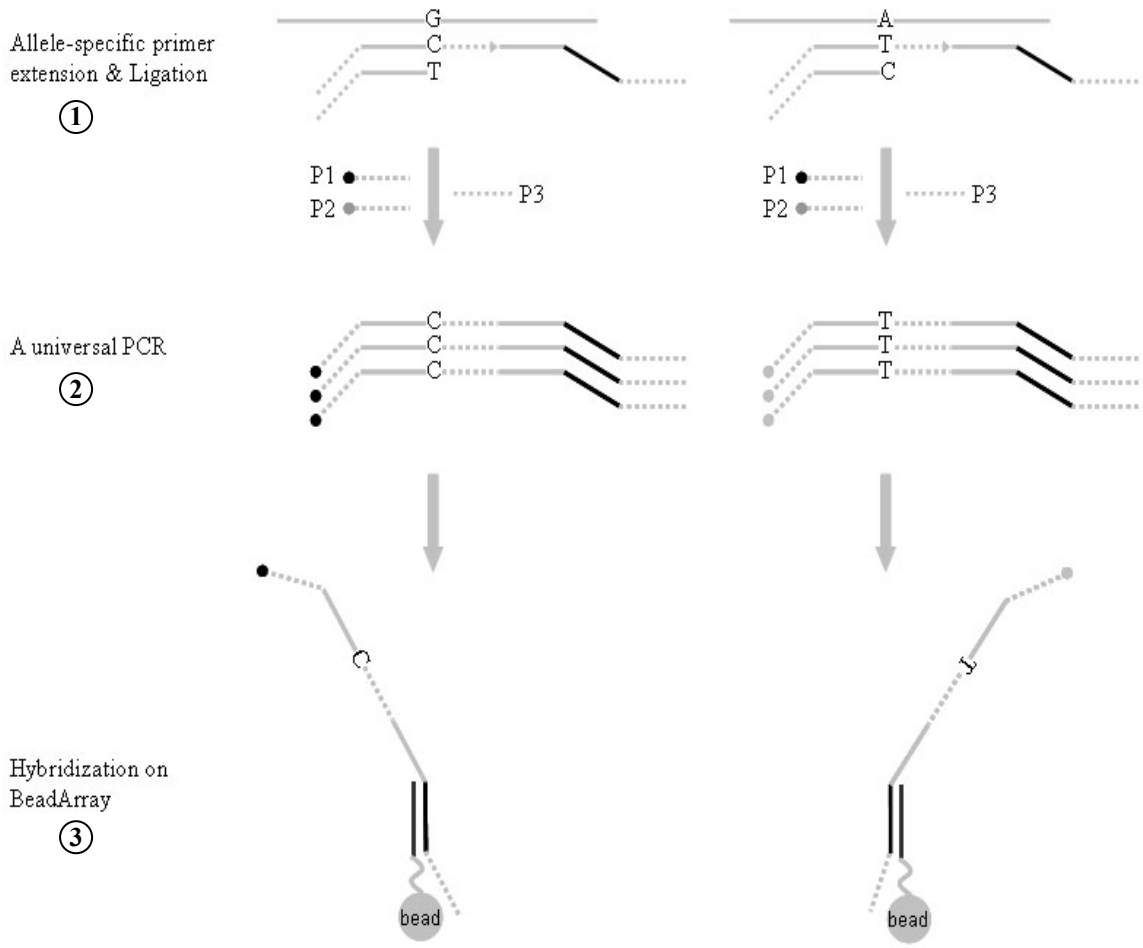

Figure 1.2 Schematic representation of ILLUMINA's GoldGate assay. Allele-specific primer extension and ligation are carried out to amplifiy genomic DNA (1), followed by a universal PCR with two upstream primers (Cy3- or Cy5-labelled) to generate sufficient targets (2) for hybridization, which is dependent on the unique sequence previously introduced in the downstream LSO (3).

Alternatively, if LSO is designed to hybridize exactly next to the SNP site, ligation can occur directly when appropriate ASO and LSO are present. As such, both allele discrimination and generation of full-length oligonucleotides can be accomplished via ligation alone, thus eliminating the use of polymerase. Indeed, no obvious advantage can be seen from using primer extension and ligation to discriminate alleles over ligation alone, if polymerase is not superior to ligase with regard to allele discrimination power.

GoldenGate permits 1,536 SNPs to be interrogated in a single reaction. Compared to MIP (12,000-plex), the multiplexing level of GoldenGate is relatively lower. This can be partially attributed to the requirement of three oligonucleotides for each SNP. Since the locus-specific sequences of ASOs differ by only one nucleotide, they may compete for hybridization to gDNA, which in return prevent efficient extension of the perfectly matched ASO. If a common upstream oligonucleotide and a downstream LSO are designed to hybridize to the regions exactly before and after the SNP site, the single-base gap can be filled by primer extension and ligation. In this manner, it is expected that the multiplexing level can be increased as each SNP requires only two oligonucleotides. More importantly, base calling can be improved because the discrimination can be double-checked by both polymerase and ligase.

c. Molecular inversion probe (ParAllele)

As another enzyme-dependent genotyping technology, MIP is very different because its allele discrimination is achieved by both polymerase-mediated primer extension and ligase-mediated ligation. As such, if misincorporation by primer extension occurs, ligation cannot proceed.

Originated from padlock probe (Nilsson *et al.,* 1994; Nilsson *et al.,* 2002), MIP probe is composed of seven regions, including two gene-specific sequences, two primer sequences common to all probes, two cleavage sites for all probes and one tag sequence specific to each SNP (Hardenbol *et al.,* 2003).

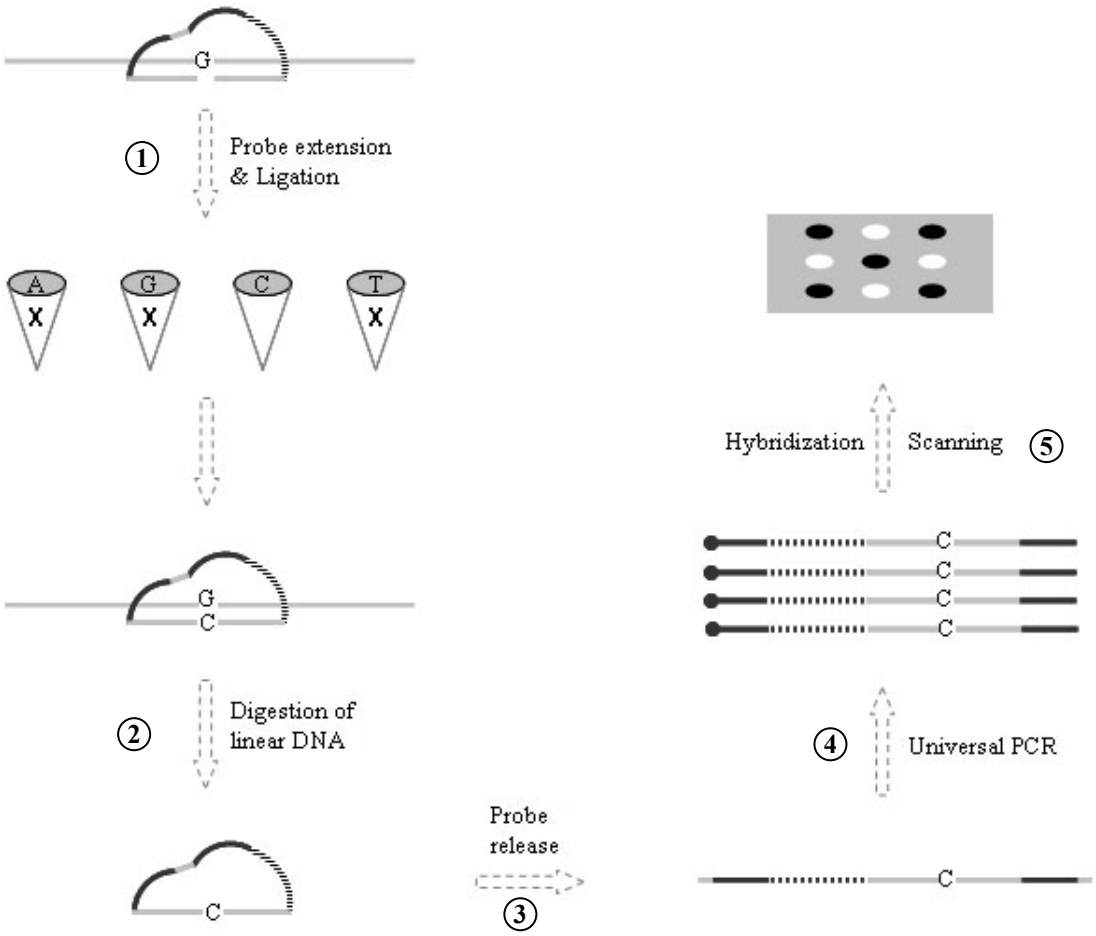

Figure 1.3 Overview of molecular inversion probe (MIP) technology. Allele-specific primer is circularized only if corresponding dNTP is present (1). In this case, the primer for G allele is closed only in the tube with dCTP. Subsequently, all liner DNA (including genomic DNA and primer) will be digested by Exo I and III (2). Following this, a specific restriction enzyme is introduced to cleave the circular probe (3) so that it can be amplified by a universal PCR (4). Finally, hybridization and scanning are performed for base calling (5).

Briefly, MIP assay can be divided into 4 steps (Hardenbol *et al.,* 2005). First, MIP probe hybridizes to a particular site of gDNA through its gene-specific sequences to form a circular structure with a single-base gap for dATP, dGTP, dCTP or dTTP. If the corresponding site of gDNA is G (Figure 1.3, adapted from Hardenbol *et al.,* 2003 and Hardenbol *et al.,* 2005), MIP probe will be extended with a single dCTP by DNA polymerase and subsequently circularized covalently by ligase under controlled conditions, a process called gap-fill. With the introduction of exonucleases I and III (Exo I and Exo III),

gDNA and linear MIP probes will be digested. Secondly, ligated MIP probes will be cleaved through a common site between the universal PCR primer sequences. Consequently, the primer sequences are seperated from each other by gene-specific sequences and the specific tag sequence. Thirdly, following this, inverted MIP probes are amplified with the common primer sequences by a universal PCR. Finally, after treatment with a restriction endonuclease to release the gene-specific sequences, all reaction products from four tubes are pooled and loaded on Affymetrix's GenFlex Tag Array to perform hybridization. Only fluorescently labeled tag sequence of MIP probes can be detected.

As only one probe is required for each SNP, MIP permits highly multiplexing to genotype 12,000 SNPs in a single reaction with a concordance rate higher than 99% (Hardenbol *et al.,* 2005). Allele detection of MIP is achieved through universal tag sequences, thus it is very flexible and can be easily adapted to any custom SNP studies. This somehow explains why Affymetrix paid $120 million to acquire ParAllele because the integration of MIP into its own universal tag array is potentially able to allow scientists to conduct virtually any experiments.

However, one limitation of MIP is the requirement of a large number of probes. Originally, MIP required four arrays for genotype calling due to its one-color strategy (Hardenbol *et al.,* 2003). To address this problem, an intermediate two-color version was developed. Under such strategy, A/C and G/T were labeled with different fluorophores. Before hybridization, two pools were prepared (one for A/G, the other for C/T). In this manner, only two arrays are required, but feature-to-feature variations between arrays are still

present. The current version of MIP uses a four-color strategy, by which each allele is labeled differentially with a spectrally distinct fluorophore. Consequently, all reaction products can be pooled together before hybridization on a single array. As such, the problematic array-to-array variations of the first two versions of MIP is completely solved.

1.4 Objectives of the present studies

In this thesis, my main aim was to evaluate and develop methodologies for highly multiplexed genotyping of single nucleotide polymorphisms using microarray technology.

The first study (Chapter 2) was an effort to adapt polymerase-mediated minisequencing to solid-phase platform (microarray).

In the second study (Chapter 3), I intended to develop a simple but versatile genotyping platform by integrating a novel two-color strategy with universal tag/anti-tag system and liquid-phase minisequencing.

The third study (Chapter 4) was an attempt to explore one unusual property of TdT, i.e., catalysis of oligonucleotide elongation in DNA template-independent manner, to develop a versatile tool capable of array fabrication and signal introduction for both genotyping and gene expression studies.

During the third study, more unusual features of TdT were observed, which thus became the topic of the fourth study (Chapter 5).

CHAPTER 2 DEVELOPMENT OF A CHIP-BASED PLATFORM FOR SIMULTANEOUS GENOTYPING OF 20 SNPS IN CORONARY ARTERY DISEASE CANDIDATE GENES

2.1 Introduction

The most common genetic variations in human genome are in the form of single nucleotide polymorphisms (SNPs). SNPs are defined as single-base substitutions, insertions or deletions in genomic DNA at a frequency of 1% or higher. They represent about 90% of all genetic variations and are estimated to occur at an average frequency of one per thousand nucleotides (Lindroos *et al.,* 2002). It is widely believed that SNPs are associated with many common diseases such as Alzheimer's, Parkinson's and cardiovascular diseases. In addition, SNPs will enable personalized drugs to become a reality in the field of pharmacogenomics. The significance of SNPs is also highlighted by the efforts to develop the third generation of genetic marker map with SNPs, which will substitute restriction fragment length polymorphism map (RFLP, first generation) and microsatellite map (second generation) (Weber 1990; Dib *et al.,* 1996; Wang *et al.,* 1998). With rapid accumulation of knowledge on the association between SNPs and multi-factorial diseases with complex inheritance patterns, the diagnosis, prevention, treatment of diseases and drug development will enter a new stage (Zhang *et al.,* 1997).

Thus far, a number of molecular technologies have come into use for rapid and accurate genotyping of SNPs (Shi, 2001). Due to the strong and specific base pairing feature of complementary DNA, DNA molecules themselves can serve as the ideal reagents to identify DNA sequences of interest. Based on this, some hybridization-based methods have been successfully developed for SNP genotyping (Khrapko *et*

al., 1991; Southern 1992). Unfortunately, the power of allele discrimination by hybridization alone is limited because it depends solely on the thermal stabilities of mismatched and perfectly matched hybrids. Alternatively, more genotyping approaches utilizes high-fidelity enzymes to access SNPs such as DNA polymerase and ligase. DNA polymerase-mediated genotyping methods are typically exemplified by minisequencing, by which the primers annealing just 5' to the nucleotide of interest on target DNA are extended with one of the four ddNTPs (Jalanko *et al.,* 1992). Another widely used method is called oligonucleotide ligation assay (OLA), which is mediated by DNA ligase (Landegren *et al.,* 1988). Genotyping SNPs by these methods have been proven to be very accurate and highly reproducible. However, they are usually carried out in liquid-phase, which does not allow genotyping of multiple SNPs simultaneously. With the advent of microarray technology, more efforts have been shifted towards development of solid-phase based genotyping approaches as they are promising to interrogate hundreds of thousands of SNPs simultaneously. This is very important for rapid identification of disease-susceptible SNPs by large-scale and genome-wide association studies.

This study is one of such efforts which attempts to adapt minisequencing to solid platform in the hope that many SNPs in candidate genes of coronary artery disease can be simultaneously genotyped for risk assessment.

2.2 Materials and methods

Amplification of genomic DNA

Genomic DNA was isolated from peripheral blood leukocytes of 10 subjects by standard phenol-chloroform method (Kunkel et al., 1977) and the concentration was

estimated by ND-1000 Spectrophotometer. Twenty SNPs were selected from some candidate genes associated with coronary artery disease previously studied by conventional methods in our group (Table 2.1). Most of these SNPs had been previously reported (Chamberlain et al., 1989; Bujo et al., 1991; Evans et al., 1993; Suzuki et al., 1994; Valdenaire et al., 1995; Wang et al., 1995; Lane et al., 1996; Weiss et al., 1996; Henry et al., 1997; Wilson et al., 1997; Carmena-Ramon et al., 1998; Cullen et al., 1998; Hibi et al., 1998; Sanghera et al., 1998; Grainger et al., 1999; Dachet et al., 2000), only two SNPs (ELN422 and ELN290) were identified in our laboratory. The primers were purchased from Alpha DNA (Canada) and dissolved to a concentration of 200μM (Table 2.1). PCR amplification was set up in a total volume of 20μl in order to generate sufficient amplicons for both sequencing and RFLP reaction. Typically, the mixture contained 0.2 units of DyNAzyme™ II DNA Polymerase (Finnzymes, Finland), 20pmols of primers, 2.5mM dNTPs and 50-100ng of genomic DNA. PCR was performed in an automated thermal cycler (Biometra) with an initial denaturation at 95°C (5 min), followed by 15 cycles of 95°C (30 sec), 63°C (1 min), 72°C (1 min) and another 20 cycles of 95°C (30 sec), 55°C (1 min), 72°C (1 min). The reaction was terminated after final extension at 72°C (10 min).

Genotyping by sequencing

The PCR products were purified using QIAquick PCR Purification Kit (Qiagen, USA). First, one volume of PCR solution was mixed with 5 volumes of PB buffer in a clean microcentrifuge tube. This mixture was loaded into the QIAquick column provided for DNA binding, followed by centrifugation (14,000rpm, 2 min) to discard the flow-through. Subsequently, 0.75ml PE buffer was introduced to wash the column and centrifugation was repeated to remove the buffer. To release DNA, 30 μl EB buffer

Table 2.1 SNPs, restriction enzymes and PCR primers of this study

SNPs		Restriction enzymes	PCR primers	
			forward	reverse
ANGR	T/G	*Dde I*	gagattgcatttctgtcagt	ataatgtaagctcatccacc
ApoAI-75	G/A	*Msp I*	cagagctgatccttgaactcttaag	ggacacctacccgtcaggaagagca
ApoAI+83	C/T			
ApoBgc	C/T	*Apa I*	ataaggcatgtggtgttgaggcgc	ccgtgcctggtgcaaacacacaag
ApoBtz	C/T	*EcoR I*	ctgagagaagtgtcttcgaaag	ctcgaaaggaagtgtaatcac
ApoE112	T/C	*Hha I*	ctgggcgcggacatggaggacgt	gatggcgctgaggccgcgctcg
ECE	C/T	*MspA1 I*	ggcctcatcaccgatcccca	tctcgtcctcagccccttcc
Eln422	G/A	*BstN I*	ggcctgggggaaatttacatcctc	acattcaagcccgatctacagtcc
Eln290	T/C			
ENOS	G/T	*Mbo I*	catgaggctcagccccagaac	ccagtcaatccctttggtgctc
F13	C/T	*BsmF I*	ttgcctgtcattatctctgg	gaaccacaccattgttagcttac
FVIIT	C/T	*Nci I*	gggagactccccaaatatcac	acgcagccttggctttctctc
GP3A	T/C	*Nci I*	tctgattgctggacttctctttg	tctctccccgcaaagagtcc
HFE63	C/G	*Dpn II*	acatggttaaggcctgttgc	gccacatctggcttgaaatt
LCAT	C/G	*Msp I*	accatcagagacgaagggaaggg	agggacgtcattcctctaagggac
LPLPvu	C/T	*Pvu II*	atggcacccatgtgtaaggtg	gtgaacttctgataacaatctc
PAI-1	G/A	*BseL I*	cacagagagagtctggccacgt	gcagccagccacgtgattgtctag
PON2	C/G	*Dde I*	acatgcatgtacggtggtcttata	gttaagttatcgcactttcatgcc
TGFb509	C/T	*Bsu36 I*	cagactctagagactgtcag	gtcaccagagaaagaggac
TNFa	G/A	*Eco130 I*	aataggttttgagggccatg	tctggaggaagcggtagtgg

Except two novel SNPs (ELN422 and ELN290) which were identified in our laboratory, all other SNPs had been previously reported.
LCAT and PAI-1 were actually single-nucleotide insertions/deletions.

was added to the column's membrane. After incubation (2 min) and centrifugation, the flow-through was collected for downstream sequencing. The sequencing reaction was carried out using BigDye™ Terminator Cycle Sequencing Ready Reaction Kit (ABI, USA). Briefly, one microliter of purified PCR product, 3.2 pmols of primers, 2µl of BigDye and 1µl of buffer were mixed and topped up with ddH_2O to a 20µl cocktail. This mixture was subjected to 25 cycles in an automated thermal cycler (Biometra), each cycle consisting of 95°C (10 sec), 50°C (5 sec) and 60°C (4 min). After precipitation with ethanol, 12µl of Hi-Di formamide was added into the reaction solution before being loaded into ABI 3100 sequencer.

Genotyping by RFLP

Typically, an aliquot of PCR products was digested with an appropriate restriction enzyme (Table 2.1) in a humidified incubator (37°C, overnight). Subsequently, digestion products were separated by an agarose gel incorporated with ethidium bromide (0.2µg/ml) in 1xTBE buffer (89mM Tris base, 89mM boric acid and 2mM EDTA). Finally, fragments of different sizes were visualized on a UV transilluminator (Mighty Bright, Hoefer Scientific) and photographed by a Polaroid MP-4 camera system.

Genotyping by chip-based minisequencing

The probes to be immobilized were 15 to 20-mer gene-specific oligonucleotides tailed with a spacer (15dT) and an amino-linker at 5' ends (Table 2.2). The immobilization was based on the Schiff's base reaction between amino and aldehyde molecules. Prior to spotting, the probes were re-suspended with ArrayIt™ Micro-Spotting Solution (TeleChem, USA) in a 384-well plate (40µM). They were spotted in quadruplicates

in a 2 X 2 format by PixSys7500 system (Cartesian Technologies, USA) on the surface of aldehyde-coated slides (CEL Associates, Inc.). On each slide, there were 5 parallel subarrays for three synthetic target pools, one sample and one negative control (Figure 2.1). After spotting, the slides were incubated in an air-tight box and placed in a humidified incubator (37°C, overnight). The washing of the slides proceeded sequentially with 0.2% SDS (2 min), water (2 min), boiling water (2 min), sodium borohybride solution (1.0g $NaBH_4$ dissolved in 300 ml phosphate buffered saline and 100 ml pure ethanol, 5 min), 0.2% SDS (1 min) and water (1 min). After air-drying, the slides were kept in the fridge at 4°C.

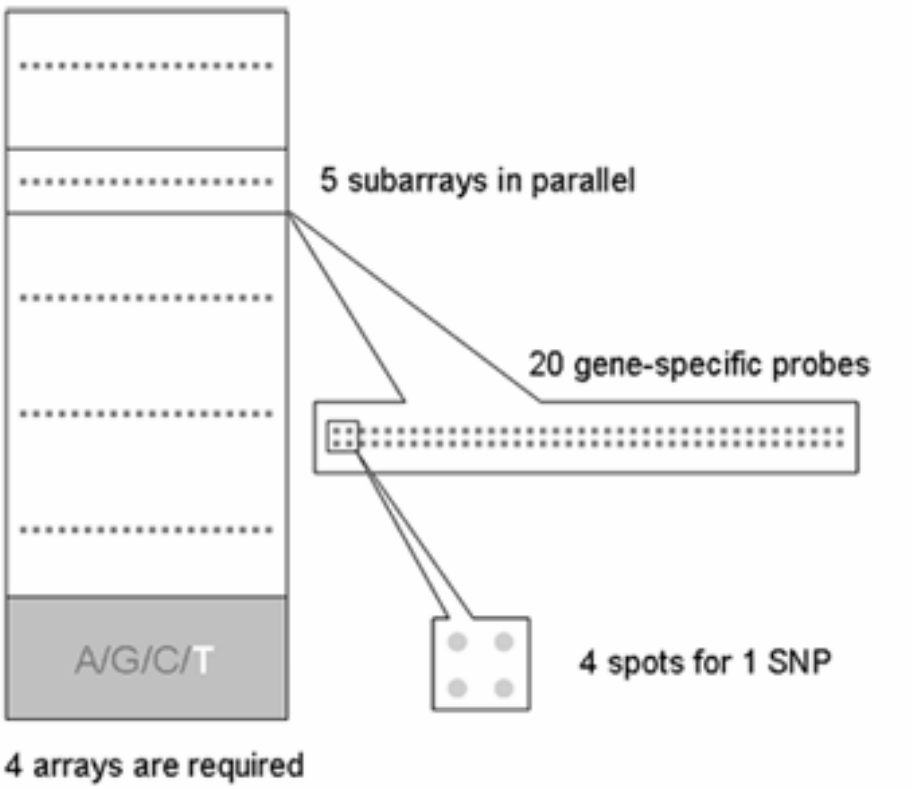

Figure 2.1 Schematic representation of array fabrication with gene-specific probes

Table 2.2 Gene-specific probes, synthetic targets and asymmetric PCR primers

SNPs	Gene-specific probes [a] (Tm, GC content)	Synthetic targets [b]	Asymmetric PCR primers	
			forward [c]	reverse
ANGR	ttttttttttttttttcaattctgaaaagtagctaa (45 °C, 30%)	gcagcacttcactaccaaatgagcC/Attagctacttttca gaattgaaggagaaaatgcattatgt	gcagcacttcactaccaaatgag	tccttcaattctgaaaagtagctaa
ApoAI-75	ttttttttttttttttcagtgagcagcaacagggcc (58 °C, 65%)	ctgggctgggaggctgataagcccagcccC/Tggccct gttgctgctcactggtcctggcaatgtggaactt	gaggctgataagcccagcc	cagtgagcagcaacagggc
ApoAI+83	ttttttttttttttttcaggccttgccccaggc (55 °C, 76%)	gggagaagacctcaggtacccagaggcccG/Agcctg gggcaaggcctgaaccttgagctggggagccagag	acctcaggtacccagaggcc	gttcaggccttgccccag
ApoBgc	ttttttttttttttttgaagaccagccagtgca (50 °C, 59%)	cagggttgaagccatacacctctttcaggG/Atgcactg gctggtcttcaggatgaagctgcagagctgggg	ggttgaagccatacacctctttcag	cctgaagaccagccagtgca
ApoBtz	ttttttttttttttttgactttcatatggaatt (36 °C, 29%)	tgtttgatggcttggtacgagttactcaaG/Aaattccatat gaaagtcaagcatctgattgactcactcat	gatggcttggtacgagttactcaa	agatgcttgactttcatatggaat
ApoE112	ttttttttttttttttgcgcggacatggaggacgtg (60 °C, 70%)	ggcctgcacctcgccgcggtactgcaccaggcggccgc A/Gcacgtcctccatgtccgcgcccagccgggcc	gtactgcaccaggcggcc	gggcgcggacatggagga

ECE	ttttttttttttttccaaggagttctcagaacacttc (54 °C, 48%)	gttcatgggtgagccaggtgggcagcG/Agaagtgttc tgagaactccttggaattggagagggagccgatga	tgggtgagccaggtgggc	ttccaaggagttctcagaacac
Eln422	ttttttttttttttgagtcgcaggtgtccct (51 °C, 65%)	acctccgactccgggaacacctccgacacC/Taggga cacctgcgactccagggatacctccgactccgaca	ccgggaacacctccgacac	ctggagtcgcaggtgtccct
Eln290	ttttttttttttttcccgtgagccttagtcaca (53 °C, 58%)	ccccatcccttctcaacccatgtccccagA/Ttgtgacta aggctcacgggaaatgccaactcccgggacac	ccttctcaacccatgtcccc	catttcccgtgagccttagtca
ENOS	ttttttttttttttctgctgcaggccccagatga (59 °C, 65%)	tcgggggggcagaaggaagagttctgggggA/Ctcatct ggggcctgcagcagcaggggcagcacgtcgaagc	gcagaaggaagagttctgggg	ctgctgcaggccccagat
F13	ttttttttttttttcaccttctacaccggggtcc (54 °C, 65%)	cgaacgtctccttcttgaattctgccttcA/Gggaccccg gtgtagaaggtgatgttggctgagagataagc	cgtctccttcttgaattctgcc	catcaccttctacaccgggg
FVIIT	ttttttttttttttcaggtaccacgtgccc (48 °C, 69%)	cagtggaggcccacatgccacccactaccG/Agggca cgtggtacctgacgggcatcgtcagctggggccag	ggcccacatgccacccac	cccgtcaggtaccacgtgc
GP3A	ttttttttttttttgtcttacaggccctgcctc (53 °C, 63%)	ccttcaggtcacagcgaggtgagcccG/Agaggcagg gcctgtaagacaggagcccaaagagaagtccagcaa	aggtcacagcgaggtgagcc	cctgtcttacaggccctgcc
HFE63	ttttttttttttttgctgttcgtgttctatgat	agttcggggctccacacggcgactctcatG/Catcatag	gctccacacggcgactct	gaccagctgttcgtgttctatg

	(48 °C, 42%)	aacacgaacagctggtcatccacgtagcccaaa		
LCAT	ttttttttttttttctcaatgtgctcttccccc (52 °C, 58%)	tactgagctcagccttgggcgtggtgtgcG/Cgggggaagagcacattgaggagccagaaggggggcggcagg	tcagccttgggcgtggtg	gctcctcaatgtgctcttcc
LPLPvu	ttttttttttttttaatgtctcttcatctttag (44 °C, 30%)	cagaagaacaacaacaaaaccccacagctG/Actaaaagatgaagagacattgaattgagaattaagcaaaa	caacaacaaaaccccacagc	ctcaattcaatgtctcttcatcttt
PAI-1	ttttttttttttttagtctggacacgtgggg (50 °C, 65%)	cggccgcctccgatgatacacggctgacT/Cccccacgtgtccagactctctctgtgcccctgagggctct	cctccgatgatacacggctg	gagagagtctggacacgtggg
PON2	ttttttttttttttccgcatccagaacattctat (51 °C, 45%)	cataaactgtagtcactgtaggcttctcaC/Gatagaatgttctggatgcggagaacctattcaggggatac	aaactgtagtcactgtaggcttctc	tctccgcatccagaacattc
TGFb509	ttttttttttttttgcctcctgacccttccatcc (55 °C, 65%)	gtgggaggaggggcaacaggacacctgaG/Aggatggaagggtcaggaggcagacacctgtaagaattgct	ggaggggcaacaggacac	ctgcctcctgacccttccat
TNFa	ttttttttttttttcaataggttttgaggggcatg (52 °C, 48%)	tgtaggaccctggaggctgaaccccgtccC/Tcatgcccctcaaaacctattgcctccatttcttttgggga	ctggaggctgaaccccgtc	aggcaataggttttgaggggc

[a] All gene-specific probes were chemically modified with amino at the 5' ends so that they could be immobilized on aldehyde-coated slide. The 15-dT tail was introduced to facilitate hybridization by reducing steric hindrance effect. Tm stands for melting temperature.
[b] Allele-specific synthetic oligonucleotides (70-mer) differed from each other by only a single nucleotide (upper case).
[c] The forward primers were designed to generate the target strands for hybridization to the probes immobilized on the slides.

Two 70-mer oligonucleotides differing by only a single base served to simulate allele-specific targets of each SNP (Table 2.2), and three pools were prepared from them to represent all genotypes. Genomic samples were first amplified by asymmetric PCR with another set of primers (Table 2.2) to generate targets of 40-50 base pairs in length. Typically, 20µl cocktail contained Taq DNA polymerase (Qiagen) (0.5U), Q-solution (4µl), dNTPs (2.5mM), 50-100ng of genomic DNA, forward primer (40pmols) and reverse primer (5pmols). This mixture was denatured at 95°C (5 min), followed by 50 cycles of 95°C (15 sec), 50°C (15 sec) and 72°C (15 sec). The amplicons were then pooled and precipitated with 95% ethanol and sodium acetate (3M, pH 5.2) (-70°C, 1 hour). After centrifugation (14,000rpm, 10 min) to remove the supernatant, 500ml of 70% ethanol was introduced and another centrifugation was performed (14,000rpm, 5 min). Finally, the pellet was re-suspended with 4µl TE and 16µl Arrayit™ UniHyb™ Hybridization Solution (TeleChem, USA).

Four microliter of pooled synthetic targets, concentrated PCR products or ddH_2O was loaded on one of the subarrays on the slide. The subarrays were separated from each other by cover slips. The slide was then transferred into a hybridization cassette and immersed in a water bath (50^oC, 2 hours). To remove redundant targets, the slide was washed with 2 X SSC+0.2% SDS (3 min), 0.2 X SSC+0.2% SDS (3 min), 0.2 X SSC (2 min) and dH_2O (2 min). After air-drying, freshly prepared minisequencing mixture was loaded on the slide, which typically contained ThermoSequenase DNA polymerase (Amersham) (3U), TAMRA-labelled ddNTP (10pmols) and the other three regular ddNTPs (10pmols/ddNTP). The slide was transferred into the cassette again to incubate at 50^oC (20 min) in the water bath, followed by washing with TENT buffer (10 mM Tris-HCl, 1 mM EDTA, 200 mM NaCl, 0.1% Tween-20) (3 min), 50 mM NaOH solution (3 min) and dH_2O (3 min). After air-drying, the slide was

scanned by ScanArray® 5000 scanner (Packard BioScience Ltd, UK). The laser power was kept constant at 95%, whereas the typical setting for the photo-multiplier tube (PMT) was 80%.

Data analysis

The images obtained were analyzed by QuantArray® 3.0 (Packard BioScience Ltd, UK), and the cut-off values were set at 1.96 x standard deviation above the mean values of the corresponding SNPs in the negative control panel. Spot intensities above the cut-off values were taken as a positive signal and those equal or below the values were taken as no signal. Correlation of spot intensities between arrays was determined by the Pearson's correlation test.

2.3 Results

Direct sequencing

The 20 SNPs of 10 genomic DNA samples were all successfully genotyped by direct sequencing (Table 2.3). These reference 200 genotypes were produced for evaluation of the performances of both RFLP and chip-based minisequencing. Figure 2.2 shows the three SNP genotypes as obtained by DNA sequencing.

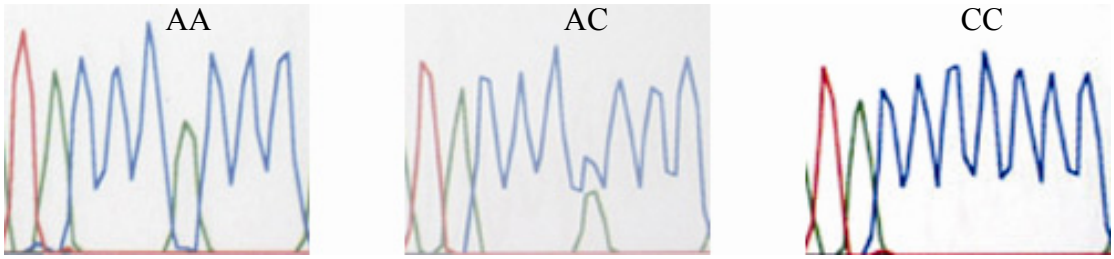

Figure 2.2 Demonstration of three genotypes called by direct sequencing

Table 2.3 The reference genotypes of 20 SNPs among 10 samples obtained from direct sequencing

SNPs	Subjects									
	HCK	BJ	JG	SL	G489	G490	G491	G493	G495	G496
ANGR	TT	TT	TT	TT	TT	TT	TT	TT	GT	TT
ApoAI-75	GA	GG	GA	GG	GA	GG	GA	GG	GA	GG
ApoAI+83	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC
ApoBgc	CC	CC	CC	CC	CC	CC	CC	CC	CT	CC
ApoBtz	CC	CC	CC	CC	CC	CC	CC	cc	CC	CC
ApoE112	CT	TT	TT	CT	TT	CT	TT	TT	TT	TT
ECE	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC
Eln422	GG	GA	GG	GG	GG	GG	GG	GG	GG	GG
Eln290	TT	TT	TT	TT	TT	CC	CT	TT	CC	CC
ENOS	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG
F13	CC	CC	CC	CT	TT	CT	CC	CC	TT	CC
FVIIT	CC	CC	CC	CT	CC	CT	CC	CT	CC	CC
GP3A	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT
HFE63	CC	CC	CC	CG	CC	CC	CC	CC	CC	CC
LCAT	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC
LPLPvu	CC	CC	TT	TT	CT	CT	CC	CC	CC	CC
PAI-1	AA	AA	AA	AA	GA	GA	GA	GG	AA	AA
PON2	CC	CC	GG	CC	GG	GG	CC	CC	CC	CC
TGFb509	CC	CC	CC	CT	CT	CC	CT	TT	TT	TT
TNFa	GG	GA	GG	GG	GG	GG	GG	GA	GG	GG

The genotype in lower case represented a wrong calling by RFLP.
Those underscored were not correctly genotyped by minisequencing.
The highlighted genotypes failed to be called by minisequencing.

RFLP

RFLP was also performed in this study as a bench mark for SNP genotyping. Compared to all reference genotypes, only ApoBtz of G493 was not correctly genotyped by RFLP (Table 2.3). Thus, the concordance rate between sequencing and RFLP, which represents the proportion that was in agreement with the reference set by sequencing, was 99.5%.

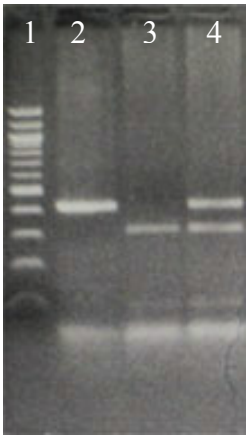

Figure 2.3 Calling of three genotypes of LPLPvu by RFLP. Lanes 1) 1kb DNA ladder; 2) TT; 3) CC; 4) TC

On-chip minisequencing

Figure 2.4 illustrates how base calling is accomplished from one of the four arrays, on which minisequencing was carried out with TAMRA-ddCTP. By using the synthetic oligonucleotides as targets, an overall concordance rate of 98.3% was obtained. In comparison, only 90.5% genotypes of the genomic samples were specifically genotyped. Four genotypes (2%) were wrongly called, and 15 genotypes (7.5%) could not be called as there were no signals at all (Table 2.3).

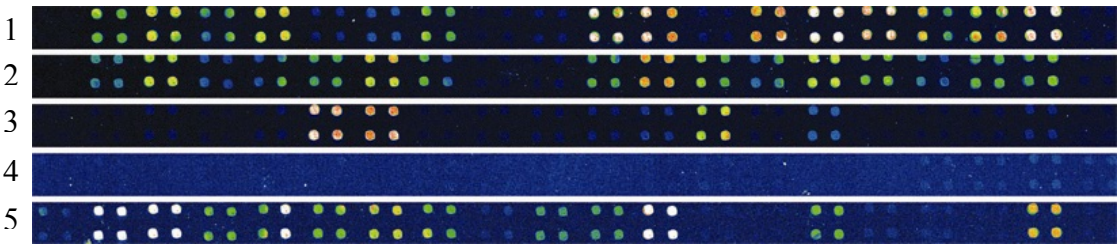

Figure 2.4 Base calling by minisequencing from TAMRA-ddCTP array. Synthetic targets composed of common homozygotes (1), heterozygotes (2) and rare homozygotes (3); Negative control (4); Genomic DNA (5).

Correlation of spot intensities between arrays

Pearson's correlation test was performed to analyze the arrays with synthetic oligonucleotides as targets. Only results from arrays with TAMRA-ddC or TAMRA-ddU were shown (Table 2.4) because the majority of the 20 SNPs contained either C, or T or both alleles. Overall, the correlation coefficients (mean) from the homozygotes were reasonably high (over 90%), except for TAMRA-ddUTP in the 10th array (highlighted). This array was consistently correlated poorly with all other arrays. In contrast, correlation coefficients of the heterozygotes were lower (78%-85%).

When the positive and negative spots of each array were analyzed separately, it was shown that their mean signal intensities and the standard deviations also showed rather consistent trend except the negative spots of the 10th array (Figure 2.5).

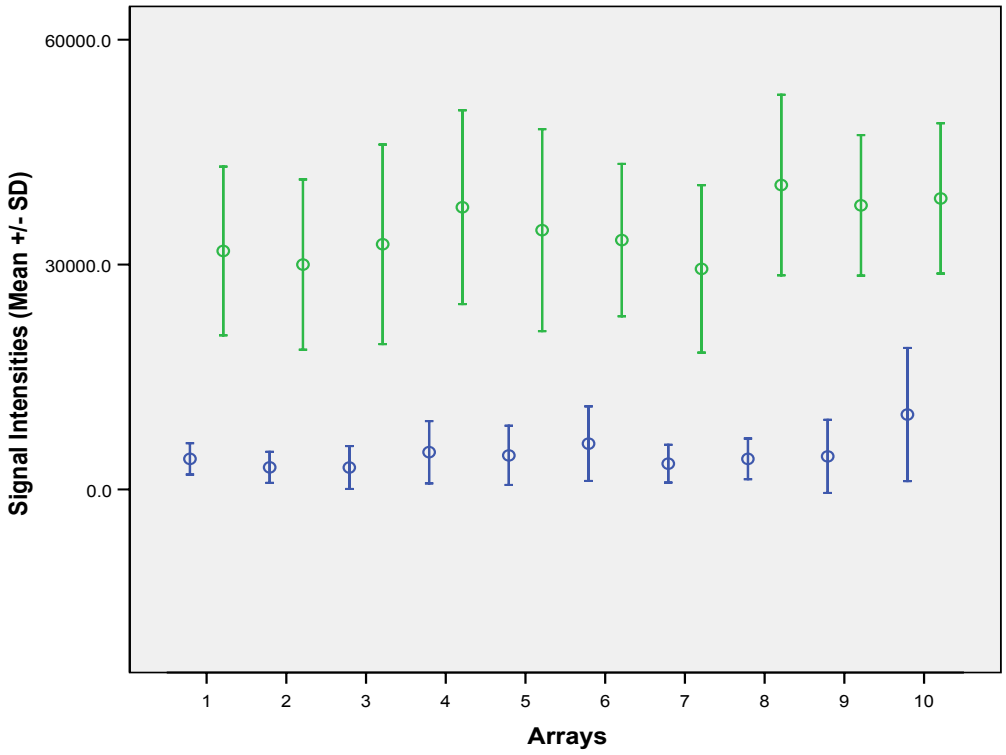

Figure 2.5 Array-to-array comparisons showing the mean +/- SD values of the positive and negative spots of each array.

Table 2.4 Array-to-array comparison using Pearson's correlation analysis

A. TAMRA-ddCTP (Homozygote, Allele 1)

Array	1	2	3	4	5	6	7	8	9	10
1	1									
2	0.9789	1								
3	0.9447	0.9402	1							
4	0.9704	0.9666	0.9427	1						
5	0.9513	0.9603	0.9590	0.9278	1					
6	0.9685	0.9579	0.8964	0.9369	0.9322	1				
7	0.8983	0.8921	0.9794	0.9022	0.9109	0.8223	1			
8	0.9785	0.9728	0.9188	0.9702	0.9517	0.9647	0.8494	1		
9	0.9538	0.9417	0.9172	0.9719	0.8947	0.9419	0.8695	0.9408	1	
10	0.9462	0.9394	0.9188	0.9449	0.8977	0.9075	0.8822	0.9319	0.9459	1

Mean: 0.9331

B. TAMRA-ddCTP (Heterozygote, Allele 1 + 2)

Array	1	2	3	4	5	6	7	8	9	10
1	1									
2	0.9024	1								
3	0.9027	0.8234	1							
4	0.8379	0.8922	0.7999	1						
5	0.8359	0.8765	0.8931	0.7503	1					
6	0.9043	0.9037	0.8417	0.7992	0.8966	1				
7	0.7929	0.8272	0.9048	0.7374	0.8910	0.7931	1			
8	0.8916	0.9391	0.7926	0.8467	0.8650	0.9379	0.7434	1		
9	0.8822	0.9072	0.7934	0.8982	0.7884	0.8108	0.7930	0.8554	1	
10	0.8344	0.9035	0.8288	0.9087	0.8347	0.7982	0.8174	0.8397	0.8740	1

Mean: 0.8487

C. TAMRA-ddCTP (Homozygote, Allele 2)

Array	1	2	3	4	5	6	7	8	9	10
1	1									
2	0.9827	1								
3	0.9752	0.9749	1							
4	0.9332	0.9342	0.9357	1						
5	0.9607	0.9484	0.9810	0.9479	1					
6	0.8095	0.8159	0.7951	0.8290	0.7876	1				
7	0.9113	0.8860	0.9224	0.9102	0.9574	0.7793	1			
8	0.9716	0.9573	0.9892	0.9322	0.9885	0.8029	0.9488	1		
9	0.8709	0.8374	0.9136	0.8392	0.9147	0.7367	0.8855	0.9239	1	
10	0.9839	0.9602	0.9716	0.9271	0.9723	0.8280	0.9408	0.9850	0.9026	1

Mean: 0.9103

D. TAMRA-ddUTP (Homozygote, Allele 1)

Array	1	2	3	4	5	6	7	8	9	10
1	1									
2	0.9509	1								
3	0.9618	0.9579	1							
4	0.9703	0.9563	0.9724	1						
5	0.9372	0.9594	0.9578	0.9146	1					
6	0.9941	0.9560	0.9538	0.9619	0.9324	1				
7	0.9778	0.9396	0.9618	0.9705	0.9499	0.9642	1			
8	0.9861	0.9476	0.9522	0.9625	0.9231	0.9817	0.9691	1		
9	0.9650	0.9275	0.9205	0.9185	0.9320	0.9700	0.9559	0.9666	1	
10	0.8934	0.8778	0.8683	0.8614	0.8678	0.9111	0.8945	0.8931	0.9474	1

Mean: 0.9421

E: TAMRA-ddUTP (Heterozygote, Allele 1 + 2)

Array	1	2	3	4	5	6	7	8	9	10
1	1									
2	0.7748	1								
3	0.8656	0.6754	1							
4	0.7083	0.9187	0.7198	1						
5	0.8366	0.7552	0.8974	0.8249	1					
6	0.8536	0.9223	0.7065	0.8084	0.8045	1				
7	0.6903	0.6152	0.6180	0.6050	0.6869	0.7180	1			
8	0.7754	0.9443	0.7349	0.9216	0.8491	0.9221	0.6956	1		
9	0.8246	0.9171	0.7240	0.8649	0.8271	0.9587	0.7004	0.9280	1	
10	0.7027	0.7444	0.6263	0.7197	0.6472	0.7731	0.6549	0.7358	0.7843	1

Mean: 0.7774

F. TAMRA-ddUTP (Homozygote, Allele 2)

Array	1	2	3	4	5	6	7	8	9	10
1	1									
2	0.8924	1								
3	0.9505	0.8942	1							
4	0.9095	0.9728	0.9314	1						
5	0.9144	0.8003	0.9400	0.8601	1					
6	0.9111	0.9676	0.8784	0.9656	0.7797	1				
7	0.8745	0.9205	0.9246	0.9508	0.9299	0.8744	1			
8	0.9256	0.9711	0.9390	0.9824	0.8263	0.9680	0.9084	1		
9	0.9343	0.9590	0.9117	0.9586	0.8045	0.9738	0.8864	0.9794	1	
10	0.6703	0.7220	0.6879	0.7189	0.5989	0.7675	0.6736	0.7199	0.7117	1

Mean: 0.8720

2.4 Discussions

In this study, we have successfully adapted minisequencing to solid phase (microarray) to perform multiplexed SNP genotyping. This format permits immediate access of the targets to the probes without diffusion, thus remarkably reducing the amount of targets required for hybridization. Significantly, this platform allows accurate interrogation of >90% of all SNPs in parallel. Moreover, it can be easily tailored to other biomedical applications as a substitute for many conventional methods. As a demonstration of principle, this study succeeded in genotyping most SNPs.

Upon hybridization and addition of polymerase, gene-specific probes will be extended by a single ddNTP, and the extended probes can be detected if the incorporated ddNTP is fluorescently-labelled. This is a procedure called minisequencing, which ensures accurate genotyping with the help of high-fidelity DNA polymerase. In this study, we used ThermoSequenase to perform single-base extension with TAMRA-labelled ddNTPs, and it demonstrated a concordance rate similar to that of RFLP when synthetic targets were genotyped.

When long PCR products were used to run minisequencing in previous studies carried out in our group, the genotype calling rate was very low. Thus, it was speculated that the target size is important for accurate genotyping by minisequencing. For this reason, this study designed another set of primers to run asymmetric PCR to generate targets of 40-50mer in length. This modification greatly improved the performance of minisequencing as most of the SNPs were correctly genotyped. Compared to long and bulky targets, small ones are less likely to form hairpins by intra-molecular base pairing, thus facilitating efficient hybridization between probes and targets. Generation of targets by asymmetric PCR in this study also simplified sample preparation, which was usually accomplished by two separate PCR reactions.

Because both hybridization and minisequencing are usually performed under one temperature, it is thus crucial to design the gene-specific probes with similar Tm values to ensure efficient hybridization and even single-base extension. As the optimal temperature of ThermoSequenase used in this study is around 60^{o}C, it is thus favourable to have a similar Tm for each probe. Unfortunately, Tm value and GC content of the probes included in this study, calculated by a web tool (http://www.basic.northwestern.edu/biotools/oligocalc.html), vary from 36^{o}C to 60^{o}C and 29% to 76%, respectively. As such, some DNA targets might have never succeeded to be captured by the probes, which explained to some extent why a higher failure rate was obtained by on-chip minisequencing. The higher failure rate could also be attributed to the failure of PCR amplification. A higher concordance rate was obtained from synthetic targets than genomic DNA, which implies that some genomic DNA might not be available for hybridization and minisequencing.

This study has demonstrated the great potential of the solid-phase minisequencing to interrogate multiple SNPs simultaneously, but it could be further improved by a few ways. For instance, as the calling of genotypes is based on a one-color strategy, it thus requires four arrays to interrogate each SNP. Other than consideration of cost and labour, signal analysis is also complicated as array variations have to be normalized before genotypes can be correctly called. The presence of array variations can be seen from Table 2.4, especially those arrays with heterozygotes. As mentioned earlier, better probe design is also potentially able to improve the performance this platform. Moreover, if the amplification of genomic DNA can be guaranteed, it would be expected that much higher calling rate could be achieved.

CHPATER 3 TAG/ANTI-TAG LIQUID-PHASE PRIMER EXTENSION ARRAY (TATLIPEXA): A FLEXIBLE AND VERSATILE GENOTYPING PLATFORM

3.1 Introduction

The common disease-common variants hypothesis postulates that SNPs in combination may render an individual susceptible to a particular complex disease (Cargill *et al.,* 1999). From a list of currently known SNPs, it is possible to identify a subset of them as useful markers for the prediction of disease occurrence, especially for complex diseases with high heritability. Additionally, SNPs are also known to contribute to differences in drug responses among individuals (Roses, 2000).

Although numerous genotyping technologies have been widely in use (Landegren *et al.,* 1998; Tsuchihashi *et al.,* 2002), only microarray appears more promising for fulfilling the need of a highly multiplexed genotyping platform. Advances in technologies such as oligonucleotide synthesis and high-precision robotic microarrayers have enabled the miniaturized solid-phase microarray to interrogate thousands of SNPs simultaneously. In general, there are two categories of microarray-based genotyping methods. One is solely dependent on sequence-specific hybridization between complementary DNA strands (Khrapko *et al.,* 1991). However, the allele discrimination power of such approach is limited (Ben-Dor *et al.,* 2000; Lindroos *et al.,* 2002). The other category consists of enzyme-mediated methods, such as DNA polymerase (Pastinen *et al.,* 1997; Lindroos *et al.,* 2003), ligase (Consolandi *et al.,* 2003) and endonucleases (Wilkins Stevens *et al.,* 2001). Among them, solid-phase primer extension (minisequencing) using DNA polymerase is the most robust one. Gene-specific probes are immobilized on the chip and will

anneal just 5' to the polymorphic nucleotide of interest on target DNA. They are then extended by fluorescent ddNTPs and detected. With this approach, remarkable improvements have been made in terms of sensitivity and specificity. However, this approach also bears a few limitations. For instance, the chips are usually custom-made by spotting gene-specific probes for the SNPs of interest. If one or a few SNPs need to be included or removed, a fresh chip has to be spotted with the new combination of probes. This inflexibility has been overcome by the introduction of the universal tag/anti-tag system (TAT), which allows 'homing' of any gene-specific sequences to particular anti-tags on chip via their artificial tags (Lovmar *et al.,* 2003). In this way, universal chips can be prepared in bulk and used for any combination of SNPs later. Another limitation of solid-phase primer extension is the reduced efficiency when many sequences are involved. From our earlier studies, we encountered many SNPs which randomly failed to be genotyped. This is most likely due to limited molecular reaction space for probe-target hybridization (steric hindrance effect), especially in the presence of a large amount of long targets. This problem has been addressed by performing primer extension in liquid phase in a cyclic manner (Lindroos *et al.,* 2002). In this way, it also obviates the need to generate single-stranded targets. In solid-phase platforms, it is usually necessary to run a second PCR with one primer to produce single-stranded targets or carry out other alternative approaches such as asymmetric PCR, or digestion of one strand by DNA exonucleases. In any case, these approaches are either not efficient, tedious or not cost-effective.

Another problem associated with conventional solid-phase primer extension is the choice of fluorophore for distinguishing the 4 ddNTPs. If only a single fluorophore is used for all four ddNTPs, four separate extension reactions would have to be

performed in a parallel manner, one for each ddNTP. Alternatively, if four fluorophores are used to label the ddNTPs, only a single reaction is required. However, the problem of crosstalking (overlapping fluorescent signals) becomes an issue because it is difficult to select fluorophores that emit at sufficiently different wavelengths or have filters in scanners that can discriminate the smaller differences in emission wavelengths. To overcome this problem, we have devised a novel two-color detection strategy and integrated it into TAT and liquid-phase prime extension systems so that it can interrogate all types of SNPs and some short insertion/deletion polymorphisms in a single reaction with only two fluorescent ddNTPs.

In essence, there are only six types of SNPs: C>T or T>C, A>G or G>A, A>C or C>A, G>T or T>G, C>G or G>C, and A>T or T>A (Figure 3.1). As C>T is the most common type, both alleles can be interrogated with one extension primer to be extended with either Cy5-ddC or TAMRA-ddU (Figure 3.1a). As for A>G, extension with Cy5-ddC and TAMRA-ddU can also proceed if the extension primer is designed from the anti-sense strand (Figure 3.1b). As for A>C or G>T SNPs, two extension primers from both strands are required so that they can be extended by either Cy5-ddC or TAMRA-ddU (Figure 3.1c, d), and both primers are tagged with the same anti-tag sequence. The challenge to our strategy is seen in genotyping C>G or A>T SNPs, although they are less commonly encountered. For instance, if detection of C allele of C>G is achieved with Cy5-ddC, G allele cannot be interrogated simply by designing another extension primer from the complementary strand because another Cy5-ddC will be introduced. One solution to this problem is to design allele-specific extension primer if the nearest neighbouring nucleotide downstream of G allele is A or T so that TAMRA-ddU can be incorporated. Another solution is to tag the extension primers differentially. Both primers will be extended with Cy5-ddC, but

they can be discriminated because the signals are addressed to different sites on the chip (Figure 3.1e). These strategies can also be applied to interrogate A>T SNPs (Figure 3.1f). Thus, all types of point mutations are covered by our two-color strategy. These strategies can also interrogate some short insertions/deletions. LCAT included in this study is indeed a single C insertion. Without this insertion, the subsequent base is G. Thus, the strategies for C>G SNPs were applied. As a proof of concept, we demonstrated this two-color strategy by genotyping a panel of 9 SNPs and evaluated it with sequencing and RFLP. Although this system is currently used for genotyping coronary artery disease candidate genes, it can be easily adapted for any genes of interest from other diseases or those involved in drug metabolism.

3.2 Materials and methods

Oligonucleotides

Five sets of oligonucleotides were used in this study for the following purposes.

Set 1 was comprised of 10 universal anti-tags initially spotted on the test chip (Table 3.1). Typically a universal anti-tag was a 20-mer artificial sequence with an amino-linker (NH_2) at its 3' terminal so that it could be immobilized through a covalent bonding with the aldehyde group on the slide. A poly (T) spacer was also included so that it could help reduce steric hindrance effect.

Set 2 contained 8 pairs of primers for PCR amplification of 9 target sequences (Table 2.1). Seven of these SNPs had been previously published (Chamberlain *et al.,* 1989; Bujo *et al.,* 1991; Suzuki *et al.,* 1994; Wang *et al.,* 1995; Hibi *et al.,* 1998; Grainger *et al.,* 1999; Dachet *et al.,* 2000). Two SNPs (ELN-422 and ELN-290) were identified in our laboratory and they were in close proximity on the same gene. All these primers were used to generate targets for both sequencing and RFLP, and the size of

the amplicons ranged from 208bp to 723bp.

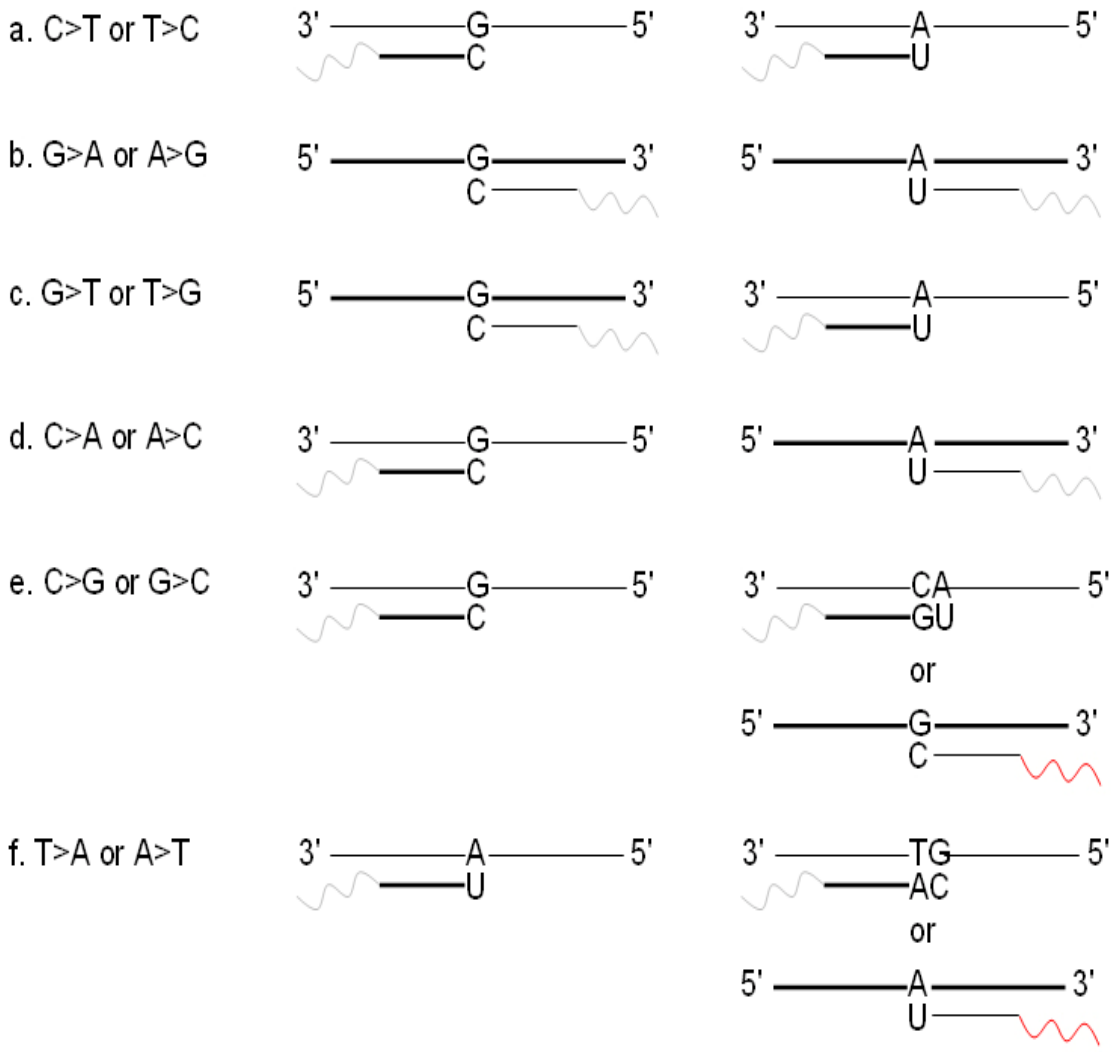

Figure 3.1 Schematic representation of two-color strategy to genotype SNPs. Sense strands are represented by thick lines and anti-sense strands by thin lines. Extension primers with the same tag are represented by wavy lines while those requiring different tags for the alternate alleles are represented by wavy lines in red.

Set 3 consisted of 18 oligonucleotides to simulate allele-specific targets of the 9 SNPs to prove the concept of TATLIPEXA. They also served as reference targets for this platform as rare genotypes might not be present in genomic DNA samples (Table 3.2). Set 4 were 9 primer pairs designed to amplify the sequences flanking the SNP site to produce targets of 40-50bp in length for liquid phase primer extension (Table 2.2). This strategy has been proven efficient in improving on-chip minisequencing by reducing the complexity of target sequences (Chapter 2).

Set 5 were 12 extension primers for liquid-phase primer extension. Some SNPs required two extension primers to introduce different fluorescence, such as CETP (A>C) and ENOS (G>T) (Table 3.2).

Table 3.1 The universal anti-tag sequences used in this study

Names	Sequences (5'→3')
PUniv1	TCGATCCAGTCACGTCGCTAttttttttttttttt-NH_2
PUniv2	TAGTACGTAGCTCGCCTTAAttttttttttttttt-NH_2
PUniv3	GTGTGGAAGCCAAGTAGACTttttttttttttttt-NH_2
PUniv4	GTGGTCCGGGATTCTCTCTCttttttttttttttt-NH_2
PUniv5	CGGTCTAGAGATTATGACGGttttttttttttttt-NH_2
PUniv6	GCTTTAATGTCGGACGACTTttttttttttttttt-NH_2
PUniv7	CTTCGTGGGTCGTAGTAGGGttttttttttttttt-NH_2
PUniv8	CCTCTTGTGGTCCGATGAGAttttttttttttttt-NH_2
PUniv9	GAAATTCAACGAGTTCGACTttttttttttttttt-NH_2
PUniv10	TTGTAGAGTGGTCAGTGGTCttttttttttttttt-NH_2

Table 3.2 The primers and synthetic targets for liquid-phase primer extension

SNPs [a]	Extension primers [b]	Synthetic targets [c]
ApoA1+83 (C/T)	AGTCTACTTGGCTTCCACACgttcaggccttgccccaggc	cagaggcccG/Agcctggggcaaggcctgaac
CETP (A/C)	TTAAGGCGAGCTACGTACTAttgatatgcataaaataactctggg	acccAcccagagttattttatgcatatcaa
	TTAAGGCGAGCTACGTACTAcaatgatctcagaggctgtataccc	aactctgggGgggtatacagcctctgagatcattg
ELN-422 (G/A)	TAGCGACGTGACTGGATCGAtccgggaacacctccgacac	gtgtccctG/Agtgtcggaggtgttcccgga
ELN-290 (C/T)	GAGAGAGAATCCCGGACCACgcatttcccgtgagccttagtcaca	tgtccccagG/Atgtgactaaggctcacgggaaatgc
ENOS (G/T)	CCGTCATAATCTCTAGACCGtgctgctgcaggccccagatgag	agttctgggggCtcatctggggcctgcagca
	CCGTCATAATCTCTAGACCGtgctgctgcaggccccagatga	gttctgggggAtcatctggggcctgcagcag
F13 (C/T)	CCCTACTACGACCCACGAAGaacatcaccttctacaccggggtcc	tctgccttcG/Aggaccccggtgtagaaggtg
LCAT (C/G)	AAGTCGTCCGACATTAAAGCgctcctcaatgtgctcttcccccc	cgtggtgtgcGgggggaagagcacattgaggagc
	GACCACTGACCACTCTACAAgctcctcaatgtgctcttcccccg	cgtggtgtgCgggggaagagcacattgaggagc
LPLpvu (C/T)	TCTCATCGGACCACAAGAGGctcaattcaatgtctcttcatcttttag	accccacagctG/Actaaaagatgaagagacatt
TGFb509 (T/C)	AGTCGAACTCGTTGAATTTCtctgcctcctgacccttccatcc	gacacctgaG/Aggatggaagggtcaggaggc

[a] Except A>T (T>A), all other forms of SNPs are included in this study. ELN-422 and ELN-290 are novel SNPs identified by our group from the elastin gene.

[b] The extension primers are chimeric oligonucleotides consisting of a universal tag (upper case) and gene-specific sequence (lower case). Only the extension primers of LCAT were tagged with two different universal sequences.

[c] Synthetic targets are short gene-specific sequences flanking the SNPs of interest.

Array fabrication

The ready-to-use CSS-100 silylated slides were purchased from CEL association, Inc. Prior to printing, the universal anti-tags (Set 1) were re-suspended to a concentration of 20μM with ArrayIt™ micro spotting solution (TeleChem, USA), sodium hydroxide (0.08M) and TE buffer (pH 8.0) according to the manufacturer's instruction. Each anti-tag was spotted on the chip in quadruplicates with a 2 x 2 format using the PixSys7500 arrayer (Cartesian Technologies, USA), and ten of these cluster of 4 spots made up an array. After an overnight incubation in a humidified chamber at 37°C, the slides were sequentially washed twice with 0.2% SDS (2 min); twice with distilled H_2O (2 min); once with distilled H_2O at 95°C (2 min); once with sodium borohybride solution (5 min); twice with 0.2% SDS (1 min) and twice with distilled H_2O (1 min). After air dry, the slides were ready to use or stored at 4°C.

Amplification of target DNA

Nine polymorphisms (Table 3.2) were selected from those that had been studied by our group for their roles in coronary artery disease. To yield sufficient amplicons for both DNA sequencing and RFLP, the PCR was set up in a total volume of 20μl. Typically the reaction mixture contained 2% DMSO, 2.5mM dNTPs, 0.2 units of DyNAzyme™ II DNA Polymerase (Finnzymes, Finland) and 20pmols of each primer. One touchdown PCR program was run to amplify all targets in an automated thermal cycler (Biometra) with initial temperature at 95°C (5 min), followed by 15 cycles of 95°C (30 sec), 63°C (1 min), 72°C (1 min) and another 20 cycles of 95°C (30 sec), 55°C (1 min), 72°C (1 min). The reaction was terminated after final extension at 72°C (10 min).

Genotyping by sequencing and RFLP

The procedures have been described in detail in Chapter 2.

Genotyping by two-color TATLIPEXA

In order to prove the concept of TATLIPEXA and optimize its reaction conditions, three synthetic target pools (Set 3) simulating the 3 respective genotypes of each SNP were prepared. Cyclic liquid-phase primer extension was performed in thermal cycler (Biometra) in multiplex format, such that all 12 extension primers were included in a single reaction. Typically the 15µl reaction mixture contained 1pmol of each extension primer, 1 unit of ThermoSequenase DNA polymerase (Amersham), 10pmols of Cy5-ddC and TAMRA-ddU (NEN Life Science). The cycling was initiated at 94°C (3 min), followed by 30 cycles of 94°C (15 sec), 50°C (15 sec) and 60°C (15 sec). Subsequently, 5 µl of the reaction mixture were directly placed on the chip to allow for hybridization between anti-tags on chip and tags of the extension primers. Hybridization was carried out in a sealed humidified cassette in a water bath (55°C, 1h). The chip was then washed once with 0.2% SDS (5 min) and twice with distilled H_2O (2 min). The fluorescence from the chip was acquired with ScanArray® 5000. The laser power was kept constant at 95%, whereas the photo-multiplier tube (PMT) was typically 80% for Cy5 (633 nm), 85% for TAMRA (575 nm). For base calling, the signal intensities were measured by QuantArray® 3.0 software.

For genomic DNA samples, amplification by PCR was first performed with primers (Set 4) to produce 40-50bp amplicons. Much less dNTPs (1mM) was used in this process because the sequences to be amplified were shorter. Otherwise, substantial amounts of shrimp alkaline phosphatase (SAP) would have to be used to ensure that excess dNTPs were completely degraded so that only Cy5-ddC or TAMRA-ddU

could be incorporated into the extension primers. Amplification was also carried out under one touchdown program, initial temperature at 95°C (3 min), followed by 20 cycles of 95°C (15 sec), 55°C (15 sec, -0.25°C/cycle), 72°C (30 sec), another 20 cycles of 95°C (15 sec), 50°C (15 sec), 72°C (30 sec) and final elongation at 72°C (5 min). After PCR amplification, 5μl amplicons of each gene were pooled and treated with 1 unit of Exo I to remove redundant primers and 0.5 unit of SAP to degrade excess dNTPs (37°C, 30 min, followed by 75°C, 15 min). The sample was now ready for cyclic liquid-phase primer extension, followed by hybridization. Both reactions were carried out under the same conditions as those of the synthetic target pools. Finally, the chip was scanned and signal intensity was measured for base calling.

Data analysis

In this study, the signal intensity ratio of Cy5-ddC against the sum of Cy5-ddC and TAMRA-ddU was calculated to set the cut-off values for three genotypes. The lower and upper limits of the cut-off value were the 95% confidence intervals (Mean ± 1.96 x standard deviation) of the mean signal intensity value for each genotype.

3.3 Results

DNA sequencing

A total of 72 genotype results were obtained by sequencing from the 9 SNPs among 8 subjects (Table 2.3), which serves as the standard for other genotyping methods.

RFLP

In comparison with the reference data from sequencing, no discrepancy was found using RFLP among all 72 genotypes (concordance rate of 100%, Table 2.3).

Two-color TATLIPEXA

As shown in Figure 3.2, all three synthetic target pools were specifically genotyped by two-color TATLIPEXA. The first pool consisted of only C or G allele of all SNPs and was genotyped with Cy5-ddC (Figure 3.2A, upper half). The second pool contained all C, G, A and T alleles. Accordingly, both Cy5-ddC and TAMRA-ddU gave signals (Figure 3.2B). The third pool included only A or T allele, as such, only TAMRA-ddU could be incorporated (Figure 3.2C, lower half). Two clusters in each subarray (Figure 3.2, columns 8 and 10) were set aside for LCAT as its extension primers were tagged with different sequences. Even though they were both elongated with Cy5-ddC, base calling could be achieved by their distinct addresses on the array.

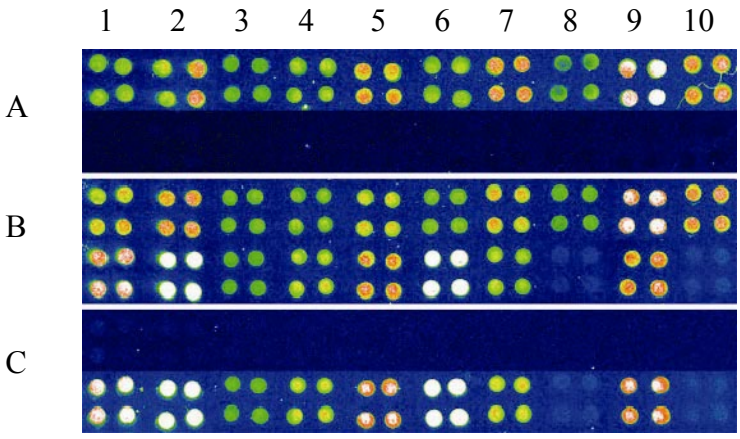

Figure 3.2 Evaluation of the specificity of TATLIPEXA using synthetic targets. A) Synthetic target pool to simulate homozygous C target of all SNPs. Only Cy5-ddC was specifically incorporated into the extension primers (upper half). B) Synthetic target pool to simulate heterozygote of all SNPs. All extension primers were extended by either Cy5-ddC (upper half) or TAMRA-ddU (lower half). C) Synthetic target pool to simulate homozygous T target of all SNPs. Only TAMRA-ddU (lower half) was specifically incorporated into the extension primers.

To assess the sensitivity of TATLIPEXA platform, only two synthetic targets simulating C and T alleles of ELN-290 were mixed in varying concentrations for cyclic liquid-phase primer extension. The ratio of C allele to T allele (C/T) ranged from 0.01 to 100 (100:1, 100:10, 100:100, 10:100, and 1:100). With as little as 10fmols of either allele, very strong fluorescence signal was produced (Figure 3.3A). When the logarithm of allele ratios (C/T, X-axis) was plotted against that of signal intensity ratios (Cy5-ddC/TAMRA-ddU, Y-axis), we found the signal intensities were proportional to the allele ratios (Figure 3.3B).

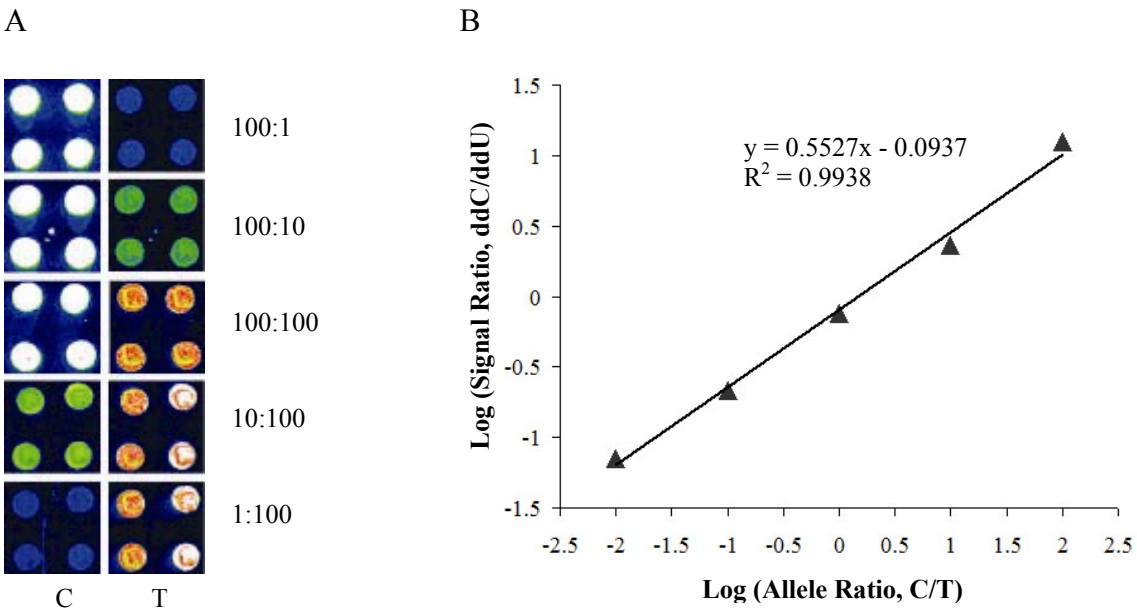

Figure 3.3 The sensitivity and quantitative feature of TATLIPEXA. A) Synthetic targets corresponding to C and T alleles of *ELN-290* were mixed for liquid-phase primer extension in the ratios of 100:1, 100:10, 100:100, 10:100 and 1:100. The pseudo-colors of the spots represent different signal intensities, with white for the highest, green for the lowest, red or orange for the intensities in between, and blue for negative. 10fmols of either allele could be detected. B) The signal intensities of images in A were analyzed and the logarithm of the signal ratios (Y-axis) was plotted against the logarithm of synthetic target concentration ratios (X-axis).

The performance of TATLIPEXA in genotyping genomic DNA samples was shown in Figure 3.4A. According to the cut-off values obtained from the signal intensity analysis, all 72 genotypes were specifically genotyped in comparison with sequencing and RFLP. The scatter plot analysis also showed three distinct clusters corresponding to the three genotypes of all SNPs (Figure 3.4B). Thus, perfect concordance rates of 100% were obtained among sequencing, RFLP and two-color TATLIPEXA.

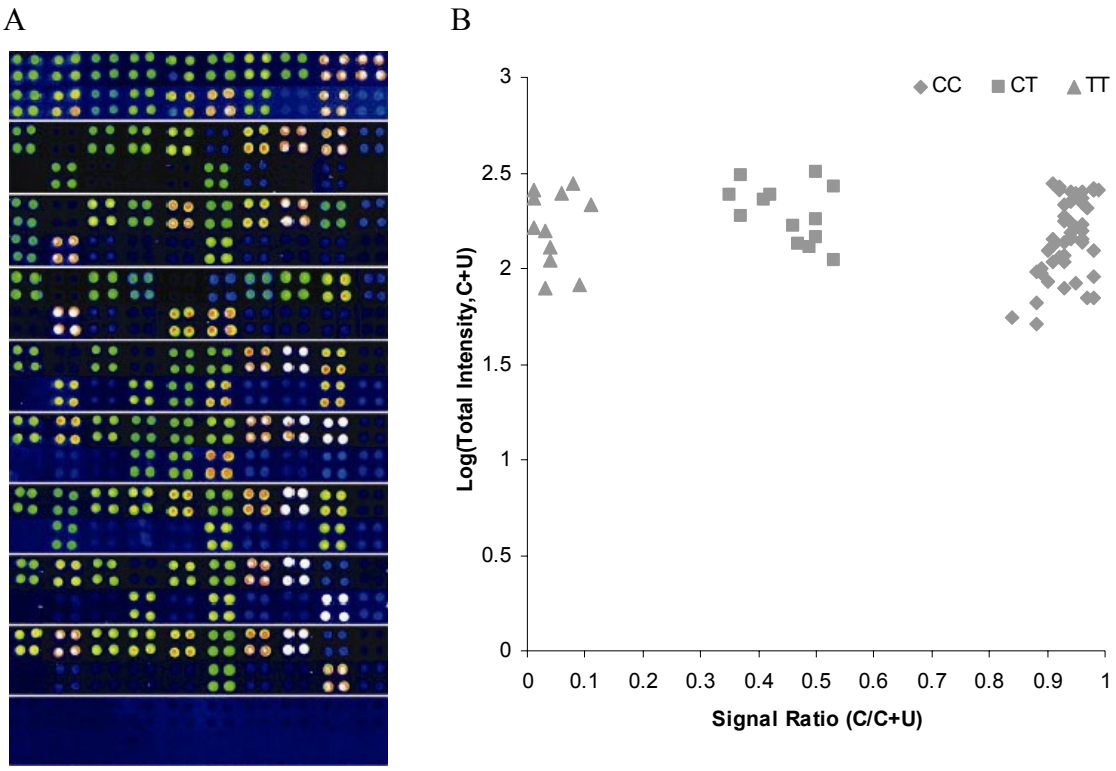

Figure 3.4 Genotyping results of genomic DNA by TATLIPEXA. A) Genotyping results by two-color TATLIPEXA. The top subarray was positive control, the bottom one was negative control, and between them were 8 genomic samples. In each subarray, the fluorescence signal of the upper row was from Cy5, and the lower one was from TAMRA. After subtraction of the background value from the signal, the signal intensity ratio of Cy5 against the sum of Cy5 and TAMRA at each site was calculated. In this study, the signal ratio for a CC genotype was >0.89; CT, 0.35–0.66; TT, <0.11. B) Scatter analysis of genotyping results from the representative 9 SNPs among 8 subjects. The X-axis was the signal ratio of Cy5-ddCTP against the sum of Cy5-ddC and TAMRA-ddU, while the Y-axis was the logarithm of the total signal intensities (Cy5-ddC+TAMRA-ddU). There were three distinct clusters corresponding to CC (48), CT (13) and TT (11), respectively.

3.4 Discussion

The TATLIPEXA platform presents a number of advantages over existing solid-phase primer extension platforms. With the introduction of tag-anti-tag system, the main consideration for chip preparation is shifted to the design of the universal anti-tags. Essentially, the choices of both the sequence and length of the anti-tags are important because higher temperature favours specific hybridization but reduce intra-molecular structure formation. Although the anti-tags used in this study were manually designed, they turned out to be very effective as no cross-hybridization was observed. Guidelines for design of universal anti-tags are now available (Gerry *et al.,* 1999) and are very helpful for array preparation.

In comparison to solid-phase primer extension, cyclic feature of liquid-phase has shown marked improvement in sensitivity as 10fmols of synthetic targets can be detected. As with solid-phase primer extension, the liquid-phase alternative could also be performed in multiplex format. However, prior generation of single-stranded targets from PCR products is not necessary in liquid-phase due to its cyclic feature. In fact, both strands of amplicons are required for many SNPs using our two-color strategy (Figure 3.1c, d, e and f). Consequently, sample preparation is significantly simplified with only conventional PCR amplification and SAP digestion involved. However, it is very critical to ensure that excess dNTPs must be completely digested. Otherwise, non-specific signal would be generated if dNTPs are incorporated into the extension primers prior to fluorescent ddNTPs. In this study, Exo I was included to digest the excess PCR primers. If not removed, these primers will consume more fluorescent ddNTPs since they could also be extended.

Theoretically it is possible to perform primer extension in a single reaction using 4 fluorescent ddNTPs (Kurg *et al.,* 2000) as four-laser scanners are commercially

available. In this way, the sample preparation and genotyping procedures can be significantly simplified. However, it is very difficult to choose four dyes for this purpose without encountering the problem of signal overlap. Moreover, it has been reported that the incorporation efficiencies of ddNTPs are different (Fortina *et al.,* 2000). Given the same amount of target, the signal intensity of Cy5-ddC was relatively stronger than that of TAMRA-ddU (Figure 3.3A). When dye-swap experiment using TAMRA-ddC and Cy5-ddU was performed, the stronger signal was still from Cy5-ddU. This implies the fluorophores can contribute to variations in signal intensities. As such, if normalization of the differential signal intensities is not carried out, false base calling may result. On the other hand, if four ddNTPs are labeled with only one fluorophore, four separate reactions would have to be performed. This is not time and cost effective, and additional computation has to be performed to correct chip-to-chip variations. Some have attempted to overcome those limitations by using two different Cyanine dyes, but their strategy was to group SNPs and interrogate them by different reactions (Fan *et al.,* 2000; Fortina *et al.,* 2000). The main novelty of this study is the two-color strategy, which allows all SNPs and some short INDELs to be interrogated simultaneously in a single reaction. As demonstrated in this study, it is possible to discriminate all types of SNPs and even INDELs with the use of only two fluorescently labeled ddNTPs. We chose Cy5-ddC and TAMRA-ddU because firstly, there is virtually no signal overlap between Cy5 and TAMRA; and secondly C>T SNPs are the most common types. As there are only two fluorescent ddNTPs involved, signal analysis is also simplified. It should be noted that this two-color strategy is completely independent of TAT system and cyclic liquid-phase primer extension. Without TAT system, this strategy can still be applied to solid-phase primer extension if the gene-specific probes are designed in the same

way as those in liquid-phase. Nevertheless, generation of single-stranded targets is necessary for some SNPs and the sensitivity may be sacrificed.

The preliminary data also suggested that TATLIPEXA could potentially provide quantitative information on allele frequencies using pooled DNA. This is especially useful when new mutations are discovered and their frequencies in the population are unknown. This is useful in large-scale association studies to identify disease-susceptible SNPs quickly. However, the signal intensities may not be in proportion to the allele frequencies if the saturation value of the confocal scanner is reached. Thus, the sample must be pooled and diluted to an appropriate range.

In conclusion, by the integration of TAT system and liquid-phase primer extension, this platform presents a few advantages over some existing approaches, which include the abilities to 1) simplify the target preparation; 2) genotype any SNP in a single reaction with only two fluorescent ddNTPs; 3) genotype some short insertion/deletion polymorphisms; and 4) estimate allele frequencies of pooled samples.

3.5 Future plans

Although TATLIPEXA has demonstrated great potential for highly parallel genotyping, it can be further improved in a few ways. For instance, the universal anti-tags used in this study were manually designed. If TATLIPEXA will be used for large-scale association studies to genotype numerous SNPs, it is desirable to have a powerful bioinformatic tool to design these probes. In this study, genomic DNA was amplified separately by multiple PCR, which was not time- and cost-effective. If multiplexed PCR with universal primers or whole genome amplification is introduced, target preparation can be remarkably simplified. Once these problems are successfully addressed, a robust platform can be established to run high-throughput

genotyping for genome-wide association studies. Thus far, TATLIPEXA has only demonstrated the principle by genotyping a limited number of SNPs among a few blood samples. As such, it is also important to apply TATLIPEXA to large-scale association studies to genotype many SNPs of a larger sample size.

CHAPTER 4 TDT-ASSISTED PROBE ELONGATION (TAPE): A VERSATILE TOOL FOR GENETIC ANALYSES ON OLIGONUCLEOTIDE ARRAY

4.1 Introduction

The completion of the whole genome sequencing through Human Genome Project (HGP) is undoubtedly one of the most marvellous achievements in human history (Lander *et al.,* 2001; Venter *et al.,* 2001). Consequently, a reference genome sequence is now available and the next big inquiry of the post-genome era would be the interpretation of this genetic blueprint.

Only 0.1% of the genome sequences between any two individuals are different. Other than insertions/deletions, tandem repeats, inversions and copy number polymorphisms, another main form of genetic variation is single nucleotide polymorphisms (SNPs). Although SNPs constitute only a very small fraction of the genome, they are widely believed to be the reasons why individuals are different from each other, why some are susceptible to certain diseases and resistant to certain drugs while the others are not (Tsuchihashi *et al.,* 2002). Thus, SNPs are greatly valued for their great potential applications in genetic studies, drug development and diagnostics. Nevertheless, there is currently a huge gap relating SNPs to complex phenotypes and therein lies the big challenge to bridge such a gap. First, five million SNPs have been validated thus far, but only a small fraction of these have potential impacts on diseases or drug responses (Rebbeck *et al.,* 2004; Riva *et al.,* 2004). Second, SNPs associated with common diseases are themselves common in both healthy and sick populations. To meet these challenges, dozens of molecular technologies have been successfully developed. However, no strategy has satisfactorily fulfilled the requirement for

highly multiplexed high-throughput genotyping until the introduction of the microarray (Southern *et al.,* 1999).

Like Southern blotting, microarray utilizes the principle of hybridization to perform genetic analyses. Typically, a microarray experiment is composed of four events. First, the array is fabricated by either *in situ* synthesis or immobilization of synthetic probes. Second, numerous targets of different sequences are sorted through the arrayed probes. Third, fluorescence is captured by a confocal scanner. Fourth, signals are analyzed by bioinformatic tool (Jain 2000). Target preparation and allele-discriminations are not listed in this workflow as they are in principle the same as those of traditional approaches. For instance, genomic DNA is usually amplified by PCR, and allele discrimination is achieved by hybridization, primer extension or ligation. The power of microarray is realized in its ability to function as a high capacity biological 'sifter' that is able to sort thousands of targets efficiently and specifically through hybridization. The application of microarray to monitor gene expression is essentially based on the same mechanism as that of genotyping, except that mRNA is the target of interest.

One of the contributing factors to the success of the microarray is the use of fluorescence to enable detection of attomoles to femtomoles of targets. Usually, fluorescence is introduced by either direct labeling of the oligonucleotides (Shi 2002; Consolandi *et al.,* 2003; Lyamichev *et al.,* 2003; Hardenbol *et al.,* 2005) or through enzyme-mediated reactions (Pastinen *et al.,* 1997; Gunderson *et al.,* 2005).

Terminal deoxyribonucleotidyl transferase (TdT) was first discovered to be involved in DNA polymerization in calf thymus almost half a century ago (Bollum 1960). However, it was soon found that TdT was very different from other DNA polymerases (Deng *et al.,* 1983). In the absence of a DNA template, it could

efficiently polymerize a single-stranded DNA (ssDNA) and is actually less efficient with double-stranded DNA (dsDNA), especially blunt ended ones. With these distinctive features, TdT has found wide applications in biomedical researches, such as detection of DNA damage and apoptosis. To exploit this distinctive feature, we have developed a versatile strategy termed TdT-assisted probe elongation (TAPE), by which array can be fabricated with unmodified oligonucleotides and fluorescently-labeled nucleotides can be incorporated to targets for detection.

4.2 Materials and Methods

Immobilization of oligonucleotides with NH_2 modifier

Commercially available oligonucleotides with NH_2-modifier from 1st BASE (Singapore) were re-suspended to 20μM with ArrayIt™ micro-spotting solution (TeleChem) prior to spotting on aldehyde-coated slides (CEL) using a PixSys 7500 arrayer (Cartesian). The quadruplicate spots were arranged in a 2 x 2 format. After incubation (37°C, overnight), the slides were washed sequentially with 2 x SSC (3 min), 0.2% SDS (3 min), boiling water (3 min) and water (3 min).

The sensitivity of TAPE

The oligonucleotide with 3'-NH_2 (TCGATCCAGTCACGTCGCTAttttttttttttttt) was immobilized as described above. The 15-dT spacer was introduced to facilitate hybridization through the capitalized sequence. In order to determine the sensitivity of TAPE, a 38-mer oligonucleotide tagged with a complementary sequence to the probe (TAGCGACGTGACTGGATCGAgggaacacctccgacacc) was serially (1/2) diluted from 40fmols/μl to 2.5fmols/μl. One microliter of these dilutions were

separately elongated by TdT (1U) with Cy5-ddC (1pmol) in a 5µl solution at room temperature (RT) for 20 minutes, followed by addition of 1µl of 0.5M EDTA (pH 8.0) and incubation at 95°C for 10 minutes to inactivate TdT. These mixtures were then loaded on the slide and placed in a humidified cassette to allow for hybridization in a water bath (50°C, 1hr). Finally, the slide was washed sequentially with 2 x SSC (3 min) and water (3 min, twice), and scanned by ScanArray® 5000 (Packard BioScience Ltd, UK) at 90% laser power and 90% PMT.

Incorporation of multiple labelled nucleotides by TAPE

Another oligonucleotide with 5'-NH_2 (ttttttttttttttttCCTATACAGTCACTTTT) was also immobilized on the slides as described above. Due to the availability of 3'-OH group of this oligonucleotide, elongation by TdT can be carried out on the slide. The objective of this experiment was to find out whether multiple labelled-nucleotides could be incorporated by TdT. For this purpose, three mixtures were prepared, containing either 0.5pmols of Cy5-ddU alone, 0.5pmols of Cy5-dU alone, or both 0.5pmols of Cy5-dU and 2pmols of normal dNTPs. Each mixture contained 1 unit of TdT and was made up to 5µl with deionized water (diH_2O). The mixtures were loaded onto three subarrays to permit elongation of the probes on the slide (RT, 20 min). Subsequently, the slide was washed and scanned as described earlier.

In addition, we also carried out elongation of the 38-mer oligonucleotide (20pmols) with 40pmols of either biotin-dU, FITC-dU or both. In this set up, it is only when both biotin-dU and FITC-dU are incorporated into the same oligonucleotide that it will be captured and detected. In each reaction, there was 20U of TdT and 400pmols of dNTPs. Following incubation (RT, 2hrs), an aliquot of the mixture was captured by streptavidin-coated beads (Dynal Biotech) (RT, 1hr). After washing, the solution

was transferred to 96-well plate to be read by Wallac VICTOR^2 V multi-label plate counter (PerkinElmer).

The time-based TAPE

The oligonucleotide immobilized through 5'-NH_2 was also used to investigate time-based TAPE because such reaction could be terminated at various time intervals by washing away active TdT and redundant Cy5-ddC from the slide as described earlier. The 5µl reaction cocktail contained 1U of TdT, 1pmol of Cy5-ddC, 1µl of TdT buffer and 2µl of diH_2O. These were loaded onto the slides and incubated at time intervals of 1, 2, 3, 4, 5, 10, 15 and 20 min. At each time point, the corresponding slide was stringently washed to remove TdT and Cy5-ddC before scanning.

3' aminoallylation of unmodified oligonucleotides by TAPE for array fabrication

A small aliquot (450pmols) of each synthetic oligonucleotide (Table 4.1) was elongated in a 15µl solution containing 20U of TdT and 1nmol of aminoallyl-dUTP (Fermentas). The mixture was incubated in a thermal cycler (37°C, 2 hrs), followed by heat inactivation (95°C, 10 min). After the concentration of the oligonucleotide was adjusted to 15µM with ArrayIt micro-spotting solution, they were spotted on the aldehyde-coated slides as described earlier. In order to find out the immobilization efficiencies of these elongated oligonucleotides, TAPE reaction with Cy5-ddC was carried out on the slide.

SNP genotyping by ASMS-TAPE

One potential application that we have exploited TAPE was SNP genotyping. This was demonstrated by integrating TAPE with allele-specific minisequencing (ASMS).

A pair of allele-specific primers has their 3' ends bearing the complementary base of the SNP. The principle behind ASMS-TAPE is that, when the SNP of interest is a homozygote, one of the two allele-specific primers will be extended by minisequencing with regular ddNTPs, while the other is still available for elongation by TAPE with a fluorescent ddNTP. In the case of heterozygote, both primers will be blocked so that no fluorescence can be introduced by the subsequent TAPE reaction. In order to prove the principle of ASMS-TAPE, five representative SNPs, C>T, C>A, C>G, G>A and G>T except A>T (as we do not have an A > T SNP in our current panel) were genotyped using synthetic oligonucleotides (Table 4.1). Three pools were prepared from these oligonucleotides to simulate three genotypes of a SNP. One pool comprised of one allele-specific target of all SNPs to simulate common homozygotes, another pool contained the other allele-specific targets to simulate rare homozygotes, and the third pool was composed of all targets to simulate heterozygotes.

Allele-specific minisequencing was performed typically in an 8µl cocktail containing 1 U of ThermoSequenase DNA polymerase, 100fmols of allele-specific primers which were complementary to the capitalized sequences of the synthetic oligonucleotides listed in Table 4.1, 20fmols of synthetic targets from one of the three pools and 20pmols of ddNTPs. The initial temperature was 95°C (5 min), followed by 30 cycles of 95°C (30 sec), 50°C (20 sec) and 54°C (30 sec). The final extension was 60°C (5 min). Following this, a 2µl mixture with 0.5U of SAP was introduced to digest redundant ddNTPs which might compete with Cy5-ddC in subsequent elongation by TdT. After heat-inactivation of SAP (95°C, 10 min), 1µl of this solution was mixed with 1 U of TdT and 1pmol of Cy5-ddC to elongate the oligonucleotides which were not terminated by ASMS (RT, 20 min). After

inactivation of TdT, the solution was loaded on the array prepared from elongated oligonucleotides with aminoallyl-dUTP. Following hybridization (50°C, 1hr) in a water bath and washing (2 x SSC, 3 min; water, 3 min, twice), the array was scanned to acquire fluorescence.

These SNPs were also genotyped by ASMS with four TAMRA-labelled ddNTPs to serve as control. Following ASMS, hybridization was directly carried out for genotype calling.

Table 4.1 The synthetic oligonucleotides used for SNP genotyping by ASMS-TAPE

SNPs	Alleles	Sequences (5'→3') *
Eln-422	G	ctggagtcgcaggtgtccctGGTGTCGGAGGTGTTCCCGG
	A	ccgggaacacctccgacacTAGGGACACCTGCGACTCCAG
Eln-290	C	ttctcaacccatgtccccagGTGTGACTAAGGCTCACGGG
	T	cccgtgagccttagtcacaTCTGGGGACATGGGTTGAGAA
ANGR	G	caattctgaaaagtagctaaGGCTCATTTGGTAGTGAAGT
	T	acttcactaccaaatgagcATTAGCTACTTTTCAGAATTG
CETP	C	tctcagaggctgtataccCCCCAGAGTTATTTTATGCAT
	A	atgcataaaataactctgggTGGGTATACAGCCTCTGAGA
ATIIID	C	acaccttccccactctcttaGGGTACAGAAAGGAGATGCA
	G	tgcatctcctttctgtaccGTAAGAGAGTGGGGAAGGTGT

* With the exception of the SNP site which is underlined, the two allele-specific oligonucleotides of each SNP are complementary to each other.
These oligonucleotides served as both probes to be immobilized on the slides and synthetic targets for allele-specific minisequencing reaction.
The capitalized sequences of these oligonucleotides are exactly complementary to their corresponding allele-specific primers for ASMS.

Elongation of RNA fragments by TAPE

The RNA fragments from frozen liver tissue of mouse were prepared according to the standard protocol of Affymetrix. In brief, total RNA was isolated using Trizol reagent (Life Technologies) and then purified with RNeasy Mini Kit (Qiagen). Following this, 10µg of purified RNA was reverse transcribed to cDNA with Superscript II (Invitrogen). Subsequently, cDNA was *in vitro* transcribed to biotinylated cRNA using the RNA transcript Labeling Kit (Affymetrix).

Two reaction mixtures were prepared and both contained 2µl of cRNA fragments and 1µl of dNTPs (100nM). 100 U of TdT were introduced into only one mixture. After topping up to 20µl with diH_2O, the mixtures were incubated in a thermal cycler (37°C, 2.5 hours), followed by separation on a 2% agarose gel for visualization of the elongation products.

4.3 Results

TAPE sensitivity

It is shown in Figure 4.1 that TAPE has a good sensitivity of 10fmols. When the amount of the target oligonucleotide was doubled to 20fmols, the signal intensity increased correspondingly but plateaued out beyond 20fmols when the saturation point of the scanner was reached. Moreover, TAPE with Cy5-dU or Cy5-dU/dNTPs generated much stronger fluorescence (2-3 times higher) than that with Cy5-ddU, indicating the occurrence of multiple-incorporation of Cy5-dU (Figure 4.2A). This was further substantiated by performing TAPE with biotin-/FITC-dU. With the presence of either biotin-dU or FITC-dU, the fluorescence was very low. When both

were present, the intensity increased more than 5 folds, implying that both biotin-dU and FITC-dU were incorporated into the same oligonucleotide (Figure 4.2B).

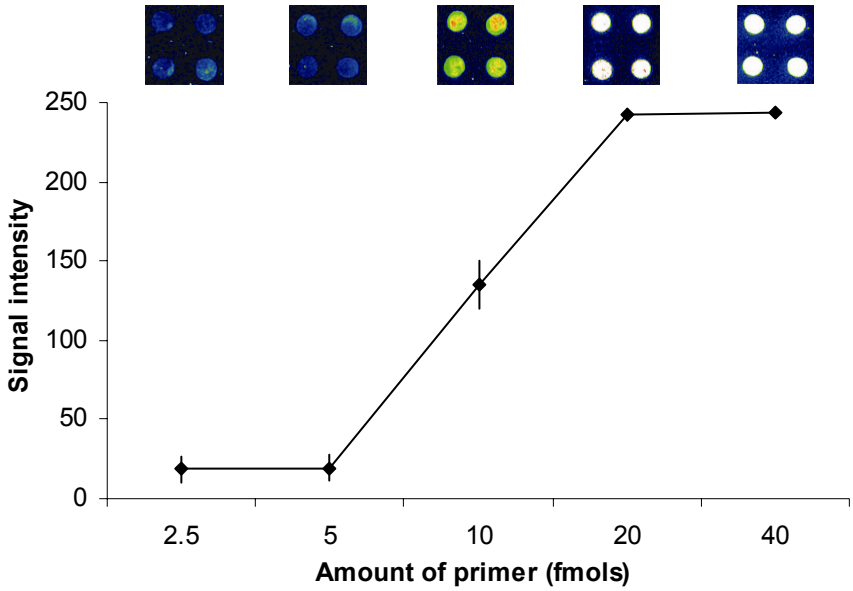

Figure 4.1 The sensitivity of TAPE with Cy5-ddCTP

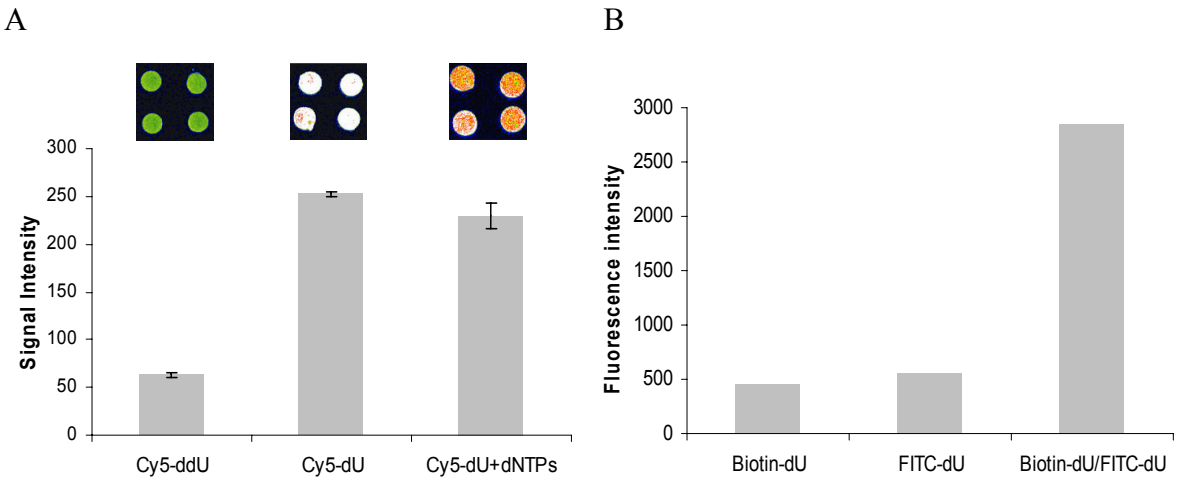

Figure 4.2 Incorporation of multiple labelled-nucleotides by TAPE. A) TAPE was carried out on the slide with Cy5-ddUTP, or Cy5-dUTP, or a mixture of Cy5-dUTP and regular dNTPs; B) TAPE was performed in solution with biotinylated dUTP, or FITC-dUTP, or a mixture of both modified dUTPs.

Time-based TAPE

Figure 4.3 shows that intense fluorescence could be obtained within one minute of elongation by TAPE with Cy5-ddC. By incubation of up to 20 minutes, however, the fluorescence intensity did not increase correspondingly. This suggests that elongation by TdT is extremely rapid.

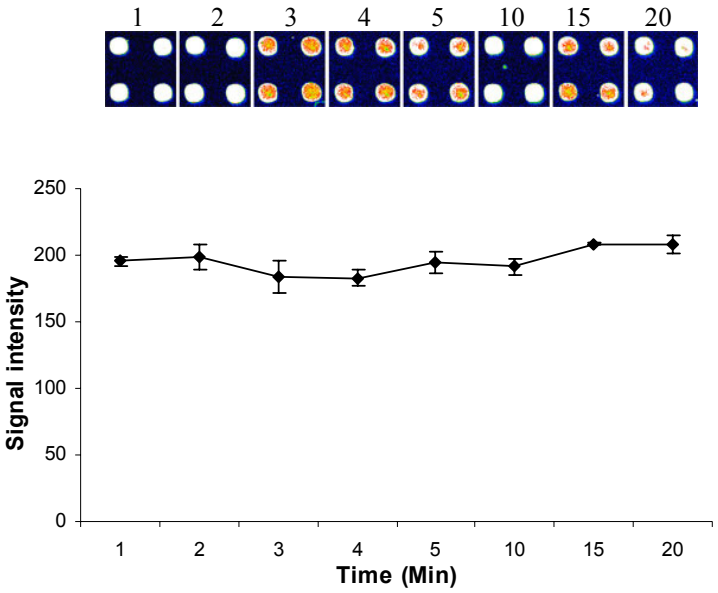

Figure 4.3 The time-based elongation by TAPE with Cy5-ddC

Array fabrication with unmodified oligonucleotides by TAPE

Intense fluorescence was observed from all spots after TAPE with Cy5-ddC was carried out on the array fabricated from oligonucleotides elongated with aminoallyl-dUTP. This implies that elongation was efficient and the amino moiety facilitated the immobilization of these oligonucleotides on aldehyde-coated slides (Figure 4.4).

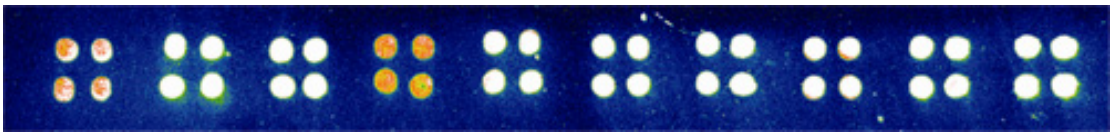

Figure 4.4 Evaluation of the attachment efficiency of elongated oligonucleotides. This array was fabricated by immobilizing regular oligonucleotides which were previously extended by TAPE with aminoallyl-dUTP, and evaluation of the quality of this custom-made array was accomplished by another TAPE with Cy5-ddC.

SNP genotyping by ASMS-TAPE

Three pools of five SNPs were all specifically genotyped by ASMS-TAPE using the custom-made oligonucleotide arrays mentioned earlier. The common homozygotes, heterozygotes and rare homozygotes were represented by subarray AA, AB and BB, respectively. It could be clearly observed in Figure 4.5A that only one allele-specific site of each SNP was fluorescently-labelled in subarray AA, and the other 5 sites were fluorescently-labelled in subarray BB, and no fluorescence were observed when two alleles were present (subarray AB). As expected, the pattern of the genotyping result by ASMS-TAPE is exactly opposite to that by ASMS with TAMRA-ddNTPs (Figure 4.5B). When the signal intensity ratios between alleles were plotted against the logarithm values of the sum of the signals, three distinct clusters were observed, further confirming the success of both ASMS-TAPE and ASMS in accurately genotyping SNPs (Figure 4.5 C, D).

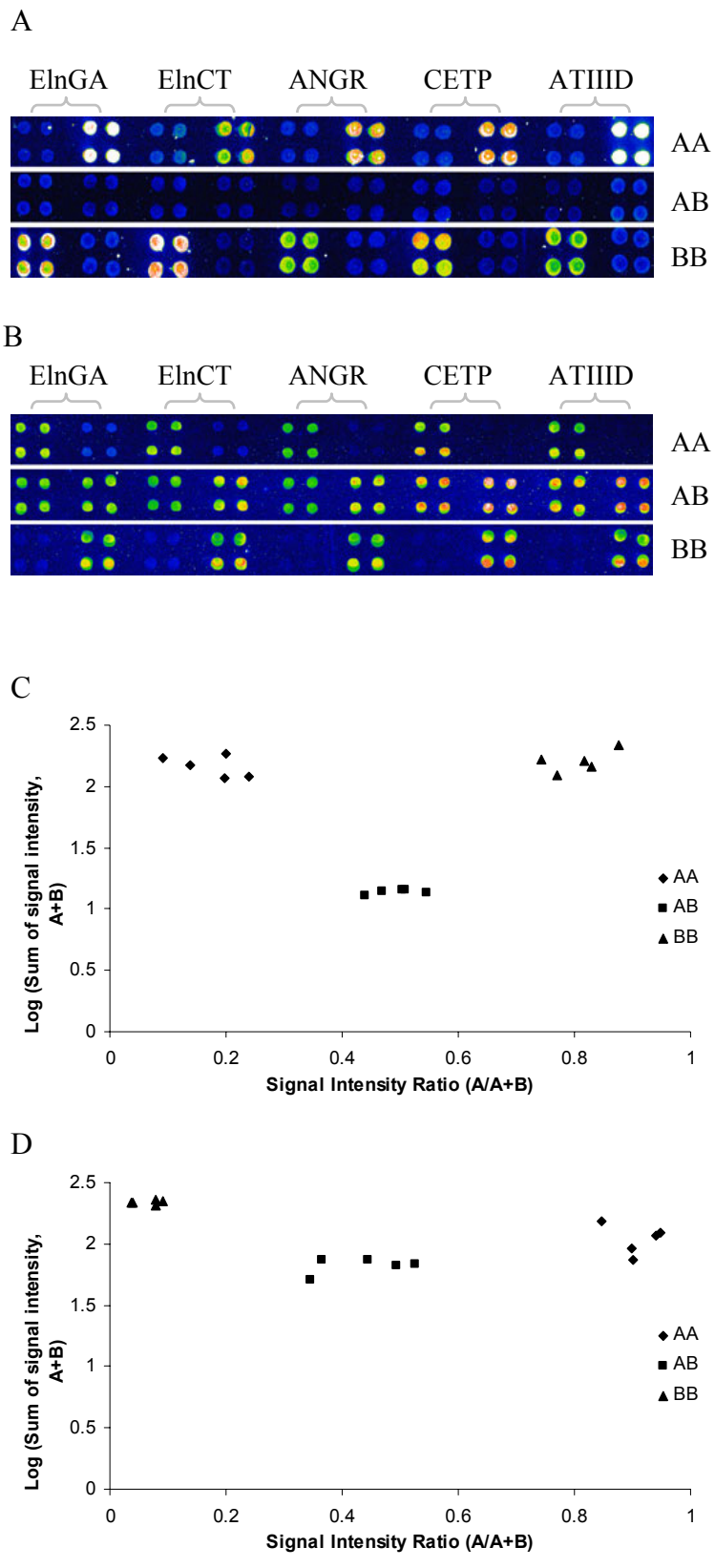

Figure 4.5 Demonstration of SNP genotyping by ASMS-TAPE strategy

Elongation of RNA fragments by TAPE

Without TdT, the majority of the cRNA fragments were 100-200 nucleotides in length (Lane 2, Figure 4.6). In the presence of TdT, however, some of these fragments were clearly polymerized, as shown by their relatively slower migration (Lane 3, Figure 4.6). This implies that RNA can also serve as substrate for elongation by TdT.

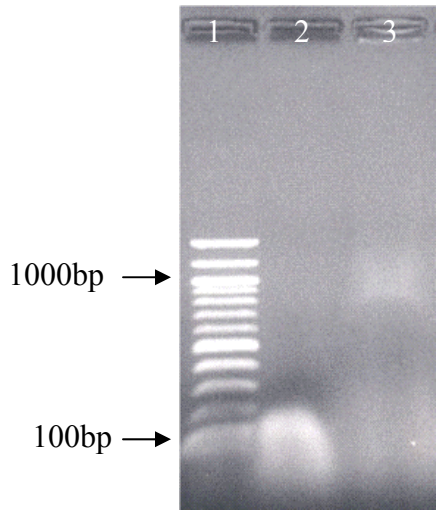

Figure 4.6 Elongation of RNA fragments by TdT with dNTPs

4.4 Discussion

This study has demonstrated the multiple applications of the rapid and simple TAPE procedure which could be carried out without any specialized laboratory equipment except for fluorescence detection. In the first application, we used TAPE to incorporate a modified nucleotide to a regular oligonucleotide. In our demonstration, we have chosen to introduce an aminoallyl-dUTP to a set of oligonucleotides for the purpose of immobilizing them onto aldehyde-coated microarray slides. Although modified oligonucleotides can be purchased commercially, the advantage of using TAPE is the flexibility that only the desired quantity of regular oligonucleotide is

modified. For smaller scale requirement, this is cost-saving instead of having to order commercially available ones that are sold in a fixed minimum quantity.

Oligonucleotide arrays are usually manufactured by either *in situ* synthesis (Fodor *et al.,* 1991) or immobilization of pre-synthesized probes by either contact printing or inkjet dispensing (Lindroos *et al.,* 2001). The former approach enables production of many arrays with consistent and reproducible features, but their price is still prohibitive to many laboratories. Immobilization method is advantageous in its ease of prototyping, rapid implementation and low cost, but it requires the probes to be chemically modified using anchors such as an amino group. This is where our method of introducing the amino group using TdT could be aptly applied. Other than gene-specific arrays, universal arrays using oligonucleotide tags can also be manufactured. When chemically synthesized oligonucleotides with 3' amino or biotin were used for as the immobilized probe in our TAPE experiments, we found that they were unable to block the elongation reaction (Chapter 5), despite the absence of the 3'-OH group. This led to a high background noise as probes in spots that were supposed to give no signals were also elongated. The noisy background could be eliminated by using TAPE to fabricate the microarray by incorporating aminoallyl-ddUTP onto 3' terminals of regular oligonucleotide probes for immobilization. Such probes are blocked at their 3' ends and do not contribute to any background noise in subsequent genotyping reactions using TAPE. Besides aldehyde-coated slides, streptavidin-coated slides or beads could also be used if the probes are extended with biotinylated nucleotides. However, since TAPE can only elongates probes from 3' hydroxyl groups, this procedure cannot be used to prepare arrays with the probes immobilized through their 5'-ends.

Interestingly, besides being used for fabricating micorarrays, TAPE can also be used further downstream in the process to incorporate fluorescent signals when custom-made arrays are used for carrying out genetic analyses. As a signal incorporating tool, TAPE has shown very good sensitivity as we have demonstrated that at least 10fmols of target DNA could be directly detected using Cy5-ddC. We have shown that TAPE reactions using Cy5-dUTP produced signals that were more than 3 times higher than Cy5-ddUTP. Currently, most molecular technologies such as molecular inversion probe and GoldenGate assay introduce fluorescence by direct labeling of the probes/primers (Oliphant *et al.,* 2002; Hardenbol *et al.,* 2005). A few methods take advantage of enzymes (mainly polymerases) to introduce fluorescent nucleotides for detection such as minisequencing (Pastinen *et al.,* 1997). It was observed that the signal obtained by TAPE was always stronger than that by minisequencing under similar conditions (Figure 4.5 A, B). This could be explained by the fact that incorporation of nucleotides by TAPE occurs at the 3' of the oligonucleotides with little steric hindrances. In contrast, minisequencing usually takes place in the middle of a target strand. In cases where the extension primers are immobilized, they are densely packed on a very small area. Consequently, some of them might not be accessible to polymerase for single-base extension and hence resulting in generating a weaker signal.

One distinctive feature of using TAPE to introduce fluorescence is that it can be accomplished by simply mixing all necessary components together and leaving the mixture on the bench at room temperature for incubation. Thus, no special laboratory equipment is required for the TAPE reaction per se but a fluorescence reader or an array scanner is required for detecting the fluorescent signals. TAPE is also advantageous in its ability to introduce fluorescent nucleotides in an exceedingly

rapid manner. We have shown that strong fluorescent signals could be detected as soon as after one minute of incubation. For some unknown reasons, however, elongation of all available oligonucleotides in a reaction could not be achieved by TAPE. This was observed when we carried out two sequential TAPE reactions. The first TAPE was carried out with the addition of excess TdT and regular ddNTPs, followed by a second TAPE reaction with 1U of TdT and 1pmol of Cy5-ddC. It turned out that intense signals were still acquired, suggesting that many oligonucleotides on the array were not terminated by the first TAPE. Thus far, it is not clear what proportion of the oligonucleotides were elongated. Additionally, we have also observed that TdT permits only a limited number of labelled-nucleotides to be incorporated simultaneously. If these limitations could be addressed, TAPE can be a powerful signal amplification tool.

In the next part of our study, we coupled the signal incorporating function of TAPE to ASMS for SNP genotyping using the TAPE custom-made arrays. All SNPs could be interrogated with the use of only one fluorescent ddNTP, such as Cy5-ddC. In contrast, direct genotyping by ASMS requires all four ddNTPs to be fluorescently-labelled. We have therefore demonstrated that the oligonucleotide array fabricated using TAPE was of sufficiently good quality for genotyping SNPs, and that TAPE is well suited as a signal incorporation tool. However, it is very crucial to optimize ASMS to terminate all the allele-specific primers with ddNTPs. Otherwise, background signal generated by TAPE can be very high. Using ASMS-TAPE, heterozygotes genotypes are called based on the absence of signals from the two corresponding sites on the array. As such, this could lead to a spurious call if the absence of signal is due to a failure of the reaction. This problem was addressed by a two-color ASMS-TAPE strategy. In brief, ASMS was first carried out with TAMRA-

labelled ddNTPs. In the first stage, all three synthetic oligonucleotide target pools were correctly genotyped by ASMS as shown in Figure 4.5B. In the second stage, TAPE was carried out with Cy5-ddC to generate another signal pattern as shown in Figure 4.5A. In this way, a single SNP can be counter-checked in a single assay. Scanning for TAMRA gives the results of ASMS while scanning for Cy5 would give the results of ASMS-TAPE. The two-colour strategy therefore provides a very confident genotype calling. It is noteworthy that, other than allele-specific minisequencing, TAPE is highly compatible with many other allele-discrimination chemistries such as Invader assay. In one of our trials, for instance, two SNPs were successfully genotyped by carrying out the first half of the conventional Invader assay (Third Wave) to generate 5'-flaps with free 3'-OH from 3'-phosphorylated signal probes. Following this, TAPE with Cy5-ddCTP was performed in liquid phase to elongate those signal probes which were successfully cleaved. Subsequently, hybridization was carried out on oligonucleotide arrays with probes complementary to the 5' universal sequences of the signal probes. In this way, Invader assay can be easily adapted to solid support to interrogate numerous SNPs simultaneously for large-scale association studies. Moreover, as a universal signal introduction procedure, TAPE is also potential to improve many commercial platforms.

For gene expression studies using high-density oligonucleotide arrays, RNA fragments are often biotinylated during *in vitro* transcription and subsequently stained with phycoerythrin (Lescallett *et al.,* 2004). In this study, we observed that RNA could be polymerized by TdT with dNTPs. It is thus expected that RNA produced by *in vitro* transcription with normal NTPs can be labelled with fluorescent nucleotides through TAPE.

We conclude from this study that TAPE is a rapid and simple reaction that can be exploited in many ways. These include 1) introduction of modified nucleotides into regular probes for immobilization on solid substrate such as aldehyde-coated slides or magnetic beads, 2) incorporation of fluorescent nucleotides onto oligonucleotides or RNA fragments, thus serving as a signal generating tool, and 3) leveraging on the ability of TdT to extend multiple fluorescently labelled dNTPs contiguously, it could be used as a signal amplification tool for detecting targets of low abundance.

4.5 Future plans

In this study, the application of TAPE in genotyping SNPs was demonstrated using synthetic oligonucleotides as targets. Thus, it is worthwhile to investigate the performance of TAPE by genotyping blood samples. As mentioned earlier, it is difficult to optimize ASMS to ensure that all primers are terminated by ddNTPs. Subsequently, the background signal can be quite high. Thus, it is favorable to integrate TAPE with a better allele-discrimination chemistry. Moreover, the application of TAPE to label RNA fragments for gene expression studies could also be evaluated.

CHAPTER 5 AN UNUSUAL PROPERTY OF TERMINAL DEOXYRIBONUCLEOTIDYL TRANSFERASE (TDT): ELONGATION OF 3' CHEMICALLY MODIFIED OLIGONUCLEOTIDES

5.1 Introduction

During PCR amplification and sequencing reactions, hybridization of primers and template DNA is required for DNA polymerases to initiate the process of nucleotide incorporation at the 3' ends of the primers. Terminal deoxyribonucleotidyl transferase (TdT) was first discovered to be involved in DNA polymerization in calf thymus almost half a century ago (Bollum 1960). Its role in the process of immunoglobulin diversity through V(D)J mechanism has been well defined (Gilfillan *et al.,* 1995). Initially, it was named calf thymus DNA polymerase. However, it was soon found that TdT was very different from other DNA polymerases (Deng *et al.,* 1983). Without the presence of two complementary DNA strands, TdT can still polymerize a single-stranded DNA (ssDNA) efficiently. Indeed, the elongation becomes less efficient if double-stranded DNA (dsDNA) is formed, and dsDNA with a protruding 3' end is preferred to that with recessed or blunt end. Thus, TdT is unique in that it does not require a DNA template and functions at a higher efficiency in the absence of a template. With this distinctive feature, TdT has found wide applications in biomedical researches, such as detection of DNA damage and apoptosis (Tornusciolo *et al.,* 1995; Piqueras *et al.,* 1996; Rohwer *et al.,* 2000).

Modification of synthetic oligonucleotide at its 3' end is currently widely used in biological studies. For example, i) 3'-modifiers are frequently introduced into oligonucleotides to prevent digestion by exonucleases , ii) Oligonucleotide probes are often labeled with non-radioactive fluorophore or enzyme at their 3' ends, iii) 3'-

biotinylated probes permit indirect introduction of signal through fluorophore-conjugated streptavidin , and iv) the oligonucleotide probes in our laboratory are modified with 3' amino group so that they can be immobilized on aldehyde-coated slides for microarray studies. Regardless of the 3'-modifier present on synthetic oligonucleotides, a common effect of these modifications is that the 3'-OH group of the last base is replaced. Consequently, it cannot be used for priming PCR amplification or sequencing reaction because the 3'-OH is required for the extension of nucleotides by DNA polymerases during DNA synthesis, which proceeds in a 5' to 3' direction *in vivo*.

While using 3'-NH_2 modified oligonucleotide probes immobilized on aldehyde-coated microarray slides for other studies, we found that these modified probes could not prevent elongation by TdT. In order to determine whether this is a common phenomenon for other 3' end modifications, oligonucleotides with other 3'-modifiers such as 3'-biotin, 3'-C3 linker and 3'-PO_4 were investigated.

5.2 Materials and Methods

Oligonucleotides

The oligonucleotides with/without 3'-modifiers that have been studied are summarized in Table 5.1, and their structures are illustrated in Figure 5.1. With the exception of 3C3Oligo (Operon, USA) and 3BioOligo (Alpha DNA, Canada), all other oligonucleotides were synthesized by 1st BASE (Singapore).

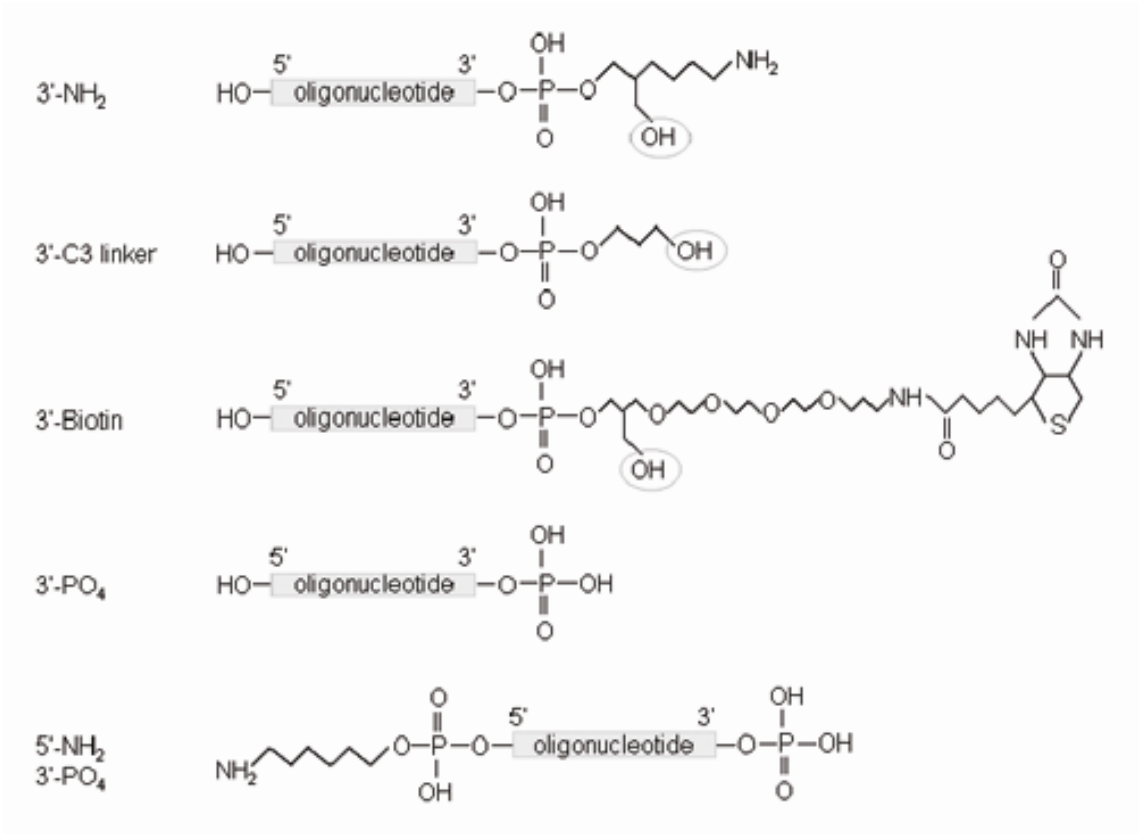

Figure 5.1 Illustration of the structures of 3'-modified oligonucleotides that were investigated in this study

Table 5.1 The oligonucleotides with/out 3'-modifiers included in this study

Oligonucleotide	Sequence and 3'-modifier*
5N3OHOligo	(NH_2) tttttttttttttttcctatacagtcactttt (OH)
3NH_2Oligo	TCGATCCAGTCACGTCGCTAttttttttttttt (NH_2)
3OHOligo	TAGCGACGTGACTGGATCGAtccgggaacacctccgacac (OH)
3C3Oligo	TAGCGACGTGACTGGATCGAttttt (C3 linker)
3PO_4Oligo	TAGCGACGTGACTGGATCGAttttt (PO_4)
3BioOligo	TAGCGACGTGACTGGATCGAttcaggtgtcctgttgccccctcc (Biotin)
5N3POligo	(NH_2) ttttttttttttttTCGATCCAGTCACGTCGCTA (PO_4)

* The underscored, capitalized sequences are complementary to the capitalized sequences without underscores.

Immobilization of oligonucleotides with NH_2-modifier

Of the probes tested in this study, only $3NH_2$Oligo and 5N3POligo can be immobilized onto the aldehyde-coated slides as they have a -NH_2 modifier at either end. Prior to being spotted, they were re-suspended to a concentration of 20μM with ArrayIt™ micro-spotting solution as described earlier. The spotting was performed by a PixSys7500 arrayer with quadruplicate spots for each oligonucleotide in a 2 x 2 format. After an overnight 37°C incubation in a humidified oven, the slides were washed sequentially with 0.2% SDS (2 min), water (2 min), boiling water (2 min), sodium borohydride solution (5 min), 0.2% SDS (2 min) and water (2 min). The $3NH_2$Oligo serves as probes for other 3' modified oligonucleotides. They have a 15-dT spacer added to make them more accessible for hybridization.

The procedure to immobilize aminoallyl-dUTP (aa-dUTP, Fermentas) was exactly the same as that of oligonucleotides with NH2-modifier. Treatment of the slides with shrimp alkaline phosphatase (SAP) was carried out with a cocktail containing 0.4U of SAP and 0.8μl of 10X buffer (RT, 30min), followed by stringent washing.

Elongation on slide

Each reaction cocktail contained 1U (0.44 pmols) of TdT (Fermentas), 1μl of TdT buffer (5 X) and 1pmol of Cy5-ddCTP (Amersham) and topped up to 5μl with deionized H2O. The relative molar ratio of enzyme to substrate was approximately 1:2. This mixture was loaded on the slide for elongation at room temperature for 20 minutes. The slide was then washed once with 2% SSC (4min) and twice with distilled H_2O (3min) before being scanned to obtain the signal of Cy5 (633 nm) by ScanArray® 5000 (Packard BioScience Ltd, UK). Both laser power and photo-multiplier tube (PMT) were set to 100%. Finally, the fluorescence intensities were

measured by QuantArray® 3.0 (Packard BioScience Ltd, UK). Following this, a reaction mixture containing 10U of Exonuclease I (Exo I, Fermentas) and 0.5µl of buffer (10 X) with a final volume of 5µl was loaded on the slide. The slide was then placed in a humidified cassette and incubated in a water bath (37°C, 30min). After washing, the slide was re-scanned to obtain the signal of Cy5.

Elongation in solution

In this set of experiments, the reaction mixture was prepared as described for elongation on slide, with the addition of 50fmols of oligonucleotides which have complementary sequences of the probes. Elongation was similarly carried out at room temperature for 20 minutes. Prior to being loaded on the slide, the reaction mixture with TdT was inactivated by both the addition of 1µl EDTA (0.5M, pH 8.0), and incubation at 95°C for 15 minutes. Hybridization was carried out in a sealed humidified cassette immersed in a water bath (50°C, 1h). The slide was subsequently washed and scanned to obtain the fluorescence signal as described above.

Elongation was also carried out on one 3'-biotinylated oligonucleotide (gtcgtgagcggct-gaggtcgatgctgaggtcgcactcaggtgtcctgttgccccctcc, 58-mer), followed by separation on 2% agarose gel for visualization of elongated products. As much more oligonucleotides are required to be visualized on a gel, each reaction mixture was given 50 pmols of each oligonucleotide and 20U of TdT. Incubation was also carried out at room temperature for one hour before gel shift assay.

5.3 Results

Elongation of 3'-NH2 modified oligonucleotide

Prior to elongation by TdT, the oligonucleotide with either 3'-NH_2 (3NH_2Oligo, 3'-NH_2) or 5'-NH_2 (5N3OHOligo, 3'-OH) was immobilized on the slides. It is clearly shown in Figure 5.2A that, in the absence of TdT, Cy5-ddCTP was not incorporated onto these oligonucleotides. When TdT was introduced, intense signals were obtained (Figure 5.2A, B), indicating that both oligonucleotides were efficiently extended with Cy5-ddCTP. This demonstrated that oligonucleotides with 3'-NH_2 can be efficiently elongated by TdT.

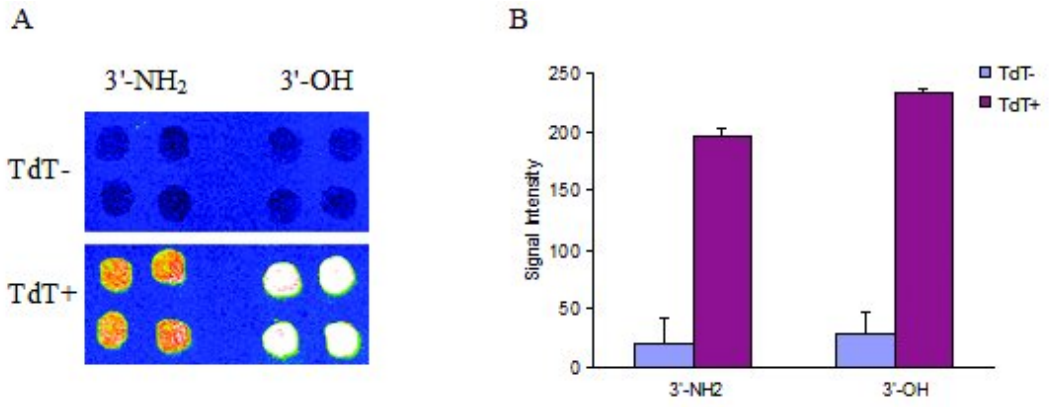

Figure 5.2 Elongation of regular and 3'-NH2 modified oligonucleotides in the absence (TdT-) and presence (TdT+) of TdT (Error bars denote standard deviation)

Elongation of oligonucleotides with 3'-C3 linker or 3'-PO4 modifier

These oligonucleotides were elongated first by TdT with Cy5-ddCTP in solution, followed by hybridization through a common tag to its complementary sequence previously immobilized on the slides. One regular oligonucleotide with free 3'-OH (3OHOligo) was included in this experiment as control. In the absence of TdT, no obvious signal could be observed from the control (3'-OH(TdT-)). In contrast, extremely intense fluorescent signal was obtained when TdT was present (3'-

OH(TdT+)), confirming that TdT is essential for the process of elongation (Figure 5.3). Besides the control, the oligonucleotide modified with 3'-C3 linker was also efficiently elongated by TdT, and the average signal intensity was only slightly lower than that of the control. However, the signal from the oligonucleotide with 3'-PO4 was very weak, suggesting that it can resist elongation by TdT (Figure 5.3).

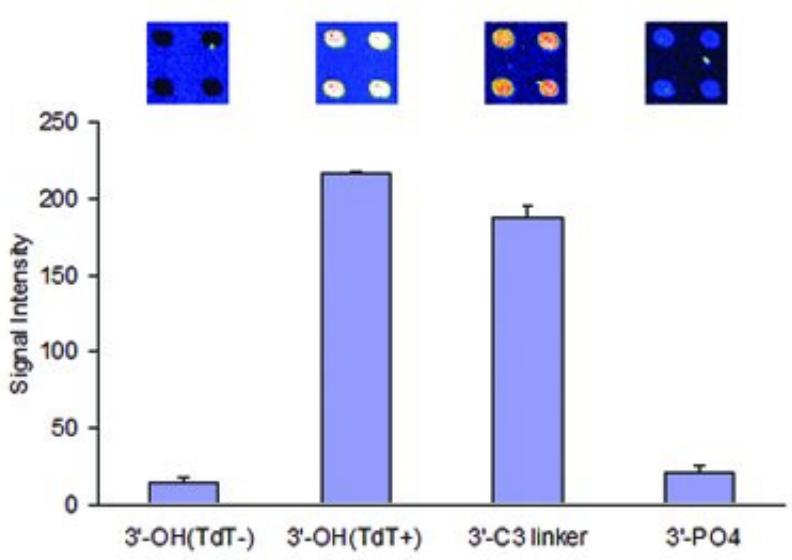

Figure 5.3 Elongation of regular, 3'-C3 linker and 3'-PO4 modified oligonucleotides

Elongation of 3'-biotinylated oligonucleotide

Incorporation of multiple nucleotides into 3'-biotinylated oligonucleotide (58-mer) by TdT was observed by gel shift assay (Figure 5.4). Without the introduction of either TdT or dNTPs or both of them, these oligonucleotides could not be elongated (Lanes 2, 3 and 4). In the presence of TdT (20U) and different amounts of dNTPs (Lane 5, 10mM; Lane 6, 100mM), however, the oligonucleotides were clearly polymerized. Obviously, the oligonucleotides were not evenly polymerized, but the majority of the extension products were longer than 200bp.

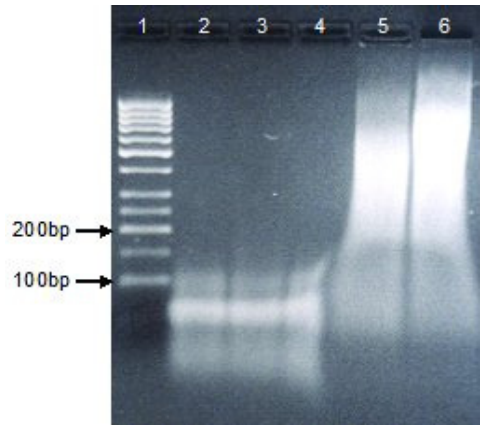

Figure 5.4 Demonstration of the elongation of 3'-biotinylated oligonucleotide by TdT. Lane 1: 50bp DNA ladder; Lane 2: 3'-biotinylated oligonucleotide; Lane 3: 3'-biotinylated oligonucleotide with TdT; Lane 4: 3'-biotinylated oligonucleotide with dNTPs; Lane 5: 3'-biotinylated oligonucleotide with TdT and dNTPs (10mM); Lane 6: 3'-biotinylated oligonucleotide, TdT and dNTPs (100mM).

Elongation of aminoallyl-dUTP

The aminoallyl-dUTP was immobilized on aldehyde-coated slide through its NH2 group before TdT-assisted elongation reaction. The oligonucleotide with 5'-NH2 (5N3OHOligo) was simultaneously spotted on the slide as control. Without treatment with SAP, intense signal was only observed from the control oligonucleotide. When the slide was treated with SAP, however, Cy5-ddCTP was efficiently incorporated into both aminoallyl-dUTP and the control oligonucleotide (Figure 5.5A, B).

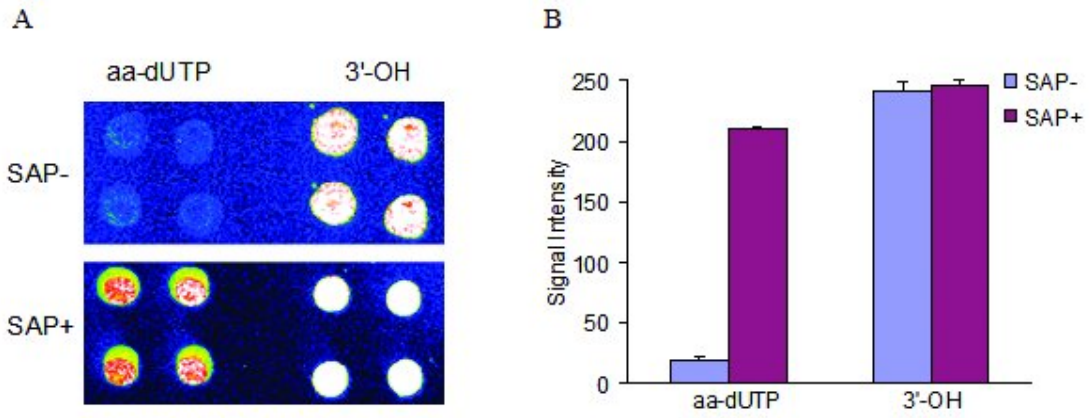

Figure 5.5 Elongation of aminoallyl-dUTP by TdT with (SAP+) and without (SAP-) prior treatment with SAP and with a regular oligonucleotide as a control

Digestion by Exonuclease I

Only $3NH_2$Oligo was used for Exo I digestion experiment after elongation by TdT because it can be immobilized on the slide, thus ensuring that any signal reduction was not due to washing away of some elongated products from the hybrid. After elongation by TdT with Cy5-ddCTP, intense fluorescence signal was obtained. However, with Exo I treatment, the signal became very weak and the signal intensity ratio (Exo I+/Exo I-) was reduced to about 20% (Figure 5.6).

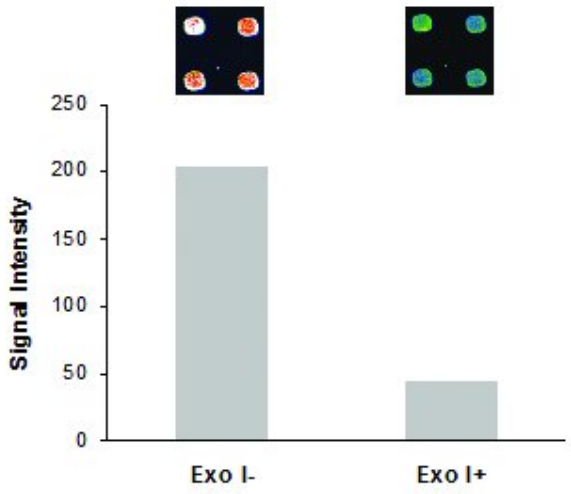

Figure 5.6 The effect of Exo I digest on 3'-NH_2 oligonucleotides that were previously elongated by TdT with Cy5-ddCTP

5.4 Discussion

Like other members of polymerase X family, TdT can polymerize DNA by adding nucleotides repetitively to its 3' terminus (Ramadan *et al.,* 2004). However, TdT is evidently unique because it can catalyze such process entirely in the absence of DNA template. A common belief is that, an oligonucleotide that is modified at its 3' end cannot be polymerized. This is because DNA synthesis proceeds in 5'→3' direction, and the free 3'-OH group of oligonucleotide is required for forming a phosphodiester bond with the subsequent nucleotide during chain elongation. Typically, in vitro

synthesis of oligonucleotide advances in an opposite direction (3'→5'). If 3' modifier is introduced, however, the terminal 3'-OH will be replaced when synthesis is complete. In the course of conducting another study, we were surprised to observe that Cy5-ddC could be incorporated onto 3' modified oligonucleotides by TdT despite the terminal 3'-OH sites being unavailable for extension. These oligonucleotides were originally modified with 3'-NH2 to permit their immobilization on aldehyde-coated slides. In this study, the slides were stringently washed to remove unbound oligonucleotides after the immobilization procedures. As such, only 3' NH_2-containing oligonucleotides could remain on the slide after the washing steps. We are therefore confident that the elongation observed was not due to the presence of any remaining unmodified oligonucleotides, but that TdT has undoubtedly elongated the oligonucleotides despite them bearing a 3'-NH_2 modifier.

One explanation for this phenomenon would be that the 3'-NH_2 modifier also carries a -OH group which might be recognized by TdT as a substitute for the 3'-OH group on the deoxyribose sugar. As shown in Figure 5.1, other than -OH moieties on the phosphate group, another -OH (circled) on the carbon chain is present along with the 3'-NH_2 modifier. This moiety is initially engaged by succinyl-long chain alkylamino (lcaa) group so that the modifier can be attached to controlled pore glass (CPG) support. When oligonucleotide synthesis is complete, succinyl-lcaa will be removed and the -OH is then exposed. To validate this speculation, a few oligonucleotides with different 3'-modifiers such as C3 linker and phosphoryl group, were also investigated. As these oligonucleotides could not be immobilized on the slide, TdT-assisted elongation was performed in solution. Subsequently, they were hybridized to complementary oligonucleotides immobilized on the slide so that fluorescence signal could be detected if elongation did occur. To prevent elongation of immobilized

oligonucleotides, the reaction cocktail was treated with EDTA and heated at 95°C for 15min to inactivate TdT before it was loaded onto the slide. It was observed that the oligonucleotide with 3'-C3 linker could also be efficiently elongated by TdT. Through gel shift assay, 3'-biotinylated oligonucleotide was found to be polymerized by TdT as well. These two modified oligonucleotides and 3'-NH_2 oligonucleotides share an extra -OH group introduced by their respective modifier, despite other structural differences in the linkers. Elongation did not occur without such a -OH group, as was the case for 3'-phosphorylated oligonucleotide. Even if a 5'-NH_2 was introduced into 3'-phosphorylated oligonucleotide (5N3POligo), Cy5-ddCTP was not incorporated. The main difference between 3'-NH_2 and 5'-NH_2 modifiers is that an extra -OH group is introduced by the former one via its linker while the latter has no such group. This further substantiates our hypothesis that this particular -OH group on the linker can be recognized by TdT. Therefore, we are confident to conclude that TdT can elongate 3' modified oligonucleotides if recognizable -OH group is introduced.

We speculate that the unusual elongation by TdT can be partially attributed to its template-independent feature. Even if an appropriate -OH group is introduced and it can be recognized by other DNA polymerases, chain elongation will still not occur because they are dependent on DNA template for incorporation of complementary nucleotides. As modification of oligonucleotides at their 3' ends is only used for in vitro experiments, the physiologically significance of this unusual elongation by TdT remains to be investigated. Nevertheless, our findings have led us to postulate that other than the 3'-OH group on the deoxyribose sugar moiety of DNA strand, some other molecules with free -OH groups might potentially be able to serve as primer for DNA polymerization by TdT. It is thus important to investigate whether DNA

polymerization can occur on 3'-modifiers alone. If it does, we would expect that oligonucleotides could be enzymatically synthesized in the complete absence of DNA molecules, except for the provision of dNTPs as building blocks. Unfortunately, the 3' modifiers available commercially are all attached to CPG supports and their free forms are currently not available for this investigation. We have yet to find a suitable chemical that could be utilized to demonstrate that TdT could incorporate nucleotides from a non-DNA primer. Thus far, we have only managed to investigate whether TdT-assisted polymerization can take place with only a single nucleotide. For this purpose, aminoallyl-dUTP was used as the single nucleotide primer. Surprisingly, Cy5-ddCTP was efficiently incorporated into this nucleotide if it was previously digested by SAP. Otherwise, it could not be elongated. Currently, it is not clear what the mechanism behind this phenomenon is, but this finding suggests that TdT can catalyze elongation with only a single nucleotide. This directly challenges another belief that TdT requires a primer of at least three bases in length (Kato *et al.*, 1967; Schott *et al.*, 1984; Delarue *et al.*, 2002).

Exo I can catalyze the removal of nucleotides (3'→5') from ssDNA strand by breaking the phosphodiester bond within DNA. In this study, treatment with Exo I was only performed on immobilized oligonucleotides following elongation by TdT. Other oligonucleotides with the various 3'-modifiers that were not immobilized were not tested. This is because they can be washed away from the immobilized probes and any decrease in signal intensity cannot be entirely attributed to the action of Exo I. Our result showed that the majority of elongated oligonucleotides were digested by Exo I, as can be seen by an 80% reduction in signal intensity (Figure 5.6). The partial digestion could be attributed to the high density of the immobilized oligonucleotides, in which case some incorporated Cy5-ddCTP could not be easily accessed by Exo I.

Nevertheless, the successful digestion by Exo I indicates that canonical phosphodiester bonds were formed when oligonucleotides with and without free 3'-OH groups are elongated by TdT. This observation does not support the hypothesis by another study that an unusual DNA structure is produced by TdT polymerization and that such structure is resistant to digestion by nucleases (Ramadan *et al.,* 2004).

It was reported that human DNA polymerase lambda, another member of polymerase X family, could elongate RNA primers (Ramadan *et al.,* 2003). Through this study, we found that TdT could also elongate RNA fragments with dNTPs (data shown in previous chapter). Hence, both DNA and RNA can be elongated by TdT. All nucleotides (dNTPs, NTPs and ddNTPs), regardless of being labelled or not, can serve as substrates for such elongation. It was also observed that elongation by TdT was extremely rapid, and very intense signal could be obtained in one minute (data shown in previous chapter). From this study and the work of others, we have strong reason to believe that TdT might have more functions to be identified and its characteristics are not well defined.

5.5 Future plans

In this study, all 3' chemically modified oligonucleotides which could be efficiently elongated by TdT bear a short DNA sequence. For this reason, it is not clear whether the unusual elongation mediated by TdT also depends on such segment. Thus, it will be of great significance to find out whether TdT can catalyze the addition of nucleotides into molecules which are completely nucleic acid-free.

CONCLUDING REMARKS

The initial effort of this thesis was to develop a high-throughput genotyping platform by adapting minisequencing onto solid support. However, it turned out that the performance of such a system was not adequate because of poor probe design, amplification failure and limitation of the one-colour strategy. To improve this situation, especially to address problems brought from one-colour strategy, a new two-colour scheme was devised. With such scheme, all kinds of SNPs and some INDELs could be interrogated simultaneously on a single array. The significance of this strategy can be seen in that it can be easily adapted to most commercial genotyping platforms. For instance, the current version of MIP is a 4-color one so that all SNPs can be interrogated on a single array. However, it requires 4 separate reactions for each allele and the upgrade of the scanner to support 4 different dyes is also compulsory. With our 2-color strategy, however, less genomic DNA is required because only two reactions are adequate for all alleles. Moreover, any confocal scanner with two different lasers is enough to read the signal. Compared to the current 4-color MIP, the only disadvantage of our 2-color scheme is that 30-40% more probes is required. In this thesis, amplification of genomic DNA was achieved by multiple PCR, which is clearly not an efficient way if a large number of SNPs need to be analyzed. Thus, a more effective way to amplify genomic DNA is now under investigation and a more efficient genotyping platform is likely to be developed.

In this thesis, a new universal signal introduction procedure was demonstrated. This method can find wide applications such as genotyping and gene expression profiling. It is also noteworthy that this procedure can be utilized to fabricate arrays with regular oligonucleotides, thus making array fabrication more flexible and cost-effective.

Another unusual property of TdT, elongation of non-standard DNA molecules with an appropriate –OH group, was observed and investigated in this thesis. This phenomenon has never been reported. However, the biological implication of this phenomenon has not been identified thus far.

In short, this study has mainly demonstrated a new 2-color strategy for SNP analysis, and a new universal procedure for signal introduction, array fabrication and qualify control. It is believed that the methods that have been developed in this study offer greater flexibility than many other existing genotyping procedures.

REFERENCES

Baner, J., A. Isaksson, E. Waldenstrom, et al. (2003). "Parallel gene analysis with allele-specific padlock probes and tag microarrays." Nucleic Acids Res 31(17): e103.

Barany, F. (1991). "Genetic disease detection and DNA amplification using cloned thermostable ligase." Proc Natl Acad Sci U S A 88(1): 189-93.

Ben-Dor, A., R. Karp, B. Schwikowski, et al. (2000). "Universal DNA tag systems: a combinatorial design scheme." J Comput Biol 7(3-4): 503-19.

Blondal, T., B. G. Waage, S. V. Smarason, et al. (2003). "A novel MALDI-TOF based methodology for genotyping single nucleotide polymorphisms." Nucleic Acids Res 31(24): e155.

Bollum, F. J. (1960). "Calf thymus polymerase." J Biol Chem 235: 2399-403.

Boule, J. B., F. Rougeon and C. Papanicolaou (2001). "Terminal deoxynucleotidyl transferase indiscriminately incorporates ribonucleotides and deoxyribonucleotides." J Biol Chem 276(33): 31388-93.

Brookes, A. J. (1999). "The essence of SNPs." Gene 234(2): 177-86.

Bujo, H., J. Kusunoki, M. Ogasawara, et al. (1991). "Molecular defect in familial lecithin:cholesterol acyltransferase (LCAT) deficiency: a single nucleotide insertion in LCAT gene causes a complete deficient type of the disease." Biochem Biophys Res Commun 181(3): 933-40.

Cargill, M., D. Altshuler, J. Ireland, et al. (1999). "Characterization of single-nucleotide polymorphisms in coding regions of human genes." Nat Genet 22(3): 231-38.

Carmena-Ramon, R. F., J. M. Ordovas, J. F. Ascaso, et al. (1998). "Influence of genetic variation at the apo A-I gene locus on lipid levels and response to diet in familial hypercholesterolemia." Atherosclerosis 139(1): 107-13.

Chamberlain, J. C., J. A. Thorn, K. Oka, et al. (1989). "DNA polymorphisms at the lipoprotein lipase gene: associations in normal and hypertriglyceridaemic subjects." Atherosclerosis 79(1): 85-91.

Chen, X. and P. F. Sullivan (2003). "Single nucleotide polymorphism genotyping: biochemistry, protocol, cost and throughput." Pharmacogenomics J 3(2): 77-96.

Clifford, R. J., M. N. Edmonson, C. Nguyen, et al. (2004). "Bioinformatics tools for single nucleotide polymorphism discovery and analysis." Ann N Y Acad Sci 1020: 101-9.

Collins, F. S., L. D. Brooks and A. Chakravarti (1998). "A DNA polymorphism discovery resource for research on human genetic variation." Genome Res 8(12): 1229-31.

Collins, F.S., M. S. Guyer, A. Chakravarti (1997). "Variations on a theme: cataloging human DNA sequence variation." Science 278:1580−81

Consolandi, C., E. Busti, C. Pera, et al. (2003). "Detection of HLA polymorphisms by ligase detection reaction and a universal array format: a pilot study for low resolution genotyping." Hum Immunol 64(1): 168-78.

Cullen, L. M., X. Gao, S. Easteal, et al. (1998). "The hemochromatosis 845 G-->A and 187 C-->G mutations: prevalence in non-Caucasian populations." Am J Hum Genet 62(6): 1403-7.

Dachet, C., O. Poirier, F. Cambien, et al. (2000). "New functional promoter polymorphism, CETP/-629, in cholesteryl ester transfer protein (CETP) gene related to CETP mass and high density lipoprotein cholesterol levels: role of Sp1/Sp3 in transcriptional regulation." Arterioscler Thromb Vasc Biol 20(2): 507-15.

De Arruda, M., V. I. Lyamichev, P. S. Eis, et al. (2002). "Invader technology for DNA and RNA analysis: principles and applications." Expert Rev Mol Diagn 2(5): 487-96.

De la Vega, F. M., K. D. Lazaruk, M. D. Rhodes, et al. (2005). "Assessment of two flexible and compatible SNP genotyping platforms: TaqMan SNP Genotyping Assays and the SNPlex Genotyping System." Mutat Res 573(1-2): 111-35.

Dearlove, A. M. (2002). "High throughput genotyping technologies." Brief Funct Genomic Proteomic 1(2): 139-50.

Delarue, M., J. B. Boule, J. Lescar, et al. (2002). "Crystal structures of a template-independent DNA polymerase: murine terminal deoxynucleotidyltransferase." Embo J 21(3):427-39.

Deng, G. and R. Wu (1983). "Terminal transferase: use of the tailing of DNA and for in vitro mutagenesis." Methods Enzymol 100: 96-116.

Dib, C., S. Faure, C. Fizames, et al. (1996). "A comprehensive genetic map of the human genome based on 5,264 microsatellites." Nature 380(6570): 152-4.

Evans, A. E., W. Zhang, J. F. Moreel, et al. (1993). "Polymorphisms of the apolipoprotein B and E genes and their relationship to plasma lipid variables in healthy Chinese men." Hum Genet 92(2): 191-7.

Fakhrai-Rad, H., N. Pourmand and M. Ronaghi (2002). "Pyrosequencing: an accurate detection platform for single nucleotide polymorphisms." Hum Mutat 19(5): 479-85.

Fan, J. B., X. Chen, M. K. Halushka, et al. (2000). "Parallel genotyping of human SNPs using generic high-density oligonucleotide tag arrays." Genome Res 10(6): 853-60.

Flickinger, J. L., G. Gebeyehu, G. Buchman, et al. (1992). "Differential incorporation of biotinylated nucleotides by terminal deoxynucleotidyl transferase." Nucleic Acids Res 20(9): 2382.

Fodor, S. P., J. L. Read, M. C. Pirrung, et al. (1991). "Light-directed, spatially addressable parallel chemical synthesis." Science 251(4995): 767-73.

Fortina, P., K. Delgrosso, T. Sakazume, et al. (2000). "Simple two-color array-based approach for mutation detection." Eur J Hum Genet 8(11): 884-94.

Gerry, N. P., N. E. Witowski, J. Day, et al. (1999). "Universal DNA microarray method for multiplex detection of low abundance point mutations." J Mol Biol 292(2): 251-62.

Gilfillan, S., C. Benoist and D. Mathis (1995). "Mice lacking terminal deoxynucleotidyl transferase: adult mice with a fetal antigen receptor repertoire." Immunol Rev 148: 201-19.

Grainger, D. J., K. Heathcote, M. Chiano, et al. (1999). "Genetic control of the circulating concentration of transforming growth factor type beta1." Hum Mol Genet 8(1): 93-7.

Griffin, T. J. and L. M. Smith (2000). "Single-nucleotide polymorphism analysis by MALDI-TOF mass spectrometry." Trends Biotechnol 18(2): 77-84.

Gunderson, K. L., F. J. Steemers, G. Lee, et al. (2005). "A genome-wide scalable SNP genotyping assay using microarray technology." Nat Genet 37(5): 549-54.

Guo, Z., R. A. Guilfoyle, A. J. Thiel, et al. (1994). "Direct fluorescence analysis of genetic polymorphisms by hybridization with oligonucleotide arrays on glass supports." Nucleic Acids Res 22(24): 5456-65.

Hacia, J. G. (1999). "Resequencing and mutational analysis using oligonucleotide microarrays." Nat Genet 21(1 Suppl): 42-7.

Hacia, J. G., S. A. Woski, J. Fidanza, et al. (1998). "Enhanced high density oligonucleotide array-based sequence analysis using modified nucleoside triphosphates." Nucleic Acids Res 26(21): 4975-82.

Hall, J. G., P. S. Eis, S. M. Law, et al. (2000). "Sensitive detection of DNA polymorphisms by the serial invasive signal amplification reaction." Proc Natl Acad Sci U S A 97(15): 8272-7.

Hall, T. S., P. Pancoska, P. V. Riccelli, et al. (2001). "Sequence context and thermodynamic stability of a single base pair mismatch in short deoxyoligonucleotide duplexes." J Am Chem Soc 123(47): 11811-2.

Hardenbol, P., J. Baner, M. Jain, et al. (2003). "Multiplexed genotyping with sequence-tagged molecular inversion probes." Nat Biotechnol 21(6): 673-8.

Hardenbol, P., F. Yu, J. Belmont, et al. (2005). "Highly multiplexed molecular inversion probe genotyping: over 10,000 targeted SNPs genotyped in a single tube assay." Genome Res 15(2): 269-75.

Hawkins, T. L., J. C. Detter and P. M. Richardson (2002). "Whole genome amplification--applications and advances." Curr Opin Biotechnol 13(1): 65-7.

Henry, M., N. Chomiki, P. Y. Scarabin, et al. (1997). "Five frequent polymorphisms of the PAI-1 gene: lack of association between genotypes, PAI activity, and triglyceride levels in a healthy population." Arterioscler Thromb Vasc Biol 17(5): 851-8.

Hibi, K., T. Ishigami, K. Tamura, et al. (1998). "Endothelial nitric oxide synthase gene polymorphism and acute myocardial infarction." Hypertension 32(3): 521-6.

Hughes, S., N. Arneson, S. Done, et al. (2005). "The use of whole genome amplification in the study of human disease." Prog Biophys Mol Biol 88(1): 173-89.

Iafrate, A.J., L. Feuk, M. N. Rivera, et al (2004). “Detection of large-scale variation in the human genome.” Nat Genet 19:1-3

Igloi, G. L. and E. Schiefermayr (1993). "Enzymatic addition of fluorescein- or biotin-riboUTP to oligonucleotides results in primers suitable for DNA sequencing and PCR." Biotechniques 15(3): 486-8, 490-2, 494-7.

Jain, K. K. (2000). "Applications of biochip and microarray systems in pharmacogenomics." Pharmacogenomics 1(3): 289-307.

Jalanko, A., J. Kere, E. Savilahti, et al. (1992). "Screening for defined cystic fibrosis mutations by solid-phase minisequencing." Clin Chem 38(1): 39-43.

Kaetzke, A. and K. Eschrich (2002). "Simultaneous determination of different DNA sequences by mass spectrometric evaluation of Sanger sequencing reactions." Nucleic Acids Res 30(21): e117.

Kato, K. I., J. M. Goncalves, G. E. Houts, et al. (1967) "Deoxynucleotide-polymerizing enzymes of calf thymus gland. II. Properties of the terminal deoxynucleotidyltransferase." J Biol Chem 242(11):2780-89.

Kennedy, G. C., H. Matsuzaki, S. Dong, et al. (2003). "Large-scale genotyping of complex DNA." Nat Biotechnol 21(10): 1233-7.

Khrapko, K. R., A. A. Khorlin, I. B. Ivanov, et al. (1991). "Hybridization of DNA with oligonucleotides immobilized in a gel: a convenient method for recording single base replacements." Mol Biol (Mosk) 25(3): 718-30.

Kunkel, L. M., K. D. Smith, S. H. Boyer, et al. (1977). "Analysis of human Y-chromosome-specific reiterated DNA in chromosome variants." Proc Natl Acad Sci U S A 74(3): 1245-9.

Kurg, A., N. Tonisson, I. Georgiou, et al. (2000). "Arrayed primer extension: solid-phase four-color DNA resequencing and mutation detection technology." Genet Test 4(1): 1-7.

Kwiatkowski, R. W., V. Lyamichev, M. de Arruda, et al. (1999). "Clinical, genetic, and pharmacogenetic applications of the Invader assay." Mol Diagn 4(4): 353-64.

Kwok, P. Y. (2002). "SNP genotyping with fluorescence polarization detection." Hum Mutat 19(4): 315-23.

Kwok, P. Y. and X. Chen (2003). "Detection of single nucleotide polymorphisms." Curr Issues Mol Biol 5(2): 43-60.

Landegren, U., R. Kaiser, J. Sanders, et al. (1988). "A ligase-mediated gene detection technique." Science 241(4869): 1077-80.

Landegren, U., M. Nilsson and P. Y. Kwok (1998). "Reading bits of genetic information: methods for single-nucleotide polymorphism analysis." Genome Res 8(8): 769-76.

Landegren, U., E. Schallmeiner, M. Nilsson, et al. (2004). "Molecular tools for a molecular medicine: analyzing genes, transcripts and proteins using padlock and proximity probes." J Mol Recognit 17(3): 194-7.

Lander, E. S., L. M. Linton, B. Birren, et al. (2001). "Initial sequencing and analysis of the human genome." Nature 409(6822): 860-921.

Lane, A., F. Green, P. Y. Scarabin, et al. (1996). "Factor VII Arg/Gln353 polymorphism determines factor VII coagulant activity in patients with myocardial infarction (MI) and control subjects in Belfast and in France but is not a strong indicator of MI risk in the ECTIM study." Atherosclerosis 119(1): 119-27.

Lescallett, J., M. E. Chicurel, R. Lipshutz, et al. (2004). "Monitoring eukaryotic gene expression using oligonucleotide microarrays." Methods Mol Biol 258: 71-94.

Lilleberg, S. L. (2003). "In-depth mutation and SNP discovery using DHPLC gene scanning." Curr Opin Drug Discov Devel 6(2): 237-52.

Lindroos, K., U. Liljedahl, M. Raitio, et al. (2001). "Minisequencing on oligonucleotide microarrays: comparison of immobilisation chemistries." Nucleic Acids Res 29(13): E69-9.

Lindroos, K., U. Liljedahl and A. C. Syvanen (2003). "Genotyping SNPs by minisequencing primer extension using oligonucleotide microarrays." Methods Mol Biol 212: 149-65.

Lindroos, K., S. Sigurdsson, K. Johansson, et al. (2002). "Multiplex SNP genotyping in pooled DNA samples by a four-colour microarray system." Nucleic Acids Res 30(14): e70.

Lipshutz, R. J., S. P. Fodor, T. R. Gingeras, et al. (1999). "High density synthetic oligonucleotide arrays." Nat Genet 21(1 Suppl): 20-4.

Liu, W. M., X. Di, G. Yang, et al. (2003). "Algorithms for large-scale genotyping microarrays." Bioinformatics 19(18): 2397-403.

Livak, K. J. (1999). "Allelic discrimination using fluorogenic probes and the 5' nuclease assay." Genet Anal 14(5-6): 143-9.

Lovmar, L., M. Fredriksson, U. Liljedahl, et al. (2003). "Quantitative evaluation by minisequencing and microarrays reveals accurate multiplexed SNP genotyping of whole genome amplified DNA." Nucleic Acids Res 31(21): e129.

Lyamichev, V., A. L. Mast, J. G. Hall, et al. (1999). "Polymorphism identification and quantitative detection of genomic DNA by invasive cleavage of oligonucleotide probes." Nat Biotechnol 17(3): 292-6.

Lyamichev, V. and B. Neri (2003). "Invader assay for SNP genotyping." Methods Mol Biol 212: 229-40.

Martin, E. R., E. H. Lai, J. R. Gilbert, et al. (2000). "SNPing away at complex diseases: analysis of single-nucleotide polymorphisms around APOE in Alzheimer disease." Am J Hum Genet 67(2): 383-94.

Matsuzaki, H., S. Dong, H. Loi, et al. (2004). "Genotyping over 100,000 SNPs on a pair of oligonucleotide arrays." Nat Methods 1(2): 109-11.

Matsuzaki, H., H. Loi, S. Dong, et al. (2004). "Parallel genotyping of over 10,000 SNPs using a one-primer assay on a high-density oligonucleotide array." Genome Res 14(3): 414-25.

Myakishev, M. V., Y. Khripin, S. Hu, et al. (2001). "High-throughput SNP genotyping by allele-specific PCR with universal energy-transfer-labeled primers." Genome Res 11(1): 163-9.

Nilsson, M., J. Baner, M. Mendel-Hartvig, et al. (2002). "Making ends meet in genetic analysis using padlock probes." Hum Mutat 19(4): 410-5.

Nilsson, M., H. Malmgren, M. Samiotaki, et al. (1994). "Padlock probes: circularizing oligonucleotides for localized DNA detection." Science 265(5181): 2085-8.

Nordhoff, E., C. Luebbert, G. Thiele, et al. (2000). "Rapid determination of short DNA sequences by the use of MALDI-MS." Nucleic Acids Res 28(20): E86.

Oliphant, A., D. L. Barker, J. R. Stuelpnagel, et al. (2002). "BeadArray technology: enabling an accurate, cost-effective approach to high-throughput genotyping." Biotechniques Suppl: 56-8, 60-1.

Olivier, M. (2005). "The Invader assay for SNP genotyping." Mutat Res 573(1-2): 103-10.

Pastinen, T., A. Kurg, A. Metspalu, et al. (1997). "Minisequencing: a specific tool for DNA analysis and diagnostics on oligonucleotide arrays." Genome Res 7(6): 606-14.

Pastinen, T., M. Raitio, K. Lindroos, et al. (2000). "A system for specific, high-throughput genotyping by allele-specific primer extension on microarrays." Genome Res 10(7): 1031-42.

Piqueras, B., B. Autran, P. Debre, et al. (1996). "Detection of apoptosis at the single-cell level by direct incorporation of fluorescein-dUTP in DNA strand breaks." Biotechniques 20(4): 634-40.

Pusch, W., J. H. Wurmbach, H. Thiele, et al. (2002). "MALDI-TOF mass spectrometry-based SNP genotyping." Pharmacogenomics 3(4): 537-48.

Ramadan, K., G. Maga, I. V. Shevelev, et al. (2003). "Human DNA polymerase lambda possesses terminal deoxyribonucleotidyl transferase activity and can elongate RNA primers: implications for novel functions." J Mol Biol 328(1): 63-72.

Ramadan, K., I. V. Shevelev, G. Maga, et al. (2004). De novo DNA synthesis by human DNA polymerase lambda, DNA polymerase mu and terminal deoxyribonucleotidyl transferase. J Mol Biol. 339: 395-404.

Rebbeck, T. R., C. B. Ambrosone, D. A. Bell, et al. (2004). "SNPs, haplotypes, and cancer: applications in molecular epidemiology." Cancer Epidemiol Biomarkers Prev 13(5): 681-7.

Rickert, A. M., H. Lehrach and S. Sperling (2004). "Multiplexed real-time PCR using universal reporters." Clin Chem 50(9): 1680-3.

Riva, A. and I. S. Kohane (2004). "A SNP-centric database for the investigation of the human genome." BMC Bioinformatics 5(1): 33.

Rohwer, F. and F. Azam (2000). "Detection of DNA damage in prokaryotes by terminal deoxyribonucleotide transferase-mediated dUTP nick end labeling." Appl Environ Microbiol 66(3): 1001-6.

Ronaghi, M., S. Karamohamed, B. Pettersson, et al. (1996). "Real-time DNA sequencing using detection of pyrophosphate release." Anal Biochem 242(1): 84-9.

Roses, A. D. (2000). "Pharmacogenetics and the practice of medicine." Nature 405(6788): 857-65.

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

Sachidanandam, R., D. Weissman, S. C. Schmidt, et al. (2001). "A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms." Nature 409(6822): 928-33.

Sanghera, D. K., C. E. Aston, N. Saha, et al. (1998). "DNA polymorphisms in two paraoxonase genes (PON1 and PON2) are associated with the risk of coronary heart disease." Am J Hum Genet 62(1): 36-44.

Schott, H., H. Schrade (1984). "Single-step elongation of oligodeoxynucleotides using terminal deoxynucleotidyl transferase." Eur J Biochem 143(3):613-20.

Sebat, J., B. Lakshmi, J. Troge, et al. (2004). “Large-scale copy number polymorphism in the human genome.” Science 305:525–28.

Shen, R., J. B. Fan, D. Campbell, et al. (2005). "High-throughput SNP genotyping on universal bead arrays." Mutat Res 573(1-2): 70-82.

Shi, M. M. (2001). "Enabling large-scale pharmacogenetic studies by high-throughput mutation detection and genotyping technologies." Clin Chem 47(2): 164-72.

Shi, M. M. (2002). "Technologies for individual genotyping: detection of genetic polymorphisms in drug targets and disease genes." Am J Pharmacogenomics 2(3): 197-205.

Southern, E. (1992). "Analyzing and comparing nucleic acid sequences by hybridization to arrays of oligonucleotides: evaluation using experimental models." Genomics 13: 1008-1017.

Southern, E., K. Mir and M. Shchepinov (1999). "Molecular interactions on microarrays." Nat Genet 21(1 Suppl): 5-9.

Stevens, P. W., K. V. Rao, J. G. Hall, et al. (2003). "Improved sensitivity for solid-support invasive cleavage reactions with flow cytometry analysis." Biotechniques 34(1): 198-203.

Storm, N., B. Darnhofer-Patel, D. van den Boom, et al. (2003). "MALDI-TOF mass spectrometry-based SNP genotyping." Methods Mol Biol 212: 241-62.

Suh, Y. and J. Vijg (2005). "SNP discovery in associating genetic variation with human disease phenotypes." Mutat Res 573(1-2): 41-53.

Suzuki, K., M. Iwata, S. Ito, et al. (1994). "Molecular basis for subtypic differences of the "a" subunit of coagulation factor XIII with description of the genesis of the subtypes." Hum Genet 94(2): 129-35.

Syvanen, A. C. (2001). "Accessing genetic variation: genotyping single nucleotide polymorphisms." Nat Rev Genet 2(12): 930-42.

Tornusciolo, D. R., R. E. Schmidt and K. A. Roth (1995). "Simultaneous detection of TDT-mediated dUTP-biotin nick end-labeling (TUNEL)-positive cells and multiple immunohistochemical markers in single tissue sections." Biotechniques 19(5): 800-5.

Tost, J. and I. G. Gut (2005). "Genotyping single nucleotide polymorphisms by MALDI mass spectrometry in clinical applications." Clin Biochem 38(4): 335-50.

Tsuchihashi, Z. and N. C. Dracopoli (2002). "Progress in high throughput SNP genotyping methods." Pharmacogenomics J 2(2): 103-10.

Tuzun, E., A. J. Sharp, J. A. Bailey, et al. (2005). “Fine-scale structural variation of the human genome.” Nature Genet 37(7):727-32.

Twyman, R. M. (2004). "SNP discovery and typing technologies for pharmacogenomics." Curr Top Med Chem 4(13): 1423-31.

Tzvetkov, M. V., C. Becker, B. Kulle, et al. (2005). "Genome-wide single-nucleotide polymorphism arrays demonstrate high fidelity of multiple displacement-based whole-genome amplification." Electrophoresis 26(3): 710-5.

Urakawa, H., S. El Fantroussi, H. Smidt, et al. (2003). "Optimization of single-base-pair mismatch discrimination in oligonucleotide microarrays." Appl Environ Microbiol 69(5): 2848-56.

Valdenaire, O., E. Rohrbacher and M. G. Mattei (1995). "Organization of the gene encoding the human endothelin-converting enzyme (ECE-1)." J Biol Chem 270(50): 29794-8.

Venter, J. C., M. D. Adams, E. W. Myers, et al. (2001). "The sequence of the human genome." Science 291(5507): 1304-51.

Wang, D. G., J. B. Fan, C. J. Siao, et al. (1998). "Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome." Science 280(5366): 1077-82.

Wang, X. L., R. Badenhop, K. E. Humphrey, et al. (1995). "C to T and/or G to A transitions are responsible for loss of a MspI restriction site at the 5'-end of the human apolipoprotein AI gene." Hum Genet 95(4): 473-4.

Weber, J. L. (1990). "Human DNA polymorphisms and methods of analysis." Curr Opin Biotechnol 1(2): 166-71.

Weiss, E. J., P. F. Bray, M. Tayback, et al. (1996). "A polymorphism of a platelet glycoprotein receptor as an inherited risk factor for coronary thrombosis." N Engl J Med 334(17): 1090-4.

Wilkins Stevens, P., J. G. Hall, V. Lyamichev, et al. (2001). "Analysis of single nucleotide polymorphisms with solid phase invasive cleavage reactions." Nucleic Acids Res 29(16): E77.

Wilson, A. G., J. A. Symons, T. L. McDowell, et al. (1997). "Effects of a polymorphism in the human tumor necrosis factor alpha promoter on transcriptional activation." Proc Natl Acad Sci U S A 94(7): 3195-9.

Zhang, H., H. Zhao and K. Merikangas (1997). "Strategies to identify genes for complex diseases." Ann Med 29(6): 493-8.