CHAPTER FIVE

DISCUSSION

5.1 DNA Extraction

Extraction of nucleic acids is crucial for many molecular biology techniques. Extracted genomic DNA (gDNA) from different biological samples is used widely in medical genetic laboratories for diagnosis of genetic diseases. A variety of DNA sources are available including buccal swabs, hair samples and dry blood on Guthrie card. However, they are often limited in their usefulness because of limited DNA yields and purity compared to whole blood.

In this study, gDNA samples were obtained from whole blood specimen, collected into EDTA tubes from 74 consenting FH patients and 77 normocholesterolaemic subjects. The quality and quantity of the gDNA obtained is of crucial importance to ensure the success of subsequent DNA-based molecular analysis. The amplification of DNA template requires extraction methods that guarantee effective recovery of nucleic acids and removal of PCR inhibitors. Therefore, a rapid, reliable and high throughput extraction method that produces a good yield of DNA is essential to process the large number of EDTA specimens collected in this study.

DNA samples from FH patients and normolipaemic subjects were extracted using paramagnetic particles method from commercial kit (Maxwell® 16 Blood DNA Purification Kit) on an automated platform (The Maxwell® 16 instrument). Initial attempts to extract the gDNA from the whole blood specimens using column-based extraction method have produced limited yield and low purity. Furthermore, the blood

specimens that have been stored at -80 ^{o}C for more than 4 years and have caused the extraction column to clog. Therefore, the extraction method using paramagnetic particles was chosen in this study as this method produced high yield of gDNA and high purity. Another advantage is that system has the ability to perform the extraction process efficiently within 30 minutes following initial step of pipetting the whole blood samples into lysis buffer. This walk-away system also required very minimal user manipulation and the extracted gDNA was collected directly into Tris-EDTA (TE) buffer without compromising the yield and purity of gDNA. The automated system used in this study was found to be very reliable and efficient in extracting gDNA from EDTA-containing whole blood that has been stored at the temperature of -80 ^{o}C for more than 4 years.

The quantity and quality of extracted gDNA were determined based on absorbance of UV light at 260 nm, as measured by spectrophotometer (NanoDrop ND-1000). The ratio of absorbance at 260 nm and 280 nm is used to assess the purity of gDNA. Generally, if the ratio is between 1.8 and 2.0, the purity of DNA is suitable for subsequent molecular analysis (Sambrook *et al.*, 1989). A ratio of more than 2.0 indicates RNA contamination while a ratio of less than 1.8 indicates the presence of protein or phenol contamination. The extracted gDNA samples obtained for this study were of good quality, yield and purity as presented in Table 4.1.2.

5.2 Amplification of LDLR gene by PCR

PCR is an in-vitro method of enzymatic amplification of specific DNA sequence, using two oligonucleotide primers that hybridize to the opposite strands and flank the region of interest in the target DNA. A number of factors can significantly affect the amplification of the target DNA sequence, including the oligonucleotide primer design,

PCR cycle parameters and the composition of the PCR mixture. Complete optimisation of the reaction conditions might require several adjustments to the annealing temperature, PCR cycle parameters and salt concentrations.

Ideal and effective oligonucleotide primers should be 17 to 30 nucleotides in length, the G+C content should be 50-60 % and the 3'-terminus of the primers should contain G and/or C (van Pelt-Verkuil *et al*., 2007). In this study, a total of 22 sets of oligonucleotide primers (Bodamer *et al*., 2002) with concentration of 0.2 µM were used to amplify the LDLR gene which spanned 45 kilobases and consists of a promoter region, 18 exons and 17 introns. For each PCR reactions, negative control samples were added to ensure that the reactions were free from DNA contamination. The PCR products obtained from PCR were ranged from 134 bp to 284 bp as listed in Table 3.1. The PCR is considered successful when a single distinctive band was observed using electrophoretic separation technique. The single band indicates that a desired PCR product has been amplified specifically and this was achieved in this study using the optimised PCR protocols.

Another important parameter that may affect the success of the PCR reaction is the annealing temperature for the oligonucleotide primer employed. The optimum annealing temperature is determined from the melting temperature (T_m) of the primers. In this study, the annealing temperature used was 57 ^oC for the promoter region and all exons except for exon 16, for which temperature of 65 ^oC was used because this primer was modified from FH website (www.ucl.ac.ul/fh). The annealing temperatures used in this study were optimised to increase PCR products yields of the promoter and all exons of LDLR gene.

Salt concentrations also affect the success of the PCR experiments. Mg^{2+} concentration can affect primer annealing to the target sequence, the T_m of the oligonucleotide-template complexes and *Taq* polymerase activity. The amount of $MgCl_2$ used in this study was 0. 2mM as this concentration was found to give clearer signal and reduced artifactual bands.

dNTPs are the essential building blocks of nucleic acid molecules, and as such are necessary components of PCR mixture. The optimum concentration of dNTPs is between 20 µM and 200 µM in a 100 µL PCR reaction volume to ensure the PCR product obtained is specific. In this study, 200 µM was used following optimisation experiments to ensure the PCR products obtained are very specific to the target regions.

Each cycle of PCR proceeds through three distinctive phases: denaturation, primer annealing and primer extension. A typical PCR program usually employs 30-50 cycles of amplification; the exact number of cycles is dependent on the initial quantity of template DNA added to the PCR mixture. In this study, a total of 35 cycles were found to be sufficient to amplify specifically all the target regions of LDLR gene.

In this study, the PCR product analysis was performed using an automated on-chip capillary gel electrophoresis system (Agilent 2100 Bioanalyzer). This approach was employed as it gave advantages against the traditional slab gel analysis in term of increased speed of analysis, ability to perform sizing and quantitation of PCR products accurately, require only 1 µl of samples of PCR product and simplify samples preparation. The system enabled 12 PCR products samples in one chip to be analysed within 20 minutes. The preloaded software of the on-chip capillary gel electrophoresis has enabled and facilitated the presentation and storage of data in multiple exportable

data formats. The PCR components and conditions employed in this study had successfully amplified the specific target regions of the promoter and all exons of the LDLR gene.

5.3 Mutations Screening by Denaturing High Performance Liquid Chromatography (DHPLC)

DHPLC is a chromatographic technique that detects mutations on the basis of mismatches (heteroduplex) between amplified DNA fragments with amplified control samples (Xiao & Oefner, 2001). In this study, the samples from definite and possible FH patients were screened and compared to the chromatogram profile obtained from the normal control subjects. Numerous optimisation experiments had to be carried out starting from PCR conditions to obtain meaningful mutation peaks. These involved the mixing of PCR amplicons from two different sources (control and patient samples) in a 1:1 ratio, heat denaturing of the fragments and cooling slowly to allow double-stranded fragments to reform. These processes allowed the original strand pairs to re-anneal and form homoduplexes and the complementary strands from the control and patient amplicons to form heteroduplexes. If a sequence variation is present between the two amplicons, the heteroduplex will contain a mismatch as per illustrated in Figure 2.6.2.

Selection of ideal column temperature is critical in partial denaturing mode of Wave System. A temperature is chosen at which double-stranded DNA containing a mismatch starts to dissociate while completely matched DNA remains double-stranded. In this study, the temperature and buffer gradient conditions were predicted using DHPLC Wave System software (WaveMaker^{TM}) based on the amplicon sequence. Temperature titrations (± 0.2 ^{o}C) were performed from the software-predicted temperature. In order

to empirically determine the optimum temperature at which to screen a particular LDLR gene regions sequence (promoter or exons), a test sample were repeatedly injected into the column by gradually increasing the column temperature until the duplex product peak begins to shift significantly towards shorter retention times. The presence of a mismatch will be clearly detected at this point by observing the appearance of one or two heteroduplex peaks eluting immediately before the homoduplex peak.

In order to make sure the column performance of the DHPLC system is working at optimum capability and capacity, three levels of mutation standard materials were injected into the column at three different temperature profiles. The successful separation of standard materials at 56 ^{o}C (low range mutation standard), 64 ^{o}C (mid range mutation standard) and 74 ^{o}C (high range mutation standard) were presented in Figure 4.5.1.

The observation of heteroduplex peaks in a chromatogram indicated the presence of sequence variants (SNPs insertion or deletion), while samples without mismatches resolve as homoduplexes. Heteroduplex peaks eluted earlier than homoduplexes and can be observed as separate peaks or subtle changes such as appearance of a fronting or trailing shoulder before the homoduplex signal. Any change in the elution profile compared to the profile of normocholesterolaemic samples required sequencing process to establish the location and nature of the variants or mutations.

In this study, distinguishable heteroduplex peaks were observed in samples from FH patients of exon 3, 5, 9, 10 and 12 when compared to the single-peak obtained from the normocholesterolaemic subjects. A total of 46 (62.2 %) FH patients were found to have

disease-causing variants when screened by DHPLC method. Out of 46 patients, 16 (34.8 %) were from Chinese ethnicity and 30 (65.2 %) were from Malay ethnicity. We found that 6 out of 46 (13 %) patients showing heteroduplex peaks on DHPLC were confirmed normal when subjected to confirmation method using Sanger sequencing. Therefore, the proportion of false positive for DHPLC in this study was 13 %. The false positive rate for DHPLC may be due to wrong interpretation of heteroduplex peaks of the DHPLC elution profile. The heteroduplex peaks found were interpreted by the presence of peaks eluted earlier than the normal control or wild type peak. But sometimes the peaks eluted earlier may be due to unspecific binding of primers during PCR. Furthermore, the PCR products used for mutational screening by DHPLC was performed on unpurified PCR products. Therefore, any difference of in the DHPLC elution profile of FH patients compared with the elution profile of NC subjects warranted to be sent for DNA sequencing to confirm.

The remaining forty out of 46 FH patients (86.9 %) showing heteroduplex peaks on DHPLC were confirmed to have presence of LDLR variants by sequencing. Therefore, the proportion of true positive for DHPLC in this study was 86.9 %. The sensitivity of 86.9 % obtained by this study was quite low compared to previous studies (Bunn *et al*., 2002; Jones *et al*., 2001). A possible explanation for low sensitivity of DHPLC found in this study, may be due to the factor that the type of mutations in FH patients are large deletion or rearrangements in LDLR gene which was common in 30% of individuals with FH (Lombardi *et al*., 1995). The remaining 28 FH patients who did not showed any heteroduplex peaks were also sequenced and confirmed to be true negative.

There are limitations to the DHPLC method: high capital cost for purchasing of the equipment, the critical need to predict the precise melting temperature for each PCR

products and the unsuitability for screening of large size DNA fragments. Screening for DNA fragments of more than 700bp using DHPLC is not recommended as the ideal size of amplicon should be between 150 – 700bp (Xiao & Oefner, 2001). Detection of any rearrangement of LDLR gene beyond the ideal amplicon sizes required alternative method such as Southern blotting (Bodamer *et al*., 2002). In order to identify and confirmed the location and nature of mutations, the samples showing heteroduplex peaks by DHPLC in this study were sequenced.

5.4 Mutations Identification of LDLR Gene by DNA sequencing

DNA sequencing is considered as the gold standard for confirmation of specific nucleotide polymorphisms (Franca *et al*., 2002). With the advancement of computer software and detection systems, DNA sequencing has become fully automated (Highsmith, 2005) with first generation of automated sequencing technology was based on the enzymatic Sanger method (Smith *et al*., 1986). Recently, a number of alternative DNA sequencing methods have been developed to offer high-throughput and cost effective sequencing process. Among them are pyrosequencing (Nyren, 2006), massively parallel signature sequencing (MPSS) (Reinartz *et al*., 2002; Brenner *et al*., 2000; Lin *et al*., 2010) and multiplex polony sequencing (Shendure *et al*., 2005).

A total of 46 (62.2 %) FH patient samples that showed heteroduplex profile by DHPLC were subjected to DNA sequencing for confirmation of the location and nature of mutations. DNA sequencing has identified 3 different types of mutations in 6 out of 46 (13%) clinically-diagnosed FH patients that showed heteroduplexes profile by DHPLC. The mutations were identified in intron 3, exons 5 and 9 while SNPs were identified in exons 10 and 12.

The mutation in intron 3 was identified as 313+1G>A in a male Chinese patient. This mutation has also been previously described by Khoo and his colleagues of the same ethnic group and is affecting the splice site domain (Khoo *et al*., 2000). This type of mutation also found in other population such as Dutch, British and Norwegian (Varret *et al*., 1997). The identified FH patient carrying this LDLR variant was found to be a homozygote who presented with highly elevated fasting serum lipid profile. The lipid profile of the affected FH patient showed TC, TG, LDL and HDL levels of 11.9, 1.9, 9.8, and 1.2mmol/L.

Mutation in exon 5 was identified as substitution of T>A at nucleotide position 763 affecting the ligand binding domain of LDLR gene. The nucleotide substitution has caused acid amino change from Cysteine (C) to Serine (S) at codon 234. The identified mutation was found in four FH patients of a Malay family. Of the four identified FH patients, we found a homozygote and three heterozygotes. The homozygote FH patient showed severely elevated TC, TG and LDL of 15.3, 1.4, 13.5 mmol/L and and HDL levels of 1.2 mmol/L, while the heterozygotes FH showed mean TC, TG, LDL and HDL levels of 8.6, 1.0, 6.7 and 1.4 mmol/L respectively. The homozygote FH patient for exon 5 mutation was presented with prominent xanthelasma, corneal arcus and xanthomata, whereas, the heterozygotes FH patients were presented with corneal arcus and xanthomata. This mutation has also been previously described (Azian *et al*., 2006) in Malaysian population.

The third mutation that has been confirmed by Sanger sequencing is the substitution of C>T of exon 9, at nucleotide position 1216 and identified as R385W (www.ucl.ac.uk/ldlr). The identified point mutation is affecting the EGF-precursor homology domain of LDLR gene and was found in a male FH patient from Chinese

ethnicity. This nucleotide substitution has caused acid amino change from Arginine (R) to Tryptophan (W) at codon 385. The identified FH patient carrying this LDLR mutation was found to be a heterozygote with elevated fasting serum lipid profile. The lipid profile of the affected FH patient showed TC, TG, LDL and HDL levels of 8.5, 2.1, 6.6, and 1.4 mmol/L. This mutation has also been previously described in Malaysian population (Azian *et al*., 2006) and among Ashkenazi Jews population (Reshef *et al*., 1996).

Other variants that have been confirmed were in exons 10 and 12. Fifteen (20.3 %) FH patients were found to have substitution G>A at nucleotide 1413 in exon 10, while 19 (52.7 %) clinically-diagnosed FH patients showed substitution T>C at nucleotide 1773 of exon 12. Both identified variants in exons 10 and 12 were found to have not caused amino acid change, Arginine at codon 450 and Proline at codon 570 respectively.

All the FH patients recruited in this study have been clinically-diagnosed either as definite FH (82.4 %) or possible FH (17.6 %) but only 62.2 % of them were found to have LDLR variants with 8% known mutations and 45.9 % SNPs. Large deletions or rearrangement in LDLR gene that may be present in the study subjects may contribute to low number of mutations detected. Other possible explanation is the presence of mutations or polymorphism in other candidate genes that were not detected in this study. This study managed to recruit FH patients from Malay and Chinese families only and this has been identified as the limitation of the study. As Malaysian population is represented by three main ethnic groups, this study should also include patients from Indian ethnicity in order to investigate the prevalence of LDLR mutations in our population.

In heterogeneous outbreed population such as in Malaysia, the presence of mutations in other candidate genes such as APOB, PCSK9, CYP7A1, SREBP-E and SCAP may possibly cause hypercholesterolaemia in the clinically-diagnosed FH patients. Mutations identification in these genes may contribute to gain better understanding of the diversity of disease-causing alleles in FH.