UNIVERSITY OF SOUTHAMPTON

Development of New Molecular Genetic Epidemiological Approaches with Application to the Human Growth Hormone, Insulin-Like Growth Factor I, and Leptin Receptor Genes

Anca-Maria Voropanov, BSc. Honours

Doctor of Philosophy

Human Genetics Research Division

Southampton Medical School

Anca Voropanov PhD

Erratum

1. Page 21, paragraph 1.3.1, line 9. The word Tanner should be removed from the end of the sentence.
2. Page 115, line 5. The statistical "analysis" should read "analyses".
3. Page 133, Results for the IGF I 2396 Marker. Line 2, "fund" should read "found".
4. Page 149, about two thirds of the way down on the left-hand side, "somatomammotrophin" should read "somatomammotropin".
5. Page 157, reference 14. The word "growht" should read "growth".
6. Page 157, reference 16 should read: "Yang, SW, Yu, JS: Pediat Int 42(1): 31-36, 2000".
7. Page 157, reference 18 should read: "Merette, C. Ott, J: Am J Med Genet 63(2): 386-391,1996".
8. Page 157, reference 21. Morton, NE: Ann Hum Genetics "20:125-134, 1955" should read, "Am Hum Genetics 7:277-318, 1955; Sequential tests for the detection of linkage".
9. Page 161, reference 59. Smith, DW: J Pediat 89: 255-230, 1976 is entitled " GH and cerebral dwarfism".
10. Page 167, reference 104. The title of this reference should read: " A mete-analysis of the association of the deletion allele of the angiotensin-converting enzyme gene with myocardial infarction".
11. Page 167, reference: 105 is a duplication of reference 107.

Abstract

Molecular genetic epidemiology study is based upon the analysis of a large number of samples. This analysis has to be achieved quickly and economically for both resources and work force. At the same time it should be accessible to small as well as large laboratories. Such studies include the testing of various polymorphisms in key genes implicated in complex diseases with late life onset, such as obesity, coronary heart disease, osteoporosis and diabetes, all of which are components of the 'metabolic syndrome'.

One class of polymorphism that is particularly suitable for this type of scanning is the insertion/deletion (I/D) polymorphism. This class is more easily distinguished from sequencing errors in the analysis of overlapping sequences. Insertion/deletion polymorphisms have, however, been rather poorly studied. I have developed a method that allows fast and efficient scanning of insertion deletion polymorphisms, method accessible to large as well as small laboratories.

It has to be considered that DNA banks are finite, and is unlikely that lymphocyte immortalisation will be possible, thus the development of highly conservative systems of DNA templates is a necessity. These systems would enable analysis of the genetic information that facilitates extensive high throughput association studies.

Long PCR is a labour saving alternative for the study of the molecular genetics of large genomic fragments. Most importantly the use of the DNA bank would be minimised thus considerably more tests can be performed on the bank DNA. This method was used and furthered in this study. The project aims and achievements are outlined next.

The project had multiple aims:

To develop economical and fast methods for the identification of insertion/deletion polymorphisms.

Use the developed methods to genotype the 3'UTR I/D polymorphism in the human leptin receptor gene in a large cohort.

To branch out and further the conservative methods of implementation of DNA banks already developed by our laboratory (DOP and L-PCR).

Use these developments to perform a cohort study of various polymorphisms in the human GH1 and IGF1 genes.

To develop ARMS assays and perform a cohort study of SNPs in the human GH1 and IGF1 genes.

To perform extensive analysis of the data available for the Hertfordshire banks. To analyse the associations, between all the markers used for the methods development, with the various phenotypic data available. The main East Hertfordshire database, which contains phenotype data for cardiovascular markers, fasting, bloods, GTT's, lung function, and anthropometrics, and relationship with each genotype found for each SNP in both GH1 and IGF1 genes was also set to be analysed. Genotype-phenotype analysis was intended to be performed for the data available for the East Hertfordshire bone database as well as for the North Hertfordshire database that contains information for ageing markers, blood pressure and anthropometrics.

To further expand the investigation of the SNP results by analysing linkage disequilibrium between the GH1 polymorphism and I/D polymorphism in the ACE gene, and LD between the investigated SNP markers in the IGF1 gene.

The following goals have been achieved:

A quick and efficient method has been developed for genotyping insertion/deletion polymorphisms, a method which meets the criteria we have set out to attain. The method developments are presented in section 3.1 with the results and further discussion being outlined in chapter 4, section 4.1.

A method that improves the resolution of bands during MADGE electrophoresis, thus allowing fast and correct scanning of heteroduplexes, was developed and the advances were published in Biotechniques 2000. (1) This method is presented in section 3.1.1 and the results and discussion are given in section 4.1.1.

The method has been successfully applied to the 3'UTR I/D polymorphism in human leptin receptor gene (OB-RB) and analysis of association with obesity was performed on a large sample of subjects in the NPHS II study. The results as well as the discussion are presented in subchapter 4.2, section 4.2.1 and 4.2.2 respectively.

Conservative methods of implementation of DNA banks were searched for, and the inferences of using such techniques were understood with regard to the analysis of various types of polymorphic markers. These methods are outlined in chapter 3, section 3.2.

A selection of diverse polymorphic markers in the human IGF1 gene and GH region were explored using such techniques. Due to excessive band stuttering, a commonly reported 'laboratory handicap' where di-nucleotide repeats are involved, we were unable to genotype the (CA) n repeat in the human IGF1 gene, but we have determined that template dilution can

be taken to great extremes without consequences to genotyping, subchapter 3.2.3. A di- and tetra-nucleotide polymorphism in the human GH region, the A1819 Alu repeat, was successfully genotyped using our new developed method. However, the discriminative binning of the results was not achieved, due to limited number of DNA templates available. The genotyping of the A1819, in the GH gene region, successfully exploited the extreme template dilution method, subchapter 3.2.2.

ARMS assays have been developed on long-PCR templates making further use of the conservative methods of implementation of DNA banks already available in our laboratory. The standard methods for the ARMS assays are presented in chapter 2, section 2.4, with the developments being outlined in section 3.4. The developments for the GH1 ARMS is presented in subchapter 3.2.2 and the IGF1 developments are presented in subchapter 3.2.3

These assays were used for genotyping various SNPs in the human IGF1 and GH1 genes.

Large amounts of data have been generated through genotyping of several SNPs markers in the GH1 and the IGF1 genes, the results being presented in chapter 5, subchapter 5.2 in section 6.2.1 and 6.2.2, respectively.

The analysis for association with the phenotypic data, available for the DNA banks used (the North and East Hertfordshire DNA banks), was performed on relative phenotypic data. The results obtained provided us with new fascinating insights into the involvement of genetic factors in the Foetal Origins of Adult Disease. The discussions of these results are presented in subchapter 5.3.

Another breakthrough discovery achieved is the 20% linkage disequilibrium found between two markers located at approximately 200 kb apart, namely the ACE insertion/deletion marker (data collected by Will Holloway) and one SNP marker in the GH1 gene (data collected by H. Patel). The results are presented in section 5.2.3 with the discussion of these results is given in section 5.3.3.

Linkage disequilibrium was also calculated between the IGF1 SNP markers. The results are shown in section 5.2.4, followed by a short discussion in section 5.3.3.

Table of Contents

List of Figures:

List of Tables:

Acknowledgements

I would like to thank first of all Professor Ian Day for giving me the opportunity of researching in his laboratories during my Ph.D. degree and for his continuous help and assistance.

The MRC and Wessex Research Trust are thanked for the funding of my work.

I would also like to thank Lesley Hinks, laboratory manager, for her constant support not only in the laboratory but also in daily life. Thanks also to Dr. Sandra O'Dell, senior research fellow, for her advice. Thanks to H.J. Patel, W. Holloway for their involvement in this research, and to Holly Sydall for the statistical analysis of the data.

Also, many thanks to Dr. Andy Collins for believing in me and offering all the support one would need through the tough months of the Ph.D. completion.

I also want to thank my mother for her devoted and everlasting support, and also to my partner Steven for his help and for enduring life with a neurotic scientist.

I dedicate this thesis to the memory of my loving father, Iura who always believed in me.

Abbreviation List:

ACTH, corticotrophin

ALS, acid labile subunit

ARMS, amplification refractory mutation system

bp, base pairs

BMC, bone mineral content

BMD, bone mineral density

CRF, chronic renal failure

DOP, degenerate oligonucleotide primer

ds, double stranded

DNA, deoxy-ribonucleic acid

FSH, follicle-stimulating hormone

GH, growth hormone (polypeptide)

GH1, growth hormone gene

GHD, growth hormone deficiency

GHR, growth hormone receptor

GTT, glucose tolerance test

IGFBP, insulin-like growth factor binding protein

IGFI, insulin-like growth factor (polypeptide)

IGF 1, insulin-like growth factor gene

IGFR, IGF receptor

IGHD, isolated growth hormone deficiency

IUGR, intra-uterine growth restriction

kb, kilo bases

KO, knock out

LD, linkage disequilibrium

LH, luteinising hormone

L-PCR, long PCR

OP, osteoporosis

p, probability

PCR, polymerase chain reaction

pHPT, primary hyperparathyroidism

PI , placental insufficiency

PIC, polymorphism information content

REE, resting energy expenditure

RFLP's, restriction fragment length polymorphism

SNPs, single nucleotide polymorphisms

SSLP's, single sequence length polymorphisms

STR, single tandem repeats

tPA, tissue plasminogen activator

TSH, thyroid-stimulating hormone

VNTR, variable number tandem repeats

vs., versus

WGA, whole genome amplification

List of Publications

1. Conservative implementation of Hertfordshire DNA banks for systematic molecular genetic epidemiological studies of foetal growth and adult disease. Poster. INM Day, A. M.Voropanov, S Ye, LJ Hinks, P Briggs, X Chen, SD O'Dell, H Bendall, CH Fall, A Aihie-Sayer, C Cooper, DI Phillips, C Osmond and DJP Barker. *Growth hormone and IGF research*, **9.** (1999)

2. Elimination of dumbbell bands and enhancement of resolution in MADGE using delayed start electrophoresis. A M Voropanov and I N M Day. Biotechniques 28:32-34 (January 2000)

3. Could Growth Hormone (GH) and Angiotensin Converting Enzyme (ACE) gene variants be acting as proxy markers for each other in relation to their phenotypic effects? A. M. Voropanov, H.J. Patel, T.R. Gaunt, W. Holloway, H. Sydall, A. Sayer, D.I.W. Phillips, C. Cooper, I.N.M. Day. Poster. Southampton University Postgraduate Conference. June, 2001.

1 Introduction

At the outset I shall introduce the unifying area of research that I have considered during my project: the human complex disease.

Complex diseases, also known as multifactorial diseases, which include coronary heart disease, obesity, osteoporosis, hypertension, diabetes and even depression, are diseases that affect large number of the global population. These conditions are thought to be caused by a multitude of factors including environmental and genetic factors. No single mutation or environmental factor can be pinpointed as the cause of such illnesses and thus the identification of the multitude of causes and predisposing factors is not a simple task. The findings of Barker and Osmond in 1986 (2), where birthweight, weight at one year, and other anthropometric and growth indices, was predictive of coronary heart disease (CHD), hypertension, insulin resistance, central obesity, and osteoporosis, stand at the base of this study. Many of these mentioned traits are components of the metabolic syndrome. (3)

Recognizing the genetic variants that predispose to common human diseases is an important goal for human genetics. The development of a dense map of single nucleotide polymorphisms (SNPs) will facilitate the identification of these variants. This will possibly aid the test for linkage in families and support an alternative strategy of a genome-wide definition of linkage disequilibrium (LD) and test for haplotype associations using simple families or large numbers of cases and controls. A map of approximately 100,000 SNPs was identified as a goal of the current five-year plan of the Human Genome Project. The plan also aims to estimate the number of SNPs required to map traits in different populations. (4)

This project has set out to investigate two different areas of the complex human diseases field: the first part has searched for association between genotype and phenotype in a candidate gene for obesity; and the second part looked at complex disease from the angle of the foetal origins of adult disease.

In order to identify genetic factors that predispose to complex diseases one has to develop the necessary techniques and methods prior to scanning large population samples to analyse the possible associations existent between genotype and phenotype.

As a first development I established a combination of heteroduplex method in conjunction with microtitre array diagonal gel electrophoresis (MADGE) potentially applicable to the wide range of small insertion/deletion polymorphisms in the human genome. The substrate for this was an association study of such a marker in the leptin receptor gene (OB-Rb) in

relation to weight in a general population sample. For this reason, I consider the leptin receptor gene briefly, and also take into account insertion/deletions in the human genome.
As a second development, I explored in detail, and further developed, possible conservative approaches for DNA bank use, in conjunction with studies of the main growth pathway genes, GH1 and IGFI, in relation to foetal origins of adult disease (FOAD) traits in a population sample from Hertfordshire. This sample contains data for birth weight, weight at one year, and late age cardiovascular disease risk traits, and also ageing and osteoporosis traits. These genes and pathways are therefore considered in same detail.

Obesity and the Leptin receptor gene

Obesity ('the second leading cause of unnecessary deaths'-American Obesity Association) has historically been seen as a trait of social cause. There has been more acceptance that growth reflects parental size but the strong heritability of these traits have been more recognised with the identification of major single gene defects causing rare severe forms of growth deficiencies and obesity in childhood.
Little is known, however, of the influence of common alleles of the many growth factors and obesity genes in general population groups. Few such studies have so far been undertaken, in part because large-scale molecular epidemiological studies remain demanding for a laboratory to achieve and because relevant alleles and assays for them have not yet been fully defined. One must consider what background information in literature is available in order to define more precisely worthwhile hypotheses, goals for development and plans of investigation.
Obesity, indicated by high body mass index (BMI), aggravates and is known to predispose to many of the components of the metabolic syndrome, notably including high blood pressure, non-insulin dependent diabetes mellitus (NIDDM), coronary heart disease, insulin resistance, and stroke, and despite intensive research in the subject and evidence for high heritability, efforts to identify human genes with major effects on body weight have not been successful so far. (5) Evidence for genes involved in the control of body weight and composition in man has come from several sources: adoption studies, twin studies and family studies. Several different genetic epidemiological methods will be discussed in more detail in sub-chapter 1.1. Although good evidence for the genetic control of body weight in man exists, little information is available on specific genes and predisposing mutations. Many candidate genes have been screened for predisposing mutations with little success. (5) In my project I have looked at a polymorphic marker in the leptin receptor gene (OB-R) and scanned a large

population sample in search for a significant association with obesity. Family studies have suggested that obesity is a highly inheritable condition although its genetic determinants remain largely unidentified. Also, due to its complex aetiology, obesity is a challenging condition to treat.

Animal studies have identified leptin as being the product of the 'obese gene' (Ob). Leptin, named after the Greek word *leptos,* which means thin, is a 16 KDa peptide hormone, which by binding to its receptor in the hypothalamus creates the sensation of satiety. Leptin is thus a hormone that regulates body weight and energy expenditure. This hormone also has peripheral actions and can influence reproductive and haematopoietic functions. Obese mice (ob/ob) have mutations in the leptin gene and administration of leptin was found to reduce their body weight. As high plasma levels of leptin are found in most obese humans, it is believed that obesity is attributed to reduced brain responsiveness to leptin, rather than to decreased production of the actual hormone itself.

The study of the insertion/deletion polymorphism in the OB-R gene used an extended number of samples from the general population from the Northwick Park Heart Study (NPHS II). The NPHS II study is a prospective study of 2500 middle-aged (45-65 years) Caucasian males from different locations within the United Kingdom. (6) Each participant is free of coronary artery disease at entry, and baseline information includes clinical and risk data such as blood pressure and smoking.

Growth Hormone and obesity:

Alterations in nutritional status, such as malnutrition and obesity, markedly influence GH secretion and its actions at tissue level. The interactions between metabolic substrates and GH secretion can be viewed as part of the overall regulation of feeding and fasting in order to maintain an adequate body weight and composition. (7)

Central and peripheral factors may contribute to the GH hyposecretion in obesity. GH/IGFI excess reduces serum levels of leptin. (8) Among the hypothalamic factors, either low growth hormone-releasing hormone (GHRH) concentrations, or defective secretion of putative endogenous ligand for GH releasing peptides (GHRP) receptors, can be involved. High levels of free IGF1 and non-esterified fatty acids (NEFA) may inhibit GH release by either acting directly on the somatotrope cells or indirectly via inhibition of GHRH (6). The latter may be enhanced by an insulin-induced sympathetic hypertone. Based on results from animal experiments, leptin in obese individuals may enhance GHRH and hence stimulate GH

secretion. However, leptin resistance might explain the lack of its stimulatory effect on GH release. (9)

The 'Old' vs. 'New' Somatomedin Hypothesis

While many hormones are involved in somatic growth, the main regulator of postnatal growth is believed to be the growth hormone (GH), which is a 22-kDa molecule (although other forms may be found) secreted in a pulsatile manner from the anterior pituitary (10). GH interaction with its receptor in target cells induces the production of somatomedins, or insulin-like growth factors (IGF's). IGF's are members of a large family of insulin-related peptides, which include insulin, IGFI and IGFII. These growth factors in plasma are found bound to a family of proteins called insulin-like growth factors binding proteins (IGFBP), of which 6 have been well characterised. These binding proteins play a role in delivering the IGF's to their target tissues, prolonging their half-life, and are able to modulate their biological responses (11). The majority of IGF's are bound to IGFBP3, which, together with a third protein known as the acid labile subunit (ALS), form a ternary complex in the serum. The somatomedins, and especially IGFI, are thought to interact with the target organs, such as growing cartilage, to induce growth. They also inhibit the pituitary production of GH through a negative feedback loop. At the target tissue the major receptor mediating the effects of both IGFI and IGFII is type I receptor, which is a tyrosine kinase receptor. IGFI receptor is primarily responsible for mediating the mitogenic responses of the cell. (11)

The 'old' somatomedin hypothesis, accepted until recently, proposed that IGFI was a hepatically derived circulating mediator of GH, and also that it is a crucial factor for normal postnatal growth and development. The. 'New' proposed version of the Somatomedin Hypothesis (12), suggests that local production of IGFI (unaffected by the absence of liver IGFI mRNA) is the critical factor that influences normal postnatal growth and development, although hepatically derived IGFI is the major source of circulating IGF1 levels. Absence of the hepatically derived IGFI in IGF1 gene-deleted mice, reduces the total circulating levels of IGFI by 75% (13) Evidence exists for the importance of autocrine/paracrine roles of IGFI in postnatal growth and development (14), although it has not yet been determined which tissues are the major contributors to the circulating levels of IGFI besides the liver. (11) However, more recent work of Yakar *et al.* on liver IGFI-deficient mouse (LID) showed that a fourfold increase in serum insulin levels is present in such conditional knockouts (KO's). (15) Also these mice have the ability to maintain normal glycaemia in the presence of apparent insulin insensitivity. The insulin insensitivity observed in the LID mice is muscle

specific. The new data provided by Yakar *et al.*, provide evidence of the role of circulating IGFI as an important component of overall insulin action in peripheral tissues.(15)

During foetal life, however, the IGFI levels seem to be under a genetic rather than a hormonal (GH) control. Serum IGFI levels were observed to rise progressively during pregnancy in both normal and GH deficient women. The relative importance of the GH in the regulation of IGF1 secretion is thus far lower in utero than postnatally. (16)

Clinical features of GH deficiency:

New-borns with GH deficiency are usually of normal length and weight. Some present micropenis or fasting hypoglycaemia in addition to their low linear growth, which becomes progressively retarded with age. In cases of isolated GH deficiency skeletal maturation is delayed in proportion to their retardation height. Clinical features often also present are truncal obesity, young facial appearance for their actual age, delayed secondary dentition, and high-pitched voice. Puberty can be delayed until late teens but normal fertility is usually achieved. (17)

The importance of growth hormone in adulthood has been only recently recognised. After the completion of linear growth, GH regulates body composition, plasma lipid concentrations, exercise capacity, and physiological well being throughout adult life. Central obesity, insulin resistance, elevated plasma lipids, are just a few of the features presented by patients with adult GH deficiency. The skin of GH deficient adults appears fine and wrinkled similar to that observed in premature ageing. Life expectancy is also reduced with cardiovascular disease being recognised as major contributor to the premature mortality.

Concomitant or combined deficiencies of other pituitary hormones, as for example luteinising hormone (LH), follicle-stimulating hormone (FSH), and thyroid stimulating hormone (TSH), in addition to GH deficiency or combined pituitary hormone deficiency (panhypopituitary dwarfism), causes more severe retardation of growth and skeletal maturation. Reduced bone mineral density has also been a reported feature of isolated growth hormone deficiency.(17)

The features of GH deficiency effects, in early as well as late life, are important to know when performing genotype-phenotype analysis involving the GH1 and IGF1 genes.

1.1 Genetic Epidemiology and Approaches to Complex Traits:

Most of the human disorders that have been genetically characterised to date are inherited in a Mendelian fashion. The molecular tools that are available have enabled their identification through positional cloning techniques. Many components of the metabolic syndrome, such as obesity, coronary heart disease, and osteoporosis, remain poorly understood due to their high degree of complexity. Although the estimated number of functional genes in the human genome, in more recent publications, is approximately 30,000, as opposed to the older figure of 100,000 (18), as functional variation may exist in anyone of them or indeed in all, the task of determining the aetiology of complex diseases is monumental. For many human traits and especially disease outcomes, family recurrence patterns strongly suggest the existence of multiple interacting loci. However, gene effects are of different magnitudes even if many of them are contributing to a complex disease.

Before 1980's the genetic risk factors for a disease were examined mainly by direct analysis of candidate genes, usually through association studies. Traits such as blood group systems (ABO, MN, and Rh) were tested directly against an array of human diseases. After scanning thousands of patients a week association between the ABO group and traits involving the gastro-intestinal tract was discovered. (19)

Linkage, Linkage disequilibrium and Heritability

Much confusion surrounds the terms linkage and linkage disequilibrium. In this section I shall describe the differences between the two, and also point out their relevance to heritability.

Two genes are genetically linked if they are found on the same chromosome and at a small physical distance from each other so that they are transmitted from a generation to another without independent segregation. **Linkage** can therefore be defined as the non-random segregation of genes from parents to the offspring. If a disease-causing gene (transmitted on average to 50% of the offspring) and a marker gene (at which the parent has two alleles) are unlinked, the offspring who inherits the disease gene will have an equal chance of inheriting either allele of the marker gene. If, however, the two genes are linked, an affected offspring will have a greater likelihood of inheriting one form of the marker than the other. If such pattern is observed, in a sufficiently large sample of cases, then the segregation is non-

random and there is evidence for linkage. One example of this case is the ABO group, which is linked to the nail-patella syndrome. When linkage was first established, the nail-patella syndrome was found to segregate with the B allele in a family. (19)

When families are large and recombinants can be counted, analysis is simple and proceeds by directly observing and counting recombinants and non-recombinants, but this ideal situation is very rarely found. Families with interesting diseases are seldom large enough for results from one family alone to reach statistical significance. Therefore the data from several families is usually combined to give sufficient power. Furthermore, the phase of linkage cannot usually be determined unambiguously. Because of homozygosity in a parent it may not be possible to determine exactly which marker allele is being transmitted with the disease. For this reason a maximum likelihood approach is required. This is achieved by using the logarithm of the odds or LOD score method. This approach estimates a recombination fraction θ, which ranges from 0 (complete linkage) to 0.5 (no linkage). Being a function of the recombination fraction LOD scores are calculated for a number of θ values and the maximum value of the LOD (Z) is estimated. The overall probability of linkage in a set of families is the product of the probabilities in each individual family and therefore LOD scores can be summed across families. (20)

Newton Morton developed the LOD score approach, introduced around the same time as nail-patella syndrome linkage was established, in 1955. (21) When a LOD score of 3 or more is achieved linkage is regarded as proven. The results obtained using such method are 1000 times more likely if the genes were linked, than if the genes were segregating randomly. Statistically the LOD score has stood the test of time with only 3% of LOD scores of this magnitude being proven to be false positives (19). Large positive LOD scores (>3) indicate evidence for linkage and negative LOD (<-2) scores, indicate absence of linkage.

Linkage disequilibrium (LD), on the other hand, is the non-random association of specific alleles at loci situated in close proximity. If there were random association between the alleles, then the frequency of each gamete type, in a randomly mating population, would be equal to the product of the population allele frequencies. The rate of approach to such random association or equilibrium is reduced by linkage and thus linkage is said to cause disequilibrium. (22) Linkage disequilibrium (or allelic association) spans short chromosomal regions (typically 30kb to 1Mb depending on many factors-the typical mapping resolution is usually 50kb). For rare, recent mutations, LD may be more extensive. A newly created allele at the time of origin is surrounded by a variety of other alleles at other polymorphic loci (a

haplotype). This corresponds to complete LD existing between the new allele and each of the surrounding polymorphisms. This implies that the new allele is 100% predictive of the alleles present at the nearby polymorphisms. Therefore the existence of linkage disequilibrium enables one allele of a certain polymorphic marker to be used as a surrogate for a specific allele of another.However, LD is not stable over time. (23) Over successive generations the level of LD between two markers will usually decline. During meiotic recombination different alleles at various loci may form new haplotypes. The closer the loci, the smaller the chance of this happening, as recombination is less likely to take place between these loci. Gene conversion events may also change the pattern of LD. Linkage disequilibrium will, in these cases, be lost in time. Other forces can act to preserve or even create LD. Specifically, random drift of haplotype frequencies may occur and thus increase LD. However this is more likely to occur in smaller populations of a stable size. Also, natural selection for or against certain sequences of DNA would drive alleles of adjacent loci to higher or lower frequencies, thus raising the total LD for that particular genomic region. (23)

The regional distribution of LD reflects the population specific demographic history, such as bottlenecks, admixture, inbreeding, migration, and assortative mating. Analysis has been inhibited by the lack of a suitably dense series of readily scorable polymorphic markers that would enable mapping in chromosome regions with sufficient resolution. This problem could be solved by the use of SNPs (single nucleotide polymorphism-see later) (24). These are very stable and abundant markers in the human genomes that are predicted to allow fast advance of this field.

The majority of human phenotype variation could be caused by genetic and environmental factors (non-genetic), as well as by a combination of these two factors in which a chance event could have intervened. Many clinical phenotypes have been shown to have a considerable genetic component. The risk of major common diseases such as cardiovascular disease, diabetes, cancer, mental illness, autoimmune diseases, are expected to be influenced by the patterns of SNPs one has in key susceptibility genes, which are still to be identified.

Genetic epidemiology data from twin, adoptees, and family studies strongly support the idea, with high heritability values, of the degree of total genetic contribution to disease, a good example being the Late Onset Alzheimer's Disease, where 74% of the risk of suffering from such disease is estimated to be genetic (25).

In my thesis, LD is considered in relation to the linked GH1 and ACE genes on chromosome 17 and their associated phenotypes (chapter 5).

1.1.1 Family single gene disorders

One example of disease caused by changes in just one gene, is growth hormone deficiency (GHD), described below.

Growth hormone deficiency

Familial GH deficiency displays a wide variety of hereditary patterns. These patterns of inheritance vary from the autosomal recessive form (Type I) to the autosomal dominant (Type II) and X-linked (Type III) (26). Congenital GHD may also be caused by mutations in the Pit 1 gene, and may also be associated with the deficiency of other pituitary hormones. Type I GHD is subdivided into Type I A and Type I B. Type I A, the most severe form of isolated growth hormone deficiency (IGHD), exhibits complete GHD secondary to gene deletion and is associated with development of anti-GH antibodies when treated with exogenous GH. In the initial stage of the treatment patients respond well to exogenous GH but soon develop antibodies. Type I B is due to splice-site mutations in the GH1 gene, patients have low levels of GH and thus do not develop anti GH antibodies. Sporadic GHD can be idiopathic (most commonly) or secondary to pathological processes at the level of the pituitary or hypothalamus. (27)

In isolated growth hormone deficiency the absence of GH, leading to severe dwarfism, is due to frame shifts, deletions and non-sense mutations. Type two isolated growth hormone deficiency (IGHD II) has an autosomal mode of inheritance and is caused by miss-sense mutations or splice site mutations with a dominant-negative effect. The severity of the IGHD II is variable. Type three IGHD (IGHD III) often associated with hypogammaglobulinemia is an X-linked disorder. (28)

Philips *et al.* found deletions of the growth hormone gene (GH1, also known as GHN) in two families with IGHD IA type, whereas in the GH1 gene, in patients with type II IGHD normal restriction patterns were found, except for two affected sibs of the six families tested who were discordant for 2 restriction sites closely linked to the GH1 cluster (29)

Goossens *et al.* described a double deletion in the GH1 gene cluster in 1986, found in cases of inherited GH deficiency. The two separate deletions were found to be responsible for a 40kb missing fragment of DNA in the region flanking the CSL gene (30) A case of son and daughter of first-cousin Italian parents was reported by Braga *et al.*(31), where isolated IGHD was the result of a 7.6kb deletion within the GH1 cluster gene that not only affected the structural gene for GH (GH1 gene) but also sequences adjacent to CSL. Both cases

developed anti-GH antibodies to treatment with human GH. However no interference with growth was observed.

In a case reported by Takahashi *et al.* (32), a boy of short stature and heterozygosity for a mutant form of GH, the growth hormone, not only could not activate the growth hormone receptor (GHR), but it also inhibited the action of wild type GH because of its greater affinity for GHR and GH-binding protein (GHBP) that is derived from the extracellular domain of the GHR. A dominant-negative effect was thus observed (32).

Illing and Prader observed a distinct form of IGHD in 1972 (33). The features of this particular form were more severe than in the majority of cases with IGHD with exaggerated tendencies to develop antibodies to administered GH that vitiates therapy. The patients show more extreme dwarfism, facial features (baby doll faces) are exaggerated, and hypoglycaemia is a conspicuous characteristic. This could be attributed to the fact that more usual forms of IGHD have some GH in circulation whereas these particular cases did not.

Laron reported four cases with severe IGHD where patients were homozygous for the deletion of the GH1 gene (GH1 or GHN). However in all four cases a good response to human pituitary hormone was observed. In 1988 Masuda (34) also reported a similar response to GH therapy in terms of the development of anti-GH antibodies in four Japanese patients with autosomal recessive GH deficiency.

Short stature is not always linked to GH deficiencies where the hormone is mostly, if not completely, absent from the circulation. Short stature children who were not deficient in GH but who presented low levels of GH-binding protein in their serum, led to the idea that low serum concentrations of GH-binding protein might indicate partial insensitivity to GH due to possible defects in the GH1 receptor gene. Indeed mutations in the region of the GH-receptor gene that codes for the extracellular domain of the receptor were found in 4 of the 14 children studied, but in none of 24 normal subjects. This observation led to the conclusion that some children with idiopathic short stature may have partial insensitivity to GH due to mutations in the GH-receptor gene.(35)

Pituitary dwarfism type II or Laron syndrome is an example of growth hormone deficiency caused by dysfunction in the growth hormone receptor (GHR) gene (10). It is characterised by marked short stature, clinical hypo-somatotropism failure to generate somatomedin (or insulin-like growth factor I, IGFI) in response to GH and with normal or increased levels of GH. Two types of GHR dysfunction are known: Type I, associated with defects in GHR, and type II, due to probable post receptor defects in signal transduction to produce IGF1, or possibly defects in IGF1 or its receptor. Two out of nine patients with Laron syndrome were

reported to have a large deletion of the extracellular, hormone-binding domain. In one Mediterranean patient with Laron syndrome, a T to C substitution causing a change from phenylalanine to serine at position 96 in the extracellular coding domain of the protein was found (36). However this mutation was not present in 7 unrelated subjects with Laron syndrome from different population groups. The GHR markers showed different haplotype background for the mutation in these families.

Many other examples of single gene disorders can be given which include, the C to T substitution at position 181 (residue 43 in exon 4), which resulted in a stop codon (TGA) associated with 2 different GHR DNA haplotypes, is thought to have arisen due to recurring mutations (36), but believed to have arisen independently. This point mutation leading to a premature termination signal [GHR, CYS38TER], the substitution [GHR, GAA180GAG] creating a cryptic splice site 24 nucleotides upstream from the normal exon 6 / intron 6 boundary, predicted to result in an abnormally spliced GHR transcript that would lead to the synthesis of a receptor protein with an 8-amino acid deletion from the extracellular domain, are other examples of the families single gene disorders.

1.1.2 Sib pair analysis

Sib pair analysis has been extensively used in human genetic mapping, in theoretical developments and more sophisticated statistical methodology, as well as being used with some success in mapping susceptibility genes for complex traits. Some examples of successful mapping in complex disease are given below.

A single base variation in the new susceptibility locus for IDDM, IDDM18, located near interleukin-12 (IL-12) p40 gene was found linked with the susceptibility locus for IDDM (T1D susceptibility locus). The linkage between these sites was determined through sib pair analysis in a sample of simplex families. (37) The role of the HLA in type I diabetes, where allele sharing by affected sib pairs has been estimated at about 73%, is one classic example. (38) One major success in mapping common diseases through linkage is the mutation in the glucokinase gene which when present in the foetus caused hyperglycaemia and reduced foetal growth and weight, whereas maternal hyperglycaemia due to the mutation in the glucokinase gene resulted in a mean increase in birth weight.(39)

Linkage has had little success in mapping complex traits and the way for the future is now thought to be linkage disequilibrium.

1.1.3 Family based association studies

Association studies, when performed in a case-control sample, present two major problems: testing a large number of markers without appropriate significance corrections, and secondly the inability to perfectly match the ethnic background of the case with the control, thus increasing the chance of detecting ethnic differences rather than a true association between a marker and a disease. Such analyses rely on linkage disequilibrium to detect association so that the major susceptibility effects will go unnoticed if the marker used is in equilibrium with the pathogenic mutation.

A family-based association method, such as the TDT (transmission disequilibrium test), represents the option to deal with such problems as the latent ethnic stratification (40). Family studies avoid the situation when false positives are obtained, due to latent ethnic stratification in population studies. One example of a false positive is the negative association between Gm haplotype Gm3; 5,13,14 and non-insulin dependent (type 2) diabetes mellitus in American Indians. By using the Gm haplotype it was shown that genetic population stratification is a variable that can confound disease association and lead to false interpretation of results even in large samples and when the statistical significance obtained is high (41).

In the TDT an analysis is made of an allele transmitted from the parent to an affected offspring compared to the un-transmitted allele. If the frequency of transmission of a given allele to an affected offspring significantly exceeds 0.5 this suggests that allele is in linkage disequilibrium with the disease allele. The un-transmitted allele represents, in this type of study, the ' ethnically matched control'. Although such a method has proven useful it has a disadvantage that the genotype of the parents has to be determined and in many cases, especially in subjects affected by late life diseases, the parents are already dead. (19) There are, however, sib TDT tests that do not need parental data, but these lack power. (19)

1.1.4 Population association studies:

In population studies a cohort of cases (individuals affected by the complex disease in question) and controls (normal subjects and who do not have a family history of the disease), are typed for a polymorphism closely linked to the candidate gene of interest. The case-control design, an allele frequency is compared in affected vs. unaffected individuals in large

samples. The most likely source of confounding is the ethnicity, whereby the allele frequency varies by ethnic origin and case and controls are not well matched in terms of ethnic origin.

False-positives are well known in the 'history' of case-control gene association studies. It is believed that high levels of false-positives obtained through this type of study are due to population stratification, although this remains to be proven. It is likely that high false positives rates result from a low prior probability that the few gene polymorphisms studied are causally related to the disease outcomes examined (42).

Choosing Candidate Genes

In order to identify the genes in complex diseases one has to screen candidate genes for genetic variation and also to evaluate the correlation between any of the variation found in such genes and any pathophysiological relevance. Correlating genotypes with phenotypic data might be an easy task when dealing with large phenotypic effects, as is the case of the low-density lipoprotein receptor mutations (LDLR).(43) When dealing with a complex disease, much of the phenotypic variance may be explained by common polymorphisms with modest individual effects, or by effects that are evident only under particular genetic backgrounds or environmental circumstances. Evaluating such variants may be very difficult since no methodology can, so far, successfully eliminate problems such as false positive association of a neutral variant with a disease or co-segregation of a variant with a disease, which may be due to tight linkage to a nearby mutation.

Genetic and phenotypic analyses are used together to find the weak genotype-phenotype correlations that are a feature of complex diseases such as obesity.

The physiological data provides a basis for evaluating positional candidates and permits the identification of candidate metabolic pathways. It improves the sensitivity of the mutation detection analysis by allowing the selection of a group of individuals most likely to present a defect in the gene of interest, as well as improving the specificity of the analysis by allowing the examination of physiological correlates of any variants found, through the study of the wider cross-sectional cohorts.

Towards genome wide association studies.

Following the recent technological and computational developments, including the sequencing of the human genome, many opportunities are opening for the field of genetic research.

The idea of a genome-wide scan for gene effects using large scale testing of single nucleotide polymorphisms (SNPs) or even polymorphisms which include small insertion/deletions and multi-nucleotide changes, as well as single nucleotide changes, has focused attention on the efficiency of different approaches used for mutation detection. The original idea of Risch and Merikangas (44), of studying coding or promoting variants with potential functional significance, is one approach. Alternatively non-coding, evenly spaced SNPs with high density could be used for identifying disease loci through linkage disequilibrium. (22) The number of SNPs needed for the latter alternative is the subject of a wide debate, mainly because the extent of LD in the human genome has not so far been determined on a large scale. Linkage disequilibrium depends heavily on the demography and social history of the populations in question. Studies of isolated populations such as Finnish, Ashkenazi Jews and Mennonites have shown extensive LD around rare disease mutations. It is not known if the same degree of LD will be found for higher frequency variants, although as a general rule the disequilibrium tends to decline with increasing allele frequency due to an older coalescence time.

As many as 500,000 evenly spaced SNPs have been stated to be necessary for the detection of LD of sufficient magnitude to allow coverage of the genome, even in population isolates. (24) Kruglyak's 3kb extent of useful LD (23) argues against Morton and Collin's estimation of 30 to 300 kb. (22) The discordant view may have risen because Kruglyak assumed a panmictic world population. Others have argued that founder populations, especially those that have remained small over time, will require fewer SNPs for mapping purposes. Smaller populations should thus improve the chances for detection of rare disease alleles (less than 5% in frequency) due to the greater LD per base pair, although the same may be true for common alleles (greater than 5% frequency). The power of association tests decreases with a decrease in LD, and as a result of discordance between the frequencies of disease and marker alleles. The greater the density of markers the greater the chances of finding a marker in LD with the disease allele. This is to say that in a genome-wide random SNP approach even with high density, many disease-causing genes would be missed.

Arguments favouring the use of SNPs in coding and promoter regions are first, that they are most likely to be of functional significance and to influence directly the trait(s) under study, and secondly that even if they are not causative in a gene such SNPs are likely to be in LD with the causative allele as its true for the random alleles as well. Prioritising SNPs based on likely functional significance would be a way of reducing the number of SNPs studied.

Furthermore, discovering the time and distribution of expression of genes, for example by using microarrays, would elucidate some of the enigma surrounding the involvement of various genes in diseases.

Proximity vs. causal markers (LD)

Linkage disequilibrium is thus a powerful tool in genetic epidemiology and despite future developments the study of human disease from an epidemiological perspective will still be important. The high–density genome scans, using evenly spaced SNPs depending on linkage disequilibrium, are merely an extension of the traditional reverse genetics techniques. Linkage disequilibrium allows the mapping of a gene to within 50kb, thus a lot 'closer' than linkage alone would allow. However, the presence of an allelic marker associated with disease does not necessarily mean that the marker is causative of the disease. Identification of a causal site within a region of LD may cause considerable difficulties. For example, human growth hormone is found at about 200 kb distance from the ACE gene, therefore linkage disequilibrium may exist between markers in these two regions. Also the possibility exists that many of the phenotypes associated with the ACE polymorphisms may be due to marker(s) in the GH1 region.

1.2 Molecular Genetic Epidemiology/Genomic Aspects/Laboratory Processes

1.2.1 Types of markers and their characteristics:

The aim of genetic mapping is to determine how often two loci are separated by meiotic recombination. If two loci are linked (inherited together more frequently than expected under random segregation), this linkage can be used as a mapping technique. Genetic markers are used in most human genetic mapping projects.

There are two general approaches in mapping using markers:

Disease marker mapping, and

Marker-marker mapping.

DNA polymorphisms provided for the first time a set of markers which were sufficiently numerous and spaced across the entire genome. DNA markers also have the advantage that they can be typed by the same technique, and their chromosomal location can easily be detected using fluorescent in-situ hybridisation (FISH) or other physical mapping methods.

The markers in this section are presented in chronological order in which they have been historically found.

Restriction fragment length polymorphism (RFLP)

RFLPs are generated whenever a mutation alters a restriction site in the genome, or inserts or deletes some part of a sequence between two restriction sites (45). The first generation of markers described, owe their existence to restriction site polymorphism (46). This type of marker presents the following limitations: firstly their informativeness as markers is limited due to the limited number of alleles formed, and secondly, the original technology (Southern hybridisation, *in situ* hybridisation) was quite laborious and required plenty of time, money and DNA. The second problem has now been resolved as RFLP's can be typed by PCR. The limited informative value of the RFLP's comes from the fact that the restriction site is either there or not, so they are effectively bi-allelic, but they are not as abundant as SNPs (discussed later in this chapter).

Large insertions/deletions (rearrangements)

Large deletions in the human genome are likely to have severe physiological effects, especially when associated with coding regions or when occurring in a gene rich region of the genome. (47). Cacciari *et al.*identified patients with extreme short stature and absence of GH secretion was identified to be homozygous for a large deletion in the hGH1 gene cluster. In two siblings the large deletion removed four of the genes in the GH1 gene cluster: hGH-N, hCS-L, hCS-A, and hGH-V. Both siblings produced high levels of anti-GH antibodies in response to replacement therapy and their growth was arrested shortly after starting the replacement therapy. (47)

An example of loss of heterozygosity comes from the analysis of DNA polymorphisms in the region close to the RB-1 locus (involved in retinoblastoma), which has revealed that normal cells differ from tumour cells in that normal cells are heterozygous at many loci whereas tumour cells are homozygous. The loss of heterozygosity may occur as a result of a structural deletion of normal retinoblastoma gene or loss of the chromosome carrying the normal retinoblastoma gene. The deletion is visible on chromosome 13 through cytogenetic analysis of the chromosomal region. (16)

Small insertion/deletions:

Small insertion/deletions (insertion/deletion polymorphisms-IDP) are estimated to be about one tenth as frequent as one base substitution (single nucleotide polymorphisms SNPs). Their structural distinctions from SNPs may confer greater likelihood of phenotypic impact and offer different possibilities for identification and genotyping.

Insertion/deletion (in/del) polymorphisms remain a relatively unexplored topic in human genetics and appear to have a frequency intermediate between single nucleotide polymorphisms and multiallelic microsatellite polymorphisms. It is also possible that insertion /deletion may have greater effect on the function of the gene than single base substitutions.

Insertion/deletion polymorphisms have several advantages as polymorphic markers: they are easier to distinguish from sequencing errors in analysis of overlapping sequences, heteroduplex analysis and sizing of the PCR product on vertical slab gels is an easy way of detecting this type of polymorphism, and the two alleles are distinguishable through various hybridisation methods. There are a few examples of positive functional or linkage disequilibrium markers involving small insertion/deletions.

The insertion/deletion Alu polymorphism in the angiotensin-converting enzyme (ACE) (287bp);

The 45 bp insertion/deletion in the 3'UTR, exon 8 of the UCP1 gene.

The CKR5 (15bp) involved in disease progression;

The ApoB signal protein [$(CTG)_3$ x 9] insertion deletion shown to be associated with elevated plasma cholesterol levels;

The stromelysin MMP3 5A/6A polymorphism in the promoter region thought to regulate the transcription of the gene.

A scan of the human genome for ins/del polymorphisms can be used for linkage disequilibrium mapping. They offer superior discrimination by hybridisation compared to SNPs. Ins/del polymorphisms can be detected in acrylamide-derived gels through the formation of heteroduplexes, which have a considerably lower mobility in the gel, compared with homoduplexes. Heteroduplex generation, by post PCR annealing, allows immediate genotyping by electrophoresis without further intervening steps (48).

Small insertion/deletions although abundant in the human genome are largely uncharacterised due to limited scanning techniques. These polymorphic markers could be highly informative therefore a fast and efficient method of genotyping such polymorphism, would be beneficiary

especially to the field of genetic epidemiology. One such marker is the insertion/deletion in the 3'UTR region of the human leptin receptor gene, a gene believed to be implicated in morbid obesity.

Single Sequence Length Polymorphisms (SSLP's) are arrays that display length variations and can be multiallelic. There are two types of SSLP's:

Minisatellites (VNTR -variable number tandem repeats), and

Microsatellites (STR- single tandem repeats) are much shorter than minisatellites, usually 2,3,or 4 nucleotide units.

Minisatellites (or VNTR's, variable number tandem repeats) are a type of genetic marker that gives a large number of alleles. Minisatellites are not evenly spread across the whole genome but they tend to be concentrated around the telomeres. The technical problems encountered during typing such markers have now been overcome by PCR typing. However, PCR amplification is quicker and more accurate with sequences less than 300 bp in length size that is exceeded by most minisatellites.

The use of microsatellites for genetic mapping presents the technical problems of Southern blotting and use of radioactive probes. Microsatellites are standard tools for linkage analysis, which were first described in 1989 (46). They are small blocks of randomly repeated DNA, in which the repeated element is usually a 2, 3, or 4 nucleotide sequence, for example (CA)n, (CAG)n, (AGAT)n. Usually microsatellites consist of 10-30 copies of repeats that are no longer than 4 bp in length, an advantage for the PCR typing. Di-nucleotide repeats, the most commonly found in all eukaryotic genomes tested, comprise 0.5% of the human genome. The number of dinucleotide repeats in the human genome was estimated at 50,000 (CA) repeats that implies that they are present at 30 kb apart (49). One valuable feature of the microsatellites is that, not only are they distributed more uniformly across the genome compared to other types of genetic markers, and also they are distributed around the centromeres. Three types of dinucleotide repeats are distinguishable according to their composition:

Perfect repeats, with no interruption in the dinucleotide repeat, e.g. (CA)n repeat,

Imperfect repeats, which have one or more interruptions in the run of the dinucleotide repeat, e.g. (CA)n-CCA-(CA)m, and

Compound repeats, with adjacent tandem repeats of different sequences, e.g. (CA)n(TG)m.

The number of repeats (but not the type of repeat) is important, and has a bearing on the level of heterozygosity. The function of these repeats is unknown although it has been suggested that they may represent 'hot spots' for recombination (49).

Trinucleotide repeats and tetranucleotide repeats are 10x less frequent than dinucleotide repeats, with spacing estimated at around 300-500kb. The most frequently found trinucleotide repeat is the (AAB) n form and for the tetranucleotides, which are rarer still, the (AAAB)n form is the most common.

The tri- and tetranucleotide repeats are more stable than the dinucleotide repeats during PCR amplification, so that the genotypes are more easily read, lacking the slippage 'ladder' of bands seen in the PCR amplification of the dinucleotide repeats. We have analysed, during the developments of conservative methods of implementation of DNA banks, a di- and tetranucleotide repeat in the human GH1 region (A1819 polymorphism) and a (CA)n repeat in the human IGF1 gene.

Single nucleotide polymorphisms (SNP)

SNPs are single base pair positions in genomic DNA at which different alleles exist in normal individuals in some populations, where the least frequent allele is present with an abundance of 1% or higher. SNPs are usually bi-allelic markers. Single base pair insertion/deletion polymorphisms would therefore not be considered to be SNPs as the base is either present at the locus or not (i.e. inserted or deleted) (50).

Single base variants in cDNA (cSNPs) are usually considered to be SNPs. This ignores the possibility that they might be the result of RNA editing processing. The typical frequency of observed single base differences in genomic DNA, from two equivalent chromosomes, is of the order of 1/1000bp. Not all of these will be polymorphisms for which the least abundant allele will be present at a frequency of 1% or higher in the population tested, i.e. it will not be at the level required for designation as an SNP (24).

SNPs are considered to be the most abundant type of human genetic variation and a resource for mapping human complex genetic traits (51). Some SNPs give rise to restriction length fragment polymorphism (RFLP's). However, most SNPs do not give rise to restriction sites. The estimated total number of SNPs found in the whole human genome is considered to be between 300,000 (23) and 3,000,000 (52). The type of polymorphism in this case can be either non-coding or coding. The advantages of the SNPs are their abundance in the human genome and the possibility of typing them without the use of gel electrophoresis. Gel

electrophoresis is difficult to automate thus abandoning such a method gives the possibility of fast and less labour intensive screening of the whole human genome.

One rapid way of screening for SNPs is DNA chip technology. (53), (54). Many SNPs can be detected in one experiment. (55) Another described method is the dynamic allele-specific hybridisation (DASH) (56), in which discrimination between the alleles is achieved by breaking down the mismatched hybrids (hetero-and homoduplexes) through an increase in temperature. The allele present in the sample of DNA is detected from the temperature at which the hybridisation-dependent fluorescent signal disappears.

We have developed ARMS assays for a number of SNPs in the human GH1 and IGF1 genes, and used these developments to genotype a large population sample. The data produced was analysed for various associations.

1.3 Foetal Origins of Adult Disease.

1.3.1 Foetal Origins of Adult Disease (FOAD) observations

Studies in animals and humans have suggested that the nutrients and oxygen that the foetus receives usually influence foetal growth.(57) Penrose in his analysis of birthweight groups within relatives, concluded that 62% of the variation between individuals was the result of the intrauterine environment, 20% due to maternal genes and 18% of foetal genes (57). These findings, supported by animal studies, suggest that the mother, rather than a genetic inheritance from both parents, controls the size at birth (57). This can be argued against if one considers Tanner charts, which show that older children fall on mid-parental centile, and if the father's centiles matter, then since he only contributed DNA, his genes (at least) must matter Tanner. Animal experiments have shown that hormones, under-nutrition, and other factors that affect development during the sensitive periods of early life, permanently program the body's tissues and systems (57). During the foetal period the body achieves its highest growth rate, during which the body is at its most susceptible to programming. The high growth rates achieved by the foetus (compared with childhood) are largely a result of cell replication. Barker *et al.* observed an increase in prevalence of a range of late onset adult diseases and traits (including CHD, hypertension, dyslipidaemia, type II diabetes, insulin resistance, central obesity, and osteoporosis) in individuals with low birth weight, one-year weight, or other related life measurements. This is known as the Foetal Origins of Adult disease (FOAD) Hypothesis (34).

1.3.2 Nutritional Hypothesis

In 1986 Barker and Osmond (2) described the association between infant mortality, childhood nutrition and ischaemic heart disease. Subsequent research into this subject has revealed that foetal birth weight, one-year weight, and other anthropometrics and growth indices, may predict occurrence of cardio-vascular disease, hypertension, dyslipidaemia, type II diabetes, insulin resistance, central obesity, osteoporosis, polycystic ovaries and chronic bronchitis in later life. Many of these findings constitute the metabolic syndrome and have been described by Barker in 1994, although the molecular mechanisms involved are not clearly understood (58). One classic example would be the observations that the Dutch famine in World War II (calorie intake of<1500kcal/day) was followed by low birth weights and subsequently cardiovascular disease in the offspring. Nutritional factors are thought influence the foetus to develop the 'thrifty phenotype' later leading to nutritional excess. Animal experiments, as well as studies of differences in human hypothalamic-pituitary system (HPS) axis according to birth weight, support that nutrition exerts such an influence. The ability to influence the health of future generations through detailed attention to maternal/foetal/newborn nutrition would potentially have enormous implications for the future, since many of the greatest causes of morbidity and mortality in later life are involved. This area is the subject of intense research (3).

1.3.3 Molecular Genetic Epidemiology and FOAD/ Genetic Hypothesis

Studies involving adopted twin children estimating heritability have been slow to examine the contributions to FOAD, limited by the study of a trajectory requiring life data at points very many years apart, and concerning not one disease but a whole array of diseases and traits.

If the FO (foetal origins) and AD (adult disease) are considered apart there are substantial heritabilities. Although birth weight shows low heritability (21) it is influenced by maternal factors and postnatal trajectory shows slow shifts toward the growth curve defined by mid parental value (59).

Each of the traits of the metabolic syndrome shows considerable heritability (60). This is reflected in the present massive sib pair linkage studies being undertaken in hypertension, NIDDM, obesity, dyslipidaemia, and even in earlier onset coronary disease. The genetic hypothesis can be formulated in parallel (the **3G** hypothesis for genotype, growth and gate

keeping) (34), that low levels of growth activation attributable to low growth genotypes in foetal life would result in poor foetal growth. Also, that in later life these genotypically poor pathways would offer disadvantageous metabolic 'gate keeping' hence risking the 'metabolic syndrome'. Thus both the above are independent events (no nutritional programming) associated simply on account of a third factor, a genotype. However, the nutritional and genetic hypotheses may perhaps interact (Fig.1).

Genetic and environmental, including nutritional factors, are known to control foetal growth and development. The non-genetic elements become increasingly important in the second half of pregnancy and contribute to the process of natural constraint on the size of the baby, therefore overriding paternal genetic influences.(61) On the other hand environmental factors could lead to foetal overgrowth. One of the most commonly found example of this being maternal diabetes mellitus. It is well known that the insulin-like growth factor (IGF) system mediates growth in response to nutritional status in the foetus, however the exact part played by the maternal IGF system is not completely understood.

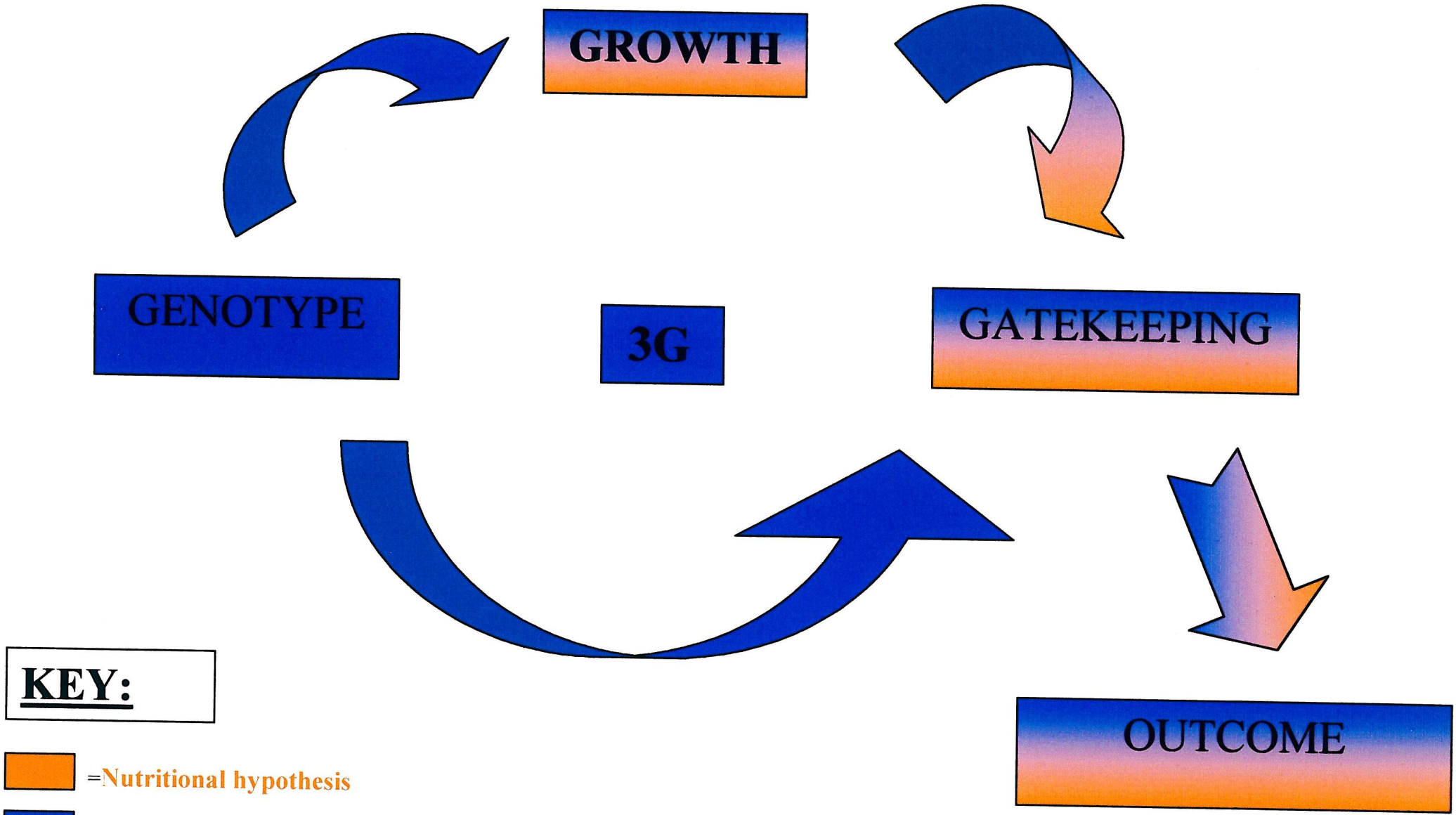


Fig.1 '3G' Hypothesis interacting with the Nutritional Hypothesis

1.4 Review of Major Candidate Genes

I shall discuss firstly the leptin receptor gene (OB-Rb) as a candidate gene in obesity.

The leptin receptor is a single transmembrane spanning receptor member of class I cytokine receptor family. It exists in several splice variants and each displays a different pattern of expression. The long isoform or the major signalling form of the leptin receptor, the Ob-Rb, is expressed in many tissues including most brain regions and it is believed to mediate the hypothalamic weight-regulatory actions of leptin. The functional domains of the leptin receptor, that are common to the other cytokine receptors, include the ligand binding serine residue domains and the transmembrane domain, as well as the intracellular domain containing sites for Jak and STAT interactions.

The leptin metabolic pathway begins in the adipocytes. Leptin is secreted into blood and reaches the hypothalamus where it binds to its receptor, the OB-RB. Different types of neurons are responsive to the actions of leptin. For my investigation I have selected the direct pathway, in which leptin binds to its receptor and has a direct effect on the brain. Through binding to this receptor leptin causes a decrease in the food intake, increases the metabolic rate, and increases energy expenditure.

At the start of my study, four amino acid variants of the coding region, four silent mutations, nine intronic variants and the 3' untranslated region (UTR) pentanucleotide insertion/deletion polymorphism, had been described in the human receptor gene.

Appendix 3 summarizes the polymorphisms found in the leptin receptor gene. It is important to remember the fact that all these studies have taken place on a small number of subjects, and that highly controlled cases of morbidly obese patients were used.

Major candidate genes in principal growth pathways:

The major candidate genes that are involved in pre- and postnatal growth pathways considered here are the growth hormone gene (GH1), GH receptor gene (GHR), insulin-like growth factor I gene (IGF1) and the insulin-like growth factor binding proteins (IGFBP) together with the IGFI type I receptor (IGF1R).

For human organ growth, the insulin-like growth factors play an important role as multifunctional growth factors, which modulate cell growth through autocrine, paracrine and endocrine routes. Insulin-like growth factors (IGF's) constitute a system of peptides that promote mitosis and growth of various organs in the human body in prenatal as well as

postnatal life. These peptides are found bound to the IGF-binding proteins which also modulate their biological activities (61).

The importance of the insulin-like growth factor (IGF) family in human foetal growth has long been recognized. Whereas IGF-II promotes proliferation and differentiation mainly during prenatal development, IGF-I is thought to influence several biological processes, which include growth, differentiation, metabolism, angiogenesis or apoptosis, mainly postnatally. Both IGF-I and IGF-II exert their actions by binding to their receptors, IGF-IR for example, receptors, which are found nearly ubiquitously. The transport of the IGF's, the mediation of their actions and the control over their availability to their receptors, is performed by the IGF-binding proteins (IGF-BP's) of which 6 have been well characterised (61).

There are many indications that deregulation of the IGF system is involved in human growth disorders, e.g. overgrowth syndromes and in the pathogenesis of some paediatric tumours, whereas data concerning the role of the IGF system in general foetal growth retardation are still conflicting (62).

Insulin-like Growth Factor I (IGFI)

IGF1 is an important growth factor and hormone, which has differentiating effects on most cell types, regulates protein turnover, as well as being a powerful mitogen.

The IGF family members are thought to be involved in endometrial development during menstrual cycle and also, to play an important role in the process of implantation. Due to their spatial and temporal expression during endometrial cycle and bearing in mind their mitogenic, differentiative and anti-apoptotic properties of IGF's and their binding proteins, it is possible that these are involved in the growth of the endometrium as well as its differentiation, apoptosis and possibly angiogenesis. The combined roles of the IGFBP1 as a binding factor for IGF's and as a trophoblast integrin ligand suggests that it may have multiple roles in the development of the endometrium as well as in the interaction between the decidua and the invading trophoblast (63). A schematic representation of the potential role of the IGFI to regulate growth is presented in Appendix 4.

Insulin-like Growth Factor 1 Receptor (IGF1R)

IGF1R and insulin receptors are members of the super family of tyrosine kinase receptors. IGFIR, known as the type I IGF receptor, binds both IGFI, IGFII, although the former with greater affinity, and insulin. A map of the IGF1 cell proliferation mediated through IGF1-R is presented in Appendix 5. The two receptors mediate different but sometimes overlapping metabolic and growth-promoting pathways in vivo. Many cells though express both type of receptors and also form a hybrid receptor between IGF1R and insulin receptor (IR). The chimeras (TIGFR and TIR respectively) were expressed in fibroblasts and adipocytes and were efficiently activated without interference from the endogenous receptors(64). In fibroblasts both chimeras showed identical capacity of DNA synthesis stimulation and both also protected against induced apoptosis. However, TIR was more effective than TIGFR in stimulating glycogen synthesis with two-fold greater increase in stimulating glucose uptake of TIR in adipocytes (64).

The differences in the abilities of IGF1R and IR to mediate some (but not all) biological responses could be due to their capacity to activate the same intracellular signalling pathways. (65)

Insulin-like Growth Factor Binding Proteins (IGFBP's)

Intense research has been conducted in the past few years on the role of IGFBP's, much of which is summarised below. Although they are regarded as possible major factors in the major growth pathways, little has been found so far. The intracellular signalling pathways have been the targets for vigorous research into finding out which are the factors that interfere with the growth processes.

IGFBP's play an important role in binding the IGF's in circulation, delivering these to the target tissues and prolonging the half-life of the IGF's. At the tissue level IGFBP's are capable of modulating the IGF's biological response.

IGFBP-3 is its involved in the regulation of cell growth and death. Breast cancer cells, for example, can develop resistance to the inhibitory effects of the IGFBP3 leading to the loss of growth regulation (66). IGFBP3 is a highly regulated gene product that potentially affects the survival and growth of multiple tissues through a host of protein-protein interactions with several novel regulators (67).

After proteolytic cleavage of IGF1b into mature Igf1 and Eb, Eb or fragments of it may compete with receptor binding for IGFBP-3 and block the growth inhibitory effects of

IGFBP-3 via common highly basic motifs. (Appendix 7) Eb or Eb fragments perform their actions via its or their own cell surface receptor. (Appendix 6) IGF1 treatment in mice caused a significant increase in body length and weight as well as an increase in the weight of many organs. The IGFBP1treatment did not cause significant changes. Co-administration of IGFI and IGFBP1 resulted in the neutralisation of the IGF1 stimulatory effects on body growth. IGFBP1 inhibited the effects of IGF1 on growth (somatic and organ weight) but had stimulatory effects on kidneys and spleen (68). The IGF family is thought to have an important role in the development of the endometrium during the menstrual cycle, and in the process of implantation. IGF2 is expressed in the trophoblast whereas IGFBP1 is heavily expressed by the maternal deciduas. Trophoblast expressed IGFII, as well as interleukin-1 β, act to inhibit IGFBP1 production by the deciduas, a fact that suggests a paracrine signalling between the two compartments. The role of IGFBP1 as an IGF-binding protein as well as the role of trophoblast integrin ligand suggests that IGFBP1 may have multiple roles in endometrial development and in the interactions between the decidua and the invading trophoblast (63).

IGFBP1 was shown to be an inhibitor of IGF1 actions and was found elevated in foetal and maternal circulation in humans with intrauterine growth restriction (IUGR) complications during pregnancy. IGFBP1 production in foetal cells (hepatocytes) is regulated by hypoxia (69). IGFBP5 is an inhibitor of the IGF-mediated cell-survival process.(70) The role of the IGFBP gene family in vivo was analysed through individual and combinatorial knockouts of IGFBP's. Most individual homozygous mutants (IGFBP-2, 3,5, and-6) were found viable and fertile with no significant changes in postnatal growth and no morphological abnormalities (71).

Human Growth Hormone gene (GH1)

The actions of the growth hormone during foetal and postnatal growth have been and are major interest point in research in both humans and animals. Placental insufficiency (PI) is a major cause of foetal growth retardation. GH action on the correction or improvement of foetal growth retardation, after the onset of PI, was studied in sheep. Several organs appeared affected by PI but restored by GH treatment. The findings were that GH treatment increases the weight of some organs but reduces others, and restores levels of IGF1 in foetuses with PI. These findings show somewhat that the GH-IGF1 axis is functional in the uterus and that this somatotrophic axis reacts to various metabolic and hormonal stimuli (72). The GH/IGF1 axis

was found to exhibit a biphasic pattern during the course of severe illness. It is believed that inflammatory cytokines and nutritional deficiencies involved in these cases are partly responsible for the mediation of such changes. A hypothesis was formulated although it hasn't so far been proven, that the high levels of GH in the plasma exert direct lipolytic, insulin antagonising and immune stimulating actions, whereas its indirect somatotrophic effects are attenuated, action that may be interpreted as beneficial and adaptive to stress (73). The controversial matter of the individual effects of GH and IGF1 on substrate metabolism remains largely unsolved. The effects of GH and IGF1 on substrate metabolism are related to the feast-famine cycle (64). Administration of GH promotes lipolysis and lipid oxidation, processes associated with reduced oxidative glucose disposal and stimulation of protein synthesis in the body. Also GH administration dramatically increases resting energy expenditure (REE) which is not only secondary to increased lean body mass. However, prolonged exposure to GH administration is associated with induced insulin resistance and hyperinsulinemia that could partly contribute to the protein anabolic effects. Administration of IGF1 has also been associated with increased lipid oxidation, which most likely represents a concomitant suppression of insulin secretion by a supra-physiological increase in the levels of free IGF1. Administration of IGF1 alone also raises the REE but to a lesser extent than GH. The following physiological feast-famine hypothesis has been formulated: a temporal dissociation exists regarding the isolated and concerted effects of GH and IGF1 on substrate metabolism. Low levels of GH are characteristic of the postprandial state as well as by hyperinsulinemia and high IGF1 bioactivity, which favour nutrient disposal. Prolonged fasting as well as late post absorptive state results in low IGF1 and insulin with a progressive increase in the GH levels, all which favour lipid oxidation (74).

Signalling and intracellular events:

Many signalling pathways and substrate interactions have been studied in an attempt to identify the responsible factors for growth and development deficiencies and pathological conditions. Receptors represent a vital cellular link to the extracellular environment. Within a cell a complex set of signalling pathways communicate receptor responses to other receptors as well as to the nuclear and cytoplasmic compartments. Autocrine and paracrine systems are often used during development and by most growth factors. This comes in addition to the classic and well-known endocrine feedback loops, which serve to maintain hormone levels within a narrow range. In many cases the distinction between the endocrine and paracrine actions is hard to make, as was recently seen in the case of the hepatic vs. local production of

IGF1. For a long time it has been accepted that hepatic production of GH increases production of IGF1 in the circulation and this increase in IGF1 was responsible for the changes in growth. Indeed the liver was again shown to be the major source of circulating levels of IGF1 (9). The new somatostatin hypothesis states that autocrine action of IGF1 is responsible for most normal growth although the liver remains major source of circulating IGF1 in the body.

A disorder can be caused by either hormone excess or deficiency, or by an abnormality in tissue sensitivity to hormone action. Receptors on the surface of the cells provide an important link in the communication between the hormone and the extracellular ligands. At the cellular level a single GH molecule binds two GHR molecules causing these to dimerise. This dimerisation process is believed to be necessary for the signal transduction, which is associated with the tyrosine kinase JAK2. It has been suggested that only one type of GHR molecule possesses different cytoplasmic domains or phosphorylation sited in different tissues and mediates the diverse effects of the GH. When the different cytoplasmic domains are activated by JAK2 they can lead to distinct phosphorylation pathways, one for growth effects and one for various metabolic effects (75).

Serum IGF1 and GH levels:

Serum concentrations of IGF1 are low in the foetal and early post-natal periods compared to adult levels, and the liver is believed to be the source of endocrine levels of it. Serum levels of IGF1 are found reduced by 75% in the liver-specific IGF1 KO animals. The low, but detectable levels of IGF1, were probably due to extra-hepatic tissue contribution to circulating IGF1 levels. However, in the IGF1-deleted animals, IGF1 levels were also reduced (50%) at the early stages of peri-pubertal growth spurt (9).

Neuroendocrine regulation of GH is achieved by two factors: a stimulatory one – the hypothalamic GHRH, and an inhibitory one, serum IGF1. Therefore, high levels of GH should also reflect low levels of IGF1 in the conditional KO mice. Indeed the levels of GH are high in liver specific KO mice. IGF1 gene expression in non-hepatic tissues is also GH dependent. However, previous studies have shown that to demonstrate this GH effect hypophysectomised animals with reduced IGF1 gene expression are required. Postnatally, IGF1 mRNA levels are similar in the gene-deleted animals to the ones in the wild type (WT) littermates; both sexes were fertile and give rise to litter of normal size. Thus in comparison to the generalised IGF1 deficient mice, where severe retardation in growth rate and organ weights is present, the effect of liver specific deletion of the IGF1 gene has little or no effect.

Liver IGF1 mRNA increases between 10-100 fold between early post-natal days and adulthood and is associated with the large postnatal increases in circulating IGF1 levels (9). It was also suggested that IGF1 production by extra-hepatic tissues play a role in growth and development during foetal, neonatal and pubertal stages. IGF's have specialised functions in various tissues including the reproductive, cardiovascular and neurological tissues, which not only produce IGF1 but also express its receptor. Also, recent studies have shown that the adipose tissue expresses high levels of IGF1 mRNA in a GH dependent manner, levels as high as the ones expressed in the liver. Bone cells synthesise and secrete IGF1, which is regulated by both paracrine and endocrine factors. The presence of IGF1 receptors on osteoblasts is strong evidence for an autocrine/paracrine role for skeletal IGF1 production. Other tissues with high relative mass contribute to the circulating IGF1 levels. An autocrine/paracrine IGF1 action is believed to be responsible for normal growth. It is therefore necessary to determine which tissues are most important for post-natal growth and development, and which tissues are contributors to the circulating IGF1. Also the role of IGF1 during this stage of development needs to be elucidated (10).

1.4.1 Gene product and function

Many point mutations and other code alterations in the growth hormone receptor gene (GHR) that cause amino acid changes and slightly disable the receptor are known to day. Most of those mutations and amino acid alterations have been reported in patients with Laron syndrome. Dwarfs with normal body proportions (ateleiotic) were first recognised in 1904 by Hastings Gilford. He also identified the sexual and asexual types of dwarfism. The two types correspond to the pituitary dwarfism type III, and I respectively. Type I has an isolated GH deficiency in recessively inherited sexual ateleiosis (76). A C to T substitution that generates a serine instead of a phenylalanine at position 96 of the extra cellular coding domain of the protein was reported in just one patient with Laron syndrome (36).

Rare severe effects

Table 1 summarises the severe effects of some mutations (and KO experimental data) in the genes discussed so far.

Table 1 The rare severe effects

Gene	Polymorphism	Association	Reference
GH1	P1 (A/T base 1663) P2 (T/G base 218) P3(G/T base 439) Arg-Cys codon 77	Polymorphism associated with GH secretion. Possible explanation of some of the variations in GH secretion and height. Extremely short stature and low levels of IGF1 IGFBP-3 and GHBP	(77) (78)
IGF1	(CA)n 192/192 192/194	192/192 Associated with low levels of circulating IGF 1 in normal as well as IOM men192/194 25% increase in levels of IGF1	(79)
IGF1 R	Classic knock-out (KO) Conditional knock-out (homologous recombination using Neo cassette and LoxP sites)	Lethal at birth due to respiratory failure Neo cassette alone caused a 60% reduction in cell surface IGF1R levels (in Hz animals) but with no changes in the birth weight compared to wt littermates. Growth deficient (Ho and Hz) during postnatal life.	(64)
IGFBP's	Combinatorial KO IGFBP-4 homozygous mutants	Small at birth Most IGFBP's have essentially no effect.	(64)

GHR

A GH receptor mutation has been identified as responsible for extreme short stature and GH insensitivity in a Bahamian genetic isolate. Individuals had normal to high GH serum levels and low IGF1 levels (characteristics found also in Laron syndrome). Also circulating levels of GH-binding protein activity were below levels of detection. SSCP analysis of exons 2-7 showed an abnormality in exon 7, a C to T transition in the third position of codon 236. Also the same mutation was found to cause a new splice donor site 63 bp 5' to the normal exon 7 splice site. This novel site was used to the exclusion of the normal site in homozygotes, whereas heterozygotes showed both normal and variant mRNA. The mutant protein lacks 21 amino acids including those defining the WS-like motif of the GH receptor extracellular domain (80).

IGF1 (CA)n repeat:

IGF1 is an important and ubiquitous polypeptide important in cell growth and differentiation. It stimulates linear growth and its plasma concentration is directly related to the pulsatile release of GH. Together with the insulin-like binding proteins, IGF1 is synthesised in the osteoblasts and GH, as well as several other endocrine and autocrine factors regulate it. It is important in bone cell differentiation, proliferation and collagen synthesis, as well as in bone resorption where it is released from the bone matrix where it recruits new osteoblasts thus aids remodelling of the bone. IGF1 is also vital for the growth and maintenance of other tissues. It is present in the circulation of post pubertal individuals, although the exact function is not yet known. Levels of IGF1 in tissues have been shown to reflect circulation levels (79). A cohort study of a highly polymorphic marker, used in the methods development part of our study, composed of variable cytosine-adenine (CA) repeats 1kb upstream from the transcription start site of IGF1, has shown that the most frequent allele (frequency reported previously by Webber and May, 1989, was the 192bp allele. Also it was shown that plasma levels for IGF1 were correlated with the homozygous (Ho) 192 and significantly lower in Ho individuals for this allele. Also 65% of the men included in the study with idiopathic osteoporosis **(IOM)**, who show low serum levels of IGF1, were reported to be homozygous for the 192 allele. It was also determined that in healthy Caucasian men and women the homozygous 192 genotype was associated with low serum levels of IGF1, whereas individuals with 192/194 (194 being another allele for the CA n polymorphism) showed a 25% increase in their plasma levels of IGF1. This is a good example in which individual

variation of IGF1 plasma levels is influenced by genetic factors and may be specifically correlated by variation at the IGF1 structural locus (79).

GH 1 polymorphism and GH secretion:

Hasegawa *et al.* reported three polymorphic sites in the GH1 gene and showed that one of these sites (P1) was associated with GH secretion and the P2 and P3 sites were found in close linkage with P1. This novel polymorphism, P1 is found in the intron 4 of the GH1 gene. The A allele of this polymorphic site was found in significantly different frequency in the subject with growth hormone insufficiency (GHI) compared to both normal adults (NA) and normal short (NS) children. AA genotype at the P1 locus was found in majority of GHI subjects. Maximal GH peaks (from GH provocative tests), the IGF1 levels as well as height were found significantly different between the AA and AT genotypes. Higher IGF1 levels as well as heights were found in the NA group for the AT genotype compared to the AA. However, the small size of the studied group was probably the reason for the lack of statistical significance in the NA group. Significant differences were found in maximal GH peaks in provocative tests, insulin-like growth factor, and height in children with the TT or GG genotypes at the P2 locus. The frequency of T allele at this locus was also found to be significantly different between pubertal short children without gross pituitary abnormalities and short children with normal GH secretion or normal adults. (77)

Height is determined by genetic as well as environmental factors. The effects of GH secretion and the height of individuals are well known. Deletions in the GH 1 gene were found in extremely short patients (33) and (29). Height during development in childhood was also shown to correlate with GH secretion. This was determined by measurements of GH levels integrated over 24 hours or serum-dependent parameters such as insulin-like growth factor I and IGF binding protein 3 (81) and (82).

Takahashi *et al.* (1996) reported a rare single mutation in the GH1 gene, which caused severe short stature in a male child (83).

This miss sense mutation was thought to convert codon 77 from Arg to Cys. Serum IGF1, IGF-BP3, and GHBP were all low or at the lower limit of the normal range. The mutant R77C growth hormone showed 6 times greater affinity to GHBP than wtGH, and showed strong antagonistic properties (inhibited tyrosine phosphorylation ten times more potently than wild type GH). Also the child did not show any increase in serum IGF1 levels in response to exogenous hGH administration.

IGF1R:

The IGF system is an important regulator of embryonic and postnatal somatic growth in vertebrates. Animal experiments where the IGF1R was completely nullified resulted in deaths at birth due to respiratory failure (64). Controlled partial invalidation of the IGF1R gene by homologous recombination using the neomycin (Neo) cassette and the Cre-LoxP system (84) were used to determine the role of the IGF1R in embryonic and postnatal development. The presence of the Neo cassette alone caused a reduction in the cell surface IGF1R levels by an astonishing 60%. However, the animals were not growth retarded at birth showing the same weight as their wild type littermates. Severe growth retardation was observed after 3 to 4 weeks of postnatal life regardless of sex. IGFI and IGFBP-2 and -3 levels in serum were found unchanged in females whereas in males IGFI and IGFBP-4 levels were significantly increased indicating an effective regulation of the IGF system in response to the partial invalidation of the type I receptor gene (11). Two RNA splicing events were observed, which could provide the explanation for the decreased levels of IGF-1R. Crosses between the IGF-1R lox mice to a Cre transgene (produces ubiquitous partial excision of floxed regions) a range between 6 and 82% invalidations were obtained (11). Although correlations between gene dosage and body weight in the young adult mice were found to be significant for both sexes, a stronger regression was observed in males than females, such that the sex-related differences in body weight essentially disappeared at 80% invalidation of the IGF1R gene. It is therefore believed that mouse vital functions can be maintained at 20% or less of intact receptor gene (64).

We have chosen two genes as part of our investigation into the association between genetic changes and phenotypic factors and their alliance with the foetal origins of adult disease. These genes are the human growth hormone (GH1) and human IGF1 genes. The reasons for choosing these two particular genes out of the whole variety of genes involved in major growth pathways (described in this chapter) are summarised below.

There is increasing evidence that changes in the foetal environment may have long term effects on metabolic and endocrine pathophysiology in adult life. This process has been termed 'programming'. It has been shown that undernutrition in the mother during the gestational period leads to programming of hyperphagia, obesity, hypertension, hyperinsulinemia, and hyperleptinemia in the off-spring (85). It is also true that intrauterine growth is influenced by genetic, environmental nutritional and endocrine (growth hormone, insulin-like growth factors, placental lactogen and/or insulin) factors, all of which are thought to be important growth regulators of foetal life. IGF1 and 2 are known to be important

growth regulators. In foetal serum IGF1 levels are lower than the adult levels and also correlate to gestational age. On the other hand it has been shown that IGF1is highly dependent on GH control (KO mouse experiments described earlier) although the IGF1 levels during pregnancy rise progressively in both normal and GH deficient women.

Several reports of positive correlations between IGF1 and birthweight exist, and the release of primary foetal hormones, involved in the regulation of growth, foetal insulin and foetal IGF1, is influenced by the delivery of nutrients (62).

1.5 Plan of investigation:

1.5.1 General Hypothesis:

More and more clinical phenotypes are found to be influenced by major genetic components. The hypothesis is that the risk of major complex diseases (cardiovascular disease, obesity, diabetes, autoimmune disease, osteoporosis, etc.) is underlined by major genetic components, such as for example, the pattern of single nucleotide polymorphisms found in key susceptibility genes. This accentuates the need for developing new, fast and efficient methods for the identification of such susceptibility loci, methods that could permit the rapid and economical scanning of a whole population.

In order to test the hypothesis I have firstly set out to develop the necessary laboratory methods and advances so that key susceptibility genes involved in complex diseases will be genotyped for various polymorphic markers. As many markers in various susceptibility genes are to be tested in our study, the need to develop conservative methods of implementation of DNA banks has arisen. The variants in genes chosen for this study were genotyped using conservative methods (heteroduplex detection method, extreme dilution, and long PCR). The data so collected were analysed for association with obesity, cardiovascular phenotypes, birthweight and weight at one year (as part of the FOAD hypothesis), and bone phenotypes.

1.5.2 Potential new developments:

Recent technological and computational developments including the sequencing of the Human Genome, open several opportunities for the field of genetic research. The possibility of a genome wide scan for gene effects using SNPs, or even polymorphisms which include small insertion/deletions, multinucleotide changes as well as single nucleotide changes, has drawn attention towards the efficiency of different approaches used for mutation detection.

New approaches in the laboratory have been developed for gathering data upon which to examine the hypothesis.

One of the cohorts used for my study, the Hertfordshire cohort, consists of 1032 subjects representing a random population sample born in East and North Hertfordshire sampled during the 1990's and with birth date range 1920-1930. The sample has 583 males (mean age 66.4 years) and 449 females (mean age 66.4 years). This cohort was used for the genotyping of several SNPs in the GH1 and IGF1 genes.

Another cohort used was the Northwick Park Heart II Study (86), consisting of 2500 middle-aged (45-65 years) Caucasian males, in which the analysis between the insertion/deletion in the 3' UTR region of the leptin receptor gene (OB-RB) has been performed using a new method of detection of such polymorphisms.

In genetic epidemiological studies one limiting parameter is the availability of large DNA banks. There is a need also for finding ways of preserving such banks.

Heteroduplex analysis, as a mean of scanning insertion/deletion polymorphisms, is a simple way of detecting such type of polymorphisms, where the two alleles can be differentiated through various methods of hybridisation. I/D are easy to scan and have superior discrimination by hybridisation compared to SNPs. In the human genome it's estimated that for every 20 SNPs there is one insertion/deletion, thus I/D may be an efficient way of scanning the genome for associations with disease.

I investigated rapid, efficient, and cost effective ways of searching for insertion/deletion polymorphism. The method developed was applied to the leptin receptor gene, a gene under much attention for its potential involvement in obesity (an increasing problem especially in the Western World).

In the search for developing conservative methods of implementation for DNA banks I have performed template dilutions and determined the extremes to which this can

be carried out, have attempted to use whole genome amplified template for microsatellite polymorphism in the GH and IGF1 genes, and also I have succeeded in using long amplicons of specific genes for SNP detection.

I have investigated highly polymorphic regions in the GH1 and IGF1 genes, regions that contained several SNPs in a physically small region of the genes. The SNPs in the GH1 and IGF1 genes are located in non-coding regions. Extensive analysis of the results has been performed.

2 Standard Materials and Methods

Molecular-genetic epidemiology studies are based on the analysis of large number of samples. Scanning of several thousand of samples should be accomplished in a relatively short period of time, be economical for both resources and work force, and should be accessible to small as well as large laboratories.

Various polymorphisms have been tested in key susceptibility genes implicated in diseases with late life onset. The studied polymorphisms were tested in the general population using an extended number of samples from the NPHS II and the North and East Hertfordshire studies.

In this chapter I shall present only the methods and reagents that were adopted from previously published material. Chapter 3, Methods Development, will describe in full all the methods used that were developed specifically for the investigations this project has undertaken.

2.1 Template amplification

Amplification of the template in all cases was achieved using polymerase chain reaction (PCR). The specific cycling conditions and reaction reagents used amplification of each marker are presented in chapter 3.

2.2 RsaI Restriction digest:

Restriction digest was initially used for the I/D polymorphism in the 5' UTR of the OB-Rb gene for the initial identification of the genotypes in our control study.

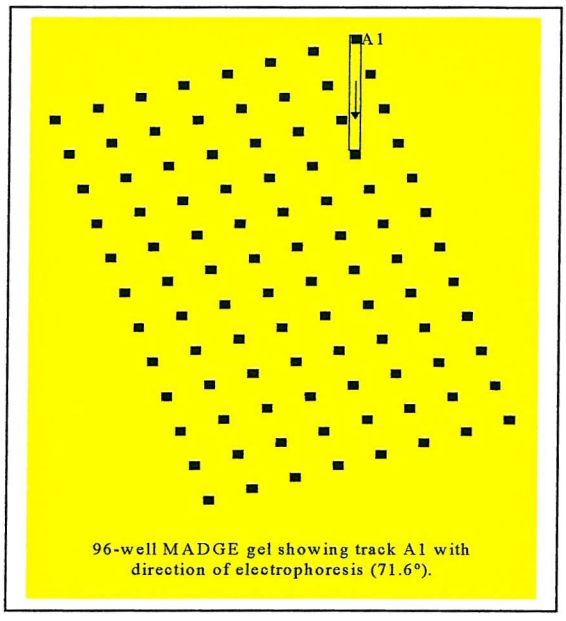
RsaI digest was performed according to the manufacturer's instructions. For each sample a 3μl digest was prepared. The digest contained 0.2μl RsaI enzyme, 0.1μl BSA (Acetylated BSA to a final concentration of 0.1mg/μl) 1μl buffer [10mM Tris-HCl (pH7.9), 50mM NaCl, 10mM $MgCl_2$, and 1mM DTT at 37°C] and 1.7μl H_2O. The digest was mixed and incubated at 37°C for 2 hours.

2.3 *Electrophoresis:*

Electrophoresis was performed on **M**icrotitre **A**rray **D**iagonal **G**el **E**lectrophoresis (MADGE) gels for the OB-R insertion/deletion analysis, as well as for the SNP analysis. Urea gels were used for the GH A1819 and IGF1 (CA)n marker analysis, as well as for the sizing of long amplicons used as templates in the SNP analysis.

MADGE system was used for all the cohort studies that had a good enough resolution on the short polyacrylamide tracks. Agarose gels do not confer high resolution for small fragments of DNA, whereas acrylamide gels offer best resolution. The MADGE system is described below emphasising the advantages and disadvantages of this type of electrophoresis system in comparison with other types of gels that have been utilised. Agarose gels have the advantage of being easily prepared in an open face format, feature that makes it convenient for horizontal gels. Acrylamide has the handicap that it does not polymerise in the presence of air. MADGE is a simple device and method to prepare and manipulate horizontal acrylamide gels.(87) The open faced horizontal matrix enables loading of arrays of wells. The device preserves the exact configuration of the industrial standard 96-well microtitre plates. (Dynatech) MADGE has its principal advantages where multiple analyses are needed (e.g. population studies, DNA banks, clone libraries, etc). The 8 x 12 array with 9 mm pitch is set on a 71.6° relative to the edge of the gel, diagonal which is parallel to the eventual direction of the electrophoresis. The array is directly compatible with 96-well plates. The former, placed horizontally, is filled with the acrylamide mix and the Sticky Silane coated plate is placed on top. After polymerisation has taken place, the glass is prized off to reveal a 96-well open face gel. The run track of the 96 well gel is of 26.5 mm, which is efficient for many post-PCR analysis.

Fig 2 Schematic representation of the MADGE system.



While heteroduplexes are generally not resolved in agarose matrices, they resolve well in acrylamide-derived matrices. Previous developments of MADGE have been mainly for single nucleotide polymorphism (SNPs) analysis (by RFLP, PIRA and ARMS) and de novo mutation scanning (Melt-MADGE). Higher resolution applications (minisatellites and tetra nucleotide microsatellite typing) have been developed (75). The convenient access to PAGE matrices in a high throughput format invites systematisation for ins/del scanning and typing. The MADGE former uses a two dimensional plastic former and a glass plate coated with gamma-methacryloxypropyltrimethoxysilane (Sticky Silane). Sticky Silane contained: 0.5% glacial acetic acid, 0.5% methacrylaxypropyltrimethoxy silane, and ethanol to volume required (usually 500 ml).

Various buffers were used for electrophoresis: TBE (10X), TAE (10X), and TE (10X). Their composition is given below.

1000 ml TBE 10X buffer contained: 108g Tris base, 55g orthoboric acid, 7.44g EDTA (or 9.3g Na EDTA) and water to volume.

1000 ml TAE 10X buffer contained: 48.4g Tris base, 11.42g Glacial acetic acid, 20ml (0.5M) pH8.0 EDTA, and water to volume.

1000 ml TE buffer 10X: 48.4g Tris base, 3.8g EDTA-$Na_2 2H_2O$(10mM) and water to volume.

Insertion/deletion polymorphism in the 3'UTR of the Leptin receptor gene

Electrophoresis:

Gel electrophoresis was performed on 7.5% acrylamide MADGE gels. A 50 ml gel mix contained: 12.5 ml acrylamide, 5 ml TBE buffer (10X) and 32.5 ml H_2O. Due to electrophoresis artefacts TAE buffer that was observed to give better resolution of the bands in our case replaced the buffer. (See elimination of 'dumbell' bands by delayed start electrophoresis, section 3.1.1) A 30 to 40 minute run at 150V (gel submerged into 1 X buffer) was found sufficient for the heteroduplexes to be resolved.

Staining was initially achieved using ethidium bromide, but where low PCR product yield was observed a more sensitive stain such as Vistra Green replaced this.

Visualisation of product was achieved by staining and imaging on Fluorimager 595 using Image Quant 5.0.

Di and tetra-nucleotide polymorphism in the growth hormone gene region (GH) and IGF1 dinucleotide repeat:

For the A1819 human GH polymorphism a combination of urea and acrylamide vertical slab gels were used. Gel electrophoresis for the A1819 marker was run on vertical slab denaturing gels that were 4 mm thick. The gel mix was poured between two vertical, unequal size plates, using spacers of 4 mm thickness (The gel company), and for the upper edge of the gels a comb of same thickness was used to provide the straight upper edge (Web Scientific). The combs were also used for the loading of the sample onto the gel. The teeth of the gel comb were spaced at 3 mm apart.

6% denaturing polyacrylamide gel were used for all samples. A 50 ml gel of such type contained: 10 ml acrylamide (30%/bis 19:1), 5 ml TBE (10X), 21.25g Urea (7M) and H_2O for up to 50 ml mix. The mix was stirred at 37°C (incubator) for 20 minutes to allow Urea to dissolve. To polymerise a mix of 50µl ammonium persulphate (APS 20%) and 50µl N,N,N,N-tetramethyl-ethylenediamine (TEMED) was used. The gels were allowed to polymerise for a minimum of 2 hours (but no longer than 16 hours).

The samples were stored at –20°C overnight prior to electrophoresis.

A loading solution containing 95% formamide, EDTA (20µM) and a combination of bromophenol blue and xylene cyanole were used to load the samples onto the gel.

5µl PCR product was mixed with an equal volume of the loading mix, and then denatured at 90°C for 3 minutes. The samples were then instantly chilled on ice for a minimum of 1 minute before loading onto the gel. A volume of 5µl of the PCR product mixed with the loading dye was loaded onto the gel for each sample. Electrophoresis was performed on long gels using TAE buffer.

The gel was pre-heated for approximately 1 hour (to a temperature of approximately 60° C before loading on the samples, and then samples were run for 2h and 45 minutes on average at a constant 30 Watt.

After electrophoresis the gels were fixed onto Whatman paper and dried at 80°C for 2 hours on a vacuum gel dryer. Autoradiography films were placed in photographic cassettes at -70°C. Autoradiography was performed for a minimum of 6 hours (or longer depending on the degradation of the radioactive material).

For better resolution and further separation of the bands electrophoresis was also performed at 60 Watt or a 30-Watt current was applied to the 6% DPG gels for a total of four hours. More detail of the electrophoresis changes is given in Methods development, chapter 3.

Handling the large and thin (0.4 mm) gels proved difficult at times with the gel ripping post-electrophoresis when separating the two plates. The protocol for handling such gels was adapted from Laywood *et al.* (88) Both plates were cleaned using detergent and water. Plates were dried and cleaned thoroughly with acetone (fume cupboard) then with ethanol. A solution of gamma-methacryloxypropyltrimethoxysilane, glacial acid and ethanol mixed by volume in the ratios, 1:10000:2000000 respectively, was made and used for the large plate (19.5 cm/36.5 cm, SA=711.75 cm^2). A total of 0.0042 ml/cm (i.e. 3 ml solution/plate) was used. The small plate (19.5 cm/33.5 cm, SA=653.25 cm^2) was treated using 0.00083-ml/ cm^2, which amounted to 0.5 ml/plate of Repelcote. The Repelcote and mixed solution of SS were spread evenly over the whole surface of the respective plate, allowed to dry in the fume cupboard for 2 minutes, then wiped using a large tissue soaked in deionised water, then dried using clean tissues. When dried the gels were poured and allowed to polymerise.

Electrophoresis of long-PCR amplified templates took was performed on 0.7% agarose gels. 100ml gel contained: 7g urea, and100ml TBE buffer (1X). The gel mix was heated for 3 minutes at high power in a 600Watt microwave, the lost volume of fluid was re-constituted using H2O, mix was allowed to cool to 60-65°C then poured and allowed to polymerise for 2 hours.

2.4 *ARMS analysis of point mutations*

The amplification refractory mutation system is a rapid, simple, and reliable method of detecting small deletions or point mutations in a DNA sample. The method is ideally suited to the detection of SNPs and variations exist for the detection of multiple mutations (Multiplex ARMS), and the phase of closely linked mutations (double ARMS). ARMS systems are used especially where high quality and accuracy is required as for example in clinical diagnosis of inherited disease. This technique has been used extensively for the analysis of polymorphisms in the HLA genes.

Taq DNA polymerase as well as other polymerases requires a primer to anneal to a complementary sequence to prime DNA synthesis. For Taq, for example, the primer has to be perfectly matched at the 3' end to allow efficient polymerisation to occur. The differences in the rate of extension between matched and mismatched templates are large, ranging between 10^3>10^6. (89) Using extra destabilising mismatches at the –2 position of the 3' end may increase this difference even further. ARMS uses this discrimination as a molecular switch that is coupled to a PCR reaction to allow specific amplification and detection. Development of ARMS therefore relies simply on determining a set of conditions whereby the yield product from the target allele is detectable whereas the yield product from the non-target allele is not. The reaction conditions must cope with the variables of the system that include sample DNA concentration and purity, melting temperature, and enzyme amount. Long primers (>28 bases) are usually chosen to minimise the difference in stability of primers annealed to the target and non-target alleles. This makes sure that the differences in extension rate rather than hybridisation rate are the ones that are generating the required specificity.

The critical factors in an ARMS assay are: the enzyme amount (high enzyme amounts tend to decrease dramatically specificity of reaction), primer quality (crude oligonucleotides preparations are generally suitable for use as ARMS primers, the reaction buffers (ensure high quality chemicals are used and aliquots are stored frozen), the template concentration (DNA amount- large amounts may lead to mispriming), DNA quality, and the thermal cyclers used for the amplification (use same model of cycler as temperature windows for ARMS are typically +/- 2° C).

3 Methods Development

In this chapter I shall cover the specific reactions, reagents and conditions used for scanning the polymorphic markers described in chapter 2. These reactions have been developed for the scanning of the various markers that the project has investigated.

3.1 Analysis of short insertion/deletion polymorphism in cohorts using MADGE:

Application to the human Leptin receptor (OB-Rb) gene:

The analysis of human short insertion deletion polymorphism in the 5'UTR region of the leptin receptor gene was performed in a cohort study using microplate array gel electrophoresis (MADGE). The MADGE system has been described in chapter 2. One of the methods developed is a combination of use of heteroduplex generator, a modified MADGE format (192-well) and a cohort 'bootstrap' which will be applicable to many ins/del polymorphisms.

Amplification of the DNA for the Insertion/deletion polymorphism in the 3'UTR of the Leptin receptor gene (OB-RB)

The amplification of the pentanucleotide insertion (CTTTA), in the 3'UTR region of the leptin receptor gene, was achieved using a 95°C, 50°C and 72°C (1 minute each), 35 cycle PCR. 5µl DNA template (of variable concentration) were used in each well of the OMNIPLATE 96 array (Hybaid Ltd.), which is fully compatible with the MADGE system used. 10-ng/µl genomic DNA template was used per amplification. The DNA templates were dried at 80° C for 15 minutes prior to PCR.

The primers used for the amplification of the 5bp insertion/deletion fragment in the 3'UTR region were the forward 5'-ATA ATG GGT AAT ATA AAG TGT AAT AGA GTA-3' and the reverse 5'-AGA GAA CAA ACA GAC AAC ATT-3' as described originally by Oksanen *et al.* (90)

The PCR reaction mix contained: 50mM KCl, 10mM Tris pH 8.3, 0.001% (wt/vol) gelatin, 0.2mM of each PCR primer, 1mM MgCl2, Taq DNA polymerase (0.2 U/10µl). TAE buffer (10x) contained: Tris-base 0.399M, sodium acetate (0.8M) $3H_2O$ (0.8M), EDTA-Na_2H_2O (40mM) and acetic acid to pH 7.2.

'Heteroduplex Generator' amplification:

After the amplification of the individual templates the 'generator' had to be identified and produced for the hybridisation step. The most common allele was determined to be the deletion (90) and therefore the homozygous genotype for this allele (DD) was chosen as the 'heteroduplex generator' in my study. Initial amplification of the 'generator' genotype (DD) was achieved using the PCR programme described above, using 5μl DNA template for each 20μl PCR. A 10-minute 95°C-denaturing step was added at the end of the original PCR programme. Instant cooling on ice of the PCR product followed this denaturing step. Re-naturation of the product was achieved at room temperature. The 'generator' was stored over time at -20°C. To ensure a continuous supply of generator, a 2nd round of PCR was performed using a 1:100 dilution of the 'generator' PCR product as a template. This second round PCR product was used for genotyping.

During the electrophoretic run gross distortion of the bands was frequently observed. As this method of genotyping was designed to allow effortless and fast genotyping of ins/del in cohort studies, this problem had to be resolved.

The method developed and subsequently used to avoid the distortion of the dumbells that presented a real impediment in genotyping heteroduplexes, is described in section 3.1.1.

Detection of the rare insertion/insertion (II) genotype.

Initial detection of the 5bp-insertion/deletion polymorphism was achieved using RsaI restriction digest of the PCR product. (90) This facilitated the identification of the genotypes of the control samples, genotypes used later for the verification of our new method. Following the initial identification of the heteroduplexes, formed in the case of the heterozygous samples, the detection of the II genotype was achieved using the DD 'generator'. 5μl of the sample PCR product were removed and mixed with an equal volume of the 'generator' PCR product.

The 10μl volume was covered with Paraffin oil to avoid evaporation during denaturing step. The mix was denatured at 95° C for 5 minutes followed by the instant cooling of the samples on ice. Re-annealing of the heteroduplexes was achieved at room temperature for 10 minutes. The time allowed for the re-annealing was tested as the stability of the heteroduplexes in time was questioned. A time-annealing test was performed for this purpose. The stability of the 'generated' heteroduplexes was tested at room temperature. The newly formed heteroduplexes

would therefore represent the insertion/insertion (II) genotype. As this genotype frequency is very small the number of new 'generated' heteroduplexes was minimal and genotyping was therefore achieved fast. The identification of the insertion/insertion genotype was achieved with the aid of the generator and the molecular hybridisation idea.

3.1.1 Elimination of 'dumbell' bands by delayed start electrophoresis

Delayed start electrophoresis was used to avoid the distortion of the bands on the MADGE gels. My development of this idea has since been published. (1) The idea of examining a delayed start electrophoresis came from the observation that when using an 8-channel pipette for loading MADGE gels, the distortion of the bands was more severe in the last set of loaded samples with little to no distortion in the first set. The samples at the top of the gel had a five minute delay in the start of electrophoresis, time that must allow considerable ionic diffusion through the walls of the wells (noticeable from marker dye diffusion), thus reducing the sample to gel ionic imbalance. I assumed this to be the factor responsible for the 'dumbell' effects (Fig.3). The solutions considered for this type of problem were the time course for sample ions to gel diffusional exchange, changing the constitution of the gel, and changing the sample composition before or after PCR. Of all the possibilities tested, one provided a quick and efficient solution to the problem. A five to ten minute delay in the start of electrophoresis was sufficient to allow optimal resolution of the bands without any loss of product (Fig. 4). Ionic composition of the loaded sample was also examined. As conditions for each PCR are optimised for PCR yield, changing sample composition would preferably take place post-PCR rather than before it. Before loading onto the gel the sample was mixed with a formamide dye. (1)

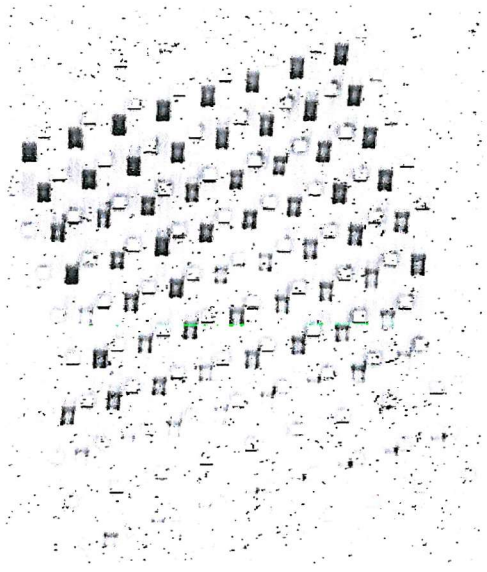
Several ratio combinations of the dye mixture with 10x TAE (where TAE gels were used) were tested (no TAE, 0.9xTAE, 1.8xTAE,and 3.6xTAE final concentration in samples loaded) (Fig. 4) TAE supplementation within a broad range, 0.9-3.6 TAE, was relevant. A combination of PCR product, dye mixture and 10xTAE in the ratio 5/1.5/1.5 (vol/vol/vol) was successful.5 µl of this mix were loaded in each of the MADGE wells (2 x 2 x 2 mm wells with a maximum volume of 8µl). The loaded samples were ionically reasonably matched to the gel so that equilibration time would take a short time. The 5µl sample undergoes dialysis through four 2mm square walls of the cubic well (and to the overlying buffer, if set up submerged rather than with wicks or direct electrode contact to gel in a dry

chamber) before the start of the electrophoresis. Small-distributed cubical wells such as those in MADGE arrays will benefit most from this approach.

Fig 3 shows the artefacts that can occur with MADGE electrophoresis.

Bands appear distorted into 'dumbell' shapes with 'skiing' on either side of the track.

Fig 3 Artefacts found with MADGE system



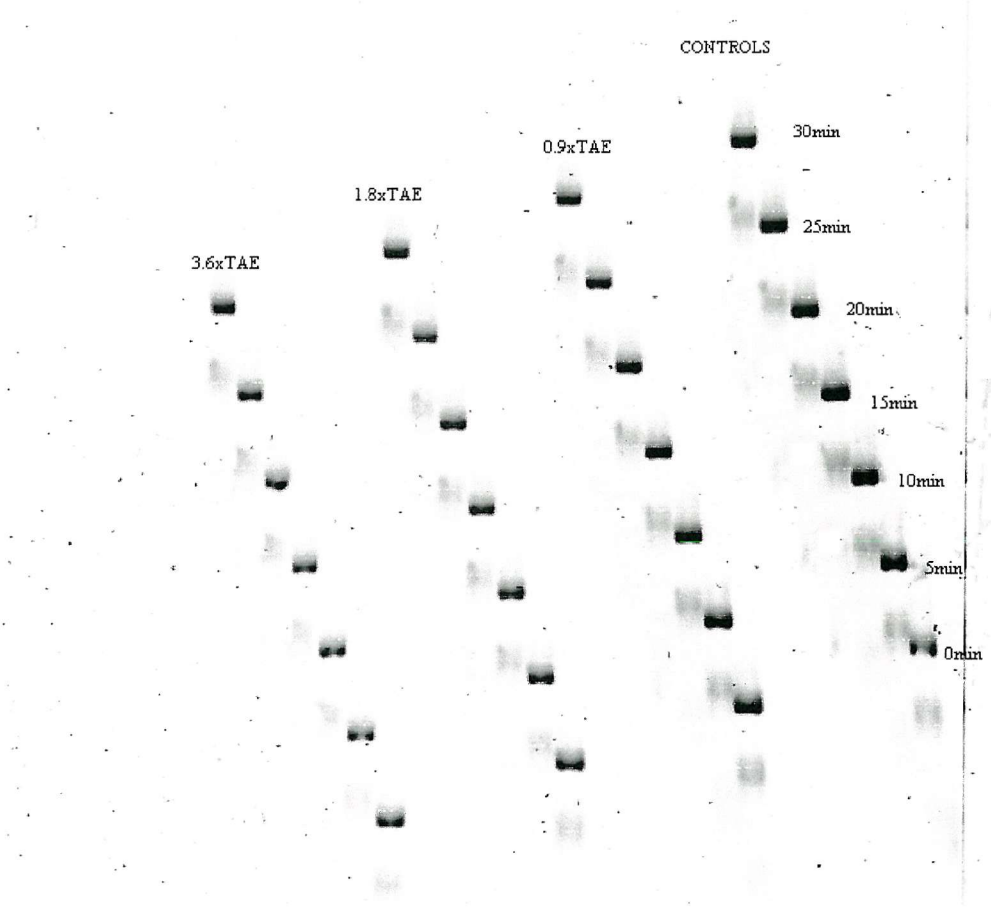


Fig 4: TAE titration and delayed start electrophoresis (delay time shown to right) time course. Duplicate wells were loaded for each time point for each TAE concentration added (see text). The right hand time increments refer to the bands on the gel.

3.2 *Implementation of conservative methods in DNA banks*

As DNA banks are finite and lymphocyte immortalisation is unlikely to be feasible, the development of systems that are highly conservative of DNA templates and cost effective for high throughput analyses was required. These systems would enable analysis of the genetic information that potentially can facilitate extensive association studies.

The developments of such systems have been performed on two separate polymorphisms in two separate genes: human GH1 (the A1819 di- and tetranucleotide polymorphism) and IGF1 (the CAn-dinucleotide repeat).

3.2.1 *Whole genome amplification using a degenerate oligonucleotide primer.*

Individual samples were amplified in the laboratory (L. Hinks) using the degenerate oligonucleotide primer technique (DOP). DOP is a type of PCR that employs oligonucleotides of partially degenerate sequences. It is a rapid and efficient technique for general DNA amplification.(91)

L. Hinks developed the method used in the laboratory for the whole genome amplification. For 50 µl reaction mix 5 µl POL MIX (10X POL MIX contains 0.5M KCl, 100 mM Tris pH8.3, 0.01% gelatin, and 2 mM dNTP's), 2.5 µl $MgCl_2$ (50 mM) to a final concentration of 2.5 mM, 0.5 µl DOP primer (200 pmol/µl) to final concentration of 2 µM, 0.16 µl TAQ DNA polymerase 5 units/µl (0.8 units), 20 ng genomic DNA and water to a final volume of 50 µl were used. The reaction mix was covered with 20 µl mineral oil before PCR.

The DOP primer used was: 5' CCG ACT CGA G NNNNNN ATG TGG 3', and the amplification was achieved using 94°C for 5 minutes (step 1), 94°C for 1min (step 2), 30°C for 1 min (step 3), 72°C for 3min (step 4), repeat cycles from step 2 for 7 more times, then 94°C for 1 (step 6), 60°C for 1min (step 7), 72°C for 3min (step8), repeat cycles from step 6 onwards for an additional 27 cycles.

3.2.2 Di and tetra-nucleotide polymorphism in the growth hormone gene region

Whole genome amplified samples (kindly provided by L Hinks) were used in the initial steps of the development of conservative methods of DNA implementation.

The amplification of the AAAG and AG repeats lying between the Alu 18 and Alu 19 of the human GH region (Genbank accession J03071) was amplified using the primers designed originally by Jeunemaitre *et al.* (92)

The initial amplification of the region was performed on 35ng/µl genomic DNA template using polymerase chain reaction. Genomic DNA templates were dried prior to the amplification. Each DNA template was amplified using 10µl of PCR reagent mix. This reagent mix (10µl) contained: 1 µl PCR buffer (10x), 1 µl W1 (1%), 0.5 µl dNTP mix (dATP 4 mM, dGTP 4 mM, dTTP 4 mM, and dCTP 2 mM), $MgCl_2$ (50 mM) 0.3 µl, 0.1 µl of each oligonucleotide (100pM/µl), 1 µl Taq DNA polymerase (GIBCO) [5U/µl], 0.1 µl alpha ^{32}P-dCTP (10uCi/µl), and H_2O up to 10µl.

The cycling conditions were: 95°C for 3min, 95° C for 30sec, 63°C for 30sec, 72°C for 30 sec, repeated for 30 cycles followed by a 72°C for 3 minutes cycle for extension. The amplicons were stored overnight at -20°C before electrophoresis run.

The dCTP amount was depleted in the dNTP mix due to the use of α32P dCTP amount used for internal labelling of the amplicon. The amount of template used for the amplification was later on diluted several times and a minimum amount required for such reaction was determined. This development is described later.

Stutter bands produced during amplification appeared to interfere with the 'reading' of the genotypes. The use of whole genome amplified (WGA) samples had previously been observed to increase the 'stuttering' during DNA amplification, thus proving impossible to identify individual genotypes. Genomic DNA templates as well as WGA template were tested.

Genomic DNA template titration was preformed in an exponential manner between 10ng/ul (original template concentration of the control samples) and 0.1ng/µl. Fluorescent end labelling (FAM) was tried for sensitivity, as means of replacing radioactive labelling.

Primers used were as follows:

Forward 5' ACT GCA CTC CAG CCT CGG AG, and

Reverse 5' ACA AAA GTC CTT TCT CCA GAG CA.

FAM modification was applied to the anti-sense strand primer. To the 5' primer sequence a miss match was introduced at the –5 position, G replaced the C in an attempt to increase specificity. Electrophoresis was performed on 6% denaturing polyacrylamide gel (DPG).

Development of a 'GH ladder' marker for the A1819 polymorphism.

In order to identify various genotypes of the A1819 polymorphism (consisting of 24 alleles previously identified by Jeunemaitre *et al.*) a sizing marker was required. A GH 'ladder' was produced to identify the various alleles.

20 'in-house' control samples were amplified for the A1819 polymorphic site and run on 6% DPG (denaturing polyacrylamide gel). The pattern band in each sample was observed. Mixes of several samples of the 'in-house' controls were performed. DNA template was mixed in equimolar concentrations, dried at 80° C for 10-15 minutes and amplified using the amplification programme described earlier.

The control samples of DNA were numbered 1 to 20. The samples were mixed as follows:

four combinations of 5 samples (e.g. sample 1-5, 6-10 and so on)

two combinations of 10 samples (1-10 and 11-20)

one combination of all 20 samples.

Individual amplification of the 20 samples showed that the yield of the PCR was not even. The amounts of each PCR product to be included in a 'GH ladder' marker had to be therefore considered such that the bands of such marker would appear of similar intensity.

Following electrophoresis of each individual amplified sample, many combinations of samples were tested on gels and four controls were chosen as the 'steps' for the marker. The samples, which were chosen, had four different allelic combinations. (Samples 1,6,8, and 12) which would create a suitable enough ladder for future reference in genotyping the A1819 polymorphism.

Control samples number 8 and 12 showed approximately double intensity bands on the gels and therefore the amount needed of each sample that constituted the 'ladder' had to be determined. A mix containing 1.5/1 (v/v) and 2/1 (v/v) samples 1 and 6 to sample 8 and 12 respectively were run electrophoresed to determine which mix gave a more even resolution of bands. It was found that a 2/1 mix (v/v) 1 and 6 to 8 and 12 samples respectively gave an even intensity of bands on gel electrophoresis.

The alleles of each of the samples tested in our study were to be 'named' according to the relative distance form the 'steps' of the DNA 'ladder' marker chosen.

For accuracy of reading and thus genotyping the larker was run at 8 sample intervals.

The bands of the 'DNA ladder' were measured from the lightest to the heaviest band (in mm of gel run) and a 'reference ruler' was set for future analysis.

The GH ladder produced in the laboratory had the following migration distances of bands at 150V for 4 hours:

Band1=0mm

Band2=43mm

Band3=58mm

Band4=74mm

Band5=106mm

Band6=123mm

Band7=160mm

In order to maintain a good quality of 'DNA ladder marker' amplification of the individual samples that form the marker had to take place at regular intervals determined by the half-life of the radioactive substance used. (i.e. ^{32}P)

Prior to loading on the gel the PCR product was denatured. The conditions under which this occurs i.e. time vs. temperature are factors that may cause the 'fuzzy' bands. Optimisation for time and temperature was carried out to determine best conditions for good resolution of bands. Three different temperatures were tested for three time intervals: T=80° C, 85°C, and 90°C (previously tested 95°C) for 1,2, and 3 minutes.

3.2.3 Insulin like Growth Hormone (IGF1) dinucleotide repeat (CA)n

The microsatellite polymorphism in the IGF1 gene was one of the types of polymorphic markers used for testing the viability of the various conservative methods of implementation in DNA banks. The polymorphic cysteine-adenine (CA)n di-nucleotide repeat, upstream of the transcription site of the IGF1 gene, was first described by Weber *et al.*. (49)

The location of the nucleotide position is 947-984 in the original human DNA sequence (Genbank accession number M12659, M77496)

10 µl PCR mix contained: 5 µl DNA template (7 ng/µl), 0.1 µl PCR mix (10x), 0.5 µl W1 (1%), 0. 5 µl dNTP mix (4 mM dATP, dGTP, dTTP and 2 mM dCTP), 0.4µl $MgCl_2$ (50 mM)/ 0.1µl (100 pmol/µl) of each IGF1 primer, 0.1µl Taq DNA polymerase (5U/µl), and 2.3 µl H_2O (sterile) and 0.1 µl alpha ^{32}P dCTP (2 µCi/µl). The amplification was performed for 30 cycles under the following conditions: 95°C for 3min, 95°C 63°C and 72°C each for 30 seconds followed by 72°C for 3 minutes. 30 cycles PCR programme was run for the

amplification of the IGF1 (CA)n repeat using WGA or genomic template. During extreme dilution the number of total PCR cycles was increased to 50. The increased number of cycles was necessary as the amount of template used for the amplification was minimal.

The initial amplification was performed on 35-ng/10μl genomic DNA template. Genomic template titration was one of the options tested for conservation of DNA bank, as whole genome amplification (WGA) was observed to produce too much stuttering, which interfered with the genotyping.

The DNA template was titrated down in an exponential fashion, from the original 10 ng/μl to 3.16 ng/μl, 1 ng/μl, 0.316 ng/μl, 0.1 ng/μl, and 0.0316 ng/μl. An additional 0.2-ng/μl template concentration was also tested. 1 μl of each template concentration was used for each 10 μl PCR reaction.

Fluorescent labelling of one of the primers was applied in an attempt to differentiate between the stutter bands and the allelic bands. (MWG Biotech) One primer was end-labelled using FAM fluorescent group at the 3'end of the primer.

The IGF1 primers were as follows:

Forward 5' GCT AGC CAG CTG GTG TTA TT 3', (later modified to carry the FAM group at the 3'end), and the anti sense or reverse primer

5' ACC ACT CTG GGA GAA GGG TA. 3'

MWG Biotech supplied all primers.

The FAM labelling was applied to the sense strand primer to aid discrimination between allelic bands and stutter bands.

3.3 *Application of conservative methods in DNA banks:*

Studies in Southampton (MRC EEU) over the past decade have demonstrated strong associations between parameters of early life growth and late life disease traits (FOAD, foetal origins and adult disease) including several facets of cardiovascular risk, diabetes and osteoporosis. The primary cohorts used in these studies were subjects born in Hertfordshire, and now in late life, for which some early life measures were available.

A relatively unexplored hypothesis to explain these observations is that the early and late life features are independent phenotypes conferred by the same genotype. Systematic examination of this hypothesis by genetic association analysis will become feasible over the next decade. A public domain genome-wide single nucleotide polymorphism (SNP) database of 300,000 markers (of an estimated 3,000,000) is being developed internationally over the

next few years with £28m support from Wellcome Trust and a consortium of pharmaceutical companies. This database will facilitate extensive association analysis, enabling studies not previously possible.

For DNA banks, DNA stocks are finite and lymphocyte immortalisation is unlikely to be feasible. Instead, we have developed (1998-9), systems that are highly conservative of DNA templates and additionally are cost efficient for high throughput analyses. A DNA bank of 1,032 Hertfordshire subjects has been established.

Genotype assays have been explored on a matched control DNA collection, for several categories of genetic diversity including SNPs, small insertion/deletions, microsatellite and minisatellite loci and rare (unknown) mutations. Robust reductions of DNA bank assays to zeptomole quantities of template have been demonstrated, with preservation of phase information for direct haplotyping of candidate genes. Such conservative systems permit a relaxation of choice of candidate from strong hypothetical justification toward systematic genome-wide measurement of the genetic contribution to the FOAD observations. We have also attended to general DNA and information management. Several assays of genetic diversity in growth factor pathways have been developed for genotype-phenotype analyses and combined epidemiological and genetic analyses will be presented. The developments of the conservative methods with application to the IGF1 gene and GH locus will be presented and discussed in this subchapter.

3.3.1 Conservative methods applied to the human IGF1 (CA)n repeat:

Whole genome amplification was initially tested as template for analysis of the microsatellite (the CA repeat) in the IGF1 region.

Extreme dilution of the template has shown that successful amplification of the region was achieved even at 0.01 ng genomic DNA for 10µl reaction mix. The dilutions of the template were performed in an exponential fashion (Fig. 26). The concentrations of genomic DNA template tested, as part of the extreme template dilution was 10, 3.16,1,0.3, 0.1, 0.03 ng/ μl.

Severe slippage was observed when using DOP template to the extent that genotyping was not possible. The slippage problems with the amplification of microsatellites have been well documented in the past, and are known to be the effect of Taq replication errors. Due to the increase in slippage bands, which interfered with genotyping, another way to conserve genomic DNA banks had to be developed.

The ability to amplify very small quantities of genomic DNA in order to increase the amount of template which can then be used for various analysis is of great interest for many research areas such as genetic disease diagnosis, genetic linkage analysis, and genetic diversity studies, to name but a few.

Whole genome amplification was previously shown to be a powerful tool for DNA analysis in single or few blastomeres and nucleated erythrocytes from maternal blood for the diagnosis of foetal inherited diseases, as well as for sperm typing, and use in tumour pathology. (93) Degenerate oligonucleotide primer (DOP) was also used for screening multiple loci in small samples of diploid and aneuploid cells. (94) We have tried to establish the viability of using such methods for analysis of microsatellites and SNP analysis. Microsatellite analysis showed a dramatic increase in the slippage bands during amplification, due to the inaccuracy of the Taq DNA polymerase. For the amplification and analysis of the (CA) n repeat in the IGF1 region the use of DOP was found not to be an appropriate method for the reasons outlined above.

Extreme dilution of the genomic DNA template has therefore been determined to give the appropriate amplicon even at the 0.1 ng DNA for the microsatellite analysis in the IGF1 region. (Fig.5) The amplicon was obtained using a 50-cycle amplification programme.

We have understood the role and limits to which template dilution can successfully produce the right amplicon and shown that it is one possibility for the preservation of DNA banks. Our findings were that microsatellite analysis could be performed thus on extreme template dilutions but not on whole genome amplification. Long PCR amplification method was not available at the time when experiments were carried out on this particular development.

3.3.2 Conservative methods applied to the human GH region -A1819 polymorphism:

The use of WGA as template for the amplification of the A1819 polymorphic region was proven unsuccessful. Genomic as well as whole genome amplified (WGA) templates were used. When using WGA as template severe slippage was observed which made differentiation between real bands and slippage bands impossible. By using genomic templates the slippage was greatly reduced although not totally eliminated.

Several dilutions of the genomic template were successfully applied.

The quality of the genomic DNA used was observed to be a critical factor for obtaining a clear resolution of bands. This was apparent in the Hertfordshire study where East Hertfordshire study presented clearer allelic bands than the North Hertfordshire (poor quality DNA) study (Fig.6). The quality of the DNA in the North Hertfordshire bank is due to extraction problems, poor quality collection of samples, and other laboratory handicaps, which have contributed to it.

Lower quality genomic DNA template resulted in a lot of background noise, with severe slippage on the track that resulted in great difficulty in 'reading' the bands.

The dyes contained in the stop solution (loading dye) were used as control size markers. Xylene cyanole in a 6% denaturing urea gel runs as 106 bp and bromophenol blue runs as 26 bp. The gels were run until the xylene cyanole reached the lower edge of the gel. (after a 2-hour run at 150V). The A1819 bands were found half way up the gel between 255 bp and 300 bp. Although the allelic bands pattern was easily identified, a marker was needed which would differentiate between the various alleles.

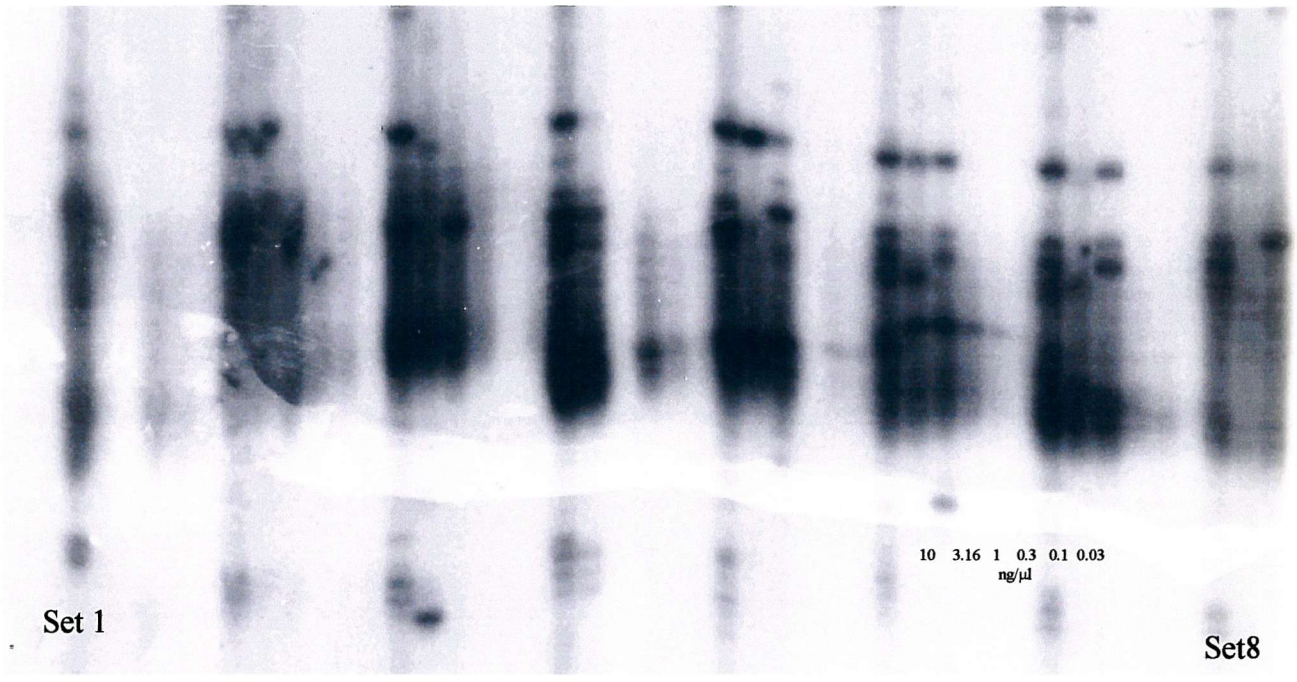


Fig. 5 The effects of extreme template dilution on IGF1 (CA) n repeat were examined. Six dilutions were performed and the set of identical samples is repeated 8 times (set 1-set 8). The template dilutions are: 10ng/µl, 3.16, 1, 0.3, 0.1, 0.03-ng/µl (from right to left in each repetition of samples). Product was observed even at the extreme dilution of 0.1ng/µl.

GH size marker (GH ladder):

A number of combinations between different control samples were tested. The marker was designed so that a high number of alleles are represented on the 'ladder' but no two samples have the same combination of alleles. The samples chosen for the DNA ladder were controls number: 1,6,8, and 12. Several combinations between these samples were tested. Final marker contained a mix of all the four samples. However, due to differential yield post PCR between the samples a mix of 2:2:1:1(v/v/v/v) of sample 1,6, 8, and 12 respectively gave the best representation of seven separate alleles on the marker (Fig.7).

The bands were measured according to the difference in distance compared with the bands in the ladder. The fastest running band on the ladder was chosen to be 0 mm, i.e. the start line for the bands. The ladder was run at every 8 samples to allow genotyping. Slippage bands were observed to have similar intensity as the allelic bands and therefore differentiation between allelic bands and slippage bands was not easily achieved.

The 20 in-house control samples were used in an attempt to discriminatively 'bin' the alleles into groups. The samples were run in order and then at random to verify the 'binning'. The alleles were identified and was found that the slippage band occurred at approximately 2.75-3.00 mm, where the distance between the allelic bands was 3 mm. Slippage band was determined by the lesser intensity (in some cases) on the photographic films. Discriminative binning has not been achieved. The graph (Fig.8) represents an attempt made to group the different size bands into discriminative allele groups ('bins'). End labelling of one of the primers did not improve resolution to the point of total discrimination between slippage bands and allelic bands although sharper allelic bands were observed.

An attempt to improve the heavy background observed in some cases consisted of testing three different temperatures for the denaturing of the samples (80°C, 85°C 90°C and 95°C for 3 minutes). Samples were also tested for stability at these 3 temperatures over a seven-day period. Bands were observed to be less sharp at 80 ° C, 85 ° C, and 90 ° C with best results being obtained at 95 ° C. Over the seven-day time lapse the bands appeared fainter and less sharp although the same trend for the melting temperatures was observed.

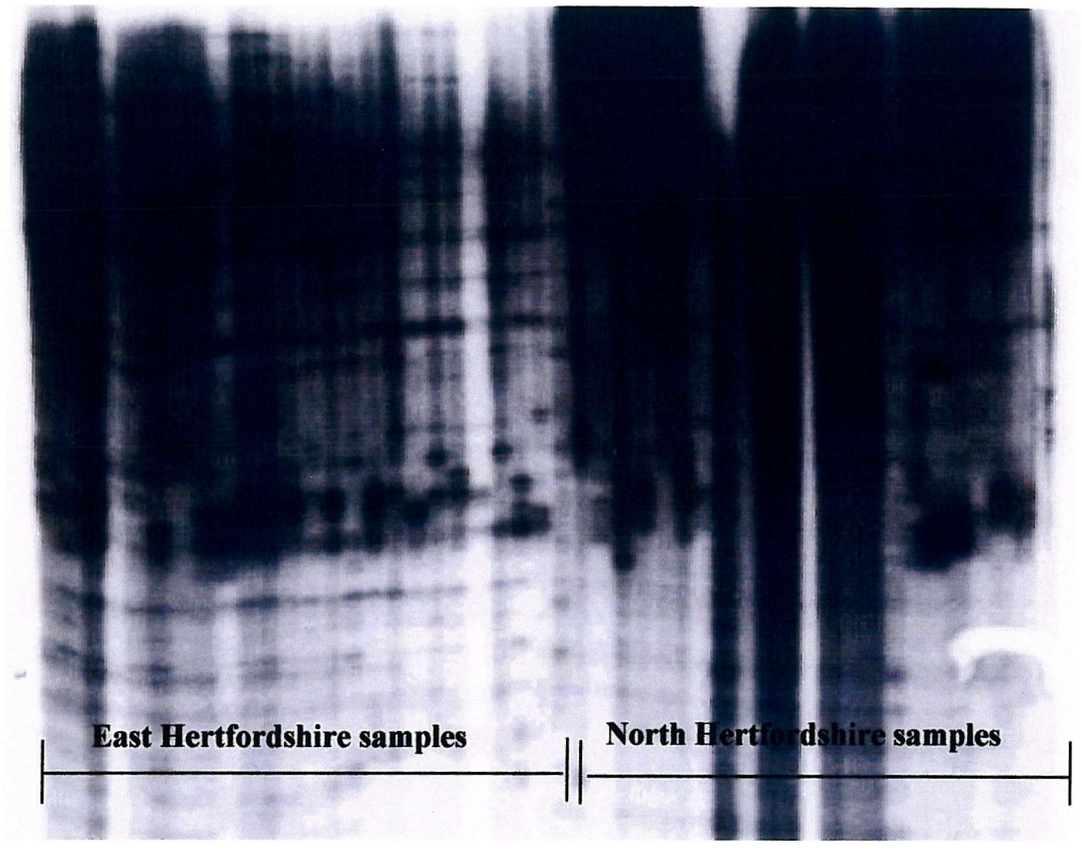


Fig. 6 East Hertfordshire (left half) DNA samples presented better resolution than North Hertfordshire(right half). The quality of the DNA template used was critical factor in resolving the genotypes. North Hertfordshire DNA was observed to increase stuttering and slippage beyond the point of discrimination.

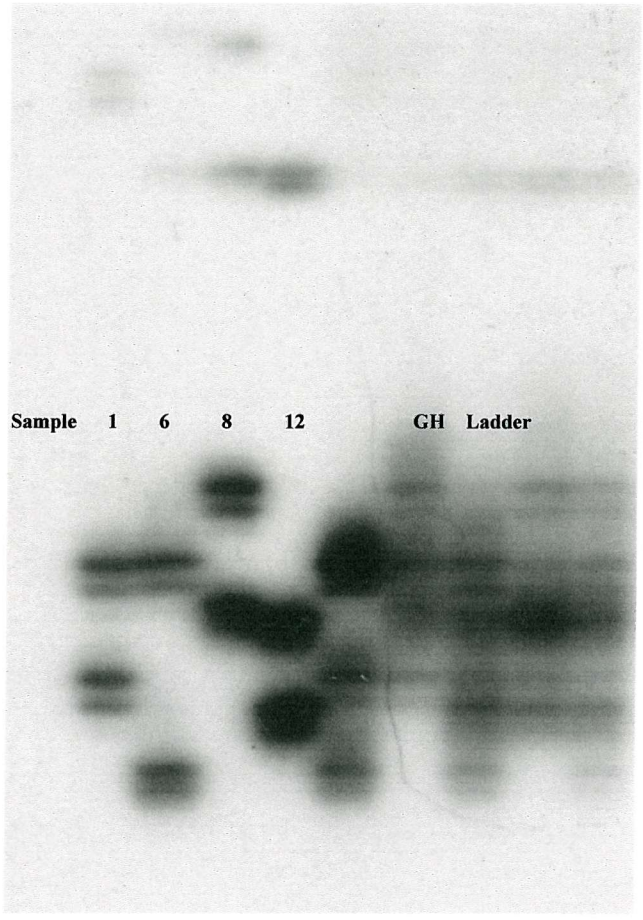
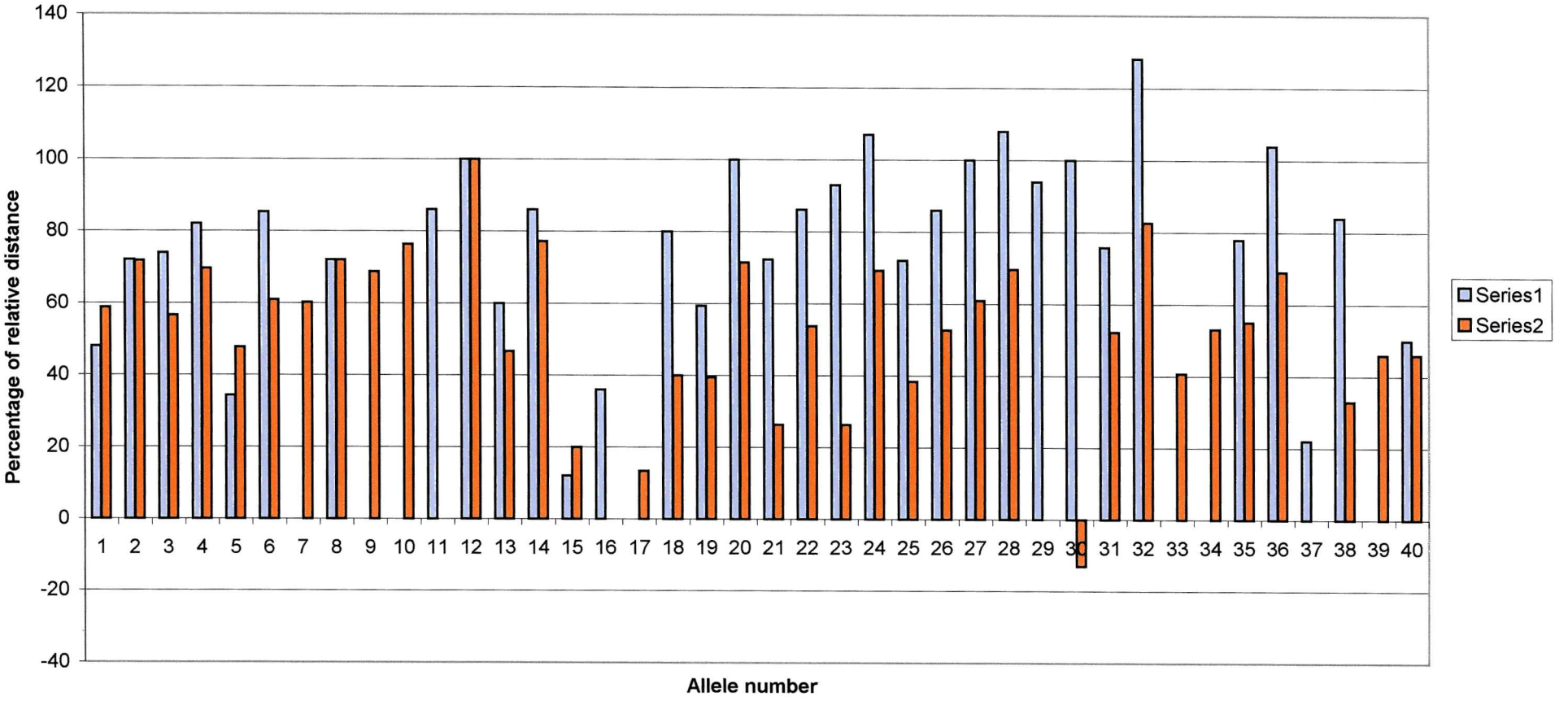


Fig. 7 GH A1819 samples 1,6, 8, and 12 were run on the gel adjacent to the ladder mix to verify the presence of the alleles in the GH ladder marker. At the left the individual DNA control samples are represented. At the right the samples were combined to obtain the GH ladder marker.

Fig. 8 GH A1819 Control samples and random samples
Series 1= randomly loaded control samples
Series 2=ordered loaded control samples



Analysis of the di- and tetra-nucleotide polymorphisms in the GH region was also tested on whole genome amplified template. After successfully amplifying the region using genomic DNA template (to verify that amplification of the right region can be achieved using the methods described) DOP template was put to the test. An increase in slippage was again observed. The intensity of the slippage bands was about the same as the intensity of the allelic bands and therefore genotyping could not have been achieved using the whole genome amplification method for this type of polymorphism. However, an attempt was made to genotype this region using low amounts of genomic DNA template. The combination of di- and tetra-nucleotide polymorphisms found in the region delimited by the Alu 18 and Alu 19 of the GH region was identified and we were able to create a way of 'reading' the alleles in the region by creating the GH DNA 'ladder'. The PCR amplification was achieved on very low amounts of genomic template, and we were able to discriminate between the allelic and stutter bands (a well recognised technical problem in genotyping such polymorphisms).

The allelic variation at the A1819 locus of the GH region was previously described. (92) The number of possible and previously observed alleles at this locus was counted to 24. Any combination of these 24 possible alleles would therefore be possibly found in a human genotype. Due to the small number of samples tested (20 'in-house' controls) in our method developments the 'discriminative binning' of the samples into definite groups was not possible. Figure 8 shows a graphic representation of the alleles found in the control samples. The number of samples tested was low (20 controls) as it was determined that a crucial factor in identifying this polymorphic site was the quality of the DNA template used for the amplification. Therefore the samples in our DNA banks would increase slippage and genotyping would not be achieved with confidence. Also miss reading the slippage bands for the allelic bands would not have produced an error free result and conclusion.

We were able to determine that high CG regions in the genome are not possible to amplify using whole genome amplification. DOP was found not to be a valid method of amplification for the IGF 11 and ApoE genes. (L.Hinks)

3.4 *Long PCR*

Long PCR is achieved by the use of usually two polymerase enzymes: a non-proofreading polymerase (usually Taq DNA polymerase), which is the main polymerase, and a proofreading polymerase (3' to 5' exo-) that is present in the reaction mix at lower concentration. Taq DNA polymerase lacks the proofreading activity that is provided in the reaction by the second enzyme, the proofreading enzyme. Small nicks in the DNA strand are 'repaired' by the proofread enzyme thus allowing Taq polymerase to continue amplification. Barnes and Cheng used the two-polymerase system. Barnes used a combination of Klentaq 1 (an N-terminal deletion variant of Taq) and Vent/Deep Vent/Pfu. (95) The best results were obtained when using Pfu and a fragment of 35kb was successfully amplified from 1ng of Lambda template. Cheng used recombinant Tth DNA polymerase and Vent. (96) The amplifications reported were of 22kb length using 100ng human genomic DNA and 42 kb from the bacteriophage Lambda.

Primers are usually 21 to 34 oligomers with balanced melting temperatures between 60 and 70°C. This allows higher annealing temperatures to enhance reaction specificity. The pH is particularly critical-a higher pH environment than the typical Tris, pH 8.3 Buffer. This offers DNA strands greater protection against denaturation and subsequent nicking during thermal cycling. Thermal conditions are carried out using annealing and extension steps at fairly high temperatures using a minute per kilo base target extension. Thermal cycler profiles vary with target lengths and primer sequences.

For the present study PWO was used as the proof reading enzyme. Titrations of the PWO enzymes were carried out (L Hinks) in the presence of Taq polymerase in order to determine optimal levels for the removal of the mismatches. The optimum conditions for Taq DNA polymerase were: 2.5 units: 0.02 units+ 125:1. Using Taq alone in the reaction gave no visible amplification (L. Hinks).

The upper limit of the size of the amplicon obtained through long-PCR has not yet been identified. The integrity of the templates represents a major key parameter as longer targets can be amplified from DNA with fewer nicks. This indicates the need for careful sample preparation. Long PCR can be a labour saving alternative for the study of molecular genetics of large genomic fragments. Most importantly the use of the DNA bank would be minimised thus infinitely more tests can be performed on the bank DNA.

3.4.1 Long PCR for SNP analysis

GH1 (5kb) and IGF1 exon 5 (6.6kb) regions were amplified. The 'long PCR' (L. Hinks and I N M Day) technique consisted of a 35 cycle modified PCR program.

The following **primers** were used for the long amplification:

GH1: forward 5' CGG GGA GGA GGA AAG GGA TAG GAT AGA GAA 3' and reverse 5' AAT GGA AGA GGG TCC CTG AGC TCC AAG TAG 3';

IGF1 (ex5): forward 5' AAC TCG GGC TGT TTG TTT TAC AGT GTC TGA 3' and reverse 5' AGG ATT GAT ATT CCT CTG CCA TAA GTG AAT TG 3'.

Primers were designed using Primer 3 program with the following sequences: GH1 Genbank Accession no. J03071 (Appendix1) and IGF1 X57025 (Appendix 2).

The amplification of the GH1 (5kb) and IGF1 exon 5 (6.6kb) region was achieved using denaturation step at 94°C for 2 minutes, then 35 cycles of 94°C for 20 seconds, 50°C for 30 seconds 68°C for 6 minutes, followed by extension at 68° C for 20 minutes.

GH1 long PCR contained (50 µl reaction mix): 5 µl Buffer pH8.5 (500mM Tris-HCl pH 8.9, 140mM Ammonium sulphate), 250 µM dNTP mix (10 mM equimolar mix of dATP CTP, GTP and TTP), 400 nmol of each of the primers, 2 mM $MgCl_2$ (50mM), 1.3 mM Betaine (5M), 2.5U Taq DNA polymerase (Gibco), 0.02U PWO, and 26.6 µl H2O. 20 ng DNA were used for each 50ul PCR reaction.

IGF1 6.6kb amplicon was amplified using the same reaction mix as described but at a final MgC_2 concentration of 3mM.

PCR product in both cases was tested on 0.7% agarose gel. Electrophoresis was performed immersed in TBE (1X) in a horizontal electrophoresis tank at 150 V for 20 minutes. Staining was achieved using Ethidium bromide. A 1kb ladder was used to determine the size of the amplicon.

SNP polymorphism in the GH1 and IGF1 genes and ARMS assay:

Cargill *et al.* originally identified the single nucleotide polymorphisms in the GH1 and IGF1 genes. (97)

GH1 SNPs were identified on the Genbank sequence J03071. (Appendix 1) 5 of 7 SNPs listed are in the 5' Flank region of the hGH1 gene. Only one SNP is found in the coding region of exon 4 (SNP 6099). All the single nucleotide polymorphic sites in the GH1 gene are transversions. GH1 map is presented in Appendix 10.

IGF1 SNPs were mapped using Genbank accession number X 57025. (Appendix 2) A map of IGF1 is given in Appendix 11.The primers were designed using a mismatch base at the 3rd position of the 3' of the primer, in order to obtain high destabilisation strength. The 5' end of each primer was also altered in an attempt to reduce the melting temperature and match them between all primers. The 3' end of one primer ended with the SNP base or with the complementary base for the reverse primers. The primers sequences for GH1 and IGF1 SNPs are given in Appendix 8. Constant common primers as well as control primers were selected using the same Genbank sequence. The ARMS primers were selected as long oligonucleotides (~28bp in length). The 3' terminal base of the mutant primer was made complementary to the mutation whereas the 3' terminal base of the wild type primer was selected complementary to the normal sequence. Deliberate additional mismatches were selected at the –2 position of the 3'-end in order to increase specificity of the primer. Strong additional mismatches were introduced to all primers using the guidelines outlined by Little, S. (89) The mismatches were as follows: G to A, C to T, and T to T. The reminder of the primer sequence was complementary to the target sequence. As the sample genotype is determined by the presence or the absence of a reaction product, PCR control reaction was essential. The control primers were selected several hundred bases away from the target sequence. The control amplicon size for the GH1 gene was chosen to be of 268bp long, over 100bp longer than the biggest allele specific amplicon. For the IGF1 gene, control primers had to be designed for each set of allele specific primers. Due to the distances (in base pairs) between the SNPs these were divided into five sets as follows:

Set 1=SNP 2350, and 2396:

Set2= SNP 2892;

Set 3=SNP 3123, 3276 and 3340;

Set4=SNP 5899; and

Set 5= SNP 6721, 6939 and 7012.

All SNPs were named after their location in the sequences indicated earlier. The mutations are found in the 3' non-coding region of the IGF1 gene.

Primers were selected manually using the same Genbank sequences, after the initial identification and position check for each SNP was performed.

Cross dimer formation and Tm alteration for all primers (in both genes) were performed using Net primer analysis (www).

Primer optimisation was achieved using a 10 DNA's mix (10ng/UL) from the 'in house' control DNA samples. (Appendix 9) The 'in house' control samples are anonymous male and female volunteer samples of DNA. The volunteers are members of the research division who donated blood for extraction of DNA, thus allowing a continuous supply of good quality DNA. The samples are anonymous and are used for reaction optimisations and control purposes.

Genomic template used for primer optimisation was replaced by long-PCR amplicon for the genotyping of SNPs. When using the long-PCR amplicon, template titration was performed and best results were obtained using 1:100 dilution. Higher dilutions such as 1:1000 gave a faint appearance of the bands.

A $MgCl_2$ concentration of 2.5 for GH1 and 3 to 4 mM for IGF1 was selected as standard and the melting temperature was chosen at the 'edges' of the stringency envelope.

Optimisation for all GH1 primers was performed on genomic DNA, whereas the IGF1 primers were optimised using long-PCR amplicon (the same 10 DNA's 10 ng/μl was amplified through long-PCR then diluted to 1:1000 and the dilution was used as template for the optimisation).

Optimisation of the primers achieved following the protocol was performed for 6 $MgCl_2$ concentrations over a temperature gradient of 55°C to 70°C. (Using the Gradient Cycler, PTC-225, MJ Research) The temperature gradient consisted of: 94°C for 4 minutes denaturation step, followed by 43 cycles of 94°C for 30 sec, 55-70° C for 30 sec, 72° C for 1 minute, and an extension of 72°C for 10 minutes. Samples were stored at -20°C. All amplifications were performed using the same make of thermocycler machine. (PTC-225, MJ Research-Gradient cycler)

Where a specific product failed to be obtained but the control bands were strongly visible, the concentration of the control primers was reduced in order to try and increase the sensitivity of the reaction. The control primers were titrated 1/5, 1/10 and even 1/20 where necessary. The annealing temperature was also decreased by 2-4 degrees for the same reason. Where

specificity was not observed (i.e. all samples became amplified for both allele specific reactions) the annealing temperature was increased by 2-4 degrees.

The ARMS reaction mix contained (10μl) used for amplification contained: 1μl PCR Buffer (10X), 0.25μl dNTP mix (10 mM), 0.5μl (or 0.4 for a final concentration of 2 mM) $MgCl_2$ (50 mM), 0.04 μl of each primer (100 pmol/μl), 0.2 U Taq DNA polymerase (5U/μl), and 8.05μl H_2O.

Genomic (10 ng/μl) vs. long PCR (1:100 dilution of the sample long amplicon) template amplification was performed using 1μl template for each 10 μl reaction mix.

The programme used for specific template amplification consisted of 95°C for 4 minutes denaturing step, 95°C for 30 sec, T= optimum temperature for 30seconds, and 72°C for 30 seconds followed by extension at 72°C for 10 min. The number of cycles for the amplification varied from 35 cycles for genomic template to 25 cycles for long PCR template.

Non-specific amplification of several regions was observed in some cases. Multiple bands appeared on gel electrophoresis. The bands not only interfered with the genotyping but it was amplified at the 'expense' of the correct amplicons (i.e. allele specific and control amplicons). Blast search (www.ncbi.nlm.nih.gov) showed a number of possible sites within the region, which were possibly, amplified using the common forward and control reverse primers. The use of long amplicon of the region under analysis eliminated the problem.

Also a region of approximately 500 bp was often found amplified at the expense of the allele specific amplicon where the common forward primer was found in the reaction mix with the reverse control primer (i.e. in all amplifications preformed 5' to 3') This was resolved by using both sense common primers in the reaction mix so that the allele specific amplification takes place internally of the common amplicon. For some reaction the unwanted common primer concentration had to be altered so that the allele specific reaction is favoured during amplification. (Fig. 9 and 10)

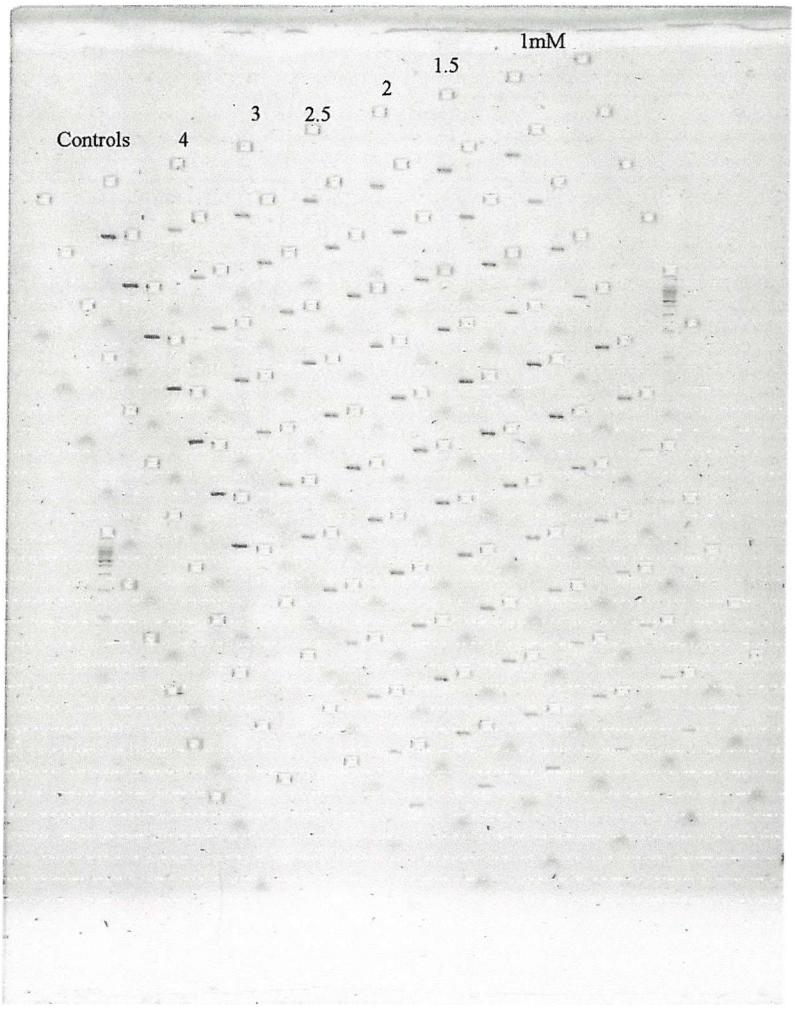


Fig. 9 Primer modification did not achieve ARMS assay specificity. In the ARMS optimisation for GH1, removing the primer's –3-end mismatch did not improve assay specificity. The amplification took place at a wide range of temperatures and salt concentrations. At 70° C and 4mM $MgCl_2$ product was still visible.

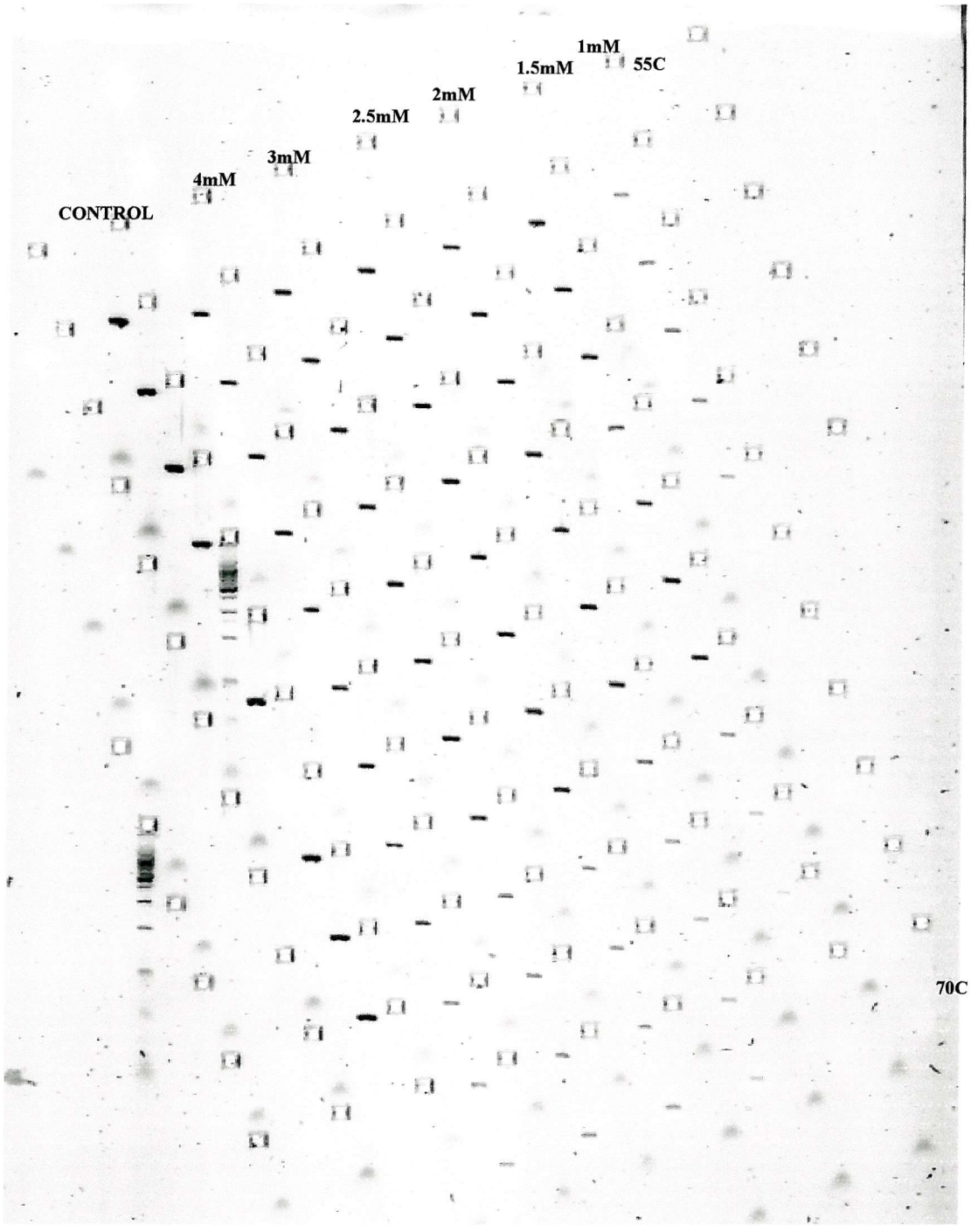


Fig. 10 Stringency envelope obtained through salt and temperature gradient. The temperature ranges between 55°C to 70°C and the $MgCl_2$ from 1mM to 4mM respectively. Stringency envelope is being observed at high (70 °C temperature) and low $MgCl_2$ concentrations.

3.4.2 Development of ARMS assays for the human GH1 gene:

We were able to successfully amplify the 5 kb region of the GH1 gene using the protocol described earlier in chapter 3 (L Hinks). Each long amplification of the original template was checked for size on 0.7% agarose gels to determine whether the product seen is of correct size.

Genomic DNA template was initially used for primer optimisation. Ten of the twenty in house control samples were pooled to a final concentration of 10ng/ul (of each of the ten pooled samples)(Appendix 9). The pooled DNA was used as template for the subsequent optimisations.

The optimum primer reaction conditions were selected from the 'edge' of the stringency envelope (Fig. 11). The melting temperature for all primers was selected such that $MgCl_2$ was kept as constant parameter in all reactions.

Multiple bands were obtained during ARMS assay in the GH1 when using genomic template for the amplification (Fig. 12). Blast search on the www.ncbi.nlm.nih.gov site revealed the presence of 4 sites with high sequence identity with the region to be amplified.

For many reactions additional product was visible on the gel. The 'misproduct' usually appeared as a heteroduplex (dual band) at approximately 500 bp mark. The presence of the misproduct did not interfered with the 'reading' of the sample amplicon therefore it was not found necessary to try and remove it form the reaction. The intensity of the miss-product 'heteroduplex' was however observed to vary in inverse proportion with the intensity of sample band(s). Therefore it was believed that miss priming elsewhere in the genome of the test sample caused the miss-product.

Control primers were observed to favour the reaction with some of the allele specific primers. 500bp product was obtained in some cases. A pair-wise Blast search (www.ncbi.nlm.nih.gov) analysis of the GH1 gene region revealed a number of other sites where amplification was possible. In such cases the control primers were removed and the common primers were introduced to the reaction. Dilutions of the 'unwanted' primer (i.e. the common primer with the same orientation as the allele specific primer) were performed where necessary to 1/10 and even 1/20. The reaction was therefore 'propped' by the common primers and a control band for the reaction was obtained without interference with the allele specific band(s).

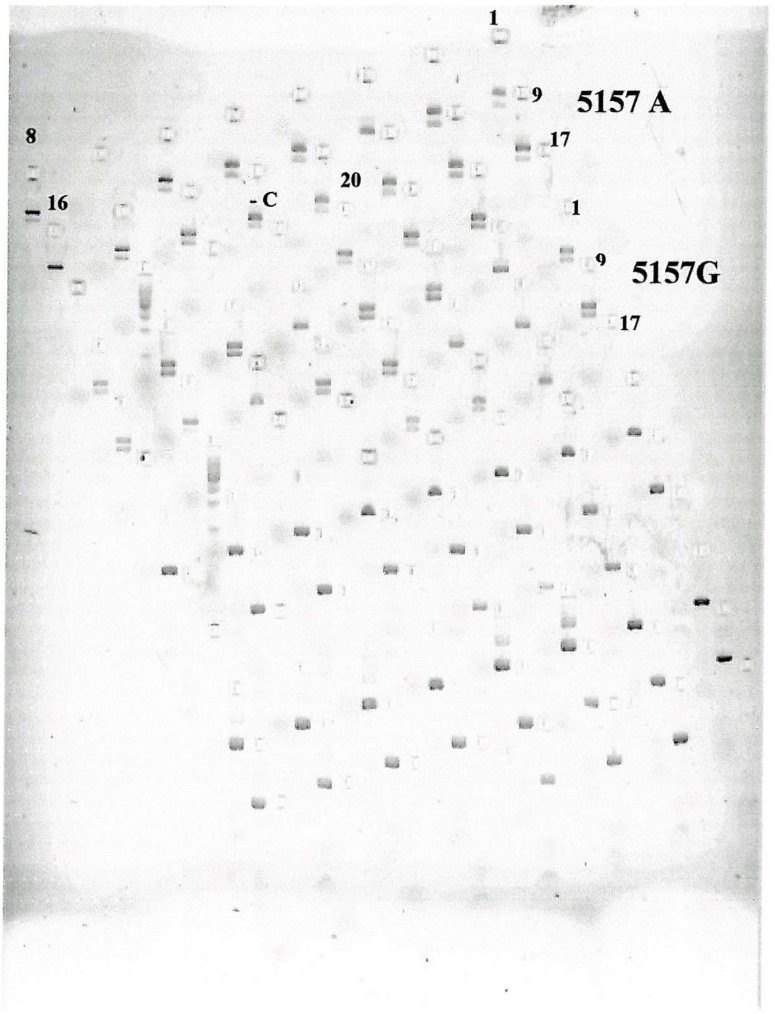


Fig. 11 SNP genotyping in hGH1 gene (long amplicon). Where the control primers are present (bottom half of the gel) no specific amplicon is observed. Specific amplification is however observed when using common primers to 'prop' the reaction (top half). For example, sample 1 is heterozygous AG for the 5157 SNP, whereas sample 2 is AA homozygous.

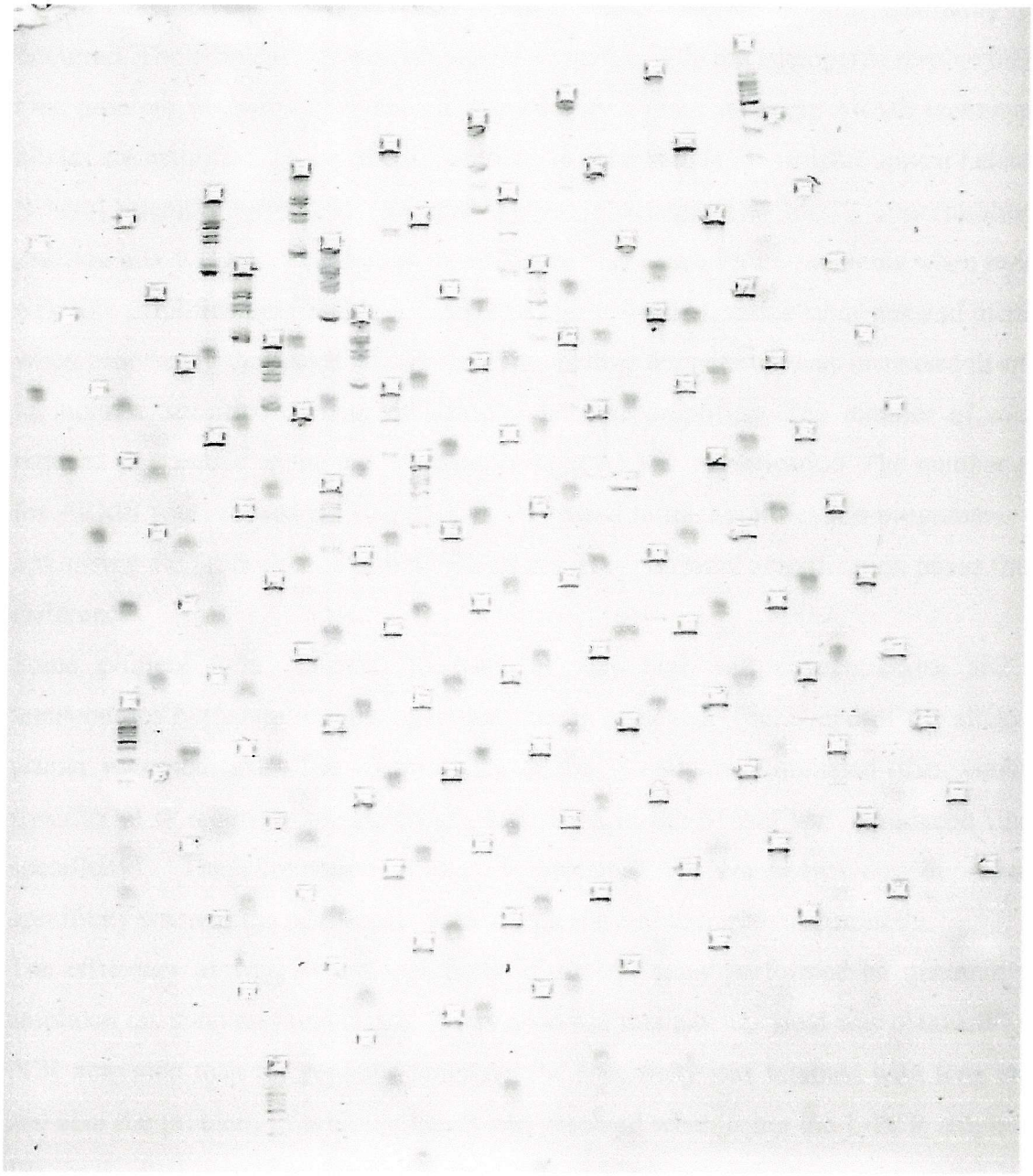


Fig. 12 Multiple bands appear after ARMS amplification of genomic DNA template. This may be due to artefacts during electrophoresis. However, BLAST search (www.ncbi.nlm.nih.gov) identified four possible loci with high sequence homology, to the control and allele specific primer sequences.

In the development of ARMS assays for the SNP analysis several laboratory problems occurred. The problems are described below together with the attempts to resolve them.

One problem when trying to specifically amplify a template using ARMS technique is that alleles are amplified, i.e. no allele specificity is observed, or all samples appear heterozygous. Several attempts were made to resolve these problems. The $MgCl_2$ concentration of the reaction mix was altered although its alteration may cause future problems when trying to do a double amplification (double ARMS) and also it would become laborious and inconvenient when genotyping thousands of samples. The melting temperature was increased in an attempt to increase specificity where all samples become amplified. The number of cycles was reduced as standard technique when performing ARMS amplification. The number of cycles for ARMS was reduced by 10 cycles as compared to the amplification programme used for optimising the primers, so that both alleles are in exponential amplification phase (maximum difference).

Some primers were observed to work at very high salt concentrations and melting temperatures but again without specificity being achieved. The 3' end of the allele specific primer was modified. The –3 mismatch of the 3' end was eliminated (that would reduce specificity) or a further mismatch at –5 position of the 3' end was introduced (increasing specificity). The elimination of the –3 mismatch had not shown any improvement in specificity whereas the additional –5 mismatch showed to improve specificity.

The efficiency as well as the specificity of ARMS assay performed on genomic vs. long amplicon (as template) was tested. It was observed that a better yield was obtained for the L-PCR amplicon than for genomic template, the specificity was retained with long amplicon, and also the problem of multiple bands was resolved when using the L-PCR amplicon (Fig. 13). Individual templates in the North and East Hertfordshire DNA banks were amplified using the long PCR protocol described and randomly checked for amplicon size. (Amplification and product checks were kindly provided by S. Diaper)

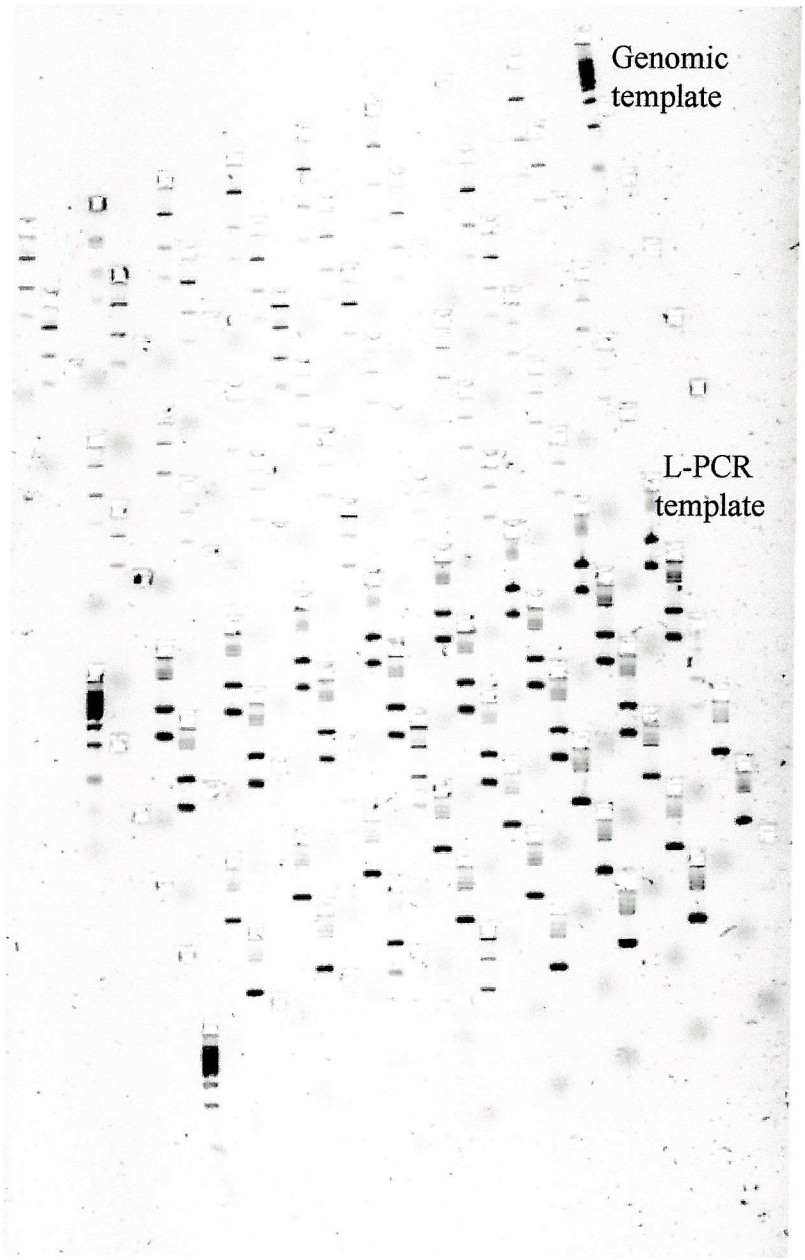
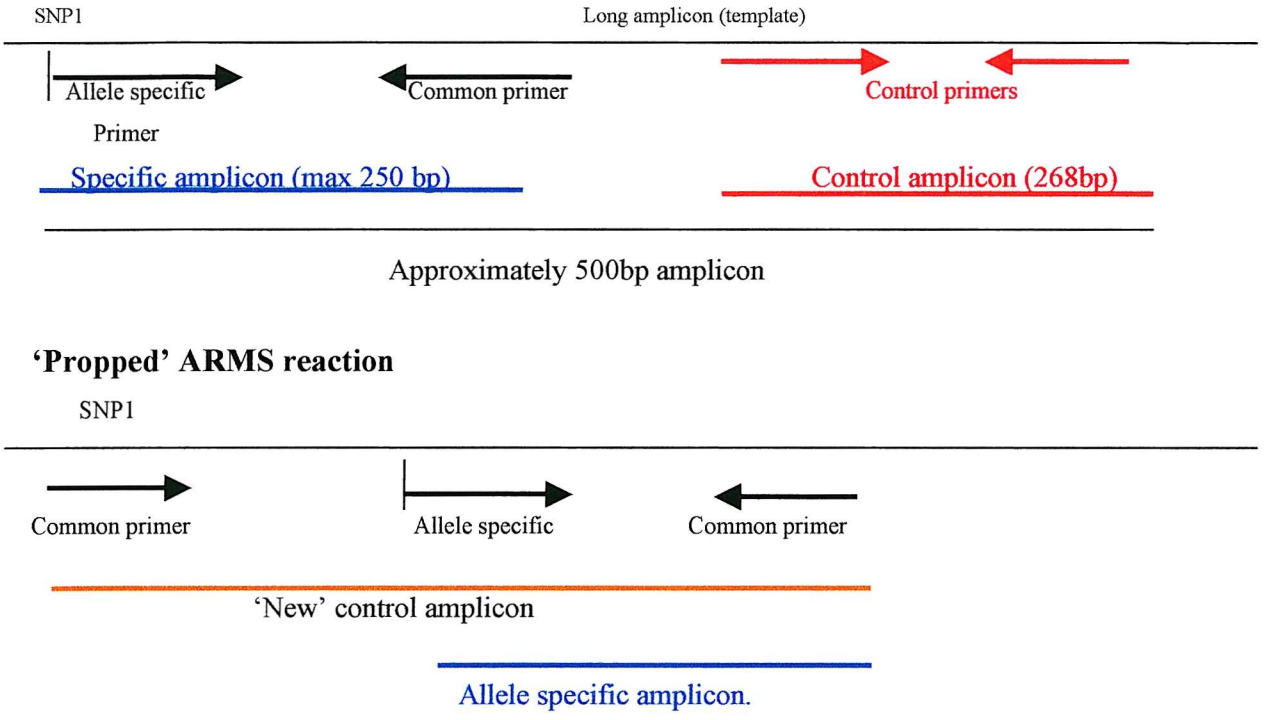


Fig 13 Advantages of using l-PCR amplified template for ARMS assay. The GH5157 SNP marker is given as an example. The upper half of the gel represents the amplification of genomic template whereas the in bottom half of the gel long amplification was achieved using the long GH1 amplicon. Bands appear more strongly where long-PCR template was used with no detriment to reaction specificity.

ARMS assays were run for the SNPs within the 5 kb long amplicon. Not all SNPs showed specific amplification. Where interaction was observed between the allele specific primer and the control primers (an approximate 500 bp amplicon size was observed without any other product) the reaction was 'propped' by the presence of the opposite sense allele specific primer. The reaction was designed so that allele specific amplification takes place internally of the control amplicon, produced in this case by the two 'common' primers. A schematic representation is shown below (Fig. 14)

Fig 14 Original ARMS assay reaction



The absence of the control primers eliminated the miss-product band (approximately 500 bp) and in many cases it has allowed the amplification of the allele specific region (Fig. 15). The common primer of the same direction as the allele specific primer produced a new control amplicon with the opposite direction common primer without interfering with the allele specific amplification. The reactions efficiency between the two common primers was reduced when necessary by diluting the common primer with the same orientation as the allele specific primer.

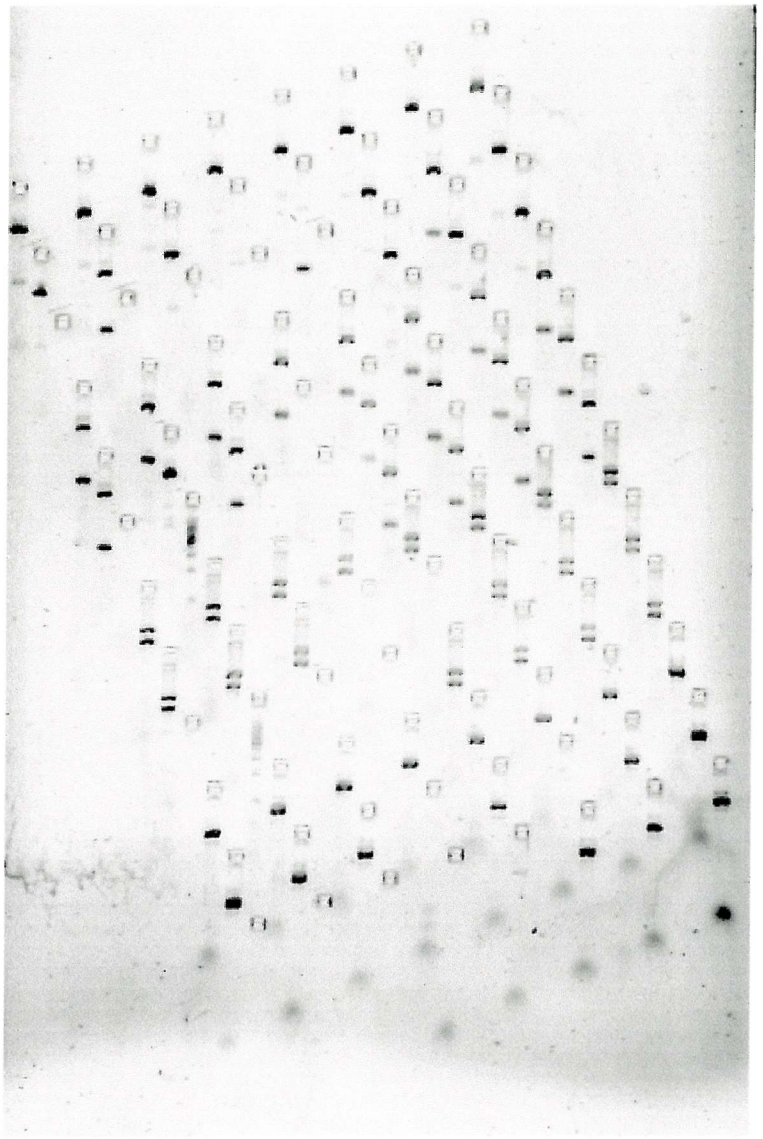


Fig 15 The 'propping' ARMS reaction exemplified during IGF1 SNP analysis. In the top half of the gel control primers were used in the reaction mix and no specific amplicon was observed. In the bottom half of the gel the 'propping' principle was used (i.e. common primers were used for the control amplicon) and specificity was achieved.

3.4.3 Development of ARMS assays for the human IGF1 gene

Initial optimisation of the IGF1 allele specific primers was achieved on long amplicon of the 10 DNA's pooled sample. (Pool consisted of the same ten samples as previously used for GH at the same 10ng/μl final concentration of each sample)

The reaction gave best results at very high $MgCl_2$ concentrations of 3 and even 4 mM.

The stringency envelope strategy identified the specific reaction conditions. A wide range of melting temperatures was observed with great variation seen even between a pair of allele specific primers.

Following optimisation allele specific amplification was tested on the individual long amplicons of the 'in-house' DNA control samples. The reactions used the 'propping' approach, described earlier, for the specific amplification.

4 Development of new insertion/deletion genotyping method: application to the human leptin receptor gene

4.1 *Heteroduplex generator combined with MADGE: a general strategy for analysis of small (5bp) insertion/deletion polymorphisms applied here to a 5bp I/D in the human leptin receptor gene (OB-RB)*

The insertion of the CTTTA fragment produces a cutting site for the RsaI restriction enzyme. Therefore the initial identification of the I/D polymorphism in the 5' UTR region of the OB-Rb gene was the one described by Oksanen (90), makes use of the RsaI restriction site generated by the insert. RsaI I restriction digest in the case of the 5' UTR I/D polymorphism investigated produced three species in the electrophoretic run:

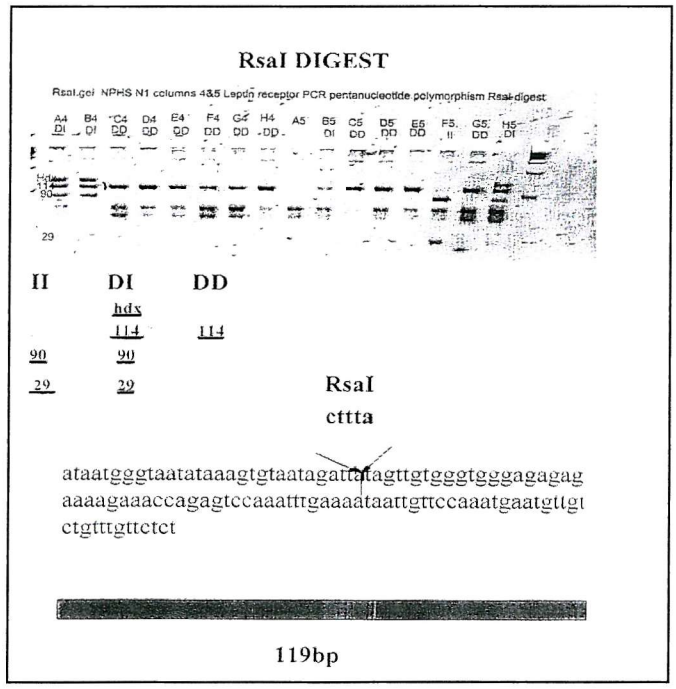
the II species, which present two bands one at 90 bp and the other at 29 bp,

the DD species with an uncut product of 129 bp,

the heterozygous DI species, which produced a total of four bands, the 114 bands, 90 bp band and 29 bp band, plus a heteroduplex band.

Figure 16 shows the three different species obtained by RsaI digest. The gel, 5% polyacrylamide was pre-stained using Vistra Green and electrophoresis was performed submerged in TAE (1X) for 20 minutes at 150 V.

Fig 16 RsaI digest in the human leptin receptor



A straightforward amplification of the fragment containing the 5bp insertion/deletion polymorphism has proven inefficient, as the yield of product was very little; the band on the polyacrylamide gel appeared diffuse and faint with no heteroduplex visible. The aim was to try and reveal naturally occurring heteroduplexes straight after the amplification of the region. Several ways of achieving this aim were tested and are described below.

$MgCl_2$ titration was performed to determine the optimum salt concentration for a higher yield. No product was obtained for 0.5 and 0.75 mM salt and 1.0 mM again produced little yield. W1 and DMSO were then tested against the same range of $MgCl_2$ concentration to determine the effects of these reagents on the yield and quality of the amplification. High amounts of primer dimer were observed and it was concluded that the primers might react with each other before the amplification had actually had a chance to begin.

To solve the problem a hot start PCR at the $MgCl_2$ determined optimum was performed to limit the primer dimer formation. Ethidium bromide staining revealed again very little amplicon presence, therefore the amplification was increased by 5 cycles (from 30 to 35 cycles). Half the reaction mix was prepared without any Taq DNA polymerase and heated at 95° C for 5 minutes after which the other half containing the full reaction amount of Taq DNA polymerase was added. Electrophoresis produced naturally occurring heteroduplexes after only 40 minutes run at 150V on a 7.5% acrylamide gel. Fig. 17 shows the heteroduplexes being visible after the straightforward amplification. The amplification of the template was achieved using 95°C denaturation, 55°C anneal, and 72° C extension.

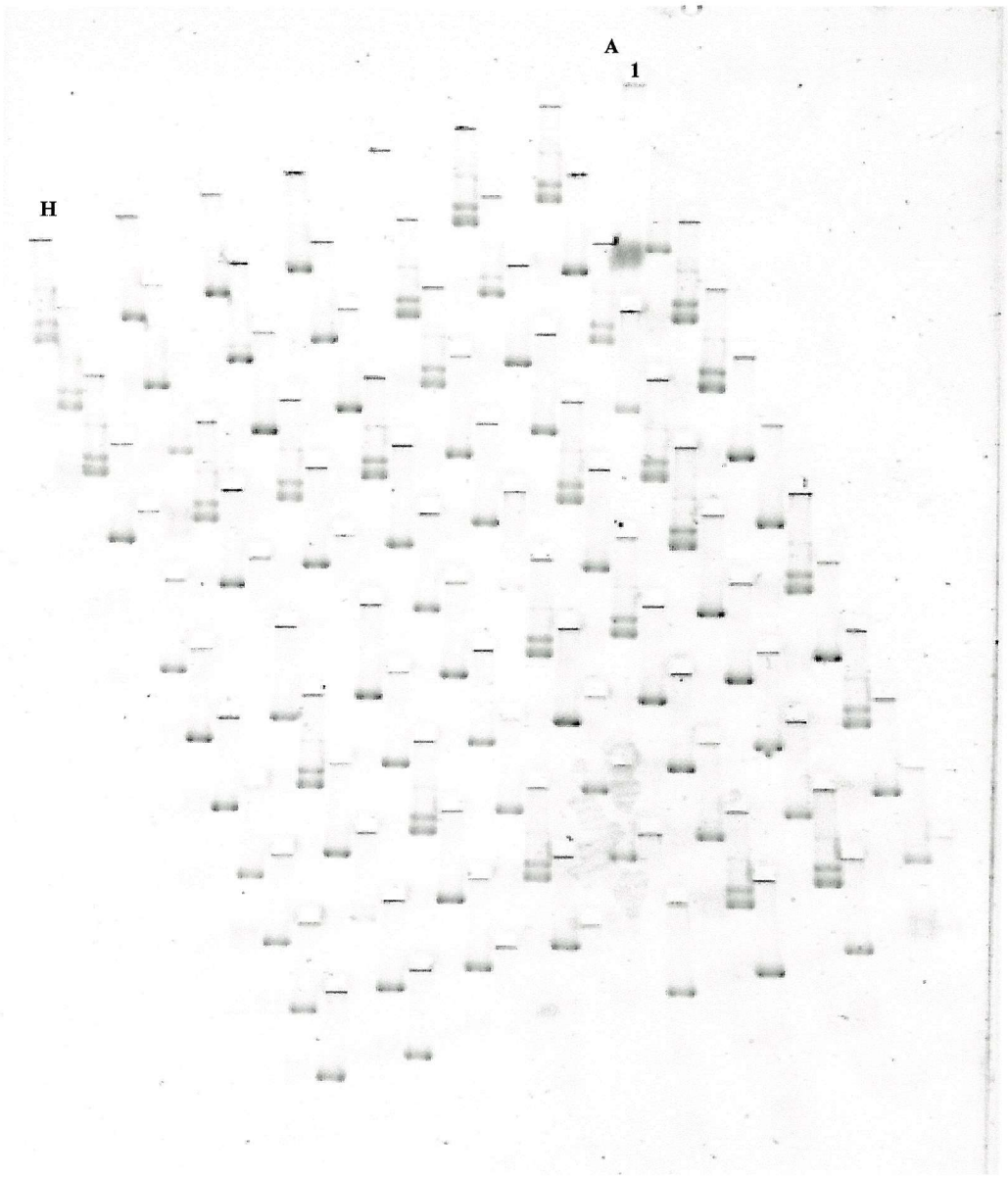


Fig. 17 Initial identification of heterozygous genotypes, which appear as heteroduplexes. Natural heteroduplexes can clearly be identified (B1, C1, H1, C2, D2, H2, A3, B3, D3, H3, A4, E4, F4, G4, B5, C5, B6, A7, C7, D7, G8, A9, F9, E10, B11, C11). The two bands of the HDx appear equally represented before the generator is added.

Using the programme described and adding an extra 'step' after the amplification an increase in the visibility of the reaction product was observed, some resolution of the naturally occurring heteroduplexes, as well as a dramatic reduction in the amount of primer dimer obtained.

However, the desired amount and quality of amplicon were not obtained. A new 'wild' type sense primer was designed where position –3 of the 3'end was matched to the wild type sequence. The removal of the mismatch at the 3'end of the primer reduced the specificity of the primer. The sequence of this sense primer was therefore changed from:

5' ATA ATG GGT AAT ATA AAG TGT AAT AGA GTA 3' to

5' ATA ATG GGT AAT ATA AAG TGT AAT AGA TTA. 3' . The wild type sense primer was proven to be unsuccessful at amplifying the template. No product was obtained when using the wild type primer.

Several steps were taken in the development of 'generator production'. The most abundant genotype was chosen as the 'generator' to limit the number of the 'generated heteroduplexes' to be identified. The DD genotype was amplified so that a further amplification would not produce large amounts of miss-product and that the amplicon obtained through re-amplification would be successful in generating heteroduplexes when annealed with the opposite homozygous genotype.

The DD genotype once identified was selected, the genomic DNA was amplified using the programme described earlier, and the amplicon was diluted prior to a new amplification. A 1/100 dilution was found to produce good amounts of template after the second amplification and therefore the original template would not only be conserved but a continuous source of 'generator' would be produced through re-amplification of the dilution of the DD amplicon obtained initially. A further 3rd round of amplification was tested on a 1/1000 dilution of the 2nd round PCR amplicon. The amplification after the third round was found to be equally successful as the second round. The generator amplicon following the second PCR run was checked on a 7.5% acrylamide gel (40 minutes, 150V, stain Vistra Green) using a 100 bp molecular weight marker. The template for the generator was kept constant to one individual only to ensure best replicability of the protocol. The DD amplicon used for the heteroduplex generation was produced in several aliquots of 10µl following, which the amplicon was pooled and concentration equalised.

Several electrophoretic stains were tested to find the best resolution possible, as the amount of amplicon obtained was small. Ethidium bromide, Vistra Green, SYBR gold and Pico Green were all tested on the acrylamide gels. Ethidium bromide is the most commonly known used dye for both DNA and RNA. It binds to ds-DNA as well as to single stranded and triple stranded DNA. Vistra Green is a sensitive in-gel stain for double stranded DNA. It is 10X more sensitive than Ethidium bromide on an UV transluminator. It works with acrylamide gels as well as with denaturing, non-denaturing and agarose gels. It is non-fluorescent until it binds to DNA and it is thought to bind non-covalently to the DNA backbone. The resulting high signal-to- noise ratio permits detection of less than 20 pg/band in both agarose and acrylamide gels. Ethidium bromide was found not to have enough sensitivity for the low yield PCR product. SYBR gold identified several bands alongside the product size bands, which were believed to be miss-product, whereas Pico Green stain did not have the required sensitivity for band detection. SYBR gold is the most powerful stain available. It is part of a new generation of green fluorescent nucleic acid stains, that detect nucleic acids on both acrylamide and agarose gels. It exhibits exceptional affinity for DNA and a large fluorescence enhancement after binding to DNA- at least an order of magnitude greater than that of Ethidium Bromide. Pico Green has high affinity for double stranded DNA and large fluorescence enhancement upon binding to DNA. It can readily detect as little as 25-pg/ml dsDNA and it is often used for quantitating DNA in solution. A combination of Cyber Gold and Pico Green was also tested without success. Pico, Cyber and Vistra combination of stain resulted in the resolution of several miss-product bands. It was decided that optimum results, for the detection of our amplicon and heteroduplex, was obtained using Vistra Green and therefore this stain was used for all the gels.

Severe skiing alongside track was observed together with the distortion of the bands.

The distortion of the bands together with the skiing effect interfered with the genotyping through band pattern recognition. (Fig. 3)

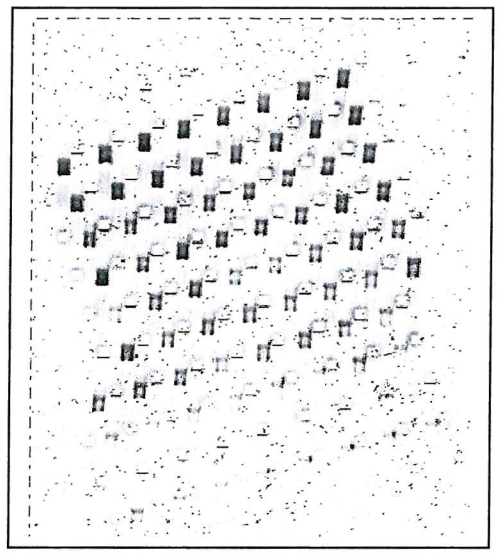


Fig 3: Artefacts that can occur with MADGE electrophoresis.
Bands appear distorted into 'dumbell' shapes with 'skiing' on either side of the track.

Several possibilities were investigated, where changes brought to the amplicon would eliminate the electrophoretic artefacts and improve resolution allowing easy genotyping through band pattern recognition. Changing the reaction mix for the PCR was not considered an option as the amplification of any template is performed at specific salt and reagent concentrations to obtain optimum yield and quality of amplification. Moreover the amplification of the template required several steps of development (described above) and was proven not to be an easy task. By not introducing pre-PCR modifications, the reproducibility of the method is ensured. The observation that the distortion of the bands is more severe in the samples loaded last (when using a 8-channel pipette on a MADGE gel) has lead to the idea that using delayed start electrophoresis may solve the problem by allowing the diffusion of the dialysis of the sample through the walls of the well and also through the surface of the sample when using immersed gels.

The gel composition was altered to resolve the distortion and skiing problem. The types of gels tested ranged between 'water gel' (where water was used instead of the TAE buffer, 0.1XTAE gel, and 1XTAE gel. The bands were found to have little mobility on the 'water' gel, with little improvement for the 0.1X TAE gel. All the nucleic acids stains mentioned above were tested in conjunction with the different types of gels in trial. Further to the 1XTAE gel a T (A) E gel was also tested. Pico Green staining in such a gel has shown better resolution of bands than Vistra Green. Best results were observed when using a 1XTAE gel

and 1XTAE buffer for the electrophoresis. Vistra Green was again found to be the best choice of nucleic acid stain, as product yield continued to remain low.

The pH of the gels was tested to determine the effect this parameter would have on band resolution. A 8.0pH TAE gel with a high ionic strength and a 8.5pH gel with the usual ionic strength were tested. The electrophoresis buffer contained the same amount of TAE at the same ionic strength as the gel electrophoresed. The pH was measured in all cases at the beginning and at the end of electrophoresis to determine the changes in H^+ concentration that take place during electrophoresis. The 8.5-pH gel at the cathode had a reading of 8.5-pH prior to electrophoresis changing to an 8.0-pH post-electrophoresis, whereas the anode pH was maintained throughout electrophoresis. Better band resolution was obtained for the higher ionic strength of the 8.0-pH gel than for the 8.5-pH one. Heterozygous samples were easily identified through the formation of heteroduplexes clearly visible even in the short run of the MADGE system. Therefore the identification of the genotypes proved straight forward, consisting in recognising a simple pattern of single or double band present in the tracks.

As the small size difference between a DD band (129 bp) and the heavy band of the II of 90bp (as the 29 bp band runs off the short track of the MADGE system) is indistinguishable with the naked eye, differentiation between II and DD genotypes could not be achieved by mobility retardation. Using one of the two genotypes as a generator and hybridising it with the opposite genotype should, in theory, produce a heteroduplex. The most commonly observed genotype, in the case of our control samples and as reported by Oksanen *et al.*, was the DD genotype. By using the DD genotype, as our 'generator' only a small number of heteroduplexes would be generated and therefore the genotyping would be achieved at a very fast rate (the I allele was reported to have a very small frequency and therefore the infrequent homozygous for this allele would generate a small number of heteroduplexes).

The denatured mix, which was allowed to re-nature at room temperature post denaturation, when run on the MADGE system, revealed the heteroduplexes formed by hybridising the DD with the II samples. Furthermore the initial DI heteroduplexes appeared distinguishably fainter than the newly generated ones, allowing a fast recognition of the II 'generated' heteroduplexes. (Fig. 17 and 18)

An initial electrophoretic run identified the naturally occurring heteroduplexes after which the PCR product denatured in the presence of the generator was run on a separate gel, and the newly formed heteroduplexes were scored as the rare II genotype.

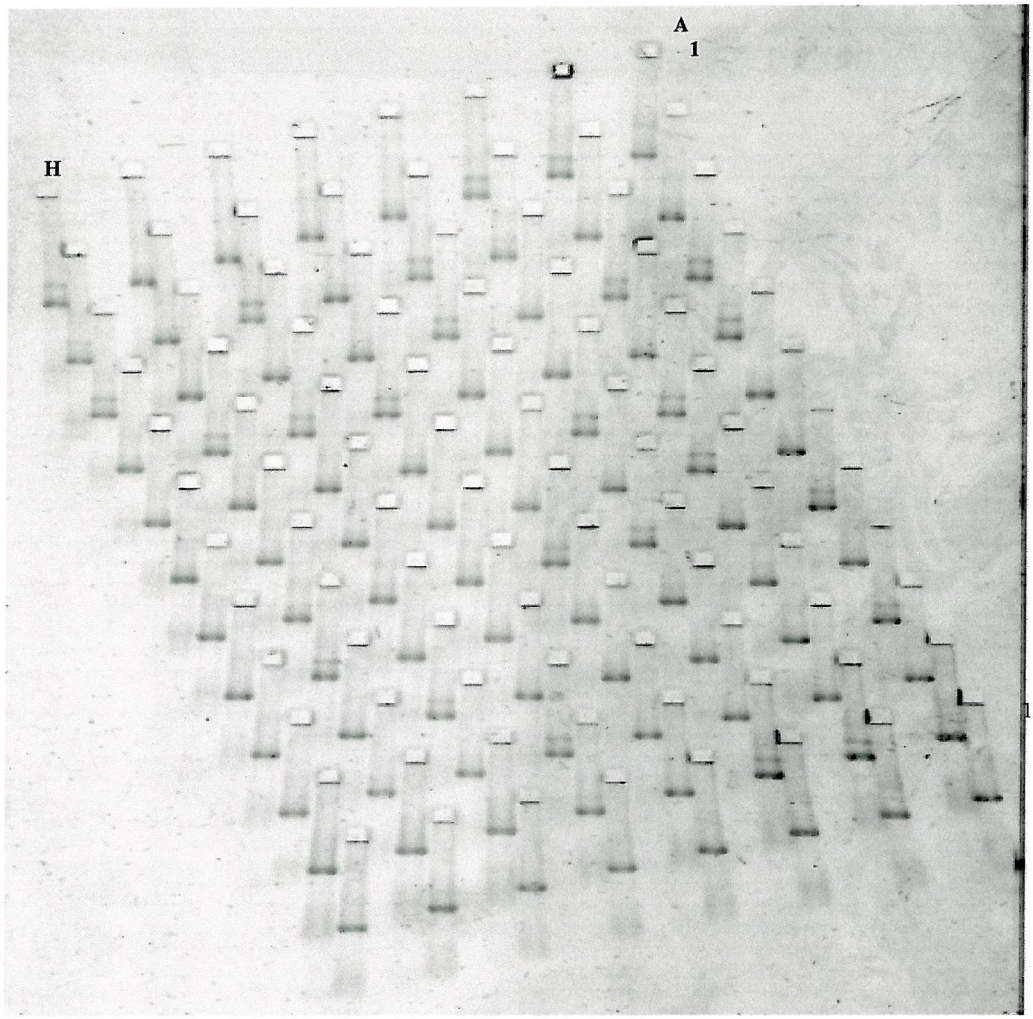


Fig. 18 Molecular hybridisation techniques were used to identify the rare II genotypes by adding the 'generator' (the DD genotype) to the PCR products. In this 'generator' step, the II genotypes will appear as heteroduplexes, whereas the DI heteroduplexes (already scored) will appear less visible due to excess of DD genotype in the mix.

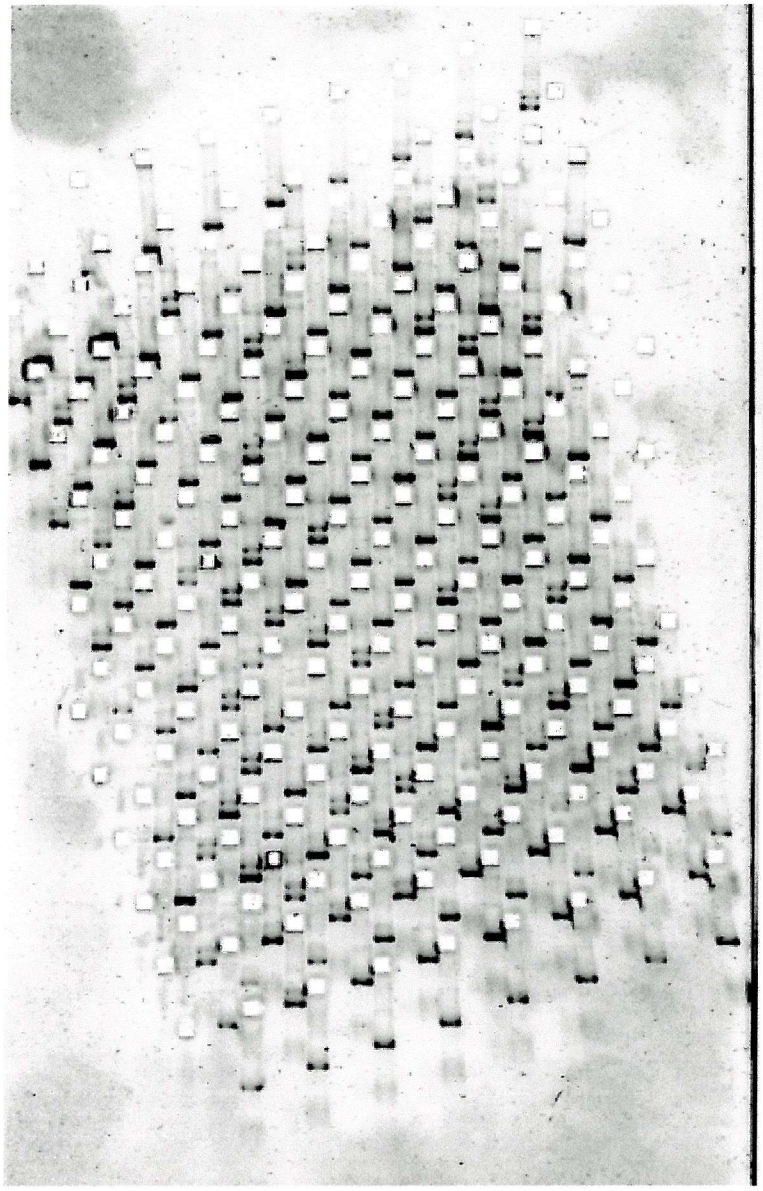


Fig. 19 The 192 well gel as means of fast genotyping. A fast way of identifying all genotypes on one gel is the 192 format gel. The raw PCR product is loaded in the first 96 well array, and the hybridised product is loaded in the second 96-well array. Instant genotyping for all genotypes using the generator technique of all 96 samples is achieved using Phoretics software. The first well on the right (top) represents the positive control for the heteroduplex generator where the II genotype has been mixed with the DD generator.

Alternatively a higher throughput method for genotyping is the 192-well format MADGE, where the PCR product is loaded in parallel with the hybridised samples, thus obtaining an instant result for all genotypes. (Fig. 19) Positive, as well as negative controls, were run on gels. One II individual, identified through RsaI digest, was amplified through PCR, and the amplicon (1/100 dilution) was used as a positive control in the 'generator step' to ensure the formation of heteroduplexes between the unidentified II and the generator DD took place. (Fig. 20 and 21)

One problem encountered during genotyping was the faint appearance of the heteroduplexes as well a distortion of the bands into 'dumbell' shapes and skiing alongside the track (Fig. 3). W1 and DMSO were tested, for the improvement of the yield and quality of the amplification, with no effect.

The faint appearance of the bands was presumed to be either a problem caused by pre-PCR contamination of samples with the DD generator, or a time vs. heteroduplex stability problem.

The stability of the generated heteroduplexes was tested over a period of time. Samples previously identified as II and DD genotypes respectively, were mixed (see generator step) and heteroduplexes were generated. Heteroduplex stability was tested for a 5 to 120 minute time interval, and then again at 24 hour, and 48-hour interval. The analysis of the intensity of the two heteroduplex bands revealed no significant changes over time. (HDx band relative to the 90 bp band, Fig. 22 and Fig. 23) Even after a time frame of four days the results have shown that heteroduplex stability, in the case of the 5 bp insertion, was not dependent of time. (Fig. 24, 25 and 26)

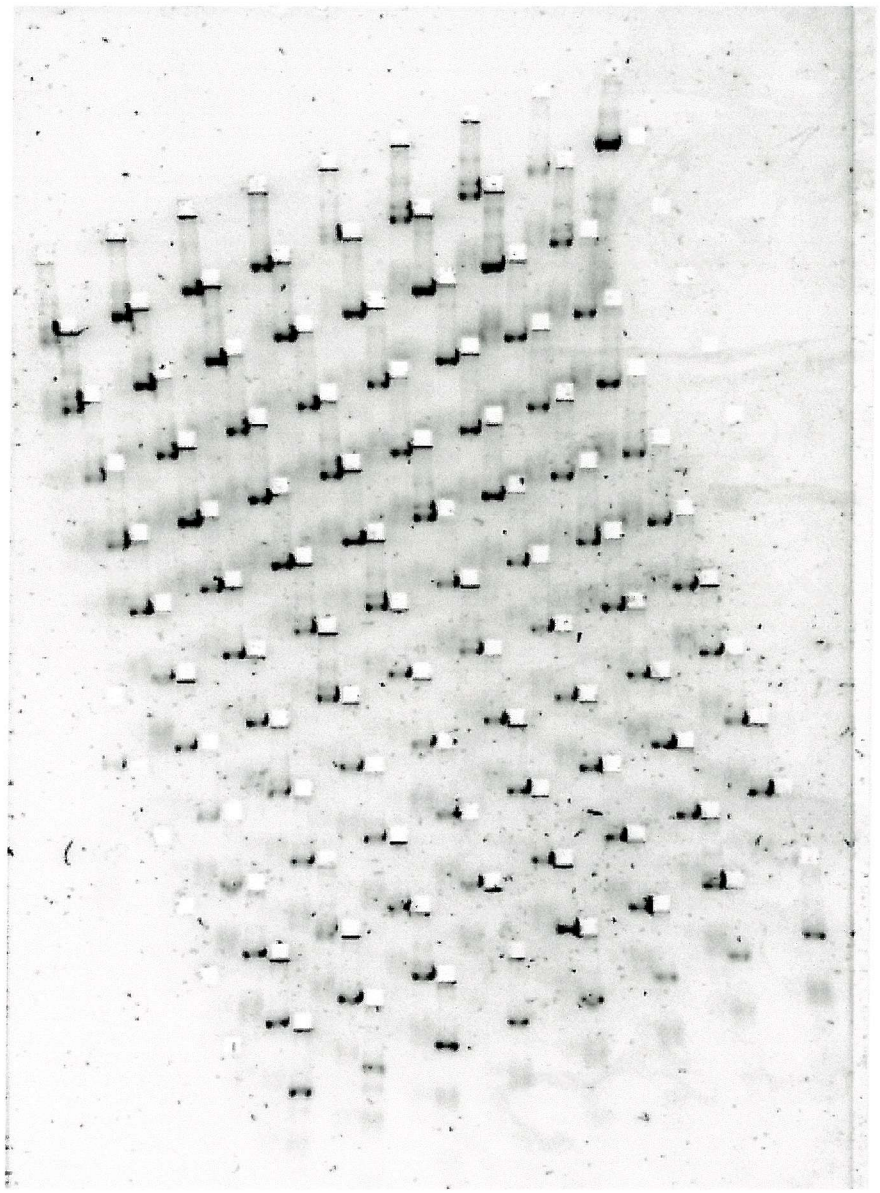


Fig. 20 Example of OB-R heteroduplex analysis on MADGE. Samples were loaded straight after PCR amplification then run on MADGE system. After electrophoresis and identification of the natural heteroduplexes (DI) the samples are subjected to the generator step. Top right side well represents the II sample (positive control for the formation of heteroduplexes after the generator step), which was used for two controls: 1.Checking that the suspected contamination with DD did not interfere with the genotyping; and 2. Positive control for the generator step.

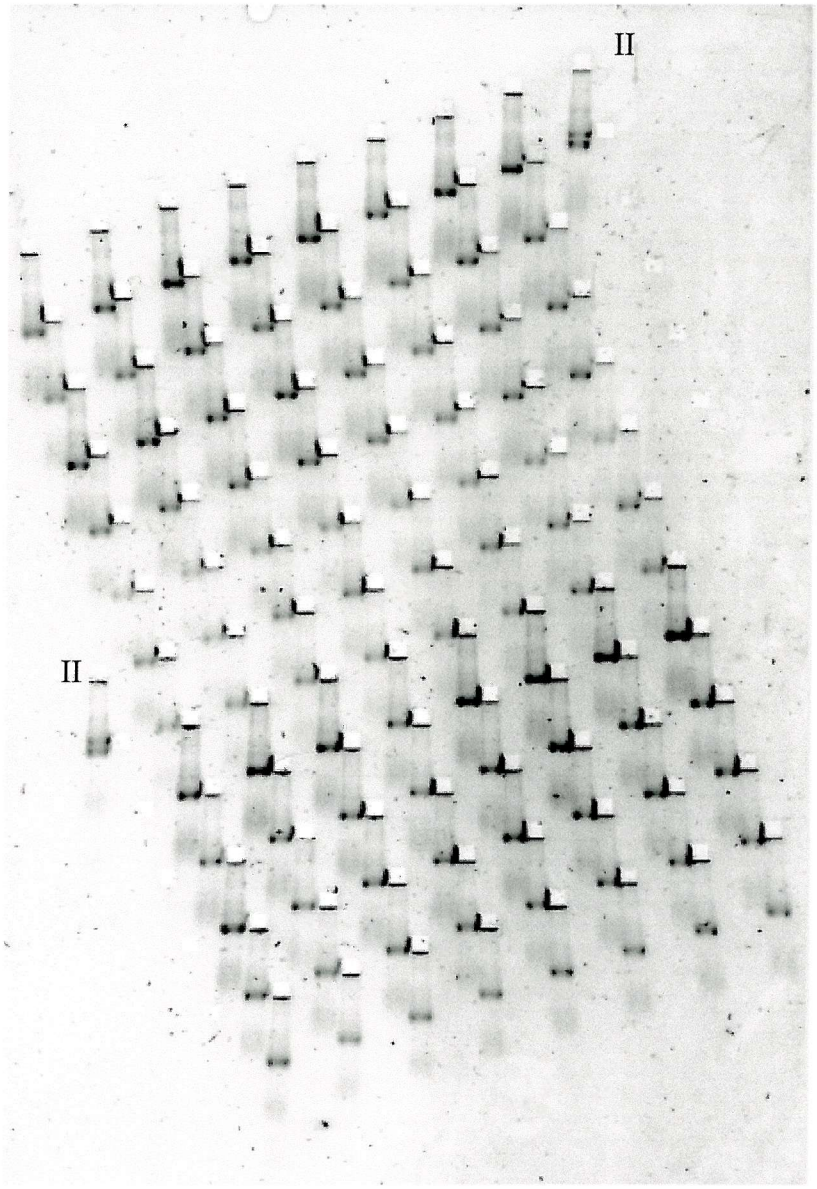


Fig. 21 Example of identifying the rare genotypes by generating new heteroduplexes. Gel represents the same samples from Fig. 20, after the addition of the 'generator' (DD genotype). The II samples used as a positive control (II) now appear as a heteroduplexes.

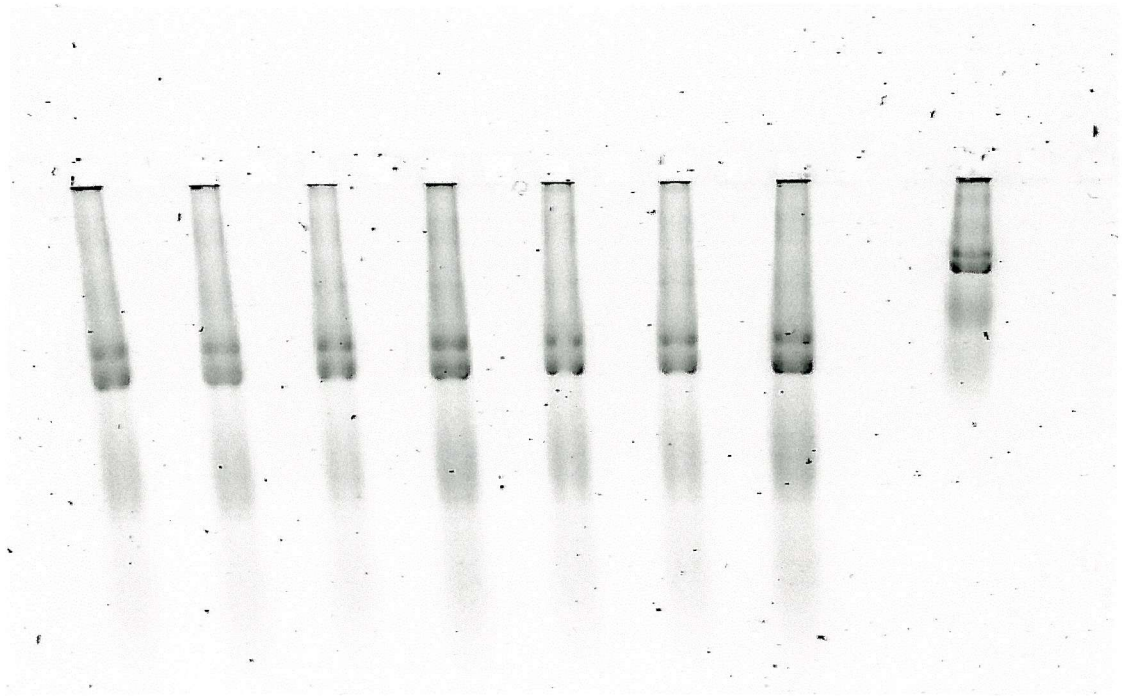


Fig. 22 Determination of heteroduplex stability. The heteroduplex formed by the generator step remain stable at room temperature for a long time. From left to right of the image, samples were loaded at 0, 5, 10, 20, 30, 40, 50 and 60 minutes respectively. No visible change in was observed. The difference in mobility between samples 1 to 7 and sample 8 is due to sample 8 being loaded several minutes after electrophoresis was started.

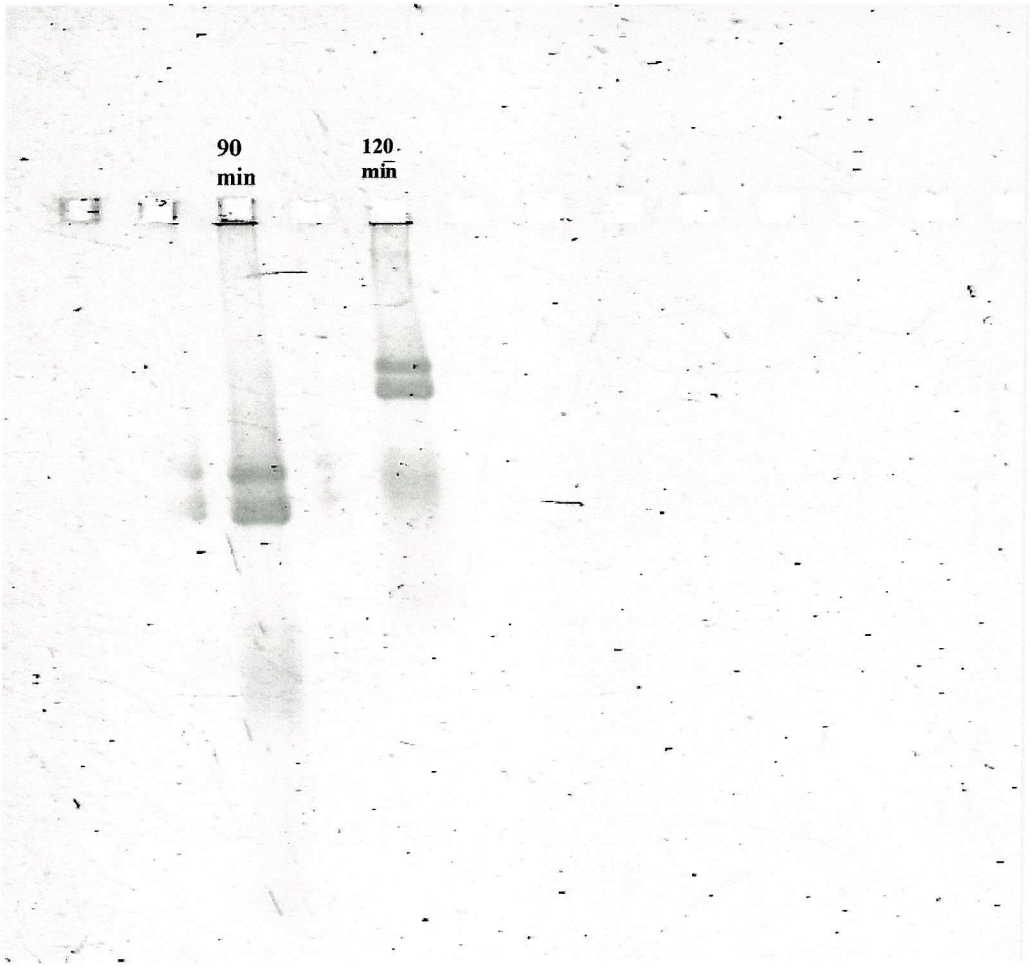


Fig.23 Heteroduplex stability after 90 and 120 minutes. The heteroduplex stability experiment was assessed for longer periods of time at room temperature. Figure shows the heteroduplex after 90 and 120 minutes standing time at room temperature. The visible difference in mobility is attributed again to loading the samples at different times during electrophoresis and thus different running times on the gel. No difference was observed.

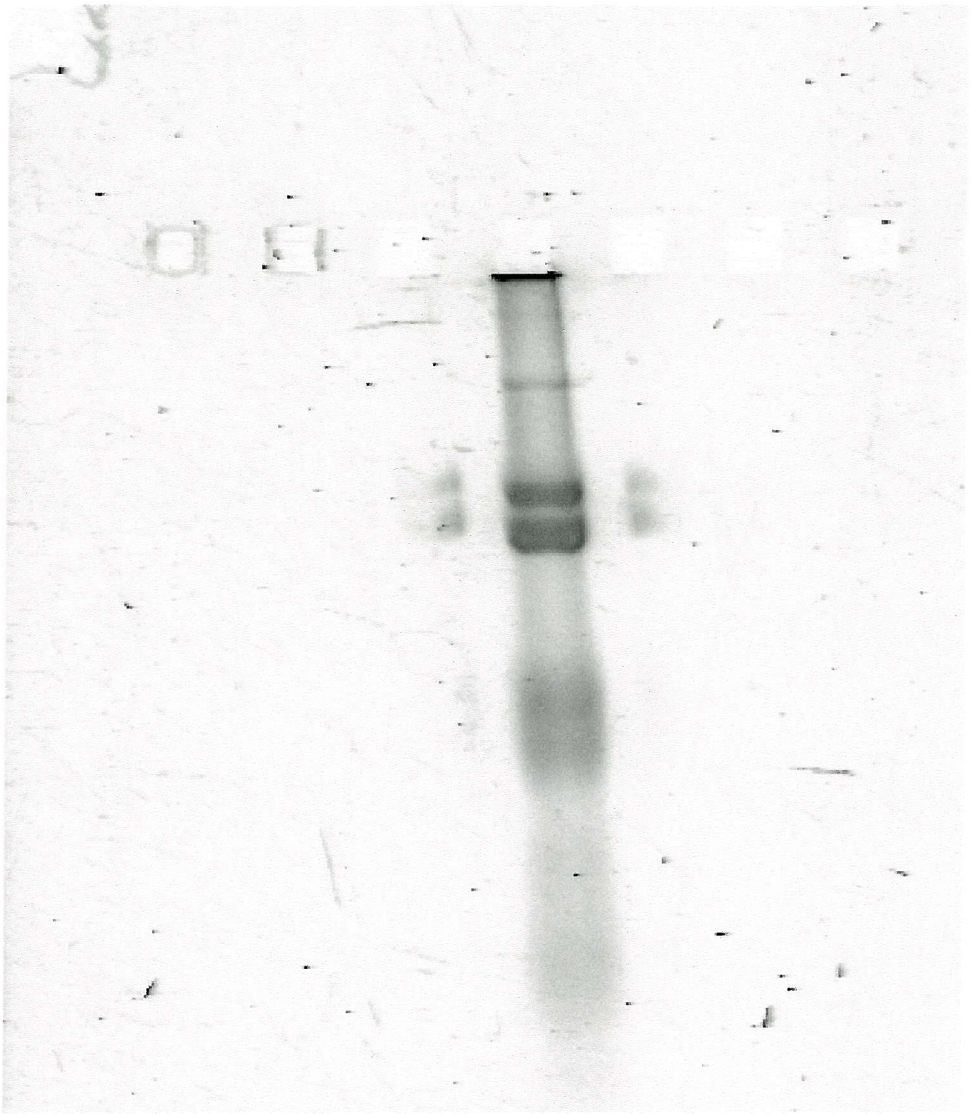


Fig.24 Heteroduplex stability after 2 days. Heteroduplex stability was again tested after 2 days. The heteroduplex stability does not seem affected by time. The two bands are still clearly and equally visible.

Fig.25 Heteroduplex stability after 3 days. Heteroduplex stability, at room temperature, was tested after 3 days. Bands appear more faint, although the heteroduplex is still present. The two bands of the heteroduplex appear equally represented.

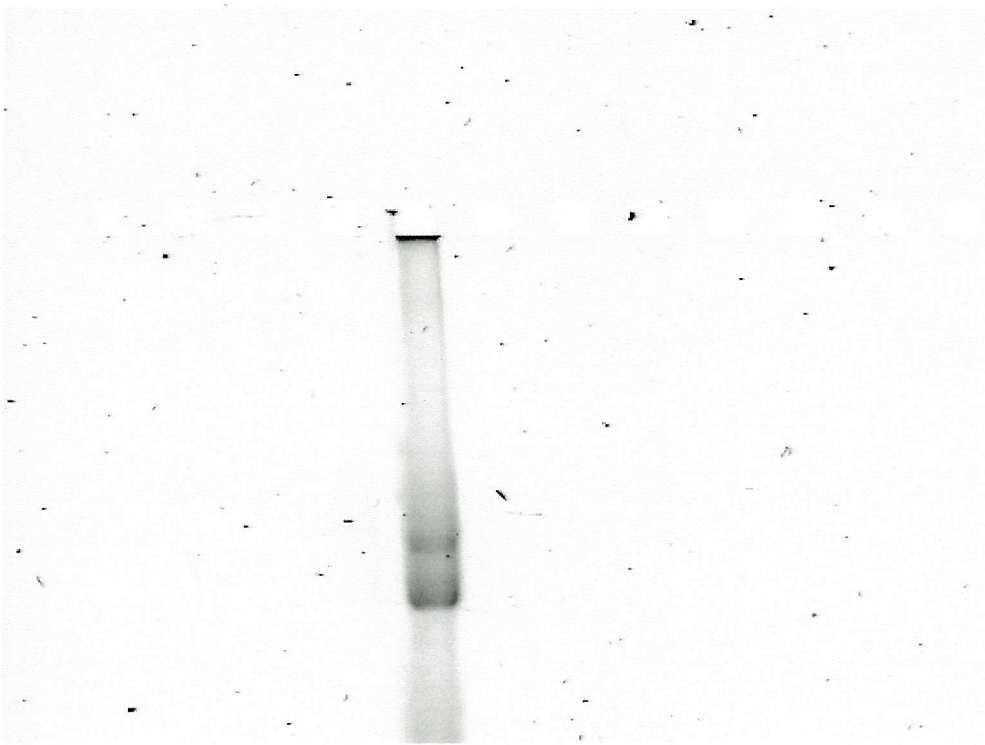


Fig. 26 Heteroduplex stability after 4 days. Heteroduplex stability was again tested after 4 days at room temperature. Bands appear more faint although the two bands are still equally represented. The heteroduplex stability was not influenced by time.

The contamination theory was tested using a probability generator for strand re-annealing, which is akin to the Hardy-Weinberg Equation in its terms, where the upper band of the heteroduplex represents the 2DI and the lower band represents the 1-2DI. (Fig. 27)

The idea of strand re-annealing in a heterozygotes is explained here. If one supposes that there are D deletion strands and D' of complementary deletion strands and I of insertion strands and I' of complementary insertion strands and that total yield is:

$$D^2+2DI+I^2=1$$

therefore:

$$Hx/Ho=2DI/1-2DI$$

Then during random re-annealing the probability generation is

$$(D+I)(D'+I')=DD'+II'+DI'+D'I,$$

which equals at total of 1.

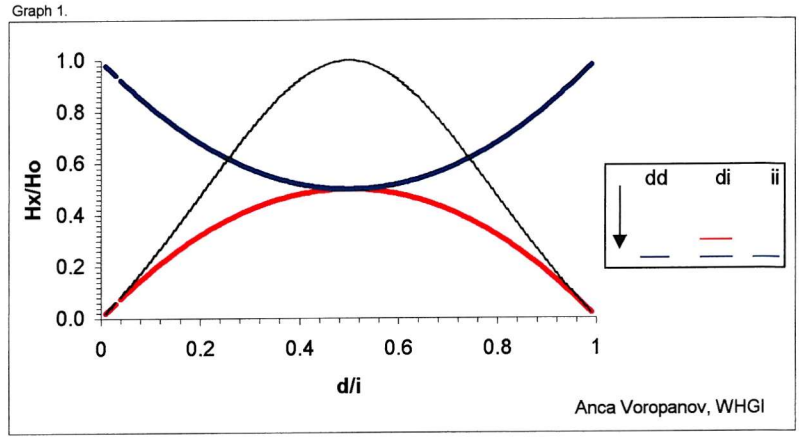
If ***DI'*** and ***D'I*** co-run and ***DD'*** and ***II'*** effectively co-run (due to very small size difference) the ***Hx/Ho= 2DI'/ I-2ID'*** (if ***D=D'*** and ***I=I'***, which is expected during any standard PCR amplification of DNA duplexes. According to this theory, very small amounts of DD contamination in the PCR mix can render the heteroduplex invisible. Therefore the faint appearance of the heteroduplexes could have been the result of a pre-PCR contamination problem. (Fig.28) The theoretical model was tested experimentally. (Fig. 29) II and DD PCR product was mixed in various ratios, from 1:9 (v/v) to 9:1 (v/v) II to DD respectively, and it was observed that the two bands of the heteroduplex became more equally represented as the amounts of DD and II in the mix balance each other.

Attempts to identify the source of contamination were unsuccessful. The reagents were tested for contamination source but the results were inconclusive. The contamination source seems to be present even when all reagents, laboratory apparatus and even laboratories were changed. The contamination did not seem to interfere with the identification of the rare II genotype.

The heteroduplex story
The effects of contamination with the DD generator
A theoretical model

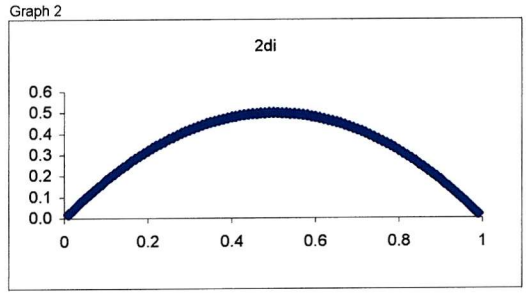
d	2di	1-2di	2di/1-2di
0.01	0.0	1.0	0.0
0.02	0.0	1.0	0.0
0.03	0.1	0.9	0.1
0.04	0.1	0.9	0.1
0.05	0.1	0.9	0.1
0.06	0.1	0.9	0.1
0.07	0.1	0.9	0.1
0.08	0.1	0.9	0.2
0.09	0.2	0.8	0.2
0.1	0.2	0.8	0.2
0.11	0.2	0.8	0.2
0.12	0.2	0.8	0.3
0.13	0.2	0.8	0.3
0.14	0.2	0.8	0.3
0.15	0.3	0.7	0.3
0.16	0.3	0.7	0.4
0.17	0.3	0.7	0.4
0.18	0.3	0.7	0.4
0.19	0.3	0.7	0.4
0.2	0.3	0.7	0.5
0.21	0.3	0.7	0.5
0.22	0.3	0.7	0.5
0.23	0.4	0.6	0.5
0.24	0.4	0.6	0.6
0.25	0.4	0.6	0.6
0.26	0.4	0.6	0.6
0.27	0.4	0.6	0.7
0.28	0.4	0.6	0.7
0.29	0.4	0.6	0.7
0.3	0.4	0.6	0.7
0.31	0.4	0.6	0.7
0.32	0.4	0.6	0.8
0.33	0.4	0.6	0.8
0.34	0.4	0.6	0.8
0.35	0.5	0.5	0.8
0.36	0.5	0.5	0.9
0.37	0.5	0.5	0.9
0.38	0.5	0.5	0.9
0.39	0.5	0.5	0.9
0.4	0.5	0.5	0.9
0.41	0.5	0.5	0.9
0.42	0.5	0.5	1.0
0.43	0.5	0.5	1.0
0.44	0.5	0.5	1.0
0.45	0.5	0.5	1.0
0.46	0.5	0.5	1.0
0.47	0.5	0.5	1.0
0.48	0.5	0.5	1.0
0.49	0.5	0.5	1.0
0.5	0.5	0.5	1.0
0.51	0.5	0.5	1.0
0.52	0.5	0.5	1.0
0.53	0.5	0.5	1.0
0.54	0.5	0.5	1.0
0.55	0.5	0.5	1.0
0.56	0.5	0.5	1.0
0.57	0.5	0.5	1.0
0.58	0.5	0.5	1.0
0.59	0.5	0.5	0.9
0.6	0.5	0.5	0.9
0.61	0.5	0.5	0.9
0.62	0.5	0.5	0.9
0.63	0.5	0.5	0.9
0.64	0.5	0.5	0.9
0.65	0.5	0.5	0.8
0.66	0.4	0.6	0.8
0.67	0.4	0.6	0.8
0.68	0.4	0.6	0.8
0.69	0.4	0.6	0.7
0.7	0.4	0.6	0.7
0.71	0.4	0.6	0.7
0.72	0.4	0.6	0.7
0.73	0.4	0.6	0.7
0.74	0.4	0.6	0.6
0.75	0.4	0.6	0.6
0.76	0.4	0.6	0.6
0.77	0.4	0.6	0.5
0.78	0.3	0.7	0.5
0.79	0.3	0.7	0.5
0.8	0.3	0.7	0.5
0.81	0.3	0.7	0.4
0.82	0.3	0.7	0.4
0.83	0.3	0.7	0.4
0.84	0.3	0.7	0.4
0.85	0.3	0.7	0.3
0.86	0.2	0.8	0.3
0.87	0.2	0.8	0.3
0.88	0.2	0.8	0.3
0.89	0.2	0.8	0.2
0.9	0.2	0.8	0.2
0.91	0.2	0.8	0.2
0.92	0.1	0.9	0.2
0.93	0.1	0.9	0.1
0.94	0.1	0.9	0.1
0.95	0.1	0.9	0.1
0.96	0.1	0.9	0.1
0.97	0.1	0.9	0.1
0.98	0.0	1.0	0.0
0.99	0.0	1.0	0.0

Graph 1.



The equall representation of the heteroduplex bands depends on the amount of DD or II in the reaction mix.
The heteroduplex bands will therefore render themselves invisible if either of the two genotypes present in reaction are in exc

Graph 2



The representation of the upper band of the heteroduplex (here denoted 2DI)
is dependent of the amounts of D and I present relative to eachother.
An equal amount of D and I (0.5 on the graph) gives best representation of the heteroduplex bands.

Fig.27 A theoretical model, based on the Hardy_Weinberg Equation, was designed to asses the effects of contamination with the DD genotype, on the heteroduplex band representation.

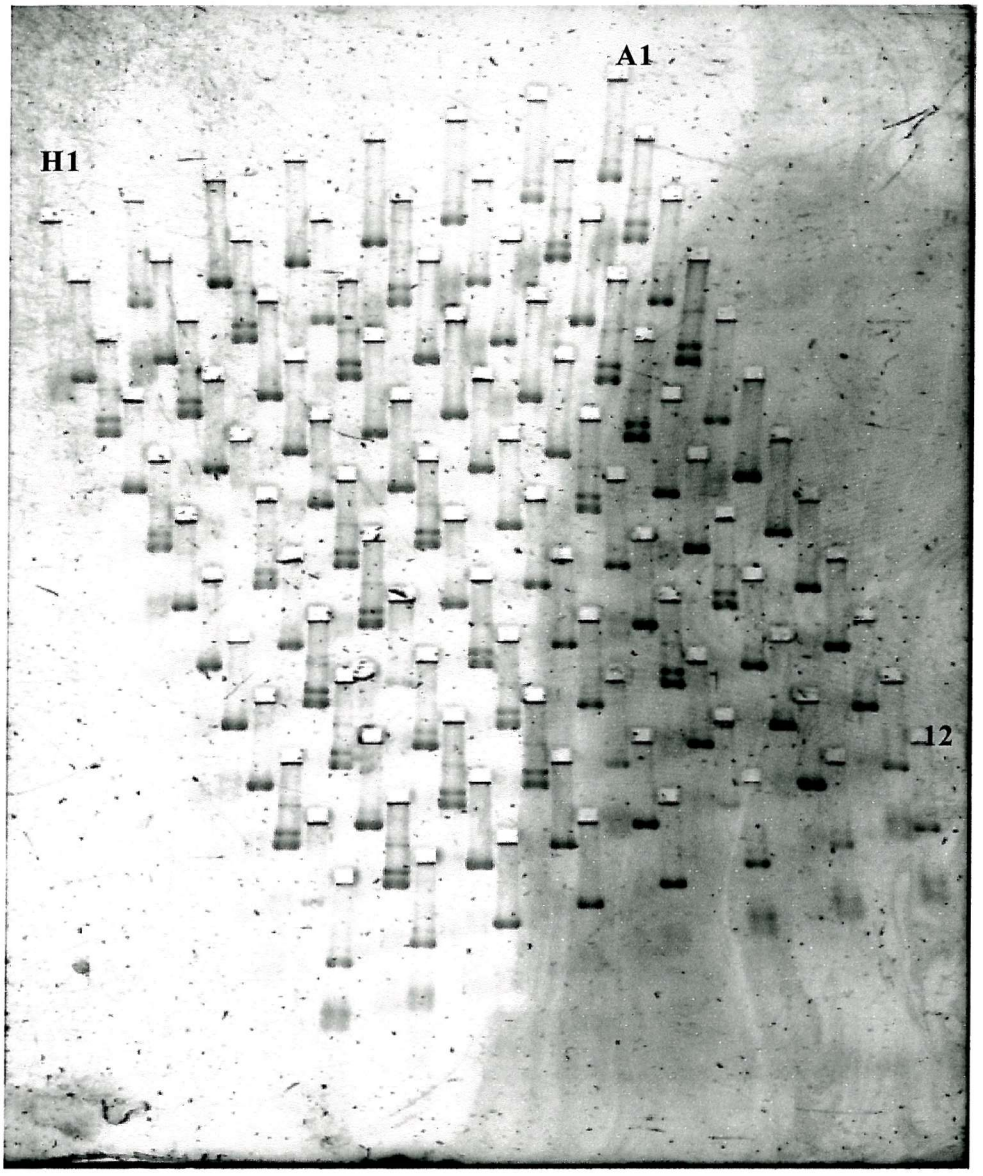


Fig. 28 The faint appearance of the heteroduplex was due to 'generator' contamination. The effects of contamination with the 'generator' genotype over the heteroduplex band representation. Natural heteroduplexes can be detected easily (A2, A4, B2, B4, B5, B8, C5, C8, D2, E3 ,E6, E8, E9, E10, F2, F6, F7, F10, G3, G6, G8, G11, H3, H10). Some heteroduplexes appear faint (A1, B1, B2, F9, G5, G9, H5). This was due to the contamination with the DD genotype (generator genotype).

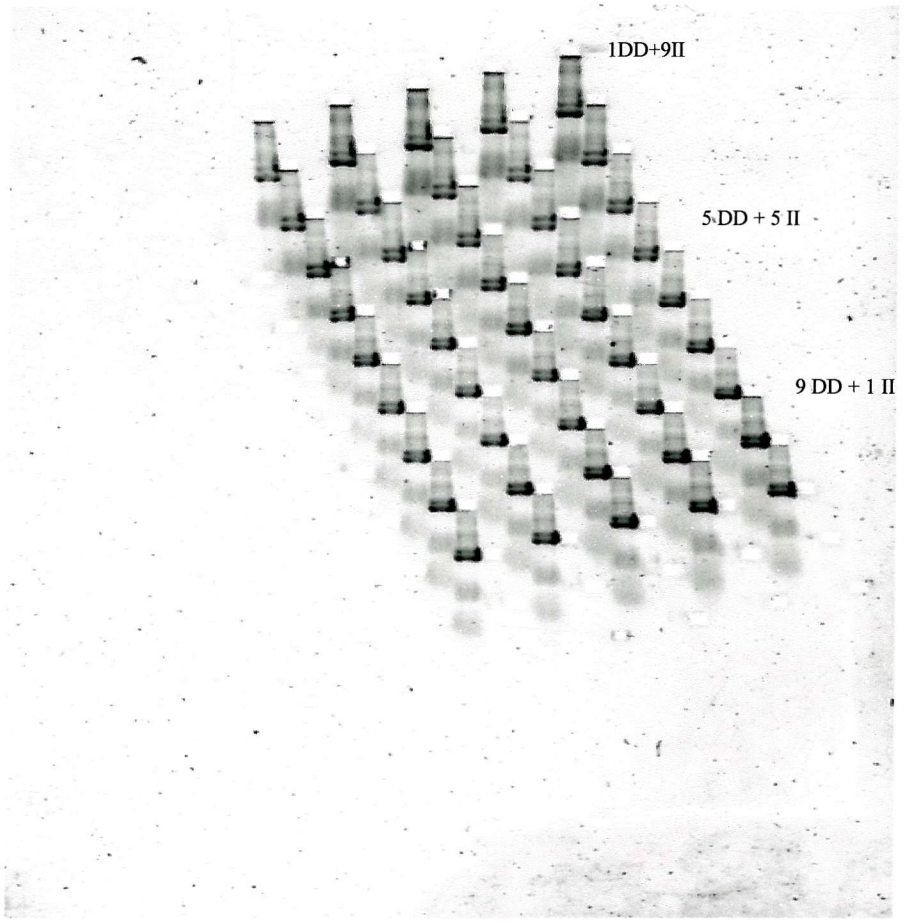


Fig. 29 Equal representation of the two bands of the heteroduplex is dependent on the amount of DD, relative to II, in the sample. This is to say that the heteroduplex bands representation is dependent on contamination (with either homozygous genotype). Row 1 represents 1 DD to 9 II (v/v) going up in DD amount to row 9, which represents 9 DD to 1 II (v/v). The bands are equally represented where equal volumes of DD and II are present in the mix.

4.1.1 Elimination of dumbell bands during MADGE electrophoresis

Ionic imbalances between the sample composition and gel/ electrophoresis buffer composition were determined to have an effect on the mobility, as well as on the appearance, of the bands. The 0.1 X TAE, 1X TAE and no TAE gels, accompanied by the 0.1 and 1 X TAE electrophoresis buffer respectively, were observed to have influenced resolution of bands in the MADGE system. Changing the ionic composition strength of the loaded sample to match that of the gel was also examined. The pH variation seen with the 8.5-pH gel was an indication that exchange of ions does take place during electrophoresis, thus indicating that 'equalising ' ions between the electrophoretic elements (gel, sample, buffer) would improve band distortion.

The time delayed, together with the sample post-PCR modification, was shown to be successful in eliminating the 'dumbelling' and skiing alongside electrophoretic tracks (Fig. 30) allowing fast and efficient genotyping. (1)

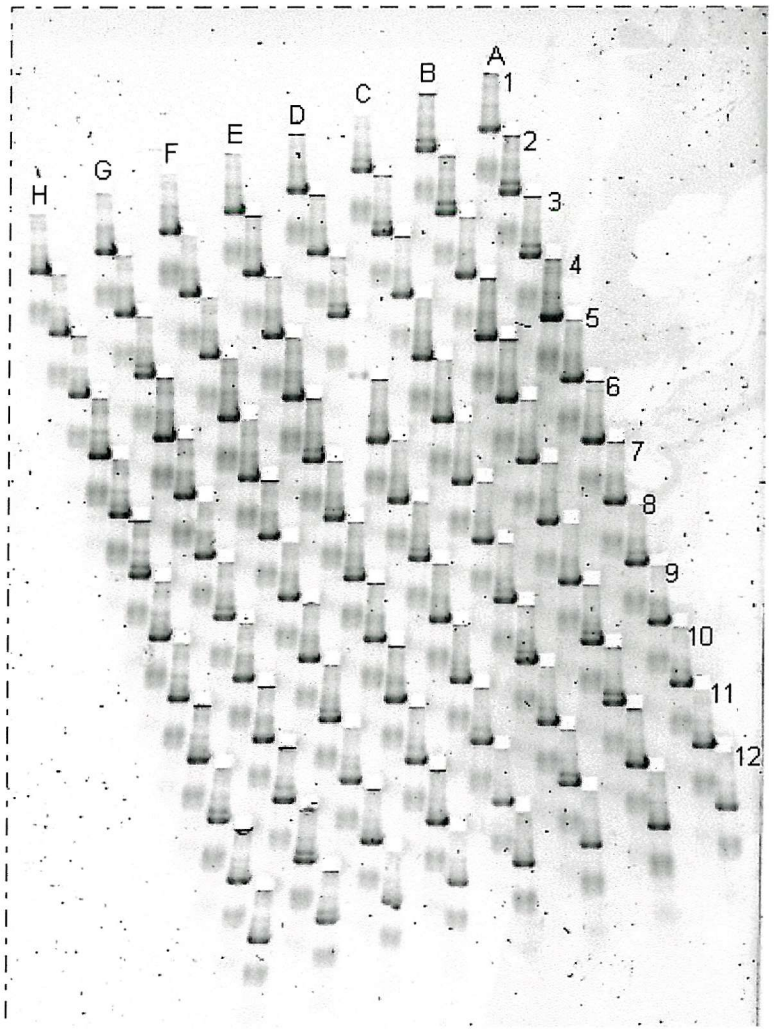


Fig 30 MADGE with delayed start electrophoresis. The bands are free of any distortion and the 'skiing' effect has been eliminated. Tracks with a heteroduplex bands, are: B1, C1, A2, B2, C2, A3, D3, G3, E5, E6, H6, D7, E7, G7, C8, B9, C9, G9, B10, H10, C11, G11, AND G12. These are clearly identifiable.

4.1.2 Consideration of the advances brought into laboratory process.

Heteroduplex generator: a general strategy for analysis of small insertion/deletion polymorphisms.

One of the major objectives of the field of human genetics is the identification of the genetic factors that predispose to complex human diseases such as coronary heart disease, obesity, osteoporosis, hypertension, and diabetes. The renewed and extended interest in genome polymorphism signifies a development in human genetics research that will have a major impact upon population genetics, drug development, forensics and genetic disease. Although insertion/deletions are quite frequently found in the human genome, there is no method to date that enables the screening for such polymorphisms unless the insertion of the fragment creates a restriction site. As molecular-genetic epidemiology studies engage the analysis of a large number of samples, the development of a new, fast and effective technique for efficiently scanning thousands of samples each day was identified as a necessity. The method developed in this project for examining I/D polymorphisms, is fast, easy, and is accessible to small and large laboratories.

Insertion/deletion (I/D) polymorphisms are more easily distinguished from sequencing errors in the analysis of overlapping sequences. Heteroduplex analysis is an easy way of detecting such polymorphisms, where the two alleles can be differentiated through various methods of hybridisation. I/D are easy to scan and have superior discrimination by hybridisation compared to SNPs and are frequently found polymorphisms in the human genome.

Whole genome insertion/deletion scanning can be applied to mapping in linkage disequilibrium, to confirmation of family structure, for the determination of haplotype risk in various populations, and to the detection of sub-microscopic insertion/deletion duplications in the genome. They offer superior discrimination by hybridisation compared with SNPs.

The method developed enables the screening of a large number of samples for a particular insertion/deletion polymorphic site by making use of microplate array diagonal gel electrophoresis (MADGE), a system created for molecular genetic epidemiological studies. Microplate array diagonal gel electrophoresis (MADGE) is fully compatible with microplates and allows the running of several hundred thousand samples a day. The development of this quick, efficient and inexpensive method, where all the genotypes present with the I/D

polymorphisms are detected, and anchoring the technique to the MADGE system is advantageous for a genetic epidemiological study. Using various hybridisation techniques, actually only used DD hybridisation, II, DD and ID genotypes can be distinguished, with a minimal expenditure of both human and laboratory resources.

The principle of the technique combines the two homozygous genotypes, melts them together and allows them to re-anneal so they hybridise and form a heteroduplex. Indeed I have shown that mixing the DD and II amplicons, melting them at 95°C, and allowing them to re-anneal (after instant cooling on ice), leads to the formation of heteroduplexes. Genotyping is quick, especially after the human eye has become accustomed to the recognition of the band pattern.

I have shown that the generated heteroduplexes are stable in time at room temperature for over four days. (Fig. 22-26) This is probably due to the small size of the insert, 5 base pairs, and instability of the heteroduplex should be expected to decline with the increase of the size of the insert.

The repeated amplification of the 'generator' (which may be any of the two homozygous genotypes but not the heterozygous one) ensures a virtually continuous resource of material for the genotyping of thousands of samples. The production of generator is cheap, easily obtainable, and it does not exploit the original genomic template beyond recovery. However, the risk of contamination increases as several amplifications of the generator are achieved. For this a specific pre-PCR room is to be preferred to avoid possible contamination of the samples with the 'generator'. A theoretical model, which estimates the effects of contamination on the illustration of the heteroduplex, is shown in Fig.27. Equal representation of the two bands of the heteroduplex was determined to be dependent on the amounts of DD and II genotypes in the reaction mix. The same observation was made with an experimental model where the heteroduplex bands representation was seen to depend on the amounts of DD and II in the reaction. (Fig.29)

4.1.3 Effects of the DD contamination:

As pre-PCR contamination with the generator was believed to be the cause of the faint appearance of heteroduplexes, a probability generator was designed. The Hardy-Weinberg Equilibrium equation was used to predict strand duplex yields and estimate the effect contamination, with the generator prior to PCR amplification, would have on the resolution of the heteroduplexes. Small amounts of DD present in the PCR mix, prior to amplification of the template, can render the heteroduplex invisible. The upper band of the heteroduplex becomes less visible as the amount of DD contamination increases (Fig. 27).

An experimental model was designed to observe the changes in the heteroduplex band representation, with increasing amounts of contaminant of both types of possible 'generator' (i.e. either II or DD could have been selected as the generator, DD locates rare II's whereas II would locate DD, i.e. most of the plate). The chosen heteroduplex generator in this case was DD that allowed the fastest genotyping. A titration of DD relative to II was performed experimentally and it was observed that the two bands of the heteroduplex became more equally represented on the gel as the amount of DD and II in the hybridising mix approach (volumetric) equality. Another reason for the faint appearance of the heteroduplexes could have been that the II samples on meeting the DD 'contaminant' will hybridise and form a heteroduplex. This heteroduplex would appear more visible on the gel as the amount of contaminant increases to a concentration closer to that of the template to be amplified.

Gel electrophoresis on a MADGE format has also presented some electrophoretic problems. The distortion of the bands into 'dumbell' shapes as well as skiing alongside the electrophoretic track reduced the ability, in some cases, to easily and quickly distinguish between heteroduplexes and homoduplexes (Fig. 3).

The distortion of bands and skiing alongside the electrophoretic tracks was resolved using a combination of delayed start electrophoresis together with a slight change in the ionic make-up of sample composition pre-electrophoresis. The electrophoretic artefacts were reduced to an acceptable minimum, which allowed complete confidence in genotyping in the short electrophoresis tracks of the MADGE system. The complete method was discussed earlier in this chapter. (1)

4.2 Insertion deletion polymorphism in the OB-RB gene:

Initially the OBR 5'UTR I/D polymorphism was identified using Oksanen's restriction enzyme digest (90), and used these samples as positive controls in the development of this method.

The restriction digest produces four bands in total: the heteroduplex band, 114 bp, 90 bp, and 29 bp band (Fig. 16).

The identification of the heterozygous individuals became a matter of pattern recognition and therefore was a quick and efficient way of genotyping. However, the problem of identifying and differentiating between the two homozygous genotypes (II and DD) remained to be resolved. The molecular hybridisation idea exploits the annealing ability of the generator DNA strand and a complementary target DNA strand to form a heteroduplex. The target DNA and the generator were melted together, at a temperature that disrupts the hydrogen bonds that holds the two complementary DNA strands together. The re-annealing of the DNA strands is then achieved under controlled conditions. The insertion/insertion genotype denatured and re-annealed in the presence of the opposite (deletion/deletion) genotype formed 'generated' heteroduplexes. The methods available to date, for scanning I/D polymorphisms, rely on the presence or the introduction of a restriction site and the identification of genotypes by restriction digests. Such methods are laborious, expensive, time consuming and do not permit scanning of several thousand samples a day.

In the new developed method of scanning I/D, using as a generator the most commonly observed genotype, the 'generated' heteroduplexes would represent the most rare genotype and thus the number of newly formed heteroduplexes would be kept to a minimum and therefore genotyping would be achieved fast.

Simply running the PCR product on the MADGE identified the naturally occurring heteroduplexes representing the heterozygous individuals. After the 'generator' step the identification of the rarest genotype is achieved quickly through another electrophoretic run. The generator amplification, achieved in a totally separate PCR step was designed to produce virtually unlimited amounts of DD product. Repeated DNA amplification however increases the chance of contamination.

4.2.1 Genotyping the Leptin receptor gene in the NPHSII Study:

The genotypes found in 1367 (1500 tested with 133 drop-outs) samples tested were not in Hardy Weinberg Equilibrium (chi-squared test, $p<0.0001$). The results of the Hardy Weinberg Equilibrium test are shown in Fig.31, 32, and 33.

Due to the low number of rare genotype II found in the entire sample tested, it was expected that the genotype frequencies found would not be the same as the expected frequencies under the Hardy-Weinberg Equilibrium. The possible reasons for this have been discussed in section 4.1.3, but the departure affects small numbers and should not substantially affect the phenotypic associations tested except to diminish possible distinctions between genotype groups slightly.

Analysis of baseline weight in relation to the OB-R genotypes in the NPHS II study showed a slight increase in the geometric mean between the DD and the II genotypes. However, the difference was not of statistical significance. The DD and DI individuals did not show any difference in weight geometric mean.

A similar trend was observed for the BMI vs. OB-R genotypes but again no statistical significance. (Fig. 32 and 33)

Stats for a 2-allele polymorphism

poly=OBR I/D

Observed

	OBR I/D	OBR I/D	OBR I/D	
Genotypes	11	12	22	Total
Frequency	908	347	5	1287

p	**q**	**p+q**
0.85	0.15	1

Expected

	OBR I/D	OBR I/D	OBR I/D	
Genotypes	11	12	22	Total
Frequency	931.6	326.7	28.6	1287

Chi^2 test and level of significance

	O	E	(o-e)^2/e	Chi^2	Sig level
11	908	931.6	0.6		
12	347	326.7	6.8		
22	5	28.6	19.5		
				27.0	2.08E-07
					Sig at 1df

Fig.31Hardy-Weinberg Equilibrium for
total eligible men in the NPHSII study.

Stats for a 2-allele polymorphism

poly=OBR I/D

Observed

	OBR I/D	OBR I/D	OBR I/D	
Genotypes	11	12	22	Total
Frequency	955	404	8	1367

p	**q**	**p+q**
0.85	0.15	1

Expected

	OBR I/D	OBR I/D	OBR I/D	
Genotypes	11	12	22	Total
Frequency	979.3	355.5	32.3	1367

Chi^2 test and level of significance

	O	E	(o-e)^2/e	Chi^2	Sig level
11	955	979.3	0.6		
12	404	355.5	6.6		
22	8	32.3	18.2		
				25.5	4.50E-07
					Sig at 1df

Fig.32 Hardy-Weinberg Equilibrium for total men in the NPHSII study.

Genotypes of Leptin Receptor Gene (OR-R ins/del polymorphism)

Total Men on file:

Genotype	Frequency	Percent	Cum
11	955	69.86	69.86
12	404	29.55	99.41
22	8	0.59	100
Total	1367	100	

Eligible European Men:

Genotype	Frequency	Percent	Cum
11	908	70.55	70.55
12	347	29.06	99.61
22	5	0.39	100
Total	1287	100	

These genotypes are not in Hardy-Weinberg Equilibrium, chi-squared test gives p<0.0001.

	Genotype	Frequency	*Geometric mean	95% CI	One way ANOVA p-value
Weight	11	908	80.1	(79.4, 80.9)	
	12	347	80.2	(79.0, 81.4)	0.32
	22	5	88.4	(77.9, 100.2)	
BMI	11	907	26.3	(26.1, 26.5)	
	12	347	26.4	(26.1, 26.8)	0.14
	22	5	29.3	(26.1, 26.8)	

* One way ANOVA was carried out on logged (base e) weight and BMI

Fig.33 Genotype-phenotype analysis for the ins/del polymophism in the OB-R gene. Population sample:NPHSII Study.

4.2.2 Analysis of the Leptin receptor gene (OB-RB) data-NPHS II Study

As discussed earlier, far less rare II genotypes have been found in the NPHS II Study. There are several reasons why the II genotype was found in far lower frequency than expected. Contamination with the DD genotype (due to several generator amplification steps) could influence the equal representation of the II genotype. Where the amount of contamination was high compared to the template concentration the generated heteroduplexes would be masked by a single band representing in fact the DD generator product. The theoretical model constructed for the evaluation of the effects of contamination vs. amount of contamination was described earlier.

Although a tendency towards higher BMI in the II genotype was observed, the lack of strong association between genotype and phenotype in the human leptin receptor gene is consistent with most work in the region so far. Appendix 3 lists a selection of references on the polymorphisms identified in the OB-RB and the lack of association with obesity. So far one association between variation at the receptor gene level and obesity was found in a small sample of 20 Pima Indians, in a case-control study (obese vs. non-obese). (98) An association between plasma insulin levels and the 3'UTR of the leptin receptor gene 5 base pair insertion /deletion was found in 30 morbidly obese patients. (90) Therefore all the associations found so far with the OB-R polymorphisms were all in small and highly controlled samples. No other positive associations between leptin receptor gene and polymorphic loci in the region have been found. All other published data on the subject describes negative associations.

5 GH1 and IGF1 population studies in the Hertfordshire cohort

5.1 *Genotype-Phenotype Associations in the Hertfordshire Study*

Using the newly developed conservative methods, previously described in chapter 3, a large amount of data has been generated for the GH1 and IGF1 genes. The data has been statistically analysed for associations with the phenotypic data available in the Hertfordshire cohorts.

The statistical analysis applied to the genotypic data from Hertfordshire database, are summarised below. The same analysis was performed for all markers genotyped.

The relationship between each continuously distributed phenotype variable and each genotype was explored using both analysis of variance (ANOVA) and linear regression models. ANOVA tests for any difference in the mean of the outcome variable between the genotype groups on one less degree of freedom than the number of genotypes in question. Regression analysis is used to test specifically for a trend in the means of the outcome variable across a genotype group. Continuous variables were $\log_e$ transformed as necessary. The means and sds are on $\log_e$ scale for transformed variables. The relationship between each binary phenotype variable and a gene was explored using cross-tabulations of frequencies and percentages, chi-square tests, and logistic regression models. The logistic regression models were used to test for any difference in outcome by genotype, and to test clearly for a trend in outcome across a genotype group. The relationship between each gene and the phenotype tested was explored without and with adjustment for the factors that have been identified as correlates of the phenotype in the earlier relevant Hertfordshire publications. Several sets of analysis were carried out. For main Hertfordshire database unadjusted associations of phenotype with genotype, as well as associations adjusted for age and BMI, associations adjusted for age, BMI, alcohol, smoking and current social class were made. For the North Hertfordshire ageing database, unadjusted associations of phenotype with genotype as well as associations adjusted for age, current social class and at birth, and height, associations adjusted for age, current social class and at birth, height and smoking and alcohol were made.

The relationship between birthweight and weight at one and each gene was also explored in each of the databases. All analyses were carried out separately for men and women, for each marker in each separate database.

The databases used were: Main East Hertfordshire data (which contains phenotype data for cardiovascular markers, fasting bloods, GTTs, and anthropometrics), the East Hertfordshire

bone database (containing phenotype data for bone area, BMC, BMD and loss rate at the lumbar spine, femoral neck and total femur), and finally the North Hertfordshire database (which contains data for ageing markers).

The table below, shows the codes for the alleles found for all the bi-allelic markers (SNPs) tested and the total number of samples that were genotyped in the Hertfordshire study.

Table 2 The bi-allelic markers genotyped in the Hertfordshire Study

Gene/Marker	Allele	Valid calls
GH1 5157	**1=A** **2=G**	924
GH1 5187	**1=C** **2=A**	1004
IGF1 6721	**1=G** **2=C**	1017
IGF1 2396	**1=C** **2=T**	886
IGF1 3276	**1=T** **2=C**	1032

WHERE: ALLELE 1= COMMON AND ALLELE 2 = RARE

5.2 Results

The results of the genotype-phenotype analysis in the human GH1 and IGF1 genes are presented in this subchapter. The results are shown firstly for the GH1 markers and then for the IGF1 marker. Each set of results corresponds to one particular marker in the respective gene. The markers are presented from the 5' position to the 3'.

5.2.1 Associations for GH1 SNP markers and the Hertfordshire study

In the main East Hertfordshire men, the rare allele of the **GH1 5157** marker showed a tendency towards higher fasting triglycerides and higher fasting proinsulin. Also, slightly lower weight at one was observed for the GH1 5157 marker in men in the main East Hertfordshire bank.

Table 3 Weight at one vs. GH1 5157 in Men (E. Herts)

Summary of Weight at one year (oz)

GH15157	Mean	Std. Dev.	Freq.
11	360.95495	40.698087	111
12	357.77206	41.082386	136
22	367.72414	47.465516	58
Total	360.82295	42.243593	305

p=0.324 on 2df

Table 4 Fasting triglycerides vs. GH1 5157 in Men (E.Herts)

GH15157	Fasting triglycerides (mmol/l) Mean	Std. Dev.	Freq.
11	.29503467	.52627373	64
12	.37675029	.54939925	108
22	.43200596	.50951943	53
Total	.36652252	.53373234	225

p= 0.168 fully adjusted

Table 5 Fasting proinsulin vs. GH1 5157 in Men (E.Herts)

Summary of Fasting Proinsulin

(pmol/l)

GH15157	Mean	Std. Dev.	Freq.
11	1.0373493	.69024291	64
12	1.0884637	.65728099	106
22	1.1600354	.76854043	52
Total	1.0904926	.69234138	222

p= 0.262

No associations were found between the cardiovascular phenotypes and the **GH1 5157** marker, in men or women. No link was found with triglycerides levels, total cholesterol, LDL:HDL ratio (in men), or with definite CHD.

Table 6 Summary of fasting triglycerides vs. GH1 5157 in Men (E.Herts)

Summary of Fasting triglycerides

(mmol/l)

GH1 5157	Mean	Std. Dev.	Freq.
11	.29503467	.52627373	64
12	.37675029	.54939925	108
22	.43200596	.50951943	53
Total	.36652252	.53373234	225

p= 0.372 on 2 df

Table 7 Summary of fasting total cholesterol vs. GH1 5157 in Men (E.Herts)

Summary of Fasting cholesterol

(mmol/l)

GH1 5157	Mean	Std. Dev.	Freq.
11	1.8911401	.18722943	64
12	1.8938142	.16183042	108
22	1.8850601	.203629	53
Total	1.8909915	.17890518	225

p= 0.959 on 2 df

Table 8 Ratio LDL:HDL vs. GH1 5157 in Men (E.Herts)

Summary of LDL:HDL ratio

GH1 5157	Mean	Std. Dev.	Freq.
11	1.3021336	.45737234	63
12	1.3923341	.40086736	105
22	1.3286688	.32556082	52
Total	1.3514558	.40229473	220

p= 0.335 on 2 df

Table 9 Definite CHD vs. GH1 5157 in Men (N.Herts)

Definite CHD

GH1 5157	0	1	Total
11	57	7	64
	89.06	10.94	100.00
12	90	17	107
	84.11	15.89	100.00
22	45	7	52
	86.54	13.46	100.00
Total	192	31	223
	86.10	13.90	100.00

p=0.66 for unadjusted test of association on 2df

Table 10 Fasting triglycerides vs. GH1 5157 in Women (E.Herts)

Summary of Fasting triglycerides

(mmol/l)

GH1 5157	Mean	Std. Dev.	Freq.
11	.32257174	.41811361	52
12	.20232167	.43842052	55
22	.28155601	.37828987	30
Total	.26531462	.41869787	137

p= 0.325 on 2 df

Table 11 Fasting total cholesterol vs. GH1 5157 in Women (E.Herts)

Summary of Fasting cholesterol

(mmol/l)

GH1 5157	Mean	Std. Dev.	Freq.
11	1.9908654	.16838504	52
12	1.9265875	.21833313	55
22	1.9500071	.14870693	30
Total	1.9561133	.18734484	137

p= 0.204 on 2 df

Table 12 Definite CHD vs. GH15157 in Women (N.Herts)

*Frequencies and %s

Definite CHD

GH1 5157	0	1	Total
11	41	10	51
	80.39	19.61	100.00
12	50	5	55
	90.91	9.09	100.00
22	25	5	30
	83.33	16.67	100.00
Total	116	20	136
	85.29	14.71	100.00

p=0.31 for unadjusted test of association on 2df

No association was determined between GH 1 5157 and BMI:

Table 13 BMI vs. GH1 5157 in Men (N. Herts)

Summary of BMI (kg/m2)

GH1 5157	Mean	Std. Dev.	Freq.
11	27.151515	3.452769	66
12	26.87037	3.1965595	108
22	27.171698	3.5056756	53
Total	27.022467	3.3339734	227

p= 0.808 on 2 df for bmi1 vs. GH1 5157

Table 14 BMI vs. GH1 5157 in Women (N. Herts)

Summary of EH1 BMI (kg/m2)

GH1 5157	Mean	Std. Dev.	Freq.
11	26.526923	3.582918	52
12	27.158182	4.6970711	55
22	26.543333	4.006347	30
Total	26.783942	4.1341696	137

p= 0.689 on 2 df for bmi1 vs. GH1 5157

For women, the rare allele had a tendency towards higher pulse rate, and a lower LDL: HDL.

Table 15 Pulse rate vs. GH1 5157 in Women (N. Herts)

Summary of Pulse rate (bp)

GH15157	Mean	Std. Dev.	Freq.
11	67.730769	10.918801	52
12	68.690909	9.0918619	55
22	71.7	11.43241	30
Total	68.985401	10.376008	137

p= 0.037 unadjusted

Table 16 Fasting LDL:HDL vs. GH1 5157 in Women (E. Herts)

Summary of Ln(LDL:HDL ratio)

GH15157	Mean	Std. Dev.	Freq.
11	1.293014	.36864922	52
12	1.1716334	.36361283	54
22	1.197641	.25930682	30
Total	1.2237806	.34753017	136

p= 0.082 on 2 df

For the **GH1 5187 marker**, the rare allele was designated the 1 allele number.

In **men** the common heterozygous (22) showed strong associations with lower Apolipoprotein A1 compared to the rare homozygous (11) and the heterozygous (12) individuals (p=0.005 fully adjusted). A tendency towards lower birthweight and weight at one was found for the common heterozygous compared to the rare homozygous and the heterozygous genotypes (6oz lower birthweight in the 22 genotype, and p=0.012 for 22 oz lower weight at one in the 22 genotype). Weaker associations were suggested between the 22 genotype and higher systolic blood pressure.

Table 17 Fasting Apolipoprotein A1 vs GH1 5187 in Men (E.Herts)

Summary of Apolipoprotein A1

(g/l)

GH15187	Mean	Std. Dev.	Freq.
11/12	.39352752	.2113068	24
22	.27523667	.22173485	210
Total	.28736907	.22316597	234

p= 0.005 on 1 df fully adjusted

Table 18 Birthweight vs GH1 5187 in Men (E. Herts)

Summary of birthweight in ounces

GH15187	Mean	Std. Dev.	Freq.
11/12	132	17.722499	24
22	125.94009	19.790128	217
Total	126.54357	19.644103	241

p= 0.152 on 1 df

Table 19 Weight at one vs GH15187 in Men (E.Herts)

Summary of weight at 1 yr in ounces

GH15187	Mean	Std. Dev.	Freq.
11/12	385.33333	45.875698	24
22	363.15668	40.355634	217
Total	365.36515	41.372589	241

p= 0.012 on 1 df fully adjusted

Table 20 Systolic BP vs. GH1 5187 in Men (E. Herts)

Summary of EH1 Systolic BP (mmHg)

GH15187	Mean	Std. Dev.	Freq.
11/12	158	19.880075	24
22	163.76959	21.494021	217
Total	163.19502	21.369744	241

p= 0.210 on 1 df

For the common heterozygous **women** an association between higher pulse pressure was found compared to the 12 and 11 genotypes and with lower fasting proinsulin.

Table 21 Pulse pressure vs. GH1 5187 in Women (E.Herts)

Summary of Pulse pressure mmHg (SBP-DBP)

GH15187	Mean	Std. Dev.	Freq.
11/12	66.909091	17.055524	11
22	74.55303	15.136687	132
Total	73.965035	15.363397	143

pp= 0.102

Table 22 Fasting proinsulin vs. GH1 5187 in Women (E.Herts)

Summary of Fasting Proinsulin (pmol/l)

GH15187	Mean	Std. Dev.	Freq.
11/12	1.551338	.5818534	11
22	1.238678	.53165861	131
Total	1.2628982	.54005166	142

p= 0.076

For the **North Hertfordshire** ageing database an association between birthweight and the common homozygous (22) genotype was found. Birthweight was lower in the 22 women compared to women of the other genotypes (p=0.05 on 1df for difference in birthweight). On average birthweight was 9.2 oz lower in the common homozygous genotype compared to the 11 and 12 genotypes in women.

Table 23 Birthweight vs. GH1 5187 in Women (N. Herts)

Summary of birthweight (oz)

GH15187	Mean	Std. Dev.	Freq.
11/12	128.28571	18.726487	14
22	119.10425	17.027131	259
Total	119.57509	17.201093	273

p=0.052 on 1df unadjusted

No adjustment has been made for multiple testing at this point.

No associations were found for BMI or definite coronary heart disease (CHD) for the 5187 marker in the human GH1 gene, in either women or men.

Table 24 BMI vs. GH1 5187 in Men (N.Herts)

Summary of BMI (kg/m2)

GH1 5187	Mean	Std. Dev.	Freq.
11/12	26.616667	2.6465181	24
22	27.037327	3.3626129	217
Total	26.995436	3.2959983	241

p= 0.554 on 1 df

Table 25 Definite CHD vs. GH1 5187 in Men (N. Herts)

*Frequencies and %s

Definite CHD

GH1 5187	0	1	Total
11/12	22	2	24
	91.67	8.33	100.00
22	180	33	213
	84.51	15.49	100.00
Total	202	35	237
	85.23	14.77	100.00

p=0.36 for unadjusted test of association on 1df

Table 26 BMI vs. GH1 5187 in Women (N. Herts)

Summary of EH1 BMI (kg/m2)

GH1 5187	Mean	Std. Dev.	Freq.
11/12	27.9	4.9390278	11
22	26.7	4.0924806	132
Total	26.792308	4.1559406	143

p= 0.359 on 1 df for bmi1

Table 27 Definite CHD vs. GH1 5187 in Women (N. Herts)

*Frequencies and %s

Definite CHD

GH1 5187	0	1	Total
11/12	9	2	11
	81.82	18.18	100.00
22	113	18	131
	86.26	13.74	100.00
Total	122	20	142
	85.92	14.08	100.00

p=0.69 for unadjusted test of association on 1df

5.2.2 Associations for IGF1 SNPs and the Hertfordshire study:

Genotyping for the N. and E. Hertfordshire studies has been achieved for a large number of samples. A total of four SNPs were genotyped. All the genotypes were found to be in Hardy-Weinberg Equilibrium. Analysis of data has been performed for cardiovascular measurements, birthweight and weight at one year.

The four SNPs genotyped, in the human IGF 1 gene, were: IGF 1 6721, 2396 and 3276. The SNPs are all representatives of clusters of SNPs in the region.

The results for the IGF1 6721 marker

The **IGF1 6721** common homozygous genotype in the main **East Hertfordshire** study, in **men** was associated with high HDL compared to the other genotypes (p=0.05 fully adjusted).

Table 28 Fasting HDL vs. IGF1 6721 in Men (E. Herts)

Summary of high density lipoproteins (mmol/l)

IGF16721	Mean	Std. Dev.	Freq.
11	.2065157	.25382013	183
12/22	.10097919	.30858559	39
Total	.1879755	.26655087	222

p= 0.045 for trend

For women the same genotype, 11, had a tendency towards association with higher insulin levels at zero time.

Table 29 Insulin vs. IGF1 6721 in Women (E. Herts)

Summary of Insulin (pmol/l)

IGF16721	Mean	Std. Dev.	Freq.
11	3.838749	.55035524	121
12/22	3.6534379	.58142173	21
Total	3.8113438	.55685558	142

p= 0.160 on 1 df

No associations were found between the IGF1 6721 genotypes and birthweight, weight at one, or with BMI. No adjustment has been made for multiple testing.

Table 30 Birthweight vs IGF1 6721 in Men (N. Herts)

Summary of birthweight in ounces

IGF16721	Mean	Std. Dev.	Freq.
11	126.21277	19.407189	188
12/22	127.775	20.906983	40
Total	126.48684	19.639814	228

p= 0.649 on 1 df

Table 31 Weight at one vs. IGF1 6721 in Men (N. Herts)

Summary of weight at 1 yr in ounces

IGF16721	Mean	Std. Dev.	Freq.
11	363.1383	41.749213	188
12/22	371.575	41.356417	40
Total	364.61842	41.71393	228

p= 0.246 on 1 df

Table 32 BMI vs. IGF1 6721 in Men (N. Herts)

BMI (kg/m2)

IGF16721	Mean	Std. Dev.	Freq.
11	27.063298	3.3167674	188
12/22	26.69	3.1073775	40
Total	26.997807	3.2774419	228

p= 0.514 on 1 df

Table 33 BMI vs. IGF1 6721 in Women (N. Herts)

BMI (kg/m2)

IGF16721	Mean	Std. Dev.	Freq.
11	26.783471	4.2103117	121
12/22	26.809524	4.0373139	21
Total	26.787324	4.1711753	142

p= 0.979 on 1 df

Associations between IGF1 SNP markers and bone phenotypes were investigated. The phenotypes looked at were: spine bone mineral density (BMD), spine bone loss rate, femoral neck BMD, and femoral neck bone loss rate, in men as well as women.

No significant associations were found between the IGF1 6721 marker and any of these phenotypes.

Table 34 Spine bone mineral density vs. IGF1 6721 in Men (E.Herts)

Summary of spine bmd

IGF16721	Mean	Std. Dev.	Freq.
11	1.0708438	.18280156	160
12/22	1.0754516	.17468483	31
Total	1.0715916	.18106707	191

p= 0.689 for trend

Table 35 Spine bone loss rate vs. IGF1 6721 in Men (E. Herts)

Summary of Annual % spine loss

IGF16721	Mean	Std. Dev.	Freq.
11	-.36093871	1.3901263	138
12/22	-.49977739	.97267684	24
Total	-.38150741	1.3349115	162

p= 0.395 for trend

Table 36 Femoral neck bone mineral density vs. IGF1 6721 in Men (E. Herts)

Summary of neck bmd

IGF16721	Mean	Std. Dev.	Freq.
11	.81469183	.12366943	159
12/22	.82051613	.12277347	31
Total	.81564211	.12321856	190

p= 0.339 for trend

Table 37 Femoral neck bone loss rate vs. IGF1 6721 in Men (E. Herts)

Summary of Annual % neck loss

IGF16721	Mean	Std. Dev.	Freq.
11	.2398782	1.6411476	137
12/22	.26996734	1.0538258	24
Total	.24436354	1.5649657	161

p=0.7051 on 1 df

Table 38 Spine bone mineral density vs. IGF1 6721 in Women (E. Herts)

Summary of spine bmd

IGF16721	Mean	Std. Dev.	Freq.
11	.91316822	.16316876	107
12/22	.93138889	.14859901	18
Total	.915792	.16071043	125

p=0.6581 on 1 df

Table 39 Spine bone loss rate vs IGF1 6721 in Women (E. Herts)

Summary of Annual % spine loss

IGF16721	Mean	Std. Dev.	Freq.
11	.3125602	1.1602893	105
12/22	.18127004	1.6024156	17
Total	.29426567	1.2242325	122

p= 0.671 for trend

Table 40 Femoral neck bone mineral density vs. IGF1 6721 in Women (E. Herts)

Summary of neck bmd

IGF6721	Mean	Std. Dev.	Freq.
11	.70514019	.10152362	107
12/22	.70800001	.09732904	18
Total	.705552	.10055142	125

p=0.9116 on 1 df

Table 41 Femoral neck bone loss rate vs. IGF1 6721 in Women (E.Herts)

Summary of Annual % neck loss

IGF16721	Mean	Std. Dev.	Freq.
11	1.1483199	1.3345565	105
12/22	.7108065	1.9133244	17
Total	1.0873549	1.4275966	122

p=0.2427 on 1 df

Results for the IGF1 2396 marker:

The second marker analysed in the IGF1 gene was IGF 2396. The 22 homozygous genotype in men was fund strongly associated with lower HDL (p=0.001 on 2df), higher LDL, and higher LDL to HDL ratio (p=0.001 on 2df unadjusted) compared with the 11 homozygote or 12 heterozygote.

Table 42 High density lipoproteins vs. IGF1 2396 in Men (E.Herts)

High density lipoproteins (mmol/l)

IGF12396	Mean	Std. Dev.	Freq.
11	.19103747	.21296228	90
12	.17236059	.27136851	83
22	-.12461152	.52781838	12
Total	.16218358	.27789993	185

p= 0.001 on 2 df unadjusted

Table 43 Low density lipoproteins vs. IGF1 2396 in Men (E. Herts)

Low density lipoproteins (mmol/l)

IGF12396	Mean	Std. Dev.	Freq.
11	1.5592314	.24339257	90
12	1.5058	.27423813	85
22	1.6694627	.11460912	12
Total	1.5420181	.25471114	187

p= 0.076 on 2 df

Table 44 LDL:HDL ratio vs. IGF1 2396 in Men (E. Herts)

IGF12396	Mean	Std. Dev.	Freq.
11	1.3681939	.33666928	90
12	1.314482	.42037425	83
22	1.7940742	.58838279	12
Total	1.3717208	.40913601	185

p= 0.001 on 2 df unadjusted

No association was found between the IGF1 2396 marker and other cardio markers.

For women the same 22 homozygous genotype was strongly associated with lower fasting insulin levels (p=0.008 on 2df unadjusted).

Table 45 Insulin vs. IGF1 2396 in Women (E. Herts)

Insulin (pmol/l)

IGF12396	Mean	Std. Dev.	Freq.
11	3.8589101	.57012353	69
12	3.8642639	.51778603	52
22	3.2917347	.49468327	10
Total	3.8177394	.56141801	131

p= 0.008 on 2 df unadjusted

No other associations were found between the IGF1 2396 marker and any cardiovascular phenotypes in the Hertfordshire study in men or women.

For the East Hertfordshire bone database no particular strong genotype-phenotype associations were found. None of the associations identified and presented below were either strong association nor were they consistent in men and women. The statistical power of the associations was also limited due to the small number of men or women in some of the genotypes.

Bone loss rate at the lumbar spine was higher for the 12 and 22 men than in the 11 genotype (p=0.05 on 2df unadjusted). The rare allele (2) was associated with increased femoral bone mineral content and density after the adjustment for weight was made (p=0.02 for trend unadjusted).

Table 46 Annual % spine loss vs. IGF1 2396 in Men (E. Herts)

IGF12396	Mean	Std. Dev.	Freq.
11	-.6754903	1.2693251	72
12	-.14963073	1.2545731	70
22	-.31924288	1.4293123	10
Total	-.4098808	1.2900182	152

p=0.0500 on 2 df unadjusted

Table 47 Summary of annual % total loss vs. IGF1 2396 in Men (E.Herts)

IGF12396	Mean	Std. Dev.	Freq.
11	-.13408201	1.6301827	72
12	.25712846	1.0055363	69
22	.79428852	1.2504387	10
Total	.10616453	1.3710652	151

p= 0.019 for trend unadjusted

For women no significant associations were found with any of the bone phenotypes.

Analysis of the IGF1 2396 against the North Hertfordshire ageing database has shown an association between allele 2 and lower prevalence of a previous heart attack in men (p=0.04

for trend unadjusted, p=0.02 fully adjusted). For women however, no associations were found)

Table 48 Frequencies and %s for heart attact vs igf1 2396 in Women (N. Herts)

IGF12396	No	Yes	Total
11	87	19	106
	82.08	17.92	100.00
12	110	14	124
	88.71	11.29	100.00
22	55	4	59
	93.22	6.78	100.00
Total	252	37	289
	87.20	12.80	100.00

for 2 allele= 0.58, p= 0.035 for trend in heart attack with igf12396

for 2 allele= 0.53, p= 0.021 fully adjusted

Results for the IGF1 3276 marker

For the IGF1 3276 marker several associations were found with the East Hertfordshire cardiovascular phenotypes.

For men, higher fasting triglycerides levels were found associated with the 12 heterozygous genotype compared to the 11 homozygotes (p=0.007 on 1df unadjusted, p=0.05 fully adjusted).

Table 49 Fasting triglycerides vs. IGF1 3276 in Men (E. Herts)

(mmol/l)

IGF13276	Mean	Std. Dev.	Freq.
11	.35023768	.51444225	206
12/22	.73722672	.53508723	14
Total	.37486426	.52315683	220

p= 0.007 on 1 df unadjusted

BMI tended to be higher in the male heterozygotes compared with the 11 homozygotes.

Table 50 BMI vs. IGF1 3276 in Men (N. Herts)

(kg/m2)

IGF13276	Mean	Std. Dev.	Freq.
11	26.905769	3.2232048	208
12/22	29.057143	3.8631823	14
Total	27.041441	3.2990187	222

p= 0.018 on 1 df fully adjusted

For women the power to detect any genotype-phenotype associations was severely limited by the low number of heterozygous females for the IGF1 3276 marker. Possible associations were found between the 12 heterozygotes and lower weight, and lower fasting triglycerides. These results were however not consistent with the associations found in men.

Table 51 Weight vs. IGF1 3276 in Women (N. Herts)

Weight (kg)

IGF13276	Mean	Std. Dev.	Freq.
11	68.9375	10.819296	136
12/22	58.375	8.0350793	4
Total	68.635714	10.872023	140

p= 0.055 on 1 df unadjusted

Table 52 Fasting triglycerides vs. IGF1 3276 in Women (E. Herts)

(mmol/l)

IGF13276	Mean	Std. Dev.	Freq.
11	.26081075	.39416477	136
12/22	-.21029582	.21323994	4
Total	.24735056	.39759354	140

p= 0.019 on 1 df fully adjusted

Birthweight and weight at one tended to be lower in the 12 heterozygotes women. These differences in weight although substantial were not statistically significant as only 4 women of the 12 genotype were present in the study for this marker.

The analysis of the associations between the IGF1 3276 genotypes and bone data showed no significant associations for men or women. As for the East cardiovascular database descriptive statistics suggested that birthweight and weight at one tended to be lower in the 12 heterozygotes and these differences were substantial, but with only 4 women with the 12 genotype present no statistical power was achieved.

No associations were found between the IGF1 3276 marker and the North Hertfordshire ageing database.

5.2.3 *Pairwise Linkage Disequilibrium Analysis for the GH1 markers*

Pairwise linkage disequilibrium measures were calculated between the GH markers and ACE insertion/deletion polymorphisms. (Himanshu J. Patel) An Expectation-Maximisation Algorithm was used to delineate un-phased genotypes.

The table below shows commonly used measures of LD.

Table 53 Commonly used measures of LD

LD measure	5157 vs. 5187	5157 vs. ACE	5187 vs. ACE
D'	0.45	0.20	0.55
R or Δ	0.105	0.16	0.096
δ	0.044	0.24	0.47
δ*	0.018	0.12	0.45
λ	0.33	1.9	0.302
Q	0.502	0.31	0.54
D	0.042	0.16	0.26

Table 54 shows the significance levels for **Δ.**

Table 54 Significance levels for Delta

Loci	Δ^{2N}	Significance Level (4df)
5157 vs. 5187	9.77	0.04
5157 vs. ACE	21.97	<0.001
5187 vs. ACE	9.02	0.06

Pairwise LD measures equations are presented below.

Expectation maximisation algorithm used to iterate phase of double heterozygotes.

Notation for estimated haplotype, marker allele and disease allele frequencies

Table 55 Pairwise Linkage Disequilibrium Measures

	SNP1 allele-1	SNP-1 allele-2	
SNP2 allele-1	p_{11}	p_{12}	p_{1+}
SNP2 allele-2	p_{21}	p_{22}	p_{2+}
	p_{+1}	p_{+2}	1

p=frequency

$\Delta =$ $p_{11}p_{22}-p_{12}p_{21} / (p_{1+}p_{2+}p_{+1}p_{+2})^{\wedge 1/2}$

D'= $p_{11}p_{22}-p_{12}p_{21} / p_{+1}p_{2+}$

$\delta =$ $p_{11}p_{22}-p_{12}p_{21} / p_{+1}p_{22}$

$\delta^* =$ $p_{11}p_{22}-p_{12}p_{21} / p_{+1}p_{2+}$

d= $p_{11}p_{22}-p_{12}p_{21} / p_{+1}p_{+2}$

Q= $p_{11}p_{22}-p_{12}p_{21} / p_{11}p_{22}+p_{12}p_{21}$

Levels of significance = chisq of $\Delta^{\wedge 2N}$ to 4 degrees of freedom

Statistical significance was achieved for two of the three pairwise analyses, with the third close to the statistical cut-off level. The GH1 5157 marker was in LD with the ACE I/D marker, but the GH1 5187 marker did not reach a statistically significant level. This has potential significance for a number of reasons. The promoter region of the human GH1 is highly heterogeneous due to promoter-shuffling and mutational events, and hence LD may not be dependent on its primary determinants of recombination fractions and meiotic events in this region. Furthermore the similarity of the GH gene cluster genes, repetitive regions and recombination hotspots predisposes the cluster to deletional and mutational events. Thus we would observe a varied picture for LD across this region and not a clear relationship between distance and LD. The GH1 5187 SNP has a minor allele frequency of $\leq$5% that may affect the statistical performance of the LD measures used.

5.2.4 Pairwise Linkage Disequilibrium Analysis for the IGF1 markers

Table 56 Various measurements of LD between the IGF1 markers.

LD measure	IGF1 2396 vs. IGF1 3276	IGF1 3276 vs. IGF1 6721	IGF1 2396 vs. IGF1 6721
D'	18.125	0.186	0.130
R or Δ	1.00	0.006	0.019
δ	1.00	0.010	0.280
δ*	1.00	0.017	0.123
λ	1.00	1.210	1.370
Q	1.00	0.095	1.165
D	0.049	0.0002	0.001

Complete linkage disequilibrium was observed between the IGF1 22396 and IGF1 3276 markers. Moderate to low LD was observed between the other t markers. IGF1 2396 and IGF1 3276 are located closely together therefore LD was expected. It is important to acknowledge the fact that a genome-wide LD map of common disease can be more informative and powerful than linkage analysis, if the appropriate numbers of markers are known. Finding cheap methods and lowering the efforts of genotyping for producing such maps is necessary for maximising the number of markers such LD map would contain. The most informative and cost effective method for LD mapping would be based on haplotypes. (99) The haplotype approach to LD mapping will be further discussed in chapter 6.

5.3 *Discussion*

The development of a dense map of SNPs will facilitate the identification of genetic variants that predispose to common human disease and support a strategy of a genome-wide search for LD using simple families or large numbers of cases and controls. The current 5-year plan of the Human Genome Project is to assemble a map of approximately 100,000 SNPs. (3)

About 90% of the human genome variation is constituted by SNPs. Attention is focused on the use of whole genome LD studies to map common disease genes. Such studies would employ a dense map of SNPs to detect association between a marker and a disease.

Several SNPs in the human GH1 and IGF1 genes have been identified (on the World Wide Web) (97). In order to analyse any genotype-phenotype associations, I had to develop in the first instance a method to genotype the markers in the cohorts available. The conservative methods of implementation of DNA banks available in our laboratory were expanded by the development of ARMS assays using long PCR amplification of the template.

The samples were amplified using the long amplification methods and dilutions of these amplifications were used as template for the SNP genotyping. Long amplification of both the human GH1 gene (5 kb) and of the exon 5 of the IGF1 gene was successfully achieved (L Hinks) for the DNA banks available. (North and East Hertfordshire) Various SNPs were identified in these regions. Long PCR was shown to be a valid way of amplifying individual DNA templates used for the analysis of SNPs in the GH1 and IGF1 regions.

The use of a long amplicon not only offered a way of conserving the DNA banks to allow multiple analyses for the same individual templates but also eliminated the problems created by the presence of highly repetitive sequences in both human GH region and IGF1 region. Due to the presence of regions of sequence repetition the ability to discriminate between amplicons with only one base mismatch primer, was very low. Blast searches on the www.ncbi.nlm.nih.gov for the sequences used revealed several different possible sequences of high homology in the GH and the IGF1 regions. However, producing a long amplicon of the region of interest, reduces the chances of mispriming due to high sequence homology. We were able to determine that ARMS analysis for the SNPs in our study was best controlled through the 'propping' technique in which the allele specific amplicon is internal to the control amplicon.

The common primers were used to obtain the control amplicon in the ARMS assay. The common primer with the same orientation as the allele specific primer was used in the

reaction to create a control amplicon in the reaction. The allele specific reaction would therefore take place internal to the common control amplicon, reducing the efficiency of the amplification of the longest (control) reaction. Although it was expected that the allele specific amplicon to perform between than the common amplicon, this was not always the case. This could be due to the fact that PCR efficiency in the later cycles of amplification is dependent on the availability of the template and primer. If after the denaturation step, in the later stages of the amplification, the single strand template is faced with an immense abundance of the strands of the already formed amplicon (the common amplicon) strand re-annealing will occur, the allele specific primer will have less available template and allele specific reaction will be less efficient than the common 'control' one. The common primer, of same orientation as the allele specific one, was sometimes diluted (from 100pMol/μl to 10 or even 5pmol/μl where no specific amplicon was obtained with other tested dilutions). Differential efficiency of the allele specific primer is one factor that causes variation between reactions.

All samples in the North and East Hertfordshire study were genotyped. The miscalls were probably due to failure in the initial long amplification of the template.

5.3.1 Data analysis for the GH1 SNP polymorphism

Before discussing the results obtained for the Hertfordshire study a short introduction to the field of statistical analysis and the statistical significance of the phenotypes analysed for associations with the genotypic data collected, shall be provided.

Statistically a population is defined as an entire collection of items that are the focus of interest. There is a critical difference between a population and a sample: with a population the aim is to identify its characteristics whereas with a sample one seeks to make inferences about the characteristics of the population from which the sample was take. Hertfordshire study is a population sample of Caucasian subjects and the aim of the study was to identify the common disease characteristics of the population to whom the subjects belong.

A finding is described as statistically significant when it can be demonstrated that the probability of obtaining such a difference by chance only, is relatively low. A significant result is obtained if theoretically only 5 out of 100 times ($p<0.05$) or less that result would be obtained when the only factors operating are the chance variations that occur whenever random samples are drawn. Therefore the word 'significant' in statistics language means

'probably true' (not due to chance).(www.surveysystem.com)The genotype-phenotype associations found in the Hertfordshire study are therefore statistically significant bearing in mind the large size of the population sample used. These results are discussed next.

In1986 Barker and Osmond demonstrated that foetal birthweight as well as weight at one and other anthropomertic measurements predict the occurrence of CHD, hypertension, dyslipidaemia, Type II Diabetes, insulin resistance, central obesity, osteoporosis many of which constitute the 'metabolic syndrome. (100)

Metabolic syndrome is strongly linked to a 'westernised' style characterised by lack of, or diminished physical activity and an abundant supply of high-fat foods.

Many medical conditions are considered part of the metabolic syndrome. These include coronary heart disease, non-insulin dependent diabetes mellitus, morbid obesity and hypertension. Several endocrine as well as biochemical abnormalities are representative features of the metabolic syndrome: hyperinsulinemia, insulin resistance, hypertriglyceridemia, low high-density lipoprotein and raised small dense low-density lipoprotein cholesterol. Obesity aggravates many of the components of the metabolic syndrome. The results obtained present one with a very interesting view over components of the metabolic syndrome.

A tendency of higher fasting triglycerides and the rare allele at the GH1 5157 locus was found. Lipid disturbances such as high triglycerides levels have been identified as a key feature of the metabolic syndrome. Several studies have shown that individuals with high levels of such particles (as well as raised small dense LDL and low high-density lipoprotein (HDL) cholesterol) have a greater risk of myocardial infarction. Obesity contributes to both hypertension and hyperinsulinemia. Indeed one other association was found, that between the GH1 5157 and higher fasting proinsulin levels in men.

For women in the East Hertfordshire study, a tendency towards association with higher pulse rate was found with the rare allele at the locus. A lower LDL: HDL ratio was also associated with the rare allele. High LDL and low HDL cholesterol is a feature of the metabolic syndrome and has been associated with increased risk of myocardial infarction. A tendency for lower weight at one for men with the rare allele at the 5157 locus was observed.

There were no other significant associations between GH1 5157 marker and BMI, total cholesterol, birthweight, weight at one year, and definite coronary heart disease.

The second marker analysed GH1 5187 SNP, has revealed much more interesting associations with phenotypic data. The common homozygous (22) genotype showed strong association with lower apolipoprotein A1 ($p=0.005$). Apo A1 is a measure of HDL level in

plasma (101). Low HDL cholesterol is another key feature of the metabolic syndrome that has been associated with increased risk of myocardial infarction. Increased levels of free fatty acid may inhibit cholesterol esterification capacity therefore decreasing acquisition of cholesterol by HDL. High levels of fatty acid leads to an increase in plasma triglycerides concentration leading to raised small dense LDL which is highly atherogenic. (100) A link between the common homozygous and higher systolic blood pressure in men was found, although the association was weaker. The association between the GH1 5187 marker with blood pressure is an interesting one. Conflicting results regarding the association between ACE I/D polymorphism and blood pressure have been reported in several publications (84,85). It is not known whether the I/D polymorphism in the ACE gene is functionally relevant or whether there is a marker in LD with a nearby functional polymorphism. Fornage *et al.* (102) proposed that genetic variation in a gene nearby ACE might explain the lack of association between the I/D polymorphism in the ACE gene despite the knowledge about the physiological involvement of ACE in blood pressure regulation found linkage between the ACE locus and diastolic blood pressure in a large population sample of men (but not women). He proposed as well that either ACE or a nearby gene is a sex specific candidate gene for hypertension. (103) My study found association between GH1 5187 marker and blood pressure, and moderate linkage disequilibrium between the ACE I/D marker and the GH1 5187 SNP. This leads to the suggestion that the GH1 5187 SNP is a good candidate for the 'nearby gene genetic variation' for hypertension. The results obtained for the GH1 5187 marker support the sex-specific candidacy for hypertension. The linkage disequilibrium found between the GH1 5187 SNP and the ACE I/D is discussed in detail later in this chapter.

The same genotype, 22, was associated with lower birthweight and lower weight at one year. The same association was indeed found in the North Hertfordshire men study where again the 22 genotype showed significant association with birthweight ($p=0.05$). As part of the foetal origins of adult disease we have Barker's nutritional hypothesis. The hypothesis states that nutritional events programme the foetus towards the development of a 'thrifty phenotype' that later in life becomes a liability if nutritional excess is present. The genetic hypothesis proposed by Day *et al.* (3), which has guided a whole range of genetic research as well as my project, was also mentioned in the introduction. The proposed genetic hypothesis states that, some of the genetic factors controlling early life growth may also be the same ones influencing late life traits. The results have shown a definite association between the GH1 5187 genetic marker and birthweight, and a slight association between GH1 5157 and weigh

at one, facts that come to support the genetic hypothesis. The genetic and the nutritional schemes may well interact (Fig.1).

5.3.2 Data analysis for the human IGF1 SNP Polymorphism

IGF1 levels are under nutritional influence and its levels seem to vary on a genotypic basis. (79) The involvement of IGF1 in foetal and postnatal growth and development is well documented. The results obtained for the IGF1 6721 marker, have shown an association between the common homozygous genotype (11) and higher levels of HDL in men (p=0.05 fully adjusted). This could mean that the 11 genotype has beneficiary properties with regards to the risk of CHD in men. It has been discussed earlier that low levels of HDL have previously been shown to correlate with high risk of myocardial infarction. Therefore having high levels of HDL associated with the 11 IGF1 6721 could be beneficial in lowering the risk of coronary heart disease. In women, however, 11 genotype was seen to associate with higher levels of insulin compared to the 12 or 22 genotypes. It is interesting to note that high plasma concentration of insulin is characteristic of insulin resistance. In insulin resistance, high insulin levels in the plasma fail to reduce the glucose levels normally. The results from the IGF1 6721 marker have not shown an association with high glucose levels. However, insulin resistance and hyperinsulinemia could cause hypertension via an increase in catecholamine activity without a change in plasma glucose concentrations. (100)

No significant associations were found between the IGF1 6721 marker and birthweight, weight at one year, or BMI, in men or women.

No associations with the bone data available in the East Hertfordshire database have been found for this marker..

For the IGF1 6721 marker, the 22 genotype, in men, was associated with lower levels of HDL (p=0.001), higher levels of LDL and a higher LDL to HDL ratio (p=0.001). The 22 genotype therefore could be considered a risk factor for CHD in men, as lower levels of HDL have been shown to associate with higher risk of myocardial infarction, as discussed earlier. It is also interesting to note the statistical significant association found between the 2 allele and previous heart attacks in men (p=0.04 for trend unadjusted, p=0.02 fully adjusted). It can be speculated that the 2 allele at this locus is a causative marker for CHD, or is in linkage disequilibrium with a causative marker(s) near by.

For women the same genotype was associated with lower fasting insulin levels. As high plasma insulin levels are characteristic of insulin resistance, the 22 homozygote for the IGF1

2396 marker could be a beneficial genotype. The 11/12 genotypes were found associated also with greater bone loss rate at the lumbar spine. It is a well accepted fact that IGF1 mediates not only soft tissue growth but also bone growth and it is a major candidate in the 3G Hypothesis (Fig.1) for finding the link between the early life traits and late life osteoporosis. One of the most affected by osteoporosis bone tissues in the body is the lumbar spine. The 12/22 genotype could represent a risk marker for osteoporosis in later life. For women, however, no associations were found with any of the phenotypes in the East Hertfordshire bone database.

High levels of triglycerides were found associated with the 12 genotype of the IGF1 3276 marker, in men. The same genotype was associated with increased body weight as well as higher BMI. It has been discussed earlier that high plasma levels of triglycerides lead to raised small dense LDL levels, which are highly atherogenic.(100) High BMI is representative of obesity, and the components of the metabolic syndrome are known to be affected by this factor. The 12/22 genotypes were associated with substantial lower weight at one year in 7 men of this genotype than in 369 men of the 11 genotype ($p=0.03$ for difference).

However, in women the opposite relationship was found between the 12 genotype and weight (i.e. lower weight was found associated with the 12 heterozygote.). Also lower fasting levels of triglycerides were also associated with the 12 genotype. These results may be due to chance as no statistical power could be conferred due to a very limited number of heterozygous women in the study. Descriptive statistics have suggested that birthweight and weight at one year tended to be lower in the 12 heterozygous women. This is an interesting factor when considering the genetic hypothesis and its involvement in the foetal origins of adult disease. However, no statistical power was achieved due to the low number of cases, as mentioned above. The same genotype in women was associated with increased prevalence of ischaemic heart disease.

No associations were found with any of the bone phenotypes available.

5.3.3 Pairwise linkage disequilibrium analysis

Between the GH1 and the ACE genes

We have examined the linkage disequilibrium between GH1 and ACE genes on the basis that some or all genotype associations might represent ACE alleles, 'proxy marking' causative alleles actually in the GH1 gene or nearby. PCR amplification of long regions of the GH1 and ACE genes, followed by internal single nucleotide polymorphic typing by allele-specific PCR, has been carried out on 1,100 Hertfordshire subjects. The results of this project showed moderate linkage disequilibrium (20% of maximum possible) between the GH1 5157 SNP marker and the ACE 287 bp Alu repeat insertion deletion calling into question the D-related aetiological pathway of all ACE associations reported in literature so far.

Angiotensin-converting enzyme (ACE or dipeptyl carboxidase) is a key enzyme in the renin-angiotensin system (RAS), which plays an important role in circulatory homeostasis. ACE hydrolyses angiotensin I into its physiologically active angiotensin II therefore being an important factor in the control of systemic blood pressure and fluid-electrolyte balance. (104) The ACE gene has been associated with susceptibility for left ventricular hypertrophy (Caucasians), cardiomyopathy, myocardial infarction in association with angiotensin 2 receptor, hypertension, (African Americans) particularly when associated with type II diabetes. (105) It is also associated with predisposition to IDDM neuropathies, towards end stage renal failure associated with deletion polymorphism in the ACE gene (allele D) or associated with deletion polymorphism in the APOE gene (allele I). (106) and (103) Recent positive associations between the insertion or the deletion of the 287-bp Alu repeat element in intron 16 in the ACE gene and coronary heart disease, the level of circulating enzyme, and risk of myocardial infarction as well as metabolic traits including insulin resistance relative to birth weight and metabolic and weight responses to exercise have been described. (98) Our findings in the study of a random sample population of 1032 subjects from North and East Herefordshire, born between 1920 and 1930's support the above conclusions.

Linkage analysis has shown that the plasma levels of ACE are influenced by a major quantitative trait locus that maps within or close to the ACE gene. Multiple variants that are in linkage disequilibrium with the 287 bp I/D polymorphism have been described and have been associated with the levels of ACE in plasma as well as with increased risk of cardiovascular disease. (105;107)

The GH gene is approximately 187kb from the ACE gene, which raises the issue of any LD effect. Amongst a number of other findings, a significant association was found between the

ACE deletion allele with increased body mass index (BMI) and impaired insulin response during the OGTT. Other studies have shown that the D allele may be associated with higher insulin insensitivity. Individuals homozygous for the D allele may have higher circulating levels of ACE and possibly more angiotensin II, which when infused exogenously, causes an increase in insulin insensitivity. (108) Furthermore, analysis of the GH5157 SNP data showed tendency for lower weight at one year of age for the common allele. Thus together, the results suggest that there may be a genetic component to the FOAD. This is also supported by the finding that the ACE D allele has been associated with attenuation of consequences of intrauterine growth retardation (reduced insulin resistance). (109)

The GH1 5157 showed 20% linkage disequilibrium with the ACE I/D marker.

The genotype-phenotype information for the ACE markers may well be an effect of LD between the GH and ACE locus. Central obesity is a common finding in GH deficient and MS patients, which could suggest that, a functionally modifying gene-variant of GH maybe in LD with the D allele of the ACE gene. Furthermore many of the SNPs are located in the GH promoter region. These could be modifying GH expression and contributing to this morbid phenotype. Many of the transcription factors of GH 1 bind conserved or degenerate recognition elements. Thus if one of these elements were subtly altered by a SNP, it may modulate binding kinetics and subunit formation of transcription factors and translation initiation complexes.

Other etiologically relevant genes surround the GH gene locus, such as the chorionic somatomamotrophin (CSL 1 and 2) that are highly expressed in the placenta, human placental lactogen (PL), and growth hormone variant gene (GHv). It is not implausible that some of these genes may also be in LD with the ACE markers and thus influence the disease-phenotype. When the GH gene locus is viewed on the ensembl contig viewer, it is apparent that much of region has not been sequenced or annotated yet. This has to be considered in interpreting any genotype-phenotype association as it could well be in LD with an, as yet unidentified, gene. Nonetheless a number of potential etiological genes have been characterised and predicted on the basis of sequence. Notable inclusions are protein kinases, ion channels and G-proteins. In particular the MAPK gene is approximately 200kb from GH, and any subtle variation in function may modulate GH1 signal transduction and/or cross talk. Thus a combination of sequence variants leading to a combination of subtly (or grossly) different biological properties may modulate gene function and the phenotype.

As shown the various measures of LD shows considerable heterogeneity, which probably results from different mapping properties. Despite this heterogeneity, the numerators in the

calculation are empirically the same. This results in all these measures being sensitive to marker frequencies (D' performance suffers at low allele frequencies). However δ and D' have more correlation with the recombination fraction than Q, Δ and d, making them more informative. (110)

Between the IGF1 markers

Analysis of LD between the IGF1 markers revealed moderate LD between only two of the markers analysed, the IGF1 2396 and IGF1 6721 (Delta = 0.33).

The analysis of linkage disequilibrium between markers of the same gene was necessary to point out the fact that a straight forward genotype-phenotype analysis does not pin point the causative marker for a specific trait or disease. Having information about the extent of LD between these markers gives information about the haplotype. Haplotype analysis is more enlightening providing information about the phase between markers and their association with a particular trait. Although associations have been found between the IGF1 markers and several phenotypes, little can be said about whether the marker associated with the trait is actually the causative one. More information is necessary about the phase between several markers in the region and possible associations with phenotypes.

In chapter 7, 'Future work', the advantages of haplotype information will be discussed, the reader will be introduced to the advances of other laboratories as well as my own.

6 Future work

Molecular genetic epidemiology studies involve the analysis of a great number of individuals. My study investigated several types of polymorphic markers in various genes in a large population sample with the aim of identifying possible associations between the markers and various components of the metabolic syndrome.

In order to do that firstly I had to develop methods that would be accessible to both large and small laboratories, and which would allow the fast, labour efficient genotyping of several thousands of samples a day.

However, these methods have to be conservative of the DNA bank in order to allow maximum number of tests of various polymorphic markers in many genes to be carried out. DNA banks consist of several thousands of samples from the general population and information from birth to present day about the subjects involved. Many of these subjects may actually have already died; thus their re-sampling is impossible. DNA banks have finite stocks and immortalisation of lymphocytes is unlikely, so far, to be possible. Therefore, the methods I have developed are conservative of the DNA template and cost effective for high throughput analysis. Genotype assays have been explored on a matched control DNA collection, for several categories of genetic markers including small insertion/deletions, microsatellites, and SNPs. I have shown that the template amount can be dramatically reduced with preservation of phase for direct haplotyping of candidate genes.

The management of DNA banks is achieved by a national committee and also by a local management group. For the DNA banks I have used in my studies the local management of DNA banks is provided locally by the Southampton Genetic Epidemiology Unit and by the MRC Centre, Clinical Research Advisory Group.

I have also identified linkage disequilibrium between the GH1 and ACE genes.

Linkage disequilibrium mapping provides a potent method for fine-structure localisation of rare disease genes, but has not yet been widely applied to common disease. (111) LD analysis is traditionally based on individual genetic markers and often gives a varied picture because the power to detect allelic associations depends on the properties of the marker analysed, such as frequency and population history. (112) LD analysis should in fact be based on the underlying haplotype structure of the human genome but this structure remains largely poorly understood. A genome-wide LD map of common diseases can be a lot more powerful than a straight forward linkage analysis, if the appropriate density of polymorphic markers is known and if the genotyping can be achieved with little cost, for human labour as well as financially. One method of LD mapping in common disease genes is the haplotype tagging and has recently been described by Johnson *et al.*.(99) The knowledge about the common haplotypes

and the SNPs that tag them has been used in this case not only to reduce the amount of genotyping but also to explain the complex patterns of LD between adjacent markers. Through haplotype tagging they also found key fine-mapping data within regions of strong LD. (99)

I have started to develop another method of direct haplotyping. The method consists of a double-ARMS reaction, where two SNP markers are amplified together, with only one control reaction for both markers. This type of reaction was applied to the GH1 5157 and GH1 5187 markers. The two rare genotypes found for these markers were also observed in the double-ARMS reaction. This type of amplification requires fine tuning due to the number of oligonucleotides used as well as for the specific reaction parameters required by each individual amplification.

The double ARMS reaction is schematically represented in figure 34.

Fig 34 The 'double' ARMS assay

Amplification of two SNPs in one reaction provides information about the haplotype at the locus. The two SNPs used for the initial trials of this method were the GH1 5157 and the GH1 5187. The amplification is performed for the four combinations of allele specific oligonucleotides, and only takes place where that particular combination of alleles is present. This haplotype analysis together with associated phenotypic data provides a more informative picture of the region.

A variation on the theme of double ARMS could also be the 'long ARMS'. For example, if a long amplicon between the GH1 5157 and a primer located some 4.5 kb distance is produced, the 'internal' SNPs found inside this long amplicon (e.g. 5165, 5187 and 5221), could be genotyped on a allele specific basis. Each long allele specific amplicon would then be

analysed as part of a haplotype formed with an internally placed SNP. Phase information about the 'internal' SNPs with regard to an allele specific amplicon would therefore be obtained.

The multiple aims of my project were achieved. Fast, efficient, cost effective and conservative methods for genotyping of various polymorphic markers have been developed and used to analyse genotype-phenotype associations with components of the metabolic disease. An insight into the foetal origins of adult disease from the angle of the genetic hypothesis has also been provided. Moderate linkage disequilibrium between the human GH1 and ACE genes has also been identified.

The findings of this project should be regarded not only as simple genotype-phenotype associations within a population sample, but mainly as a direction for the future analysis that should be performed in the genes investigated. The haplotype analysis methods started here should be further investigated and developed with the aim of possibly constructing a genome-wide linkage disequilibrium map of common disease genes.

7 Reference List

1. Voropanov,AM, Day,IN: Elimination of dumbbell bands and enhancement of resolution in MADGE using delayed start electrophoresis. *Biotechniques* 28:32-34, 2000

2. Barker,DJ, Osmond,C: Infant mortality, childhood nutrition, and ischaemic heart disease in England and Wales. *Lancet* 1:1077-1081, 1986

3. Day, I. N. Foetal Origins of Adult disease- the 3G Hypothesis-Research Grant Proposal. 1998.
Ref Type: Personal Communication

4. Boehnke,M: A look at linkage disequilibrium. *Nat.Genet* 25:246-247, 2000

5. Naggert,J: *Current Opinion in Genetics and Development*398-404, 1997

6. Meade,TW, Mellows,S, Brozovic,M, Miller,GJ, Chakrabarti,RR, North,WR, Haines,AP, Stirling,Y, Imeson,JD, Thompson,SG: Haemostatic function and ischaemic heart disease: principal results of the Northwick Park Heart Study. *Lancet* 2:533-537, 1986

7. Dieguez,C: *Trends Endocrinol Metab* 6:55-59, 1995

8. Miyakawa,M: Effect of growth hormone (GH) on serum concentrations of leptin: study in patients with acromegaly and GH deficiency. *J Clin Endocrinol Metab* 83:3476-3479, 1998

9. Scacchi,M: *In J obesity* 23:260-271, 2001

10. Textbook of Endocrine Physiology. New York, Oxford University Press, 1996, p. 244-259

11. Yakar,S, Liu,JL, Stannard,B, Butler,A, Accili,D, Sauer,B, LeRoith,D: Normal growth and development in the absence of hepatic insulin-like growth factor I. *Proc.Natl.Acad.Sci.U.S.A* 96:7324-7329, 1999

12. Le Roith,D, Bondy,C, Yakar,S, Liu,JL, Butler,A: The somatomedin hypothesis: 2001. *Endocrinology Rev* 22:53-74, 2001

13. Sjogren,K, Liu,JL: Liver-derived IGF1 is the principal source of IGF1 in blood but is not required for postnatal body growth. *Proc.Natl.Acad.Sci.U.S.A* 96:7088-7092, 1999

14. Yakar,S, Liu,JL, Le Roith,D: The growht hormone/insulin-like growth factor-I system: implication in development. *Pediatr Nephrol* 14:544-549, 2000

15. Yakar,S, Liu,JL, Fernandez,AM, Wu,Y, Schally,AV, Frystyk,J, Chernausek,SD, Mejia,W, Le Roith,D: Liver-specific igf-1 gene deletion leads to muscle insulin insensitivity. *Diabetes* 50:1110-1118, 2001

16. Won,S: *Pediatrics International* 42:31-36, 2000

17. Human Growth Hormone. www . 1999.
Ref Type: Internet Communication

18. Ott,J: *PNAS* 97: 2000

19. Genetics of Common Diseases. Bios Scientific Publishere Limited, 1997,

20. Strachan T, Read AP: Human Molecular Genetics. Bios Scientific Publishers Limited, 1996,

21. Morton,NE: *Ann Hum Genet* 20:125-134, 1955

22. King RC, Stansfield WD: A dictionary of genetics. New York, Oxford University Press, 1997,

23. Collins,A, Lonjou,C, Morton,NE: Genetic epidemiology of single-nucleotide polymorphisms. *Proc.Natl.Acad.Sci.U.S.A* 96:15173-15177, 1999

24. Brookes,AJ: The essence of SNPs. *Gene* 234:177-186, 1999

25. Emahazion,T, Jobs,M, Howell,WM, Siegfried,M, Wyoni,PI, Prince,JA, Brookes,AJ: Identification of 167 polymorphisms in 88 genes from candidate neurodegeneration pathways. *Gene* 238:315-324, 1999

26. Abdul-Latif,H: Growth hormone deficiency type IB caused by cryptic splicing of the GH-1 gene. *J.Pediat.Endocr.Metab.* 13: 21-28, 2000

27. Akinci,A: Isolated growth hormone (GH) deficiency type IA associated with a 45-kilobase gene deletion within the human GH gene cluster. *J.Clin.Endocr.Metab.* 75: 437-441, 1992

28. Phillips JA: Heterogeneity in the molecular basis of familial growth hormone deficiency (IGHD). (Abstract). *Am.J.Hum.Genet* 33, 1981

29. Phillips,JA, Hjelle,BL, Seeburg,PH, Zachmann,M: Molecular basis for familial isolated growth hormone deficiency. *Proc.Natl.Acad.Sci.U.S.A* 78:6372-6375, 1981

30. Goossens,M: Isolated growth hormone (GH) deficiency type 1A associated with a double deletion in the human GH gene cluster. *J.Clin.Endocr.Metab.* 62:712-716, 1986

31. Braga,S, Phillips,JA: Familial growth hormone deficiency resulting from a 7.6 kb deletion within the growth hormone gene cluster. *Am.J.Med.Genet.* 25:443-452, 1986

32. Takahashi,Y, Shirono,H: Biologically inactive growth hormone caused by an amino acid substitution. *Clin.Invest.J* 100:1159-1165, 1997

33. Illig,R, Budliger,H, Kind,HP, Fanconi,A, Prader,A: Effect of synthetic human, porcine, and salmon calcitonin on calcium and phosphorus in immobilized children. *Helv.Paediatr.Acta* 27:225-238, 1972

34. Masuda,N, Watahiki,M, Tanaka,M, Yamakawa,M, Shimizu,K, Nagai,J, Nakashima,K: Molecular cloning of cDNA encoding 20 kDa variant human growth hormone and the alternative splicing mechanism. *Biochim.Biophys.Acta* 949:125-131, 1988

35. Goddard,AD, Covello,R, Luoh,SM, Clackson,T, Attie,KM, Gesundheit,N, Rundle,AC, Wells,JA, Carlsson,LM: Mutations of the growth hormone receptor in children with idiopathic short stature. The Growth Hormone Insensitivity Study Group. *N.Engl.J.Med.* 333:1093-1098, 1995

36. Amselem,S, Goossens,M: [Hereditary dwarfism. Recent advances in molecular genetics of the growth hormone receptor]. *Presse Med.* 19:1699-1701, 1990

37. Morahan,G, Huang,D, Ymer,SI, Cancilla,MR, Stephen,K, Dabadghao,P, Werther,G, Tait,BD, Harrison,LC, Colman,PG: Linkage disequilibrium of a type 1 diabetes susceptibility locus with a regulatory IL12B allele. *Nat.Genet.* 27:218-221, 2001

38. Concannon,P, Gogolin-Ewens,KJ, Hinds,DA, Wapelhorst,B, Morrison,VA, Stirling,B, Mitra,M, Farmer,J, Williams,SR, Cox,NJ, Bell,GI, Risch,N, Spielman,RS: A second-generation screen of the human genome for susceptibility to insulin-dependent diabetes mellitus. *Nat.Genet.* 19:292-296, 1998

39. Hattersley,AT, Beards,F, Ballantyne,E, Appleton,M, Harvey,R, Ellard,S: Mutations in the glucokinase gene of the foetus result in reduced birth weight. *Nat.Genet* 19:268-270, 1998

40. Lander,ES, Schork,NJ: Genetic dissection of complex traits. *Science* 265:2037-2048, 1994

41. Knowler,WC, Williams,RC, Pettitt,DJ, Steinberg,AG: Gm3;5,13,14 and type 2 diabetes mellitus: an association in American Indians with genetic admixture. *Am.J.Hum.Genet.* 43:520-526, 1988

42. Lockhart,DJ, Winzeler,EA: Genomics, gene expression and DNA arrays. *Nature* 405:827-836, 2000

43. Day INM, Humphries SE: Genetics of common diseases-future therapeutic and diagnostic possibilities. Bios Scientifica Publishers, 1999,

44. Risch,N, Merikangas,K: The future of genetic studies of complex human diseases. *Science* 273:1516-1517, 1996

45. Rasko I, Downes CS: Principles of identification of human genetic disorders. In Genes in medicine. Oxford, Chapman and Hall, 1995, p. 147-171

46. Dear PH: Genome mapping- a practical approach. -ISBN . published in. NY USA, Oxford University Press, 1999,

47. Cacciari,E, Pirazzoli,P, Gualandi,S, Baroncini,C, Baldazzi,L, Trevisani,B, Capelli,M, Zucchini,S, Balsamo,A, Cicognani,A, .: Molecular study of human growth hormone gene cluster in three families with isolated growth hormone deficiency and similar phenotype. *Eur.J.Pediatr.* 153:635-641, 1994

48. Day,IN, Humphries,SE: electrophoresis for genotyping:microtitre array diagonal gel electrophoresis on horizontal polyacrylamide gels, hydrolink, or agarose. *Anal Biochem* 222:389-395, 1994

49. Weber,JL, May,PE: Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. *Am.J.Hum.Genet.* 44:388-396, 1989

50. Weber, J. L. and Che, J. Identification and analysis of human short insertion/deletion polymorphisms. 2001.
Ref Type: Personal Communication

51. Marth,GT, Korf,I, Yandell,MD, Yeh,RT, Gu,Z, Zakeri,H, Stitziel,NO, Hillier,L, Kwok,PY, Gish,WR: A general approach to single-nucleotide polymorphism discovery. *Nat.Genet.* 23:452-456, 1999

52. Kruglyak,L: Prospects for whole-genome linkage disequilibrium mapping of common disease genes. *Nat.Genet.* 22:139-144, 1999

53. Gilles,PN, Wu,DJ, Foster,CB, Dillon,PJ, Chanock,SJ: Single nucleotide polymorphic discrimination by an electronic dot blot assay on semiconductor microchip. *Nat Biotechnol* 17:365-370, 1999

54. Sapolsky,RJ, Hsie,L, Berno,A, Ghandour,G, Mittmann,M, Fan,JB: High-troughput polymorphism screening and genotyping with high-density oligonucleotide arrays. *Genet Anal* 14:187-192, 1999

55. Wang,LQ, Chen,LZ, Jin,H, Du,RF: [The genetic polymorphism of serum cholinesterase in Yi, Tibetan and Hani ethnic groups of China]. *Yi.Chuan Xue.Bao.* 15:138-143, 1988

56. Pennisi,E: A closer look at SNPs suggests difficulties. *Science* 281:1787-1789, 1998

57. Penrose,LS: Some recent trends in human genetics. *Caryologia* 6:521-529, 1954

58. Barker DJP: Mothers, babies and disease in later life. BMJ Publishing Group, 1994,

59. Smith,DW: *J Pediat* 89:225-230, 1976

60. Edwards,KL, Newman,B, Mayer,E, Selby,JV, Krauss,RM, Austin,MA: Heritability of factors of the insulin resistance syndrome in women twins. *Genet.Epidemiol.* 14:241-253, 1997

61. Holmes,RP, Holly,JM, Soothill,PW: Maternal insulin-like growth factor binding protein-1, body mass index, and fetal growth. *Arch.Dis.Child Fetal Neonatal Ed* 82:F113-F117, 2000

62. Debertin,AS, Korwitz,S, Tschernig,T, Kleemann,WJ, Gotz,W: Tissue distribution of components of the insulin-like growth factor system in sudden infant death and controls. *Forensic Sci.Int.* 113:199-204, 2000

63. Giudice, L. C. The IGF family at the maternal-foetal interface in human gestation. 9. 2001. Growth Hormone and IGF research.
Ref Type: Conference Proceeding

64. *Growth hormone and IGF research* 9: 1999

65. Siddle K. (1999) *Growth hormone and IGF research*, **9**. 9. 1999.
Ref Type: Conference Proceeding

66. Baxter, R. C. Regulation of cellular sensitivity to IGFBP-3 in breast cancer. 9. 1999. Growth Hormone and IGF research.
Ref Type: Conference Proceeding

67. Cohen P. Novel partners and functions of IGFBP-3. 9. 1999. Growth Hormone and IGF research.
Ref Type: Conference Proceeding

68. Van Buul-Offers SC. Human IGFBP-1 inhibits IGF1-stimulated body growth but stimulates growth of the kidney in snell dwarf mice. 9. 1999. Growth Hormone and IGF

research.
Ref Type: Conference Proceeding

69. Popovici R.M. Hypoxia and insulin regulation of IGFBP-1, in human foetal hepatocytes in primary culture: molecular mechanisms underlying elevated IGFBP-1 levels in foetal growth restriction. 9. 1999. Growth Hormone and IGF research.
Ref Type: Conference Proceeding

70. Tonner, E. Role of IGFBP-5 in cell death and plasminogen activation in the mammary gland. 9. 1999. Growth Hormone and IGF research.
Ref Type: Conference Proceeding

71. van Kleffens,M, Groffen,CA, Dits,NF, Lindenbergh-Kortleve,DJ, Schuller,AG, Bradshaw,SL, Pintar,JE, Zwarthoff,EC, Drop,SL, van Neck,JW: Generation of antisera to mouse insulin-like growth factor binding proteins (IGFBP)-1 to -6: comparison of IGFBP protein and messenger ribonucleic acid localization in the mouse embryo. *Endocrinology* 140:5944-5952, 1999

72. Bauer, M. K. GH treatment of foetal sheep with placental insufficiency restores foetal IGFI concentrations but has different actions on selected organs. 9. 1999. Growth Hormone and IGF research.
Ref Type: Conference Proceeding

73. Van Der Berghe, G. The GH/IGF axis in the catabolic state of critical illness. 9. 1999. Growth Hormone and IGF research.
Ref Type: Conference Proceeding

74. Jorgensen, J. O. L. The effects of GH and IGF I on substrate metabolism are related to the feast-famine cycle. 9. 1999. Growth Hormone and IGF research.
Ref Type: Conference Proceeding

75. Phillips,JA, Cogan,JD: Genetic basis of endocrine disease. 6. Molecular basis of familial human growth hormone deficiency. *J.Clin.Endocrinol.Metab* 78:11-16, 1994

76. Rimoin,DL, Merimee,TJ, McKusick,VA: Sexual ateliotic dwarfism: a recessively inherited isolated deficiency of growth hormone. *Trans.Assoc.Am.Physicians* 79:297-311, 1966

77. Hasegawa,Y, Fujii,K, Yamada,M, Igarashi,Y, Tachibana,K, Tanaka,T, Onigata,K, Nishi,Y, Kato,S, Hasegawa,T: Identification of novel human GH-1 gene polymorphisms that are associated with growth hormone secretion and height. *J.Clin.Endocrinol.Metab* 85:1290-1295, 2000

78. Takahashi,Y, Kaji,H, Okimura,Y, Goji,K, Abe,H, Chihara,K: Short stature caused by a mutant growth hormone with an antagonistic effect. *Endocr.J.* 43 Suppl:S27-S32, 1996

79. Rosen,CJ, Kurland,ES, Vereault,D, Adler,RA, Rackoff,PJ, Craig,WY, Witte,S, Rogers,J, Bilezikian,JP: Association between serum insulin growth factor-I (IGF-I) and a simple sequence repeat in IGF-I gene: implications for genetic studies of bone mineral density. *J.Clin.Endocrinol.Metab* 83:2286-2290, 1998

80. Baumbach,L, Schiavi,A, Bartlett,R, Perera,E, Day,J, Brown,MR, Stein,S, Eidson,M, Parks,JS, Cleveland,W: Clinical, biochemical, and molecular investigations of a genetic isolate of growth hormone insensitivity (Laron's syndrome). *J.Clin.Endocrinol.Metab* 82:444-451, 1997

81. Kristrom,B, Carlsson,B, Rosberg,S, Carlsson,LM, Albertsson-Wikland,K: Short-term changes in serum leptin levels provide a strong metabolic marker for the growth response to growth hormone treatment in children. Swedish Study Group for Growth Hormone Treatment. *J.Clin.Endocrinol.Metab* 83:2735-2741, 1998

82. Blum,WF, Albertsson-Wikland,K, Rosberg,S, Ranke,MB: Serum levels of insulin-like growth factor I (IGF-I) and IGF binding protein 3 reflect spontaneous growth hormone secretion. *J.Clin.Endocrinol.Metab* 76:1610-1616, 1993

83. Takahashi,Y, Kaji,H, Okimura,Y, Goji,K, Abe,H, Chihara,K: Brief report: short stature caused by a mutant growth hormone. *N.Engl.J.Med.* 334:432-436, 1996

84. Liu,JL, Yakar,S, Le Roith,D: Conditional knockout of mouse insulin-like growth factor-1 using the Cre-lox system. *Proc Soc Exp Biol* 223:344-351, 2000

85. Vikers M.H,IBABBH: *Endocrinology* 142:3964-3973, 2001

86. Meade,TW, Mellows,S, Brozovic,M, Miller,GJ, Chakrabarti,RR, North,WR, Haines,AP, Stirling,Y, Imeson,JD, Thompson,SG: Haemostatic function and ischaemic heart disease: principal results of the Northwick Park Heart Study. *Lancet* 2:533-537, 1986

87. Day,IN, Spanakis,E, Palamand,D, Weavind,GP, O'Dell,SD: Microplate-array diagonal-gel electrophoresis (MADGE) and melt-MADGE: tools for molecular-genetic epidemiology. *Trends Biotechnol.* 16:287-290, 1998

88. Laywood,A, Whittall,R, Gudnason,V, Humphries,SE, Day,IN: Handling of large (300 x 400 mm), thin (0.4 mm) polyacrylamide gels and recovery as dried gels. *Biotechniques* 17:850-854, 1994

89. Little S: ARMS analysis of point mutations. CRC Press,Inc, 1997, p. 45-51

90. Oksanen,L, Kaprio,J, Mustajoki,P, Kontula,K: A common pentanucleotide polymorphism of the 3'-untranslated part of the leptin receptor gene generates a putative stem-loop motif in the mRNA and is associated with serum insulin levels in obese individuals. *Int.J.Obes.Relat Metab Disord.* 22:634-640, 1998

91. Telenius,H, Carter,NP, Bebb,CE, Nordenskjold,M, Ponder,BA, Tunnacliffe,A: Degenerate oligonucleotide-primed PCR: general amplification of target DNA by a single degenerate primer. *Genomics* 13:718-725, 1992

92. Jeunemaitre,X, Lifton,RP, Hunt,SC, Williams,RR, Lalouel,JM: Absence of linkage between the angiotensin converting enzyme locus and human essential hypertension. *Nat.Genet.* 1:72-75, 1992

93. Dietmaier,W, Hartmann,A, Wallinger,S, Heinmoller,E, Kerner,T, Endl,E, Jauch,KW, Hofstadter,F, Ruschoff,J: Multiple mutation analyses in single tumor cells with improved whole genome amplification. *Am.J.Pathol.* 154:83-95, 1999

94. Kilger,C, Paabo,S: Direct exponential amplification and sequencing (DEXAS) of genomic DNA. *Biol.Chem* 378:99-105, 1997

95. Cheng,S, Fockler,C, Barnes,WM, Higuchi,R: Effective amplification of long targets from cloned inserts and human genomic DNA. *Proc.Natl.Acad.Sci.U.S.A* 91:5695-5699, 1994

96. Cheng,S, Chang,SY, Gravitt,P, Respess,R: Long PCR. *Nature* 369:684-685, 1994

97. Cargill,M, Altshuler,D, Ireland,J, Sklar,P, Ardlie,K, Patil,N, Shaw,N, Lane,CR, Lim,EP, Kalyanaraman,N, Nemesh,J, Ziaugra,L, Friedland,L, Rolfe,A, Warrington,J, Lipshutz,R, Daley,GQ, Lander,ES: Characterization of single-nucleotide polymorphisms in coding regions of human genes. *Nat.Genet.* 22:231-238, 1999

98. Thompson,DB, Ravussin,E, Bennett,PH, Bogardus,C: Structure and sequence variation at the human leptin receptor gene in lean and obese Pima Indians. *Hum.Mol.Genet.* 6:675-679, 1997

99. Johnson,GCL: Haplotype tagging for the identification of common disease genes. *Nat.Genet* 29:233-237, 2001

100. Han,TS, Lean,MEJ: *Medicine*27-29, 1998

101. Bansch,D: Basal growth hormone levels in women are positively correlated with high-density lipoprotein cholesterol and Apolipoprotein A-I independently of insulin-like grwth factor 1 or insulin. *Metabolism* 47:339-344, 1998

102. Fornage,M, Amos,CI, Kardia,S, Sing,CF, Turner,ST, Boerwinkle,E: Variation in the region of the angiotensin-converting enzyme gene influences interindividual differences in blood pressure levels in young white males. *Circulation* 97:1773-1779, 1998

103. O'Donnell,CJ, Lindpaintner,K, Larson,MG, Rao,VS, Ordovas,JM, Schaefer,EJ, Myers,RH, Levy,D: Evidence for association and genetic linkage of the angiotensin-converting enzyme locus with hypertension and blood pressure in men but not women in the Framingham Heart Study. *Circulation* 97:1766-1772, 1998

104. Samani,NJ: *Circulation* 94:708-712, 1996

105. **Keavney,B.** *Hum.Mol.Genet* 7:1-11, 1998

106. Sheu,WH: *Am Heart Journal* 136:125-131, 1998

107. Keaveney,B: Measured haplotype analysis of the angiotensin-I converting enzyme gene. *Hum Mol Genet* 7:1745-1751, 1998

108. Ryan,AS, Nicklas,BJ, Berman,DM, Ferrell,RE: The insertion/deletion polymorphism of the ACE gene is related to insulin sensitivity in overweight women. *Diabetes Care* 24:1646-1652, 2001

109. Cambien F: Angiotensin-I-converting enzyme gene polymorphism modulates the consequences of in-utero growth retardation on plasma insulin in young adults. *Diabetes* 27:470-475, 1998

110. Devlin B: A comparison of linkage disequilibrium measures for fine scale mapping. *Genomics* 29:311-322, 1995

111. Horikawa,Y: Genetic variation in the gene encoding calpain-10 is associated with type 2 diabetes mellitus. *Nat.Genet* 26:163-175, 2000

112. Daly,MJ, Rioux,JD, Schaffner,SF, Hudson,TJ, Lander,ES: High-resolution haplotype structure in the human genome. *Nat.Genet* 29:229-232, 2001

Appendices

Appendix 1

GH 1 Sequence

GENE BANK J03071

1gaattcctgggcctggggctgtggcagctgcctcgtcccttcacctcctggcttattctc
61tcccttcatatcttagcaatttcttcatgggaatgtcccccaattagaaatttctattata
121ccattatattaccaacatatatatatatacctggccgggcgcggtggctcatgcctgtaa
181tcccagcattttggtaggccaaggcgggtcggatcacctgaggtcaggagttcgagggcc
241agcctgatgaccatggtgaaaccccatctctactaaaaatacaaaattaatcgggcatgg
301tggcacatgcctgtaatcccagctactcgggaggctgaggcaggagaatcgcttaaaccc
361aggaagtggaggtttcagtgagctgagattgtgccattgcactccagcctgggtaacaag
421agcaaaactccatcaaaaaaaaataatatatgtatatatatattacaattttatatatata
481tacacattatgtaatttttttttaccatttatatatatacattacgtaatggtaaatgtt
541ttctctgccccccagtagattgttagctccagaagagaaggatcatgtctttggtttatc
601tagatatgtccatcggcctggtacagtctctggcccatgttataggcaacaactacttgt
661agaatcggtgaatgcatgaatagaagaatgagtgaatgaatgaatagacgaaaggcagaa
721atccagcctcaggccgggcgtgggtggctcacacctgtaatcccagcactttgggaggcc
781gaggcgggtggatcacaagatcaagagatggagaccaacctggctaacacagtgaaaccc
841cgtctctactaaaaatacaaaacattagccgggcatggcggtaggctcctgtagtccgag
901ctactagggagcctgaggcaggagaatggcatgaacccagggggcggacgttgcagtga
961gccaagatcgcgccactgcactccagcctgggcgacagagcaagactcggtctcaaaaaa
1021aaaaaaaaagaaatccagcctcaaagagcttacagtctggtaagaggaataaaatgtct
1081gcaaatagccacaggacaggtcaaaggaaggagaggctatttccagctgagggcacccca
1141tcaggaaagcacccagacttcctacaattactagacacatctcgatgctttcacttct
1201ctatcaatggatcgtctccctggagaataatccccaatgtgaaattacttagcacgtcaa
1261gttaggtagatccttgtgtacttcttggttgttcagagatcatcaccagtgcaacatccc
1321ccatcatacacagcagtgtcttgcccctctcctccccaagccttccgaggccctttcctcc
1381gtgcctgaaccccctggacatatcatgtggcaaactgaaggtccaacgagatacaggaa
1441gtgaaacacgatgtacactgaaacgtgcaatacaaatatgcagcatgaagtgcctcggtt
1501cactaacccgagctatgctgggtgcttctttctaccacttttccttaatgcctatggaca
1561cctcattctgtggctgaagttccttgtgttcaattccccccatcttcattgaacatcctc
1621tgtgcagggacttgacccctgtcctgctagctttgcactgaggcaagttttgtccatgcc
1681tagtagtgccacatctttactagatgaggtttctaaagagcttggcatggaaggaaagcc
1741tgggggccttagaagccatcacttaggaactgggagagctccaggcaagccacctttatc
1801tctttgggcctcagtatcctcactcacatccccatggtcagtagagtgccaagcacatgg
1861tggacacgcagaagtattgatcatcttcctcacctcccctaggagagccctggaatccc
1921catataatattctctgaagacctcaatccaatgagccgagtaattgactcggtagcaaac
1981tcatggaaaggacattagaagccaaaaggaggtagagtatgtcccagtacaataaccagc
2041ctctgatctcaaggaagaaagaacagagctgcaggtgagaagtgtgcgcctcaaatcacc
2101aaagtgagagtggaaggagcacccaccagtcccttggaggcgatccctaagaccggtgag
2161aatggcttcgaaaaatgtgatcgtcctcaaccccacagtcttgaacctaacaggagatct
2221tgaagcctgaaagacaccacttttaggatcaacagcagatctgtgacttttccacagcaggc
2281acacagaaccctagagttagtcgaggttggacataggaaggcttccaaacacacagaga
2341aattcatggatcctaaattataaagggaggttctttaaaagaaccaagatgattctgaga
2401tttatcctgagctttttttttttttttttttggatgcagagtctcgctctgtcgccca
2461ggctggagtgcagtggcgtgatctcggctaactgcaacctccacctcgcgggttcaagca
2521attcctgcctcggcctcctgaatagctgggattacaggcgcctgccaccacacctggct
2581aatttttgtatttttagcagagacgggatttcaccatcttggccaggctggtcttgaact
2641cctgatctcatgatccacccacctggcctccctaagtggtgggattacaggcgtgtgcc
2701caacttttcctgagccttttgaggctgacaccagaggtagaagcccagcctctcccact
2761ggccatgtggggagaggctccagcctgcagcaaccagggatctggcctcaagtgatgccc
2821caacagtgggcgacttcccagtactgttaggagaatcccaagtctaattcaaagttgatt
2881ttttactagcaattaatgatagacatggtctccattgctcaaggctctgggaagatctag
2941actagagaaaacgatcaccgacttctaccacacctgtgggcctcagttcttccctctggt
3001ccatggttaccacagtaacccctttgtaaaggtgtttccccaggggtccctagagtcctct
3061gagtcaccataattctgggtcaacagaaatggagaggtaagaggagacactccctgccc
3121tggctggtcccacgtgttcatggcagtaacctgtctggggagcctcgccaccccctgtgtc
3181cacctgcagagttatgcagtggtcccccaactagggcccctgccaccctcattcctaaggg
3241aggctggagacttcttccatggccgaaaatccacatctaagtcccccggcactagcaaaaa
3301ggccctgtcatgggggctccctgccccagttgagattggaggaacatggaagaagctaaa
3361ataactaaataactcagtagcatcacactacctgctctggagaaaaaaaaatttaatatt
3421attatttttgttttttgagatggagttctgctctgctgcacaggctggagtgcagtggc
3481gcaatcttggctcacggcaacctctgcctcctgagggttcaagtgattctcccgcctcag
3541cctcccgagtagttgggattacagctcatgccaccacgcccagctaatttttgtacttttt
3601agtagagatggagttttgccatgttggctagtctggccttgaactcctgacctcaagtga
3661tccacccacctcaaagccacccaaagtttggggattacaagcgtgagccactgtgtccgg
3721cctggagaaaggactttaaatgacgcaatgtaggaagagcaaggttgtggagatctgctg
3781ccctggctgaggtagctcatgcaatcagtctctctgagccacagtctcttgatctgtgaa
3841atcggaagaaaataatacctccttcacaagacaagtggcaggtcagatgtgagaatgcac
3901aggcaggccctcggcaactggaaaagctctatacagatctgaaaaggaggaggagaaaaa
3961agaggagggcttccatggctggacagggcatctttcttttctttttctttttttttt
4021ttttttttttgaggtggagtcttgctctgttgccaaggttggagtgcagcagcacgat
4081ctccgctcactgcaagctctgcctcccggattcacgccattctcctgcctcagcctcccg
4141agtagctgggaatacacggtccgccactacgcccagctaacttttttgcatttttagtac
4201agagtggatttcaccttggttagccaagatggtcttgatctactgacctcgtgatcccgcc
4261gcctcggcctcccaaagtgctgggattacaggcatgagccaccgcgcccagcctgataga
4321gcatcttcggcgtgatgtgttctgagttccaaagctgaggaagagactcaaatcttcaa
4381gagctcttctaacttgagattctctgatggtttcagggctatgggaggaagagcttgtg
4441gtccgtgtctgctcccgggatttctgtttcttggtttgtgtctctgctgcaagtccaagg
4501agctggggcaataccttgagtctgggttcttcgtccccagggacctgggggagccccagc
4561aatgctcagggaaaggggagagcaaagtgtgggttggttctctctagtggtcagtgttg
4621gaactgcatccagctgactcaggctgacccaggagtcctcagcagaagtggaattcagga
4681ctgaatcgtgctctacaaccccccacaatctattggctgtgcttggccccttttcccaacac
4741acacattctgtctggtgggtggaggttaaacatgcggggaggaggaaagggataggatag
4801agaatgggatgtggtcggtaggggtctcaaggactggctatcctgacatccttctccgc
4861gttcaggttggccaccatggcctgcggccagagggcacccacgtgaccctaaagagagg
4921acaagttgggtggtatctctggctgacactctgtgcacaacctcacaacactggtgacg
4981gtgggaagggaaagatgacaagccaggggcatgatcccagcatgtgtgggaggagcttc

5041ta**aattatccattagcacaagcccgtcag**tggccccatgcataaatgtacacagaaacag

5101gtgggggcaacagtgggagagaaggggccaggtataaaaagggcccacaagagacc**a**gct

5161caag**g**atcccaaggcccaactccccga**a**ccactcagggtcctgtggacagctcacctagc

5221**g**gc**a**atggctacaggtaagcgcccctaaaatccctttgggcacaatgtgtcctgaggggga

5281gaggcagcgacctgtagatgggacgggggcactaaccctcaggtttggggcttctgaatg

5341tgagtatcgccatgtaagcccagtatttggccaatctcagaaagctcctggtccctggag

5401gg**atggagagagaaaaacaaacagct**cctggagcagggagagtgctggcctcttgctctc

5461cggctccctctgttgccctctggtttctccccaggctcccggacgtccctgctcctggct

5521tttggcctgctctgcctgccctggcttcaagagggcagtgccttcccaaccattccctta

5581tccaggctttttgacaacgctatgctccgcgcccatcgtctgcaccagctggcctttgac

5641acctaccaggagtttgtaagctcttggggaatgggtgcgcatcaggggtggcaggaaggg

5701gtgactttcccccgctgggaaataagaggaggagactaaggagctcagggtttttcccg**a**

5761agcgaaaatgcaggcagatgagcacacgctgagtgaggttcccagaaaagtaacaatggg

5821agctggtctccagcgtagaccttggtgggcggtccttctcctaggaagaagcctatatcc

5881caaaggaacagaagtattcattcctgcagaaccccccagacctccctctgtttctcagagt

5941ctattccgacaccctccaacagggaggaaacacaacagaaatccgtgagtggatgccttc

6001tccccaggcggggatgggggagacctgtagtcagagccccgggcagcacagccaatgcc

6061cgtccttcccctgcagaacctagagctgctccgcatctccctgctgctcatccagtcgtg

6121gctggagcccgtgcagttcctcaggagtgtcttcgccaacagcctggtgtacggcgcctc

6181tgacagcaacgtctatgacctcctaaaggacctagaggaaggcatccaaacgctgatggg

6241ggtgagggtggcgccaggggtccccaatcctggagccccactgactttgagagctgtgtt

6301agagaaacactgctgccctctttttagcagtcaggccctgacccaagagaactcaccttа

6361ttcttcatttcccctcgtgaatcctccaggcctttctctacaccctgaaggggagggagg

6421aaaatgaatgaatgagaaagggagggaacagtacccaagcgcttggcctctccttctctt

6481ccttcactttgcagaggctggaagatggcagccccccggactgggcagatcttcaagcaga

6541cctacagcaagttcgacacaaactcacacaacgatgacgcactactcaagaactacgggc

6601tgctctactgcttcaggaaggacatggacaaggtcgagacattcctgcgcatcgtgcagt

6661gccgctctgtggagggcagctgtggcttctagctgcccgggtggcatccctgtgacccct

6721ccccagtgcctctcctggccctggaagttgccactccagtgcccaccagccttgtcctaa

6781taaaattaagttgcatcattttgtctgactaggtgtccttctataatattatggggtgga

6841gggggtggtatggagcaaggggcaagttgggaagacaacctgtagggcctgcggggtct

6901attcgggaaccaagctggagtgcagtggcacaatcttggctcactgcaatctccgcctcc

6961tgggttcaagcgattctcctgcctcagcctcccgagttgttgggattccaggcatgcatg

7021accaggctcagctaatttttgtttttttggtagagacggggtttcaccatattggccagg

7081ctggtctccaactcctaatctcaggtgatctacccaccttggcctcccaaattgctggga

7141ttacaggcgtgaaccactgctcccttccctgtccttctgatttttaaaataactataccag

7201caggaggacgtccagacacagcataggctacctgccatgcccaaccggtgggacatttga

7261gttgcttgcttggcactgtcctctcatgcgttgggtccactcagtagatgcctgttgaat

7321tcctgggcctagggctgtgccagctgcctcgtcccgtcaccttctggcttcttctctccc

7381tccatatcttagctgttttcctcatgagaatgttccaaattcgaaatttctatttaacca

7441ttatatatttacttgtttgctattatctctgcccccagtagattgttagctccagaagag

7501aaaggatcatgtcttttgcttatctagatatgcccatctgcctggtacaatctctggcac

7561atgttacaggcaacaactacttgtggaattggtgaatgcatgaatagaagaatgagtgaa

7621tgaatgaatagacaaaaggcagaaatccagcctcaaagaacttacagtctggtaagagga

7681ataaaatgtctgcaaatagccacaggacaggtcaaaggaaggaggggctatttccagctg

7741agggcaccccatcaggaaagcaccccagacttcctacaactactagacacatctcgatgc

7801ttttcacttctctatcaatggatcgtctccctggagaataatccccaaagtgaaattact

7861tagcacgtccagttaggtagatccttgtgtacttcttggttgttcagagatcatcaacca

7921gtgcaaacaatccccccatcaatacacagcagtgcctgcccctctccccccgaggtcttc

7981cgaggcccttcctccgtgcctgaaccccctggacatatcatatggcaaactgaagtgtcc

8041aacgagatatataggaagtgaaacacgatgtacactgaaacgtgcaatacaaatatgcagca

8101tgaagtgcctcggttcactaacccgagctacgctgggtgcttcttttctaccactttcct

8161taatgcctatggacacctcattctgtggctgaagttccttgtgttcaattcccccatct

8221tcactgaacatcctgtgtagggacctcacccctgtcctgctagctttgcactgaggcaag

8281ttctgtccatgcctagtagtgccaccacctttactagatgagacttctaaagagcttggc

8341atggaaggaaagcccgggggccttggaagccatcacttagaaactgggagagctccaggc

8401aagccacctcatctctttgagcctcagtatcctcactcacatccctatggtcagtagagt

8461gccaagcacatggtggacacgcagaagtattgatcatcttcctcacctcccctaggagag

8521cccttgggaatcccatataatattctctgaagacctcaatccaatgagccaagtaattg

8581actcggtagcaaactcatggaaaggacattagaagccaaaaggaggtagagtatgtccca

8641gtacaataaccagcctctgatctcaaggaagaaagaacagagctgcaggtgagaagtgtg

8701cgcctcaaatcaccaaagtgagagtggaaggagcacccaccagtcccttggaggcgatcc

8761ctaagaccggtgagaatggcttcgaaaaatgtgatcgtcctcaaccccacagtcttgaac

8821ctaacaggagatcttgaagcctgaaagacaccactttaggatcaacagcagatctgtgac

8881tttccacagcaggcacacagaaccctagagttagtcgaggttggacataggaagggcttc

8941caaacacacagagaaattcatggatcctaaattataaaggcaggttctttaaaagaacca

9001agatgattgtgagattacgtagagttttttttgtttttcttgatacagagtctcgctct

9061gttgcccaggcaggagtgcagtggcatgatctcagctcgctacaacctccacctcccggg

9121ttcaagcaattccctgccacagccacctcgagtagctgggattacaggcacctaccacca

9181tgcgcagctaatttttctatttctagtagagacgggatttcaccatcttgaccaggttgg

COMMON FORWARD

SNP'S 5157, 5165, 5187, 5221, 5224

COMMON REVERSE

CONTROL AMPLICON (263BP)

APPENDIX 2
IGFI exon5 sequence ref.
X57025
Sequence from: 569-7236

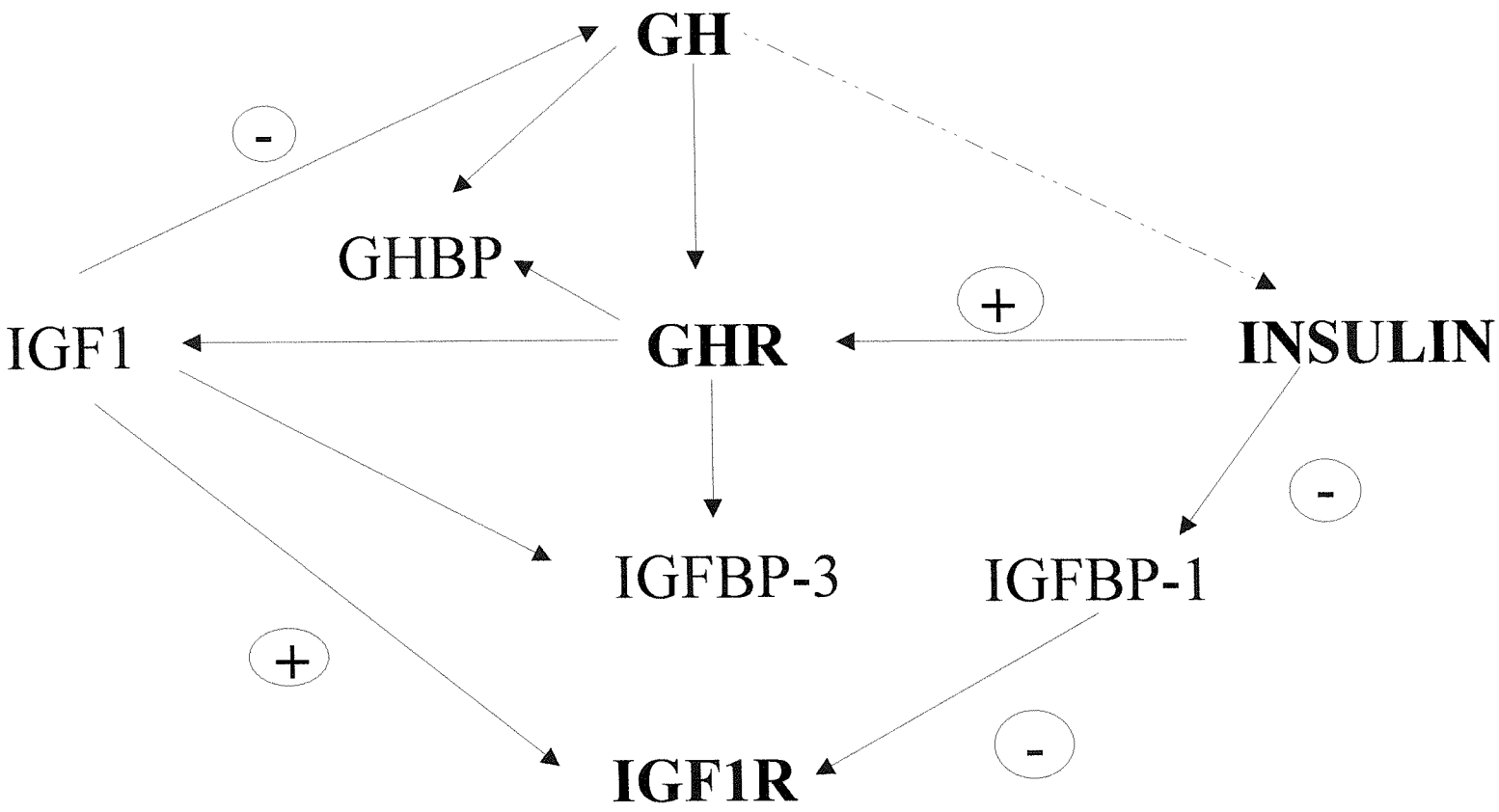
569gaagaacgcaagtagagggag
601tgcaggaaacaagaactacaggatgtaggaagaccctcctgaggagtgaagagtgacatg
661ccaccgcaggatcctttgctctgcacgagttacctgttaaactttggaacacctaccaaa
721aaataagtttgataacatttaaaagatgggcgtttccccccaatgaaatacacaagtaaac
781attccaacattgtctttaggagtgatttgcaccttgcaaaaatggtcctggagttggtag
841attgctgttgatcttttatcaataatgttctatagaaaagaaaaaaaaatatatatatat
901atatatcttagtccctgcctctcaagagccacaaatgcatgggtgttgtatagatccagt
961tgcactaaattcctctctgaatcttggctgctggagccattcattcagcaaccttgtcta
1021agtggtttatgaattgtttccttatttgcacttctttctacacaactcgggctgtttgtt
1081ttacagtgtctgataatcttgttagtctatacccaccacctcccttcataacctttatat
1141ttgccgaatttggcctcctcaaaagcagcagcaagtcgtcaagaagcacaccaattctaa
1201cccacaagattccatctgtggcatttgtaccaaatataagttggatgcattttatttag
1261acacaaagctttatttttccacatcatgcttacaaaaaagaataatgcaaatagttgcaa
1321ctttgaggccaatcatttttaggcatatgttttaaacatagaaagtttcttcaactcaaa
1381agagttccttcaaatgatgagttaatgtgcaacctaattagtaactttcctcttttatt
1441ttttccatatagagcactatgtaaatttagcatatcaattatacaggatatatcaaacag
1501tatgtaaaactctgttttttagtataatggtgctattttgtagtttgttatatgaaagag
1561tctggccaaaacggtaatacgtgaaagcaaaacaataggggaagcctggagccaaagatg
1621acacaaggggaagggtactgaaaacaccatccatttgggaaagaaggcaaagtcccccca
1681gttatgccttccaagaggaacttcagacacaaaagtccactgatgcaaattggactggcg
1741agtccagagaggaaactgtggaatggaaaaagcagaaggctaggaattttagcagtcctg
1801gtttctttttctcatggaagaaatgaacatctgccagctgtgtcatggactcaccactgt
1861gtgaccttgggcaagtcacttcacctctctgtgcctcagtttcctcatctgcaaaatggg
1921ggcaatatgtcatctacctacctcaaaggggtggtataaggtttaaaaagataaagattc
1981agatttttttaccctgggttgctgtaagggtgcaacatcagggcgcttgagttgctgaga
2041tgcaaggaattctataaataacccattcatagcatagctagagattggtgaattgaatgc
2101tcctgacatctcagttcttgtcagtgaagctatccaaataactggccaactagttgttaa
2161aagctaacagctcaatctcttaaaacacttttcaaaatatgtgggaagcatttgattttc
2221aatttgattttgaattctgcatttggttttatgaatacaaagataagtgaaaagagagaa
2281aggaaaagaaaaaggagaaaaacaaagagatttctaccagtgaaaggggaattaattact
2341ctttgttagcactcactgactcttctatgcagttactacatatctagtaaaaccttgttt
2401aatactataaataatattctattcattttgaaaaacacaatgattccttctttctaggc
2461aatataaggaaagtgatccaaaatttgaaatattaaaataatatctaataaaaagtcaca
2521aagttatcttctttaacaaactttactcttattcttagctgtatatacatttttttaaaa
2581agtttgttaaaatatgcttgactagagtttcagttgaaaggcaaaaacttccatcacaac
2641aagaaatttcccatgcctgctcagaagggtagccccctagctctctgtgaatgtgttttat
2701ccattcaactgaaaattggtatcaagaaagtccactggttagtgtactagtccatcatag
2761cctagaaaatgatccctatctgcagatcaagattttctcattagaacaatgaattatcca
2821gcattcagatctttctagtcaccttagaacttttggttaaaagtacccaggcttgatta
2881tttcatgcaaattctatattttacattcttggaaagtctatatgaaaaacaaaaataaca
2941tcttcagtttttctcccactgggtcacctcaaggatcagaggccaggaaaaaaaaaaaag
3001actccctggatctctgaatatatgcaaaaagaaggccccatttagtggagccagcaatcc
3061tgttcagtcaacaagtattttaactctcagtccaacattatttgaattgagcacctcaag
3121catgcttagcaatgttctaatcactatggacagatgtaaaagaaactatacatcatttttt
3181gccctctgcctgttttccagacatacaggttctgtggaataagatactggactcctcttc
3241ccaagatggcacttctttttatttcttgtccccagtgtgtacctttaaaattattccct
3301ctcaacaaaactttataggcagtcttctgcagacttaacatgttttctgtcatagttaga
3361tgtgataattctaagagtgtctatgacttatttccttcacttaattctatccacagtcaa
3421aaatccccaaggaggaaagctgaaagatgcaactgccaatattatctttcttaacttt
3481tccaacacataatcctctccaactggattatataaataaattgaaaataactcattatacca
3541attcactatttttatttttttaatgaattaaaactagaaaacaaattgatgcaaaccctgga
3601agtcagttgattactatatactacagcagaatgactcagatttcatagaaaggagcaacc
3661aaaatgtcacaaccaaaactttacaagctttgcttcagaattagattgctttataattct
3721tgaatgaggcaatttcaagatatttgtaaaagaacagtaaacattggtaagaatgagctt
3781tcaactcataggcttatttccaatttaattgaccatactggatacttaggtcaaatttct
3841gttctctcttgcccaaataatattaaagtattatttgaacttttaagatgaggcagttc
3901ccctgaaaaagttaatgcagctctccatcagaatccactcttctagggatatgaaaatct
3961cttaacacccaccctacatacacagacacacacacacacacacacacacacacacaca
4021cacacattcaccctaaggatccaatggaatactgaaaagaaatcacttccttgaaaattt
4081tattaaaaaacaaacaaacaaacaaaaagcctgtccacccttgagaatccttcctctcct
4141tggaacgtcaatgtttgtgtagatgaaaccatctcatgctctgtggctccagggtttctg
4201ttactattttatgcacttgggagaaggcttagaataaaagatgtagcacattttgctttc
4261ccatttattgtttggccagctatgccaatgtggtgctattgtttctttaagaaagtactt

4321 gactaaaaaaaaaagaaaaaaagaaaaaaaagaaagcatagacatatttttttaaagtat
4381 aaaaacaacaattctatagatagatggcttaataaaatagcattaggtctatctagccac
4441 caccacctttcaacttttatcactcacaagtagtgtactgttcaccaaattgtgaatttt
4501 gggggtgcaggggcaggagttggaaatttttaaagttagaaggctccattgttttgttg
4561 gctctcaaacttagcaaaattagcaatatattatccaatcttctgaacttgatcaagagc
4621 atggagaataaacgcgggaaaaaagatcttataggcaaatagaagaatttaaaagataag
4681 taagttccttattgatttttgtgcactctgctctaaaacagatattcagcaagtggagaa
4741 aataagaacaaagagaaaaaatacatagatttacctgcaaaaaatagcttctgccaaatc
4801 ccccttgggtattctttggcatttactggtttatagaagacattctcccttcacccagac
4861 atctcaaagagcagtagctctcatgaaaagcaatcactgatctcatttgggaaatgttgg
4921 aaagtatttccttatgagatgggggttatctactgataaagaaagaatttatgagaaatt
4981 gttgaaagagatggctaacaatctgtgaagatttttgtttcttggttttgttttttttt
5041 tttttttactttatacagtctttatgaatttcttaatgttcaaaatgacttggttcttt
5101 tcttctttttttttatatcagaatgaggaataataagttaaacccacatagactctttaaa
5161 actataggctagatagaaatgtatgtttgacttgttgaagctataatcagactatttaaa
5221 atgttttgctatttttaatcttaaaagattgtgctaatttattagagcagaacctgtttg
5281 gctctcctcagaagaaagaatctttccattcaaatcacatggctttccaccaatattttc
5341 aaaagataaatctgatttatgcaatggcatcatttatttttaaaacagaagaattgtgaaa
5401 gtttatgcccctcccttgcaaagaccataaagtccagatctggtaggggggcaacaacaa
5461 aaggaaaatgttgttgattcttggttttggattttgttttgttttcaatgctagtgttta
5521 atcctgtagtacatatttgcttattgctattttaatattttataagaccttcctgttagg
5581 tattagaaagtgatacatagatatctttttttgtgtaatttctatttaaaaaagagagaag
5641 actgtcagaagctttaagtgcatatggtacaggataaagatatcaatttaaataaccaat
5701 tcctatctggaacaatgcttttgttttttaaagaaacctctcacagataagacagaggcc
5761 caggggatttttgaagctgtctttattctgcccccatcccaacccagcccttattatttt
5821 agtatctgcctcagaattttatagagggctgaccaagctgaaactctagaattaaaggaa
5881 cctcactgaaaacatatatttcacgtgttccctctctttttttcctttttgtgagatgg
5941 ggtctcgcactgtcccccaggctggagtgcagtggcatgatctcggctcactgcaacctc
6001 cacctcctgggtttaagcgattctcctgcctcagcctcctgagtagctgggattacaggc
6061 acccaccactatgcccggctaatttttggatttttaatagagacggggttttaccatgt
6121 tggccaggttggactcaaactcctgaccttgtgatttgcccgcctcagcctcccaaattg
6181 ctgggattacaggcatgagccaccacacccctgcccatgtgttccctcttaatgtatgatt
6241 acatggatcttaaacatgatccttctctcctcattcttcaactatctttgatggggtctt
6301 tcaaggggaaaaaaatccaagctttttaaagtaaaaaaaaaaaaagagaggacacaaaa
6361 ccaaatgttactgctcaactgaaatatgagttaagatggagacagagtttctcctaataa
6421 ccggagctgaattacctttcactttcaaaaacatgaccttccacaatccttagaatctgc
6481 ctttttttatattactgaggcctaaaagtaaacattactcattttatttgcccaaaatg
6541 cactgatgtaaagtaggaaaaataaaaacagagctctaaaatccctttcaagccacccat
6601 tgaccccactcaccaactcatagcaaagtcacttctgttaatcccttaatctgatttgt
6661 ttggatatttatcttgtacccgctgctaaacacactgcaggagggactctgaaacctcaa
6721 gctgtctacttacatcttttatctgtgtctgtgtatcatgaaaatgtctattcaaaatat
6781 caaaaccttcaaatatcacgcagcttatattcagtttacataaaggccccaaataccat
6841 gtcagatctttttggtaaaagagttaatgaactatgagaattgggattacatcatgtatt
6901 ttgcctcatgtatttttatcacacttataggccaagtgtgataaataaacttacagacac
6961 tgaattaatttcccctgctactttgaaaccagaaaataatgactggccattcgttacatc
7021 tgtcttagttgaaaagcatatttttttattaaattaattctgattgtatttgaaattatta
7081 ttcaattcacttatggcagaggaatatcaatcctaatgacttctaaaaatgtaactaatt
7141 gaatcattatcttacatttactgtttaataagcatattttgaaaatgtatggctagagtg
7201 tcataataaaatggtatatctttctttagtaattac

Appendix 3 Polymorphism in the human leptin receptor gene (OB-R)

Polymorphism	Association	Reference	Comment
Ser 675Thr	Negative association with obesity	Roth et al. 1998	
Gln223Arg	Negative association with obesity	Gotoda et al. 1997	
SilentT>C at 343	Negative association	Gotoda et al. 1997	
C>T at 986	Negative association	Gotoda et al. 1997	
Silent G>A 1019	Negative association with juvenile onset obesity	Echwald et al 1997 Gotoda et al. 1997 Thompson et al. 1997	
A>T intron 16 2sites	Positive association with morbid obesity	Thompson et al. 1997	Association was found between variation a the leptin receptor gene level and obesity. Small sample of 20 Pima Indians
A>C intron 19	Positive association with morbid obesity	Thompson et al. 1997	Small sample of 20 Pima Indians
C>T intron 19	Positive association with morbid obesity	Thompson et al. 1997	Small sample of 20 Pima Indians
Lys 109Arg	Negative association to juvenile onset obesity	Thompson et al. 1997 Echwald et al 1997 Gotoda et al. 1997	
Lys 204Arg	Negative association to juvenile onset obesity	Echwald et al 1997	Mutation was found in only one obese subject
Lys656Asn	Negative association with obesity	Gotoda et al. 1997 Thompson et al. 1997	
Ser492Thr	Negative association with obesity	Gotoda et al. 1997 Thompson et al. 1997	

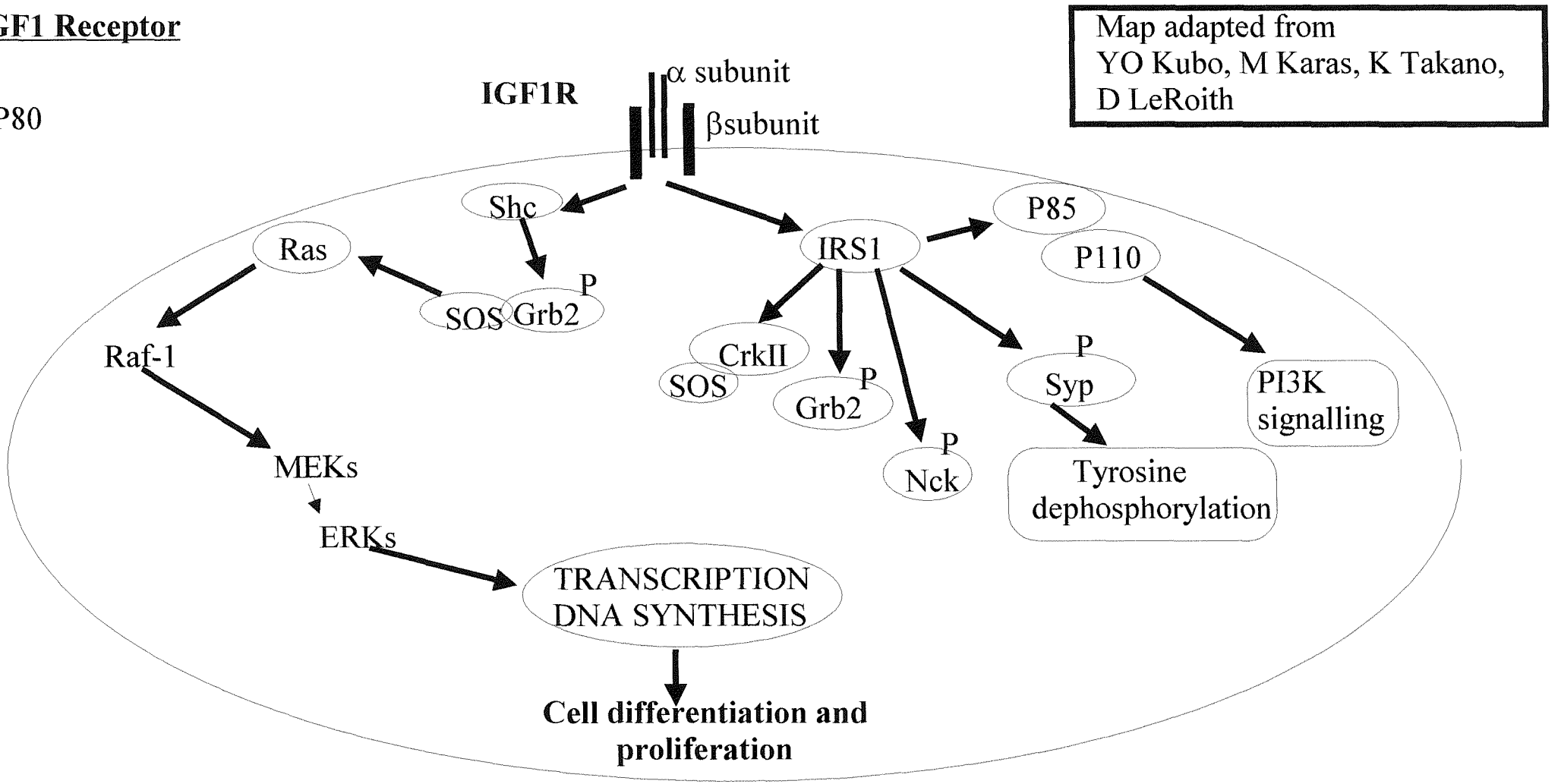
Polymorphism	Association	Reference	Comment
Ala976Asp	Negative association with obesity	Gotoda et al. 1997 Thompson et al. 1997	
IVS16DS	Indication for linkage with extreme obesity	Clement et al. 1996	
G>A intron 16	Indication for linkage with extreme obesity	Clement et al. 1996	
3'UTR 5bp ins/del	Association with serum insulin levels in obese individuals	Oksanen et al. 1998	Small sample. Only 30 morbidly obese patients were screened
MaeII intron 5	Negative association with obesity	Rolland et al. 1998	



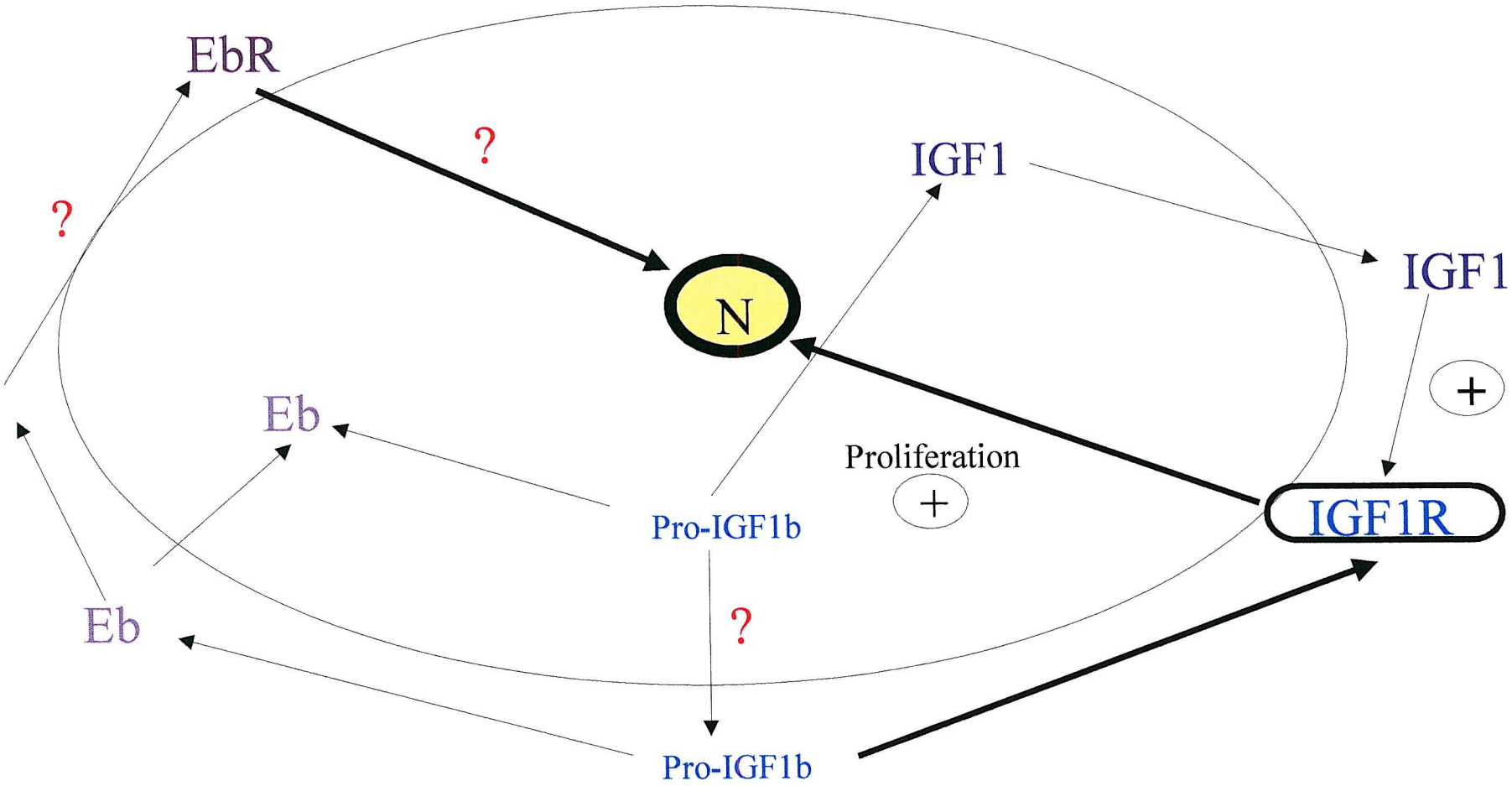
Appendix 4: IGF1 potential to stimulate regulation of growth

Map adapted from D. Dangar et al.

IGF1 Receptor

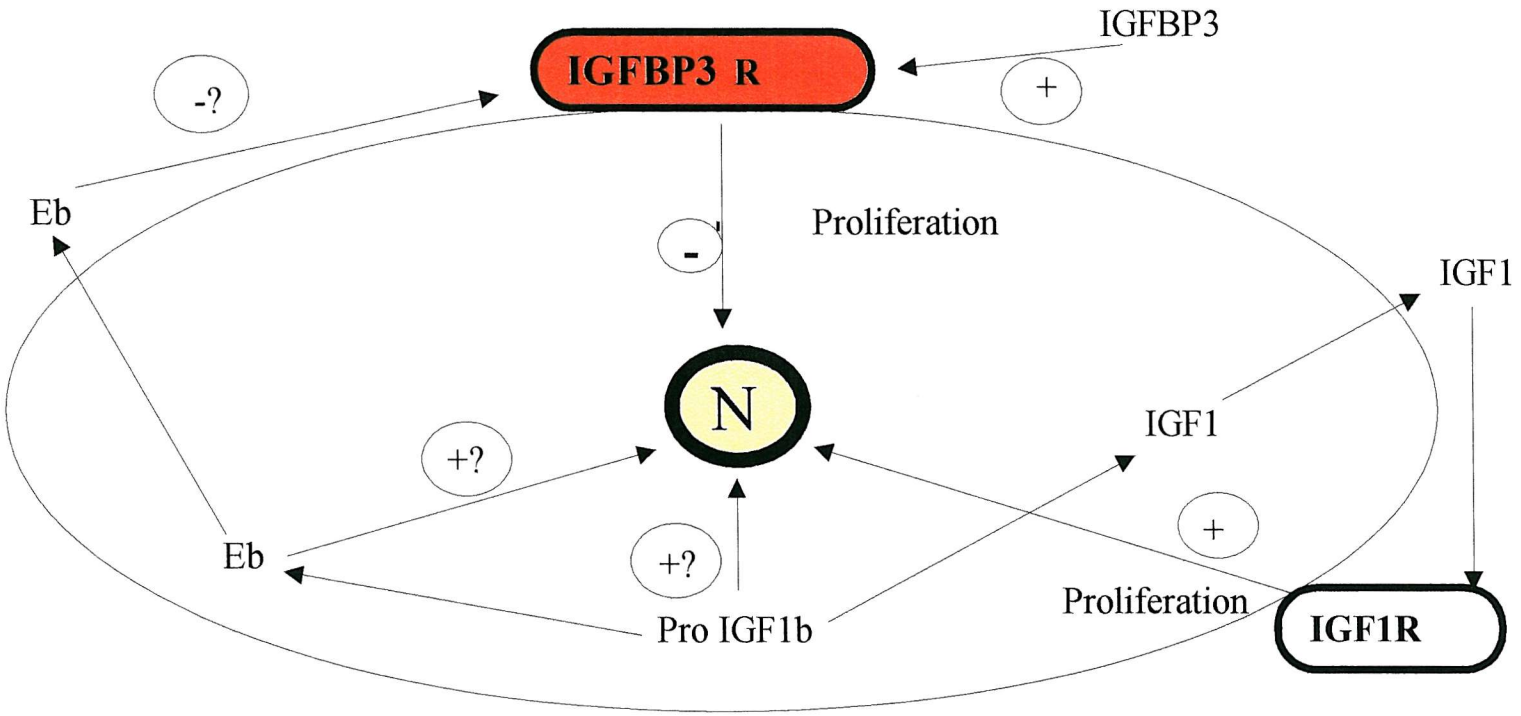


Appendix 5: Cell proliferation by IGF1 in CRK-II overexpressing cells



Appendix 6 Role of IGF1b Hypothesis

After proteolytic cleavage of IGF1b into mature IGF1 and Eb, Eb or Eb fragments may be secreted from the cell via its/their own cell surface receptor



Appendix 7:

Role of IGF1b Hypothesis 2: Eb or Eb fragments may compete with receptor binding for IGFBP3 and block the growth inhibitory effects of IGFBP3 via common highly basic motifs

Map adapted from J Garcia-Aragon *etal.*

SNP	Sequence
GH1	
5157G-F	ATTTATAAAAATGGCCCACAAGAGATCG
5157A-F	ATTTATAAAAATGGCCCACAAGAGATCA
5187C-F	ATTATTCCAAGGCCCAACTCCCAGC
5187A-F	ATTATTCCAAGGCCCAACTCCCAGA
5187G-R	TAGTTGTCCAACAGGACCCTGAGTGTTG
5187T-R	TAGTTGTCCAACAGGACCCTGAGTGTTT
IGF1	
3123G-F	CATTATTTGAATTGAGCACCTCAAGAAG
3123T-F	CATTATTTGAATTGAGCACCTCAAGAAT
3276C-F	TACTTCTTTTTATTTCTTGTCCCCAGC
3276T-F	GCACTTCTTTTTATTTCTTGTCCCCAGT
3276G-R	GAGAGGGAATAATTTTAAAAGGTACTCG
3276A-R	GAGAGGGAATAATTTTAAAAGGTACTCA
6721A-F	ATTACAGGAGGGACTCTGAAACCTCTAA
6721G-F	ATTACAGGAGGGACTCTGAAACCTCTAG
6721T-R	ACACAGATAAAAGATGTAAGTAGACTGT
6721C-R	ACACAGATAAAAGATGTAAGTAGACTGC

Appendix 8:

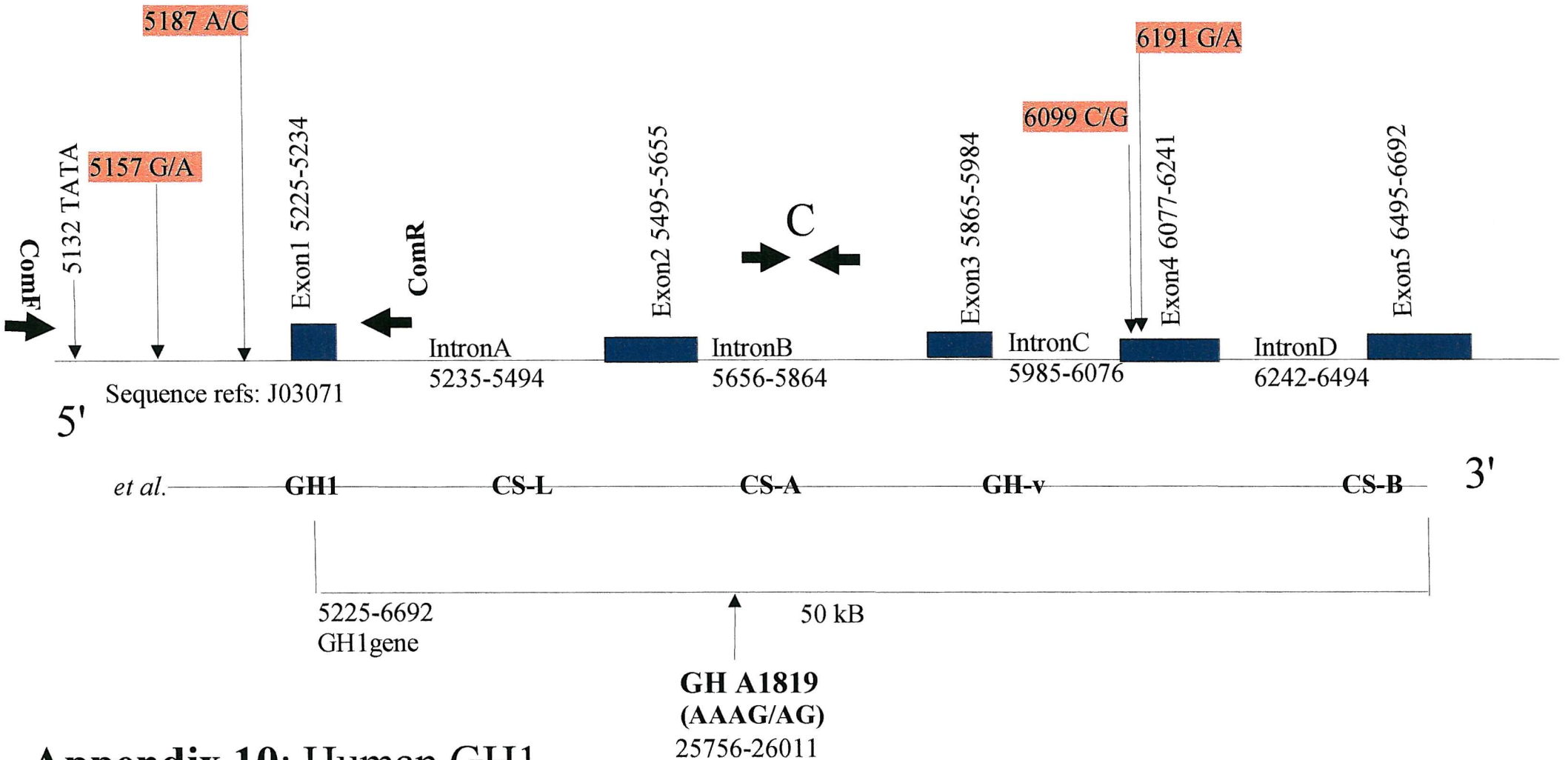
Sequences of allele specific primers used for the human GH1 and IGF1 SNPs.

Appendix 9: Control DNA

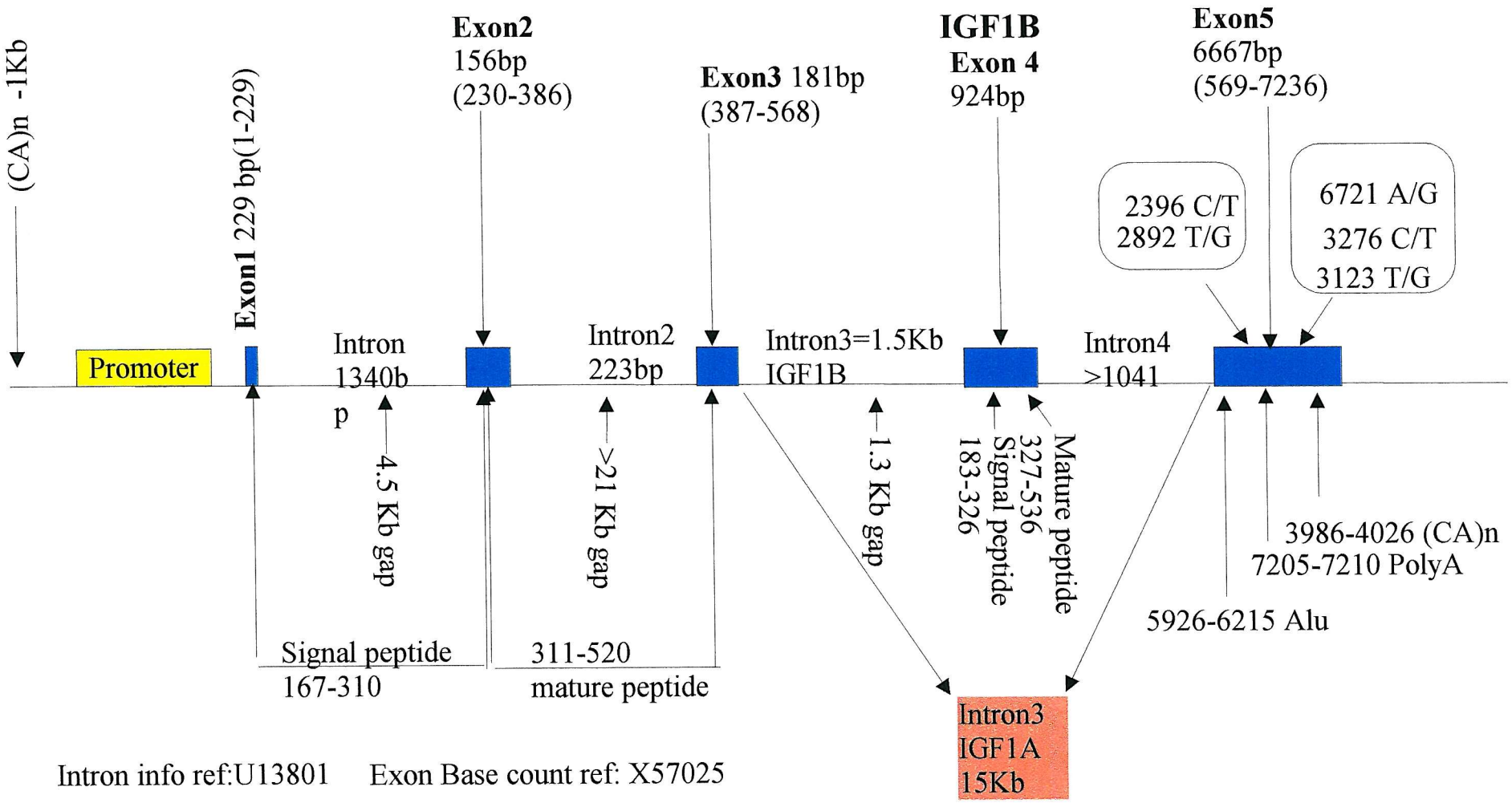
DNA samples prepared by P.Briggs

Stock DNA		100ng/ml	100ng/ml for each control Total volume 1ml final concentration 10			Control Number
OD ratio 260/280	conc. mg/ml	diln.	µl sample	µl water		
1.8	620	1/6.2	16.1	83.9		17
1.81	496	1/4.9	20.2	79.8		16
1.78	465	1/4.6	21.5	78.5		1
1.78	446	1/4.4	22.4	77.6		13
1.74	441	1/4.4	22.7	77.3		15
1.78	427	1/4.2	23.4	76.6		2
1.8	398	1/3.9	25.1	74.9		14
1.82	371	1/3.7	27	73		19
1.77	360	1/3.6	27.8	72.2		7
1.78	347	1/3.5	28.8	71.2		10
		Total =	235	765		

NOTE:
All controls were disolved in Tri/EDTA buffer pH 7.5
Working controls of 10ng/ml were diluted in sterile water.



Appendix 10: Human GH1



Intron info ref:U13801 Exon Base count ref: X57025

Alternative splicing accounts for the type A or B of IGF

Appendix 11:Human IGF1

Conservative implementation of Hertfordshire DNA banks for systematic molecular genetic epidemiological studies of fetal growth and adult disease

INM Day [1*], A Voropanov [1*], S Ye[1], LJ Hinks[1], P Briggs[1], X Chen[1], SD O'Dell[1], H Bendall[2], CH Fall[2], A Aihie-Sayer[2], C Cooper[2], DI Phillips[2], C Osmond[2] and DJP Barker[2].

[1] Research Division of Human Genetics, Southampton University Medical School, Duthie Building (MP808), Southampton University Hospital, Tremona Road, Southampton SO16 6YD, UK.
[2] Research Division of Fetal Origins and Adult Disease, Southampton University Medical School, MRC Environmental Epidemiology Unit, Southampton University Hospital, Tremona Road, Southampton SO16 6YD, UK.
*Poster Authors

Studies in Southampton (MRC EEU) over the past decade have demonstrated strong associations between parameters of early life growth and late life disease traits (FOAD, fetal origins and adult disease) including several facets of cardiovascular risk, diabetes and osteoporosis. The primary cohorts used in these studies were subjects born in Hertfordshire, and now in late life, for whom some early life measures were available.

A relatively unexplored hypothesis to explain these observations, is that the early and late life features are independent phenotypes conferred by the same genotype. Systematic examination of this hypothesis by genetic association analysis will become feasible over the next decade. A public domain genome-wide single nucleotide polymorphism (SNP) database of 300,000 markers (of an estimated 3,000,000) is being developed internationally over the next two years with £28m support from Wellcome Trust and a consortium of pharmaceutical companies. This database will facilitate extensive association analysis, enabling studies not previously possible.

For DNA banks, DNA stocks are finite and lymphocyte immortalisation is unlikely to be feasible. Instead, we have developed (1998-9), systems which are highly conservative of DNA templates and additionally are cost efficient for high throughput analyses. A DNA bank of 1,100 Hertfordshire subjects has been established (April 1999).

Genotype assays have been explored on a matched control DNA collection, for several categories of genetic diversity including SNPs, small insertion/deletions, microsatellite and minisatellite loci and rare (unknown) mutations. We have demonstrated robust reduction of DNA bank assays to zeptomole quantities of template, with preservation of phase information for direct haplotyping of candidate genes. Such conservative systems permit a relaxation of choice of candidate from strong hypothetical justification toward systematic genome-wide measurement of the genetic contribution to the FOAD observations. We have also attended to general DNA and information management. Several assays of genetic diversity in growth factor pathways have been developed so far (May 1999) for genotype-phenotype analyses and combined epidemiological and genetic analyses will be presented.

Acknowledgements. The University of Southampton and Wessex Medical Trust are thanked for support.

COULD GROWTH HORMONE (GH) AND ANGIOTENSIN CONVERTING ENZYME (ACE) GENE VARIANTS BE ACTING AS PROXY MARKERS FOR EACH OTHER IN RELATION TO THEIR PHENOTYPIC EFFECTS?

Voropanov A.M.[1], Patel H.J.[1], Gaunt T.R[1], Holloway W.[1], Syddall H.[2], Saeyr A.A.[2], Phillips D.I.W.[2], Cooper S.[2], Day I.N.M[1]

1.Human Genetics,Southampton General Hospital, Southampton.

2.MRC EEU Southampton General Hospital, Southampton.

GH is believed to be the main regulator of postnatal growth and development and has recently been shown to have a major role in adult life.

In adult life GH regulates body composition, plasma lipid concentrations, exercise capacity and physiological well being. Central obesity, insulin resistance, elevated plasma lipids, are just some of the features presented by patients with adult GH deficiency. Life expectancy is reduced with cardiovascular disease being recognised as major contributor to the premature mortality.

Angiotensin converting enzyme ACE, is an enzyme that plays an important role in the control of systemic blood pressure and fluid-electrolyte balance by hydrolysing angiotensin I into the physiologically active angiotensin II. Associations between the insertion or the deletion of the 287-bp Alu repeat element in the intron 16 in the ACE gene and coronary heart disease, the level of circulating enzyme, and risk of myocardial infarction have been described. A literature of 600 papers has recently extended to the positive relationships between ACE genotype and metabolic traits including insulin resistance relative to birth weight, and metabolic and weight responses to exercise. Studies in the Hertfordshire cohorts (1032 subjects representing a random population sample 583 males (mean age 66.4 years) and 449 females (mean age 66.4 years), born in East and North Hertfordshire sampled during the 1990's and with birth date range 1920-1930) reinforce and amplify these findings. (Above authors, unpublished)

Linkage disequilibrium is the non-random distribution into the gametes of a population of the alleles of genes that reside on the same chromosome. The distance over which disequilibrium extends has been the subject of considerable debate, with Kruglyak's computer model estimating it to 3 kb, Collins estimation to over 100 kb and Reich's recent experiment reporting LD at a distance of 60 kb.

The growth hormone gene (GH1) is located in a 50 kb cluster of homologous genes including GH1, CSHP1, CSH1, GH2, and CSH2 (GH2, growth hormone variant encodes placenta-specific growth hormone, CSHP1 is a CHS pseudogene, CSH1 CSH2 encode the chorionic somatomammotropin hormone 1, and 2 respectively).The GH1 gene is sited approximately 200 kb away from the ACE gene.

We have therefore undertaken preliminary examination of linkage disequilibrium between GH and ACE genes, on the grounds that some or all genotype associations might represent ACE alleles 'proxy marking' causative alleles actually in the GH gene or nearby. Long PCR amplification of the GH genes and ACE genes followed by internal single nucleotide polymorphic typing by allele-specific PCR has been undertaken on 1,100 Hertfordshire subjects.

We have identified moderate (20% of maximum possible) linkage disequilibrium between the GH1 5157 SNP and ACE I/D calling into question, the presumed aetiological pathway of all ACE associations reported in the literature.