# Genetics of General Cognitive Ability

By
Dragana Turic





ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

# TABLE OF CONTENTS

# ACKNOWLEDGEMENTS

I would like to thank Prof Mike Owen for giving me the opportunity to work in the Department of Psychological Medicine and be involved in an exciting research project that culminated in the creation of this manuscript. I would also like to thank my dear colleagues at the Department for all the support, advice and guidance that proved invaluable for my professional development. Thanks to Varuni, Gillian and Mark who helped me through my difficult times by being true friends and believing in me. Thanks to Nigel who taught me how to enjoy science and for remaining "on my side" all these years. Thanks to Paul Fisher, Hywel and Nadine. Last, but not least, I thank my husband Darko, my parents, and my dearest daughters Mattea and Anya as all this would not be possible or have any meaning without them.

# ABBREVIATIONS

| | |
|---|---|
| AD | Alzheimer's Disease |
| AIP | Allele image pattern |
| AMPA | α-amino-3-hydroxy-5-methyl-4-isoxazolepropionate |
| APOE | Apolipoprotein E |
| AST | Allele specific test |
| bp | Basepair |
| CaMKII-α | Calcium/calmodulin dependent protein kinase II alpha |
| cM | Centimorgan |
| DHPLC | Denaturing High-Performance Liquid Chromatography |
| DNA | Deoxyribonucleic acid |
| dNTP | Deoxynucleotide |
| ddNTP | Dideoxynucleotide |
| DZ | Dizygotic |
| EEA | Equal environment assumption |
| EEG | Electroencephalography |
| GRR | Genotype relative risk |
| HRR | Haplotype relative risk |
| IQ | Intelligence quotient |
| kb | Kilobase |
| LD | Linkage disequilibrium |
| LOD | Logarithm of odds |
| LTD | Long term depression |
| LTP | Long term potentiation |
| MRI | Magnetic resonance imaging |
| MZ | Monozygotic |
| NCV | Nerve conduction velocity |
| NMDA | N-methyl-D-aspartate |
| PCR | Polymerase chain reaction |
| QTL | Quantitative trait locus |
| RFLP | Restriction fragment length polymorphism |
| RNA | Ribonucleic acid |
| SNP | Single nucleotide polymorphism |
| STR | Short tandem repeat |
| UTR | Untranslated region |

# SUMMARY

General cognitive ability (*g*) is a general mental ability to reason, solve problems, comprehend complex ideas, think abstractly, learn quickly and learn from experience. Currently used IQ tests are excellent predictors of *g*. Heritability estimates for *g* range between 0.30 and 0.75 making it one of the most heritable human behavioural traits. Many behavioural phenotypes, including *g*, can be described as complex traits. Inheritance of such traits is governed by a mixture of genetic and environmental factors. Genetic factors contributing to total variance in *g* are likely to be numerous and additive in nature. In order to identify some of the genetic loci contributing to the total variance in *g* two approaches were employed. First, a genome-wide association study and second, candidate gene study. Genome-wide association study involved testing 1847 microsatellite markers with an average spacing of 2cM. Markers were initially screened on "original" DNA samples. This was followed by testing all positive findings on an independent "replication" sample set. Only one marker, D4S2460, was significant when all the stages of the study were completed. Investigation of candidate genes involved testing of known Apolipoprotein E (ApoE) promoter polymorphisms and Calcium/calmodulin dependant protein kinase II alpha (CaMKII-$\alpha$) polymorphisms identified in our laboratories through mutation detection techniques. None of the polymorphisms in either of the two genes showed statistically significant association with a general cognitive ability.

# Genetics of General Cognitive Ability

## 1 INTRODUCTION

*"Of the high importance of the intellectual faculties there can be no doubt, for man mainly owes to them his predominant position in the world."*[8]

### *1.1 What is intelligence ?*

It would be appropriate to begin with a scientific definition of intelligence. However, finding a single definition applicable to all times, places and cultures proved to be a difficult task for several reasons. First, intelligence is unlikely to be a single process with easily observable boundaries, detached from other human faculties such as learning and memory [18]. Second, different cultures have different conceptions of intelligence; designating as "*intelligent*" sets of cognitive, social, and behavioural attributes valued as adaptive to the requirements of living in those cultures [21]. Third, intelligence tends to be defined differently across different scientific disciplines, and the issue becomes even more complicated by the introduction of theories of multiple intelligences [22].

So, to answer the question "what is intelligence?" and by doing so to lay the foundation for this study, it seemed reasonable to choose a definition regarded as

mainstream among leading researchers on intelligence. The mainstream statement, first published in 1994, says that intelligence is a general mental capability that involves the ability to reason, plan, solve problems, think abstractly, comprehend complex ideas, learn quickly and learn from experience [23].

## *1.2 History of intelligence research*

### 1.2.1 Early beginnings

Variation in intelligence became the subject of scientific study in the last half of the 19th century. Charles Darwin's theory of evolution by natural selection inspired most of the work carried out in those early days of intelligence research. Darwin introduced the idea of advancement in intellectual capacity through natural selection and transmission of inherited intelligence as a key step in human evolution. His younger cousin, Sir Francis Galton, set out to prove this by collecting data on the intellectual capacities of successful families in Britain at that time. Galton's first impressions on the subject and strong belief that "talent is transmitted by inheritance in a very remarkable degree" were published in 1865 [24], but most of the data and evidence that intellectual superiority runs in families were presented in "*Hereditary Genius*" [25]. This book opened a permanent and sometimes heated debate over the relationship between intelligence and heredity.

Galton was fully aware of the difficulties of distinguishing between the effects of heredity (Nature) *versus* those of the environment (Nurture) and chose a powerful tool – twins. His early studies were based on a substantial collection of anecdotal stories about twins [26], but recognising the weakness of the approach he decided to devise a

test that would objectively measure differences in mental abilities. This test was meant to probe the sensitivity of perceptions by measuring sight, hearing, sensitivity to slight pressure and response to simple stimuli. Even though such measures of intellectual ability were soon discredited, Galton's achievements in the field continued to influence research on intelligence to this day.

In a similar attempt to assign exactness and consistency to measurements of intellectual abilities, James McKeen Cattell suggested a series of ten mental tests aimed at measuring individual differences in fundamental mental processes [27]. There are certain attributes that a person must have to be a successful university student and intelligence is surely one of those. However, failing to predict university students' grades, Cattell's tests were found to be an unreliable tool for measuring intelligence [18].

### 1.2.2 20th Century

The beginning of the 20th century marked another turning point in intelligence research mostly owing to Alfred Binet and his contribution to mental test design. He believed that the ability to "judge well, to comprehend well and to reason well" together with the ability of "adapting one's self to circumstances" are essential to natural intelligence and that mental tests should be designed to probe for these rather than sensory acuity [28]. The Binet scales, as we know them today, formed the basis of modern IQ tests [29].

Most of the further development of IQ tests was carried out by Henry Goddard, Lewis Terman and Robert Yerkes. In 1916 Terman produced a revised version of Binet's test, the Stanford-Binet test, that became a standard against which all later IQ tests were judged. This test is administered individually by a trained person, which is not only time consuming but has a potential problem of introducing variation in scoring.

Yerkes saw the advantage in having a test that could be administered to a large number of people in a short space of time and introduced the first version of it for the purpose of selecting US Army recruits upon America's entry into the First World War. By the 1920s such IQ tests were widely used in primary and secondary schools as well as by some universities as an entrance requirement.

Intelligence research entered a difficult period during the 1920s when misuses of IQ testing started to occur. Goddard, Terman and Yerkes were firm believers that feeble-mindedness was hereditary and that mating between such individuals should not be allowed:

*"Feeble-mindedness is hereditary and transmitted as surely as any other character. We cannot successfully cope with these conditions until we recognise feeble-mindedness and its hereditary nature, recognise it early, and take care of it."* [30].

In response, Virginia, followed by a number of other American states, passed a 1924 Eugenical Sterilization Act based on a Model Eugenical Sterilization Law published a decade earlier [31]. Sterilisation of people in institutions for the mentally ill and retarded continued through the mid 1970s resulting in more than 60,000 Americans enduring involuntarily sterilisation on the grounds of mental retardation [32].

Racial and ethnic differences in IQ scores were yet another sensitive issue. Controversy started after Goddard discovered that a high proportion of immigrants from eastern and southern Europe had IQ scores below average and used this as evidence that differences in intelligence between races are hereditary. This period in American history was difficult, both economically and politically, and politicians readily embraced Goddard's argument to support their own oppressive ideas. In an attempt to reduce the

number of immigrants, particularly those from "black-listed" countries they speeded up the legislation of the Immigration Act of 1924. Critics responded immediately pointing to the cultural bias of the tests that placed immigrants to the United States at a distinct disadvantage.

The Second World War marked another period of advance in test design. Even though it was well known that tests were capturing individual differences in intellectual capacities, no-one was attempting to use testing results to promote discriminatory laws. Attacks on testing faded.

Environmentalism seriously hampered intelligence research during the 1960s. This was the period when leading psychologists fiercely argued against the role of genes in the origins of individual differences in mental abilities. The reaction to Arthur Jensens's article (1969) on genetic influence on ethnic differences in IQ scores was immense. Jensen's monograph lead to unjust criticism of all behavioural work carried at that time bringing intelligence research almost to a halt. One beneficial consequence of the criticism was that it induced a series of bigger and better designed studies that used family, adoption and twin designs. The large amount of good quality data from such studies contributed to a final shift in scientific thought that occurred in the 1980s. It brought the acceptance of the idea that differences in intelligence are significantly associated with genetic differences between individuals [33].

## *1.3 General cognitive ability*

### 1.3.1 Galton's attempt

The prevailing belief during the 19th century was that the mind is a combination of a number of separate functions, called *faculties*. Faculty psychology assumed the existence of a distinct faculty for each and every mental activity [10]. Unlike most, Galton firmly believed that differences in general metal ability, rather than in faculty psychology, underlie individual differences in mental performance.

He hypothesised that differences in reaction times to auditory and visual stimuli would provide explanation of individual differences in higher and more complex mental functioning. Galton had measured the reaction times of thousands of subjects in an attempt to provide support for the hypothesis. Unfortunately, the large amount of data failed to show correlation with any criteria of intellectual distinction and his rejection of faculty psychology remained based mainly on general impressions rather than empirical evidence.

Charles Edward Spearman (1863-1945), much impressed by Galton's hypothesis, was puzzled by his inability to provide scientific evidence to support the idea. After re-analysing Galton's data by modern statistical methods he managed to show that the main reason for Galton's failure to prove correlation between simple and complex mental processes was in essence measurement error. Without reliable measures it was impossible to show any substantial correlation.

Probably the greatest of Spearman's contribution to intelligence research was the invention of a powerful statistical tool, factor analysis, which led to the discovery of *g* and publication of his two-factor theory in 1904. Proper explanation of the meaning of *g*, as defined by Spearman, requires some understanding of factor analysis, the principles of which will now be outlined.

### 1.3.2 Principles of factor analysis

For any given set of variables it is possible to construct a correlation matrix. A *matrix* is an array of numbers arranged in rows and columns. If these numbers are correlations of each variable with every other variable, the matrix is called a *correlation matrix.* When all correlations in the matrix are positive, it is called a *positive manifold* [10]. The technique of factor analysis is a procedure aimed to simplify complex correlation matrices with a large number of variables in order to allow us to see if there is any emerging pattern in correlations otherwise hard to detect in raw data [18].

Spearman generated one of the first matrices of this kind from data on students' ranks based on ratings in five subjects: Classics, French, English, Mathematics and Music plus a measurement of pitch discrimination (carried out by Spearman himself). He reasoned that if only one common factor can explain all the correlations among variables, the matrix will show a hierarchical order and the correlation coefficients will decrease in size the further away they are from the upper left corner (Table 1) [10].

| **Variable** | **Classics** | **French** | **English** | **Math** | **Pitch** | **Music** | ***g* loading** |
|---|---|---|---|---|---|---|---|
| **Classics** | | 0.83 | 0.78 | 0.70 | 0.66 | 0.63 | *0.958* |
| **French** | 0.83 | | 0.67 | 0.67 | 0.65 | 0.57 | *0.882* |
| **English** | 0.78 | 0.67 | | 0.64 | 0.54 | 0.51 | *0.803* |
| **Math** | 0.70 | 0.67 | 0.64 | | 0.45 | 0.51 | *0.750* |
| **Pitch** | 0.66 | 0.65 | 0.54 | 0.45 | | 0.40 | *0.673* |
| **Music** | 0.63 | 0.57 | 0.51 | 0.51 | 0.40 | | *0.646* |
| **Mean *r*** | *0.720* | *0.678* | *0.628* | *0.594* | *0.540* | *0.524* | |

**Table No. 1:** Spearman's correlation matrix and $g$ loadings [10]

Spearman proposed a method of *vanishing tetrad differences* as means of testing whether the single factor is responsible for all the correlations observed. A simple equation applied to each of the tetrads within the matrix produces a series of results which should approach zero (from either the positive or negative side) if the variables have only one factor in common (Figure 1) [10].

| **tetrad a** | | **tetrad b** | | **tetrad c** | | **tetrad d** | |
|---|---|---|---|---|---|---|---|
| 0.78 | 0.67 | 0.67 | 0.64 | 0.83 | 0.66 | 0.79 | 0.63 |
| 0.70 | 0.67 | 0.57 | 0.51 | 0.57 | 0.40 | 0.45 | 0.40 |

a) (0.78 x 0.67) – (0.70 x 0.67) = 0.054

b) (0.67 x 0.51) – (0.57 x 0.64) = -0.024

c) (0.83 x 0.40) – (0.57 x 0.66) = -0.044

d) (0.70 x 0.40) – (0.45 x 0.63) = -0.003

**Figure No. 1:** Calculating tetrad differences [10]

The fact that correlation between different variables is not the same means that variables contain the common factor to various degrees, in other words they have different *factor loadings*.

### 1.3.3 Extraction of Spearman's *g*

When all the variables used in the correlation matrix can be characterised as mental abilities (as in Table 1), the single factor they have in common Spearman termed the ***general factor, general intelligence*** or ***g*** [34].

Testing how well *g* can explain the observed correlation between two mental ability test scores is performed by reconstituting their correlation coefficients. This is done by multiplication of tests' *g* loadings. Subtracting reconstituted correlations from initially observed correlations for the whole matrix gives us a *residual matrix*. The residual matrix shows what is left (correlation wise) after the *g* factor is removed (Table 2) [10].

| Variable | Classics | French | English | Math | Pitch | Music |
|---|---|---|---|---|---|---|
| **Classics** | | -0.01 | 0.01 | -0.02 | 0.01 | 0.01 |
| **French** | -0.01 | | -0.04 | 0.01 | 0.06 | 0.00 |
| **English** | 0.01 | -0.04 | | 0.04 | 0.00 | -0.01 |
| **Math** | -0.02 | 0.01 | 0.04 | | -0.05 | 0.03 |
| **Pitch** | 0.01 | 0.06 | 0.00 | -0.05 | | -0.03 |
| **Music** | 0.01 | 0.00 | -0.01 | 0.03 | -0.03 | |
| **Mean r** | *0.001* | *0.002* | *0.000* | *0.000* | *-0.004* | *-0.002* |

**Table No. 2:** Residual correlation matrix after g is removed [10]

If no other factor/factors in addition to *g* can be extracted, as was the case with Spearman's matrix, the residual matrix will have values approaching zero in all fields. Any significant residual correlation would mean that more than one factor is implicated in explaining correlation between variables.

### 1.3.4 Spearman's two-factor theory

After calculating the size of the *g* contribution to the total variance in six variables examined, it became obvious to Spearman that only about two thirds of it could be explained by *g*. The rest he attributed to *specific factors*.

Spearman's famous two-factor theory states that individual differences on any measurement of any mental ability are due to only two factors uncorrelated with each other: a general factor, *g*, common to all mental abilities and a specific factor, *s*, specific to each and every mental ability measurement [34]. So, tests for specific mental abilities do measure those abilities, but they all reflect *g* to varying degrees as well. Years later, in an attempt to explain the meaning of *g* and *s*, Spearman compared *g* to the power generator and *s* to the different engines, each of which is powered by *g* but also has its own level of efficiency independent of *g* [35].

How well the two variables (*i.e.* scores on two different mental ability tests) will correlate is highly conditional on the level of their *g* loadings. The more complex the cognitive process measured is, the greater the correlation between the test and *g* and its contribution to overall *g* score [35]. Therefore, if one wants to asses a person's level of *g* it is more efficient to choose a highly *g* loaded test.

Spearman invested a substantial amount of work in developing a test that could potentially be administered to any individual older than three, regardless of cultural and educational background. The test was further developed by his student, John Raven,

and is now known as Raven's Progressive Matrices. Of all the tests measuring mental abilities in man, this test has the highest *g* loading (around 0.80) and is often used as a marker test for Spearman's *g* [10].

### 1.3.5 *g* and *s* and something else

It was not long after the publication of Spearman's "two factor theory" that its critics started to emerge. Sir Cyril Burt (1883-1971) was one of the leading psychologists of the time who argued that a general factor and specific factors cannot explain all the correlations between various mental ability tests. He suggested that in addition to two extremes, *g* and *s*, there are factors that are common to some but not all variables [36]. Those were later termed *group factors*.

In support to his argument Burt offered a substantial amount of data collected on a number of different tests and showed that the existence of group factors was the rule rather than the exception. Other psychologists agreed that Burt's introduction of group factors in addition to specific factors and *g* was more acceptable than Spearman's quite rigid two-factor theory. This marked the birth of the first multiple factor theory and the start of new era in intelligence research.

### 1.3.6 Models and characteristics of *g*

When performed on correlation matrices composed of a variety of tests of cognitive ability, factor analysis may give different explanations of the relationship between *g* and

the other factors. These different ways of looking at those relationships are termed "models of *g*" [10]. There are four basic models of *g*:

(a) The Spearman Model

It has the simplest structure of the four. The general factor, *g*, is present in all variables and is the only factor they have in common. If the method of vanishing tetrads shows that more than one factor can be extracted, the model will give an incorrect value for *g*. See Figure No. 2.

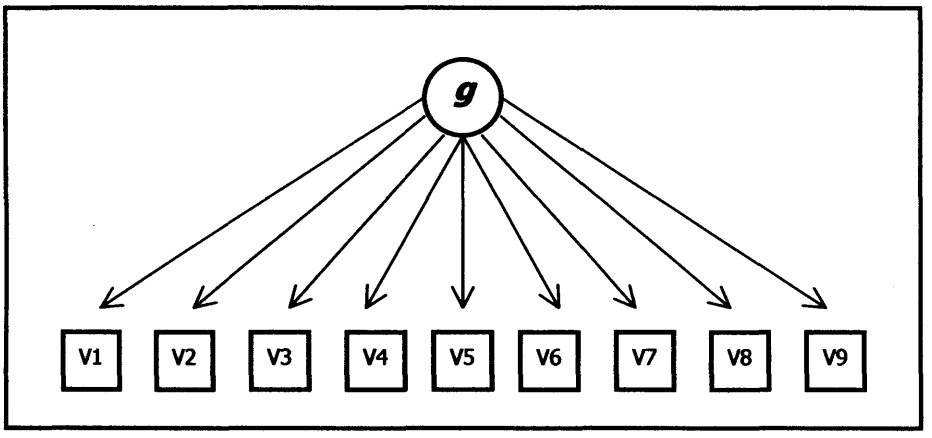



**Figure No. 2:** The Spearman model of *g*

(b) The Thurstone Model

This multifactor model proposes existence of group factors, but no general factor common to all the variables. The model never held up even when applied to mental tests, devised by Thurstone himself, aimed to measure primary mental abilities as purely as possible. All the variables correlated with each other because they all measured *g* in addition to primary mental abilities [10].

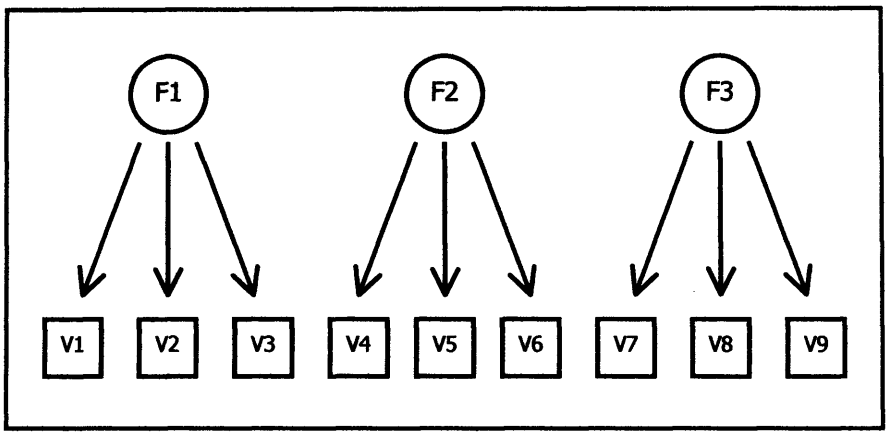



**Figure No. 3:** The Thurstone Model of *g*

(c) The Bifactor model

The model is characterised by existence of both Spearman's *g* and Thurstone's group factors. Variables' correlations with group factors and *g* are independent from each other.

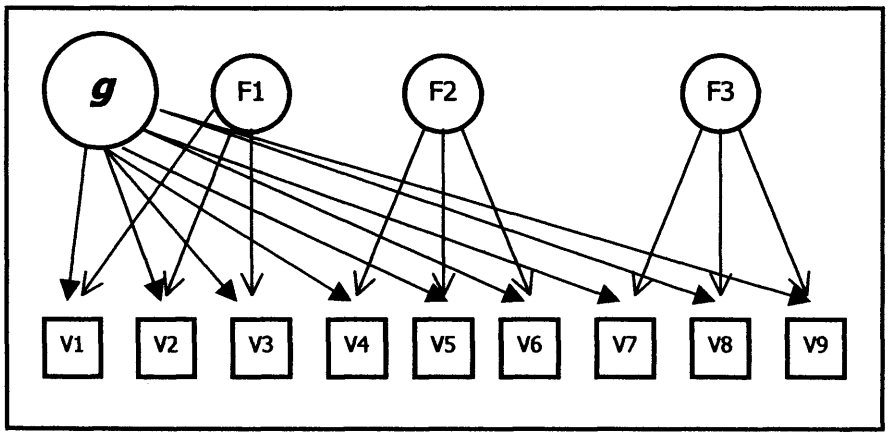



**Figure No. 4:** The Bifactor model of *g*

(d) The Orthogonalised Hierarchical Model

This is the preferred model of *g*. It consists of variables, at the bottom, first-order factors above and a second-order factor, *g*, at the top (Figure No. 5). Correlation between variables gives rise to first-order factors (group factors), while correlation between group factors (F) produces *g*. Typically, *g* accounts for a larger proportion of the total variance than all of the other factors combined. Each of the variables is also characterised by uniqueness (u), which is the variance not explained by any of the higher-order factors [10].

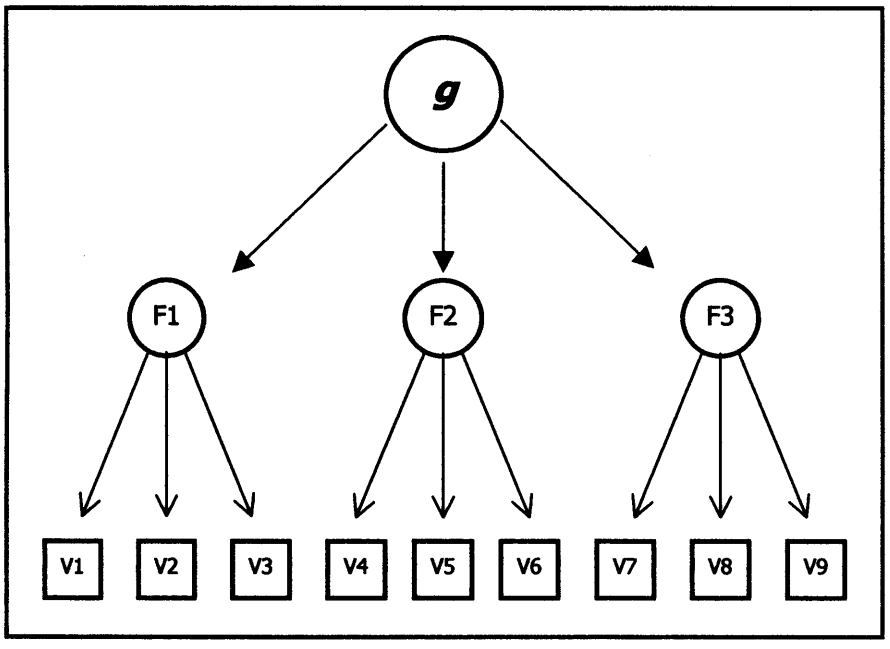



**Figure No. 5:** The Orthogonalised hierarchical model of *g*

### 1.3.7 Consistency of *g*

The basis of *g*, as described above, is that the correlations between a number of cognitive ability tests are all positive. How invariant *g* will be depends highly on the collection of tests used and the method of factor analysis applied to the correlation matrix. In other words, no method of factor analysis will produce exactly the same value of *g* when applied to different test batteries, and there are no two methods of factor analysis that will give the same value of *g* even when applied to the same set of mental tests [10].

This can be explained by the fact that when creating a set of tests we are not able to achieve a perfect representation of all possible cognitive tests. Instead, we will have a limited sample of tests that will yield somewhat different *g* than another limited sample. It has to be stressed, though, that these variations in *g* are very small and that a very similar *g* emerges from most collections of quite diverse cognitive tests. This observation is of great importance, because it gives a substantial weight to the test constructors' claims that their tests do measure something (general intelligence) that is essential for good performance on a variety of mental tests [18].

Another interesting question is whether the *g* loadings of tests are the same for subjects who differ in general intelligence. Detterman and Daniel (1989) reported a study that showed how average correlation among the tests declines as the general ability level increases [37]. Their data indicated that the amount of variance in intelligence attributable to *g* varies inversely with general intelligence [37].

### 1.3.8 Relationship between *g* and IQ

The intelligence quotient (IQ) refers to a score on a mental ability test. Characteristically, raw scores obtained are standardised within narrow age intervals on a large random sample of a specified population. This produces IQ scores that will have same mean (100 points), standard deviation (15 points) at every age level, and approximately normal distribution in the standardised population. Approximately 95% of the population has scores within two standard deviations of the mean, between 70 and 130 (Figure No. 6) [38].

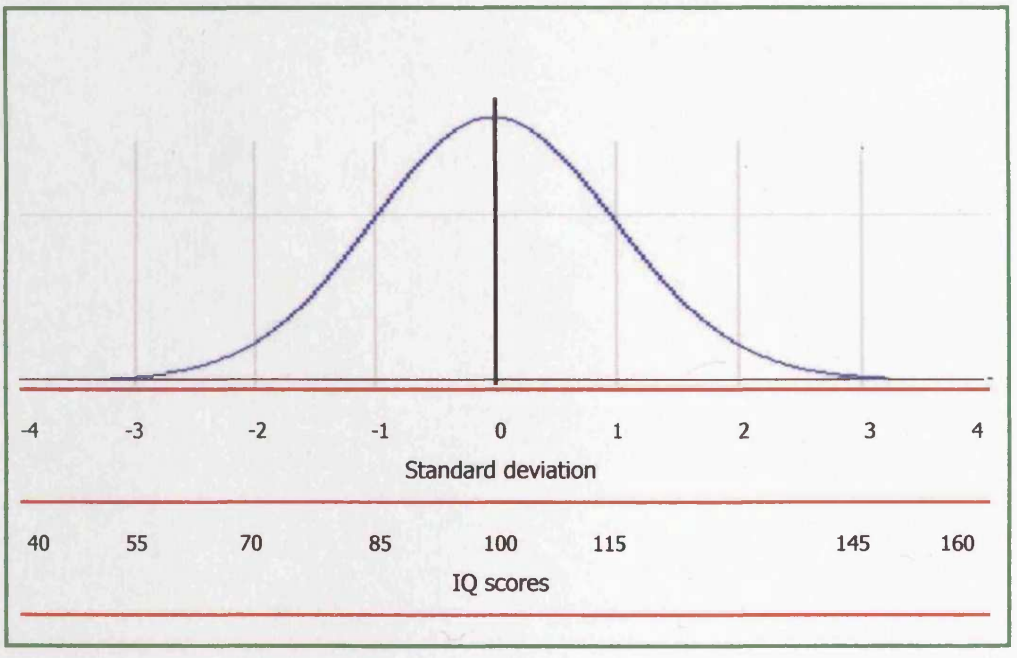



**Figure No. 6:** Population distribution of IQ scores

The main characteristic of all IQ tests is that they are all highly *g* loaded. The most *g* - rich items in the test will be those that involve some sort of deductive reasoning, spatial visualisation, quantitative reasoning and verbal reasoning [10]. The best predictors of *g* will be tests with the highest *g* loading (*i.e.* Raven's Progressive Matrices) and very small loadings on less general factors (*i.e.* group factors) [18].

### 1.3.9 *g* and educational attainment

Intelligence tests do predict school performance fairly well. Correlations between IQ scores and academic performance usually range around 0.50 [39]. A few studies have investigated the relationship between intelligence and academic achievement [40, 41, 42, 43]. They provided strong evidence for the hypothesis that all tests of intelligence measure intellectual achievements that are determined by the ability to acquire knowledge and are therefore tests of achievement [39]. The nature of correlation between intelligence and educational attainment is such that IQ scores at one age predict educational success at a later age, as shown by independent longitudinal studies performed in Britain [44] and America [45, 41].

So, it seems reasonable to say that mental ability tests measure something that is related to success in one aspect of our lives where intelligence has traditionally been thought to be relevant [18].

### 1.3.10 The biology of cognitive ability

An important approach to studying intelligence is to try to understand the biology behind it. Intellectual activities surely involve brain functioning and differences in such activities are likely to be reflected in differences observable in the brain. It is difficult to say which one of the number of biological theories best accounts for individual differences in intellectual performance. Quite possibly, a true biological explanation of intelligence will be a combination of several theories.

#### 1.3.10.1 Early biological theories

The first biological theories started to emerge during the 1950s. Hebb, the most influential psychologist of that time, proposed the existence of two types of intelligence: A, a person's innate potential and B, a product of actual brain development [21]. He also suggested that learning, which is an important basis of intelligence, is achieved through cell assemblies and through creating more and more complex connections between neurons [21].

#### 1.3.10.2 Brain activity

More recent theories deal with more specific aspects of neuronal functioning. Two main approaches dominate the field: (1) correlation between electroencephalography (EEG) traces and intelligence, and (2) speed of processing.

The first method examines the relationship between intelligence and brain activity as recorded by electrodes positioned on the head. A common finding in such studies is a negative association between brain activity under cognitive load and intelligence [46]. In other words, intellectually competent individuals when confronted with complex tasks are more likely to exclude irrelevant neural networks, thereby showing lower levels of brain activity than intellectually less competent individuals. Another interesting observation is that individuals with higher IQs display higher brain activity in resting conditions which is usually explained by the ability of such individuals to engage in complex mental activities when resting and/or dealing with less demanding tasks [46].

1.3.10.3 Speed of processing

The second method is based on a suggestion that high cognitive ability is associated with faster information processing as a result of more rapid and efficient central nervous functioning [47]. Speed of processing is usually estimated through measurements of either peripheral nerve conduction velocity (NCV) or brain nerve conduction velocity (reaction time – RT).

Several studies have found peripheral NCV, the speed at which an impulse travels along the nerve, to be positively correlated with general intelligence (0.20 [48] and 0.46 [49]). Interpretation of these findings requires much caution since there have been several failures to replicate [50, 51].

Reaction time (RT) measures the speed at which subjects respond to visual stimuli by pressing a button on a response console. RT has been found to be moderately negatively correlated with IQ scores (-0.43 to –0.66), that is higher IQ scoring individuals have shorter RTs [49, 51]. Unlike the correlation between NCV and IQ, RT is

more consistently correlated with IQ and this has been regarded as a more important factor in explaining individual differences in intellectual ability [50].

#### 1.3.10.4 Glucose metabolism

Several studies have investigated the relationship between cortical glucose metabolic rates and mental activities of individuals with different IQ scores. Haier (1988) showed that glucose metabolic rates during performance in complex mental tasks were lower in individuals with higher IQ scores [52]. This suggested that more intelligent people have to invest less effort than less intelligent ones to solve reasoning problems.

Interpretation of these and similar findings may not be straightforward since it would be difficult, at this moment, to distinguish whether the difference in metabolic rate is a cause or an effect, or whether both high IQ and low glucose utilisation are related to a third variable [21].

#### 1.3.10.5 Brain size and structure

Several major studies using magnetic resonance imaging have shown a positive correlation (0.40 is taken to be the best estimate to date) between *in vivo* brain size and intelligence [53, 54]. Despite some failures to replicate in studies with monozygotic twins [55] this correlation is generally thought to be genuine [56]. Moreover, an interesting sex difference has been observed in MRI brain studies: female brain size correlates more highly with verbal ability, while correlation figures in males are higher for spatial ability [50].

What is it about the size of the brain that is related to an individual's mental ability? Miller (1994) suggested that larger brains in more intelligent individuals can be explained by greater thickness of the myelin layer surrounding each nerve cell [56]. Myelination is known to occur during early childhood which is consistent with the improvement in intellectual functioning that occurs as infants mature. Also, less myelin can lead to more cross-talk along neurons, which could be a source of errors in signal transduction leading to lower intelligence [57]. Myelin degradation is associated with the normal ageing processes and is a possible explanation for general slowing in cognitive abilities with age [56].

Another structural feature that is genetically regulated and implicated in intellectual functioning is the distribution of grey matter across the cortex. Thompson *et al.* (2001) examined 20 MZ and 20 DZ twin pairs and found that the quantity of frontal grey matter is more similar in individuals who are genetically alike ($r=0.9$, $p<0.0001$ after correction for multiple testing) and that individual differences in brain structure are tightly correlated with individual differences in IQ scores ($p<0.0176$ after correction for multiple testing) [58].

### 1.3.11 Genetic models of cognitive abilities

Nearly all genetic research on intelligence is based on the so-called psychometric model [33]. This holds that cognitive abilities are positioned at three levels that are related to each other and organised in an hierarchical fashion (Figure No. 7). The psychometric model is based upon, and provides an explanation for, phenotypic relationships between various mental abilities. In terms of genetics, three different

versions of the model have been proposed: 1) *the bottom - up*, 2) *the top - down* and 3) *the level-of-processing* model [33].

The first model, bottom–up, assumes that there are genes that influence elementary abilities, the effects of which then feed into specific cognitive abilities that finally converge on general cognitive ability. The top-down model assumes that genes that influence specific and elementary abilities do so by affecting some general cognitive process like *g*. Finally, the level-of-processing model proposes that different genes are involved at different levels of processing within the hierarchy, but it also allows for the possibility that there are genes common to all cognitive levels within the pyramid.

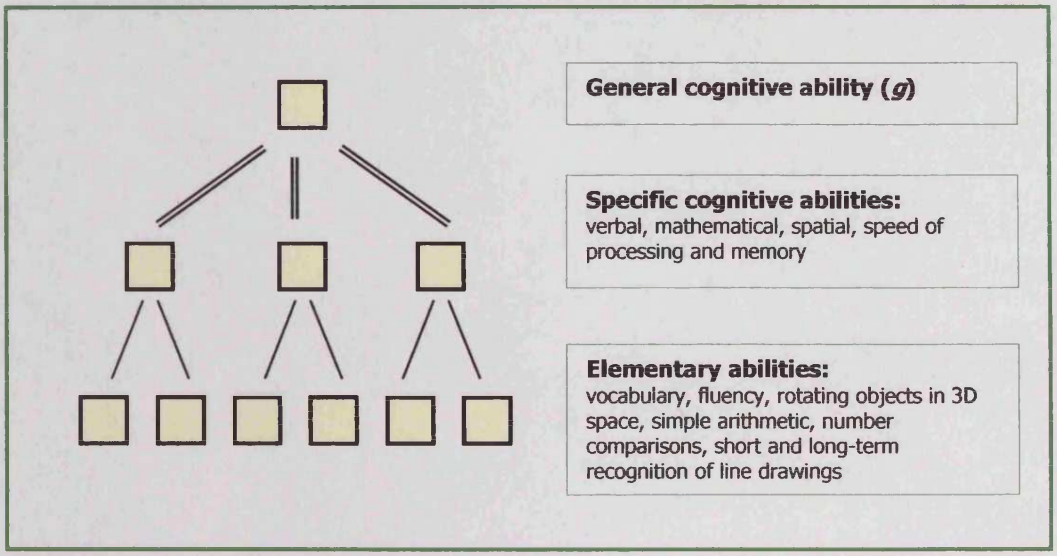



**Figure No. 7:** Hierarchical organisation of cognitive abilities

Multivariate genetic analyses in the field of intelligence research mainly focus on the relationship between general and specific cognitive abilities, searching for the source of their covariance. So far, the results imply that, to a considerable degree, genetic effects on cognitive abilities are general, whereby genes that affect one cognitive ability also

affect other cognitive abilities (top–down model). There is also evidence that some genetic effects are unique to each specific cognitive ability, as assumed by the level-of-processing model [59, 60]. The results of most comprehensive analyses incorporating all three levels of processing (general, specific and elementary) lean more towards the level-of-processing model [61].

Genetic multivariate analysis of available data thus predicts that when genes influencing *g* are identified, most of these genes will be associated with specific and elementary abilities as well, while some will be specific to certain cognitive abilities.

## *1.4 General cognitive ability as a genetically complex trait*

### 1.4.1 Complex traits

Many behavioural and neurological phenotypes of interest, including general cognitive ability, can be described as complex traits. The term "complex trait" refers to any phenotype that does not exhibit classic Mendelian inheritance, either dominant or recessive, that can be attributable to a single genetic locus [62]. A defining feature of complex phenotypes is that no single locus contains alleles that are either necessary or sufficient to produce a particular phenotype. Inheritance of such traits is governed by a mixture of genetic and environmental factors (Figure No. 8).

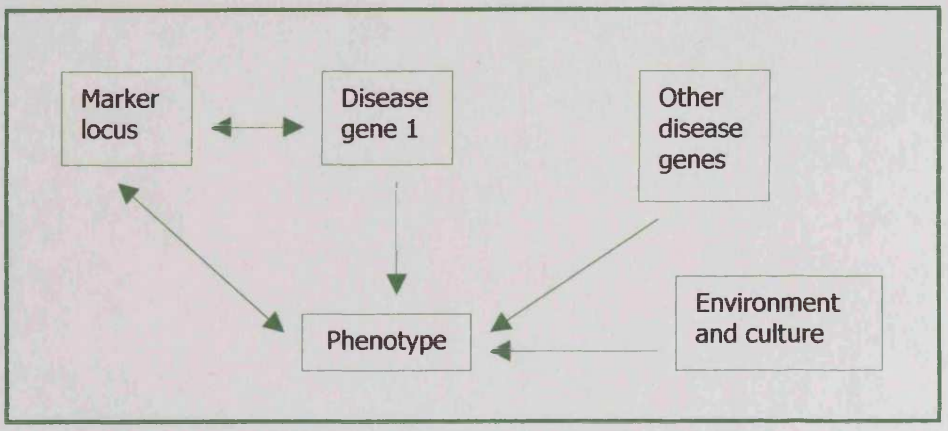



**Figure No. 8:** A simplified model of the etiological factors predisposing to complex phenotypes [4].

To some extent, even a simple genetic disease can be described as complex when closely examined [63]. However, such rare Mendelian forms are usually considered separately from more common "complex" forms that aggregate in families, but occur as a result of a combination of multiple genetic effects and environmental factors (multifactorial inheritance).

Different genetic mechanisms can lead to complex inheritance. Broadly speaking, effects on a trait can either be additive or non-additive. Additive genetic effects occur when alleles at different loci add up to affect behaviour, although the size of each allele's contribution can be highly variable [64, 63].

Alternatively, the effect of an allele can depend upon the presence of other alleles and when such interaction between alleles occurs it is called *epistasis*. *Locus heterogeneity* (mutations in any one of several genes resulting in an identical phenotype), *allelic heterogeneity* (multiple alleles existing at a single locus, different combinations of which give rise to different phenotypes), *parent of origin effects* (genomic imprinting) and *incomplete penetrance* (reduced probability that a particular genotype will result in specific phenotype) are additional mechanisms that can

contribute to the complexity of a trait. Finally, additional modes of genetic inheritance such as *genetic anticipation*, a phenomenon of increasing severity and/or decreasing age of onset in successive generations, and *mitochondrial inheritance* can further complicate determination of a precise mechanism operating in any particular trait [62, 65].

### 1.4.2 Quantitative traits

Traits that require the simultaneous presence of alleles in multiple genes are also known as polygenic. They can be further classified as *discrete* or *categorical*, with a specific outcome (*i.e.* development / no development of type I diabetes or asthma), or *quantitative*, measured by a continuous variable (*i.e.* blood pressure, height, weight, IQ scores) [62].

Each locus contributing to genetic variance in a quantitative trait, termed a quantitative trait locus (QTL), is neither necessary nor sufficient for development of a trait [66]. QTLs denote multiple genes of varying effect size so that, broadly speaking, a trait can be influenced by a few oligogenes with a moderate effect on the phenotype, or by many polygenes each with very small effect, or by a combination of the two [33].

It would be a mistake to assume that all the loci contributing to a quantitative trait such as general cognitive ability, no matter how many or few of them, are of equal effect size [64]. The magnitude of an effect exerted by any one locus can vary from major, to modest, to small. Various statistical models have shown that the most likely scenario usually involves few modest effect loci explaining much of the phenotypic variance, while the rest of it is attributed to larger number of small effect loci plus environment [63, 67, 68].

Obtaining accurate estimates for the overall number of genes that affect a given character is difficult for several reasons. First, only those genes with a sufficiently large

phenotypic effect will be detected. Second, the probability of detecting QTLs is a direct function of the distance between a QTL and the nearest genetic marker. Third, locating QTLs is highly dependent on the size of the population analysed: the larger the population, the greater the likelihood of detecting loci of lesser effect. Fourth, it is likely to be more difficult to detect QTL for traits with low heritability. Finally, using stringent criteria in declaring a significant effect lowers the chances of reporting spurious findings, but at the same time reduces the possibility of detecting small effect size QTLs. All these factors account, to a certain extent, for the underestimation of the total number of contributing loci.

### 1.4.3 The phenotypic distribution of quantitative traits

In categorical traits, it is thought that a combination of multiple genetic loci only produces a character when a threshold of susceptibility is reached [69]. For a continuous trait, the more genetic loci contribute to overall variance in phenotype, the more closely will the phenotypic distribution resemble the typical Gaussian (normal) distribution curve. In general, if there are n loci then there will be 2n + 1 possible phenotypic values (Figure No. 9) [9]. General cognitive ability with its bell-shaped distribution of IQ scores is therefore likely to be a character influenced by a larger number of genes in combination with environmental factors.

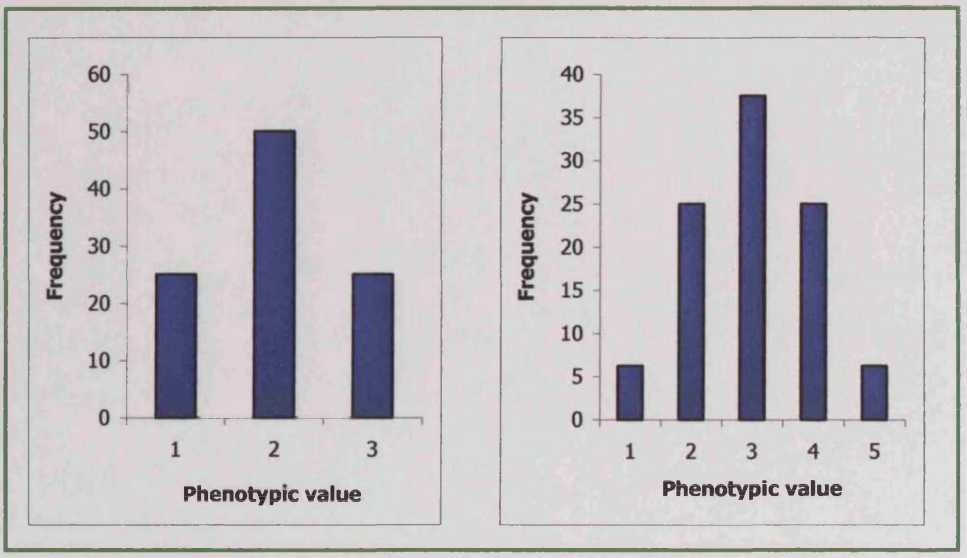



**Figure No. 9:** As the number of QTLs contributing to a trait increases, the more closely its distribution in the population approximates to a normal distribution [9].

As the number of loci increases, the simple relationship between genotype and phenotype disappears [70]. This is due not only to the larger number of possible genotypes, but also to the fact that alleles underlying quantitative traits are rarely either dominant or recessive. Instead, what is observed is the whole range of gene effects from complete dominance to complete recessiveness, and the dominance/additivity statistic (d/a) is used to describe it. A situation where d/a equals 0 means that a heterozygote is intermediate between two homozygotes and that the genetic contribution to the phenotype is entirely additive [71].

## *1.5 Mapping genes for complex traits*

### 1.5.1 Demonstrating a genetic component to a trait

Genetic epidemiological studies are required to provide evidence for the involvement of genes in complex traits. The first clue to the operation of genes often comes from evidence of familial aggregation. However, family studies do not allow the effects of shared environment to be distinguished from those of genetic factors. This requires twin and adoption studies.

#### 1.5.1.1 Family studies

Family studies examine the degree of familial clustering of a trait which can be expressed as $\lambda_r$, the ratio of the risk for relatives of patients to the population prevalence of trait (Figure No. 10) [72].

$$\lambda_r = \frac{\text{Prevalence of the disease in a relative "r" of an affected person}}{\text{Population prevalence of the disease}}$$

**Figure No. 10:** Calculating relative risk ratio

The risk ratio $\lambda_r$ decreases with the degree of relationship between proband and relatives [73]. Traits with an environmental component will show a steeper drop in $\lambda_r$ with decreasing relatedness than that predicted for a purely genetic character.

Risch (1990) suggested the use of risk ratio as a direct indicator of the mode of inheritance underlying the trait. For monogenic traits $\lambda_r$ decreases by a factor of two with each degree of relationship [73]. The same is true for traits where several gene loci are involved in its aetiology as long as they do not interact epistatically. More rapid drops in $\lambda_r$ are suggestive of multiple genetic loci and epistasis [73].

#### 1.5.1.2 Twin studies

Twin studies attempt to distinguish genetic from environmental sources of familial resemblance by comparing the phenotypic resemblance of monozygotic (MZ) twins who share 100% of their genes to the resemblance of dizygotic (DZ) twins who are on average only 50% similar genetically. In addition, both MZ and DZ twins share the same uterus, age and aspects of their early and later environment. This allows the variation of traits to be separated into genetic and environmental components.

If a trait is influenced by several genes with purely additive effect, and if there is no environmental involvement, then the trait should be twice as correlated between MZ twins than between DZ twins [33]. This relationship will remain the same if shared environment also plays a role, as long as both MZ and DZ twins experience the same level of shared environmental component [74].

The twin method relies upon the *equal environment assumption* (EEA) which assumes that environmental similarity is roughly the same for both types of twins reared in the same family [33]. Deviation from this simple relationship should be expected

for traits with an environmental component as well as those reflecting more complex gene interactions [68].

Critics of the method argue that the equal environment assumption is wrong and that, apart from being genetically more similar than DZ twins, MZ twins are often environmentally more similar [75, 74]. This, in their view, makes classical twin studies of doubtful value as indicators of genetic influence on psychiatric disorders and behavioural traits [75, 76].

In an attempt to prove the sceptics wrong, a number of studies have been conducted with the aim of testing the equal environment assumption. One of the earliest such studies examined the effect of physical resemblance of identical and fraternal twins (206 twin pairs) on the ratings for four personality traits [77]. The data suggested that greater resemblance in appearance in identical twins does not make them more similar in personality. Thus, although greater similarity in appearance may result in unequal environments for the two types of twin pairs, it does not appear to bias twin studies in the direction of inflated heritabilities [77].

Parents may treat MZ twins more similarly than DZ twins, thereby violating the EEA. To test the validity of the hypothesis Kendler *et al.* (1994) conducted a study in which they examined whether parents' beliefs about their twins' zygosity influenced twin resemblance for common psychiatric disorders. In about 20% of cases parents' beliefs about their twins' zygosity, *perceived zygosity*, disagreed with *true zygosity* of the twins [78]. The authors found that even these 'mis-labelled' twins were as similar behaviourally as correctly labelled ones, and that differential treatment of MZ and DZ twins by their parents is unlikely to cause violation of EEA [78].

Further evidence in support to the traditional twin method was offered by Kendler *et al* (1993) and Hettema *at al* (1995) who investigated the impact of physical similarity on phenotypic resemblance in five common psychiatric disorders: major depression, generalised anxiety disorder, phobia, bulimia, and alcoholism [79, 80]. They examined 1030 and 882 female-female twin pairs respectively, and concluded that the level of

physical resemblance did not have significant effect on concordance for any of the five disorders, thereby providing support for the validity of equal environment assumption [79, 80].

Similar treatment of MZ twins on the basis of their zygosity alone is, therefore, not a threat to the validity of the equal environment assumption as applied to twin studies of psychiatric disorders and behavioural traits [81, 82].

#### 1.5.1.3 Adoption studies

Adoption studies are another way of disentangling genetic from environmental causes of phenotypic variation. Adoption creates pairs of genetically related individuals who do not share a common family environment and pairs of unrelated individuals who do share the same environment. In the absence of selective placement resemblance between adopted children and their adopted parents directly assesses the environmental contribution to the parent-offspring resemblance [33].

It is true that significant selective placement has taken place in some adoption projects where correlation in IQ scores between adopted children and their adopted parents was 0.34 even before children started to live with their adopted families [83]. One way of measuring selective placement is by looking at correlation between adoptive parents and biological mothers, or between biological children of adoptive parents and biological mothers of adopted children. Any significant correlation could not be explained by shared genes, but rather the similar circumstances between biological and adoptive families [18]. If, however, correlation between biological mother and her adopted-away child is significantly higher than correlation between her and biological children of adoptive family, the only reasonable conclusion will be that resemblance

between biological mother and her child cannot be entirely explained by selective placement [18].

There are three main designs of this type of a study, each addressing a different question [18].

a) **Adoptive families *vs.* biological families**: this design is used for comparing the magnitude of correlation between children and their parents in adoptive *versus* biological families. If correlations are significantly higher between members of biological families, it suggests that they share genetic composition as well as environment.

b) **Correlations in adoptive families**: correlations in adoptive families significantly greater than zero are suggestive of common family environment being an important contributor to similarities in phenotype examined.

c) **Adopted children *vs.* biological parents**: this type of design should provide an answer to question of whether children resemble their biological parents more than adopted parents? A positive answer suggests genetic influence on a trait.

In practice, it is often possible for a single study to generate enough data necessary to provide answers to all of these questions at once [84, 85, 86].

The main obstacles in such studies are the lack of information on biological parents and potential selective placement of children into adopted families. Nevertheless, adoption studies are still the most powerful method for checking whether a character is genetically determined [70].

### 1.5.2 The heritability of *g*

#### 1.5.2.1 Defining heritability

Quantitative traits, such as *g*, tend to be normally distributed in the general population. The phenotypic variance ($V_p$) observed is due to combined action of genetic variance ($V_g$) and environmental variance ($V_e$). Both, $V_g$ and $V_e$ can be broken further down. $V_g$ is composed of additive and non-additive genetic effects, while $V_e$ is a product of shared and non-shared (special) environmental effects (Figure No. 11).

$V_P = V_G + V_E$

$V_G = V_A + V_D$, therefore $V_P = V_A + V_D + V_E$

$h^2_{broad} = V_G / V_P$

$h^2_{narrow} = V_A / V_P$

$V_P$ = phenotypic variance
$V_G$ = genetic variance
$V_E$ = environmental variance
$V_A$ = variance due to additive genetic effects
$V_D$ = variance due to dominant genetic effects
$h^2_{broad}$ = broad heritability
$h^2_{narrow}$ = narrow heritability

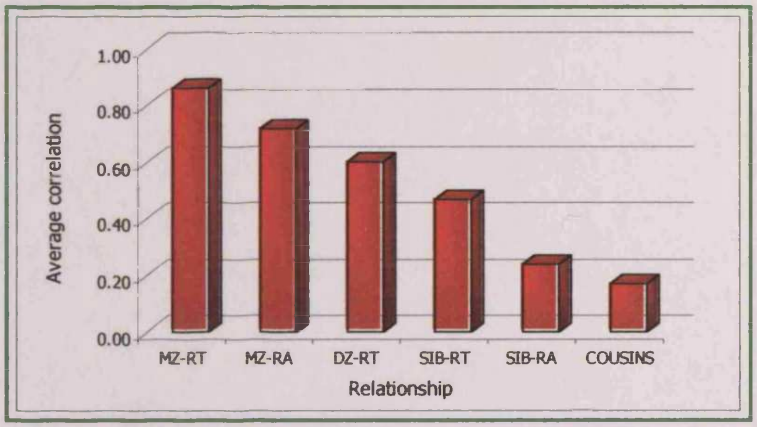
**Figure No. 11:** Variance components and heritability

The heritability ($h^2$) of a trait is the proportion of total population variation in a given characteristic that can be explained by genes [70]. It is not immutable, but can vary over time and between different populations and environments. The heritability of a trait depends greatly on the environmental variance. The lower the $V_e$, the more of the total variation in phenotype can be explained by genetic variation and the more heritable the trait will be.

### 1.5.2.2 Evidence for a genetic contribution to variance in *g*

Darwinian theory implies that members of one population vary in many characteristics and that much of that variation is inherited. Galton's work on twins was the first attempt to show that this general truth can be applied to human intelligence as well [26]. However, it took more than hundred years for a sizeable amount of good quality data to be generated.

In order to estimate the heritability of a trait one can look at the degree to which variation in environmental or genetic relatedness influences phenotypic resemblance. Bouchard and McGue published an influential review of genetic epidemiological studies of *g* up until 1981 [87]. They compiled correlation data from 111 carefully selected studies clearly showing that the higher the proportion of genes two family members have in common, the higher the average correlation between their IQ scores (average correlation for MZ twins reared together was 0.86, for MZ twins reared apart 0.72, for DZ twins reared together 0.60; a larger list of reported correlations is presented in Figure No. 12). The evidence marshalled and reviewed in this paper has often been used to support the idea that there is a substantial genetic contribution to overall variance in *g*.




**MZ-RT**: monozygotic twins reared together

**MZ-RA**: monozygotic twins reared apart

**DZ-RT**: dizygotic twins reared together

**SIB-RT**: siblings reared together

**COUSINS**

**Figure No. 12:** Dependence of average correlation in IQ scores on genetic relatedness between individuals.

### 1.5.2.3 Estimating the heritability of *g*

If MZ and DZ correlations are the same, heritability is estimated to be zero. A correlation between MZ twins of 1.0 and between DZ twins of 0.5, suggests that genetic differences among individuals completely explain their phenotypic differences. Heritability can be estimated approximately from studies involving both MZ and DZ twins. Because MZ twins share 100% of their genes and DZ twins on average only 50%, the difference in their correlations reflects half of the genetic effect. Doubling the correlation difference, the *Falconer heritability formula*, produces a heritability estimate for the trait (Figure No. 13) [2]. The equation is based on the equal environment assumption and on the assumption that all genetic influence is additive in nature [2].

$$h^2 = 2\,(r_{MZT} - r_{DZT})$$

**Figure No. 13:** Falconer formula for calculating heritability [2].

Estimating heritability from correlation data requires quantitative modelling and analysis [18]. A good model is based on several assumptions that make it more realistic. In 1990 Chipuer *et al.* provided such a model [88]. The model acknowledges a distinction between additive and non-additive genetic effects, it assumes that twins share more similar experiences than other siblings and acknowledges the possible importance of assortative mating [18].

Model-fitting analysis of the correlations reported by Bouchard and McGue [87] produced heritability estimates for *g* of about 50 percent [88]. Even though most theorists

agree with the 0.50 heritability estimate [84, 88, 89], there are some who prefer a higher figure of around 0.70 [90].

Studies involving MZ twins adopted away are important in respect that correlations observed between such individuals provide direct estimates of heritability levels for a particular behaviour dimension [33]. The Minnesota study of twins reared apart examined 45 pairs of monozygotic (MZ) twins adopted away and reported correlation between their IQ scores of 0.78 [91]. Figures between 0.60 and 0.80 for heritability of general intelligence were also reported by others [7, 92, 93]. Teasdale and Owen (1984) reported the highest IQ heritability of 0.96 [94] using siblings reared together and apart.

Pedersen *et al.* argued that higher broad heritability estimates for IQ are due to epistasis or dominance, and that a model with all non-additive genetic variance fits their data best [7]. However, other major studies on intelligence did not find evidence of non-additive genetic variance and concluded that such model is an unlikely model for general cognitive ability [94, 84, 91, 88].

Taken together, it is reasonable to say that the broad heritability of IQ in modern industrialised societies is somewhere between 0.30 and 0.75, since currently available data and quantification models do not justify much greater precision [18].

#### 1.5.2.4 Developmental changes in heritability of *g*

Most people mistakenly believe that the heritability of *g* does not change from the moment of conception. Longitudinal studies are designed to follow individuals for a longer period of time making it possible to analyse changes in the genetic contribution to phenotypic variance. The results of several such studies [95, 84, 91, 7] show that the heritability of general cognitive ability increases during the life-span, while the importance of shared environment diminishes in favour of non-shared experiences.

The Colorado Adoption Project was the first large-scale longitudinal study of behavioural development. 245 children were tested at 1, 2, 3, 4 and 7 years of age. Analysis of data provided evidence for increasing heritability of general mental ability from 9% at 1 year of age to 36% at 7 years [95].

Another change in heritability levels appears to occur at the time when most individuals leave home and start to function as adults on their own, as observed in studies following older subjects [3]. Finkel *et al.* (1995) combined data from the Minnesota [91] and Swedish studies [7], and showed that for the age periods 27–50 and 50–65 estimated IQ heritability is approaching 0.80 (Fig. No. 14) [96].

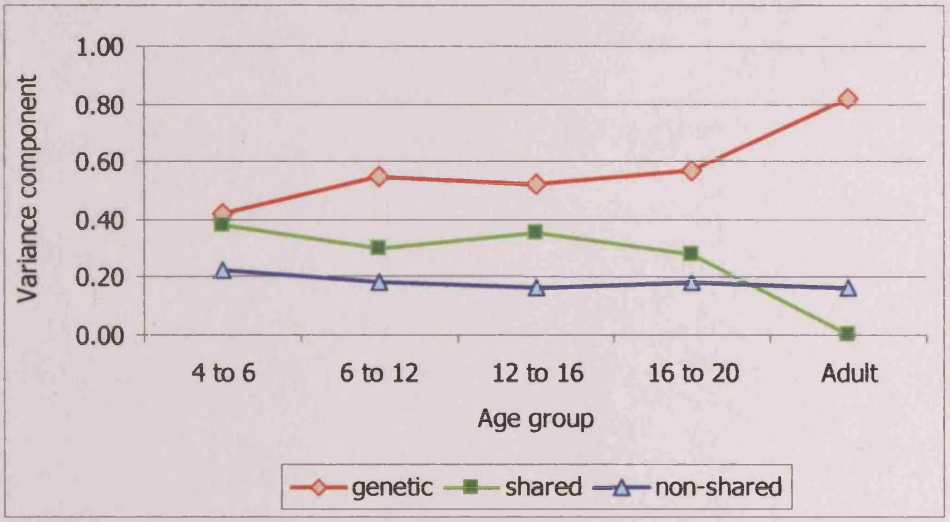



**Figure No. 14:** Effect of age on genetic and environmental sources of variation in *g* [3].

Yet, inferences have to be made with great caution. Genetic relatedness can cause similarity in experiences, which feeds back to produce similar IQ scores; so-called *gene-environment correlation* [33]. There are three types of genotype-environment correlation: passive, reactive (evocative) and active [97].

**Passive:** Children inherit environments from their parents that are correlated with their genotypes; for example, if musical talent is heritable, then musically gifted children are likely to have musically gifted parents that will provide environment necessary for development of musical ability.

**Reactive:** Individuals provoke reactions from other people on the basis of their genotypes; for example, musically talented children may be recognised at schools and given special opportunities for further development of the ability.

**Active:** Individuals seek or create environments that are correlated with their genotypes; for example, musically gifted children may seek out musical environments.

Since such effects might be cumulative, this might account for the increased correlation between IQ scores among twins with time [18]. So, children showing high correlation in IQ scores with their adoptive parents early in childhood will start creating their own environments and experiences as they grow older, driven by their genetic backgrounds. This will cause their and their adoptive parents' and siblings' IQ scores to start diverging and go further apart as they are approaching adulthood [84].

Gene–environment correlation is not clear cut. Therefore, estimating the heritability of *g* is bound to be a difficult task that would require disentangling of environmental from genetic sources of variation in *g*. Nevertheless, one thing can be concluded form all the evidence gathered so far, is that general cognitive ability is a trait substantially influenced by genes. Even if a more conservative figure of 0.50 is accepted as an estimate of heritability, it still makes *g* one of the most heritable dimensions of human behaviour and a reasonable target for genetic research.

### 1.5.3 Relating a trait to variations in DNA sequence

Having demonstrated that a trait has a genetic component, the next challenge is to identify the genetic loci involved and the way in which these determine the phenotype. Broadly speaking, there are two ways of doing this: linkage studies and association studies. Both have advantages and limitations in respect of their application to complex traits. Before considering each individually, some general points on strategies for mapping complex traits will be addressed.

#### 1.5.3.1 Molecular genetic approaches

There are three main approaches to mapping genetic variants involved in a disease/trait: functional cloning, the candidate gene strategy and positional cloning [98].

Functional cloning is based on the information about the function of an unknown gene and identification of an underlying protein leads to localisation of the gene. This approach has been useful in only few cases such as cloning of cDNA for Fanconi's anaemia (FACC gene) [99] and human coagulation factor IX (F9 gene) [100].

A candidate gene can be suggested without any prior knowledge of its chromosomal location. Candidates are selected on the basis of phenotypic resemblance with a trait for which the gene is known. Also, if the molecular pathogenesis is known it might suggest that the gene is the member of a known gene family [70].

Finally, in positional cloning the general approach is to try to isolate a gene for which only a subchromosomal position is known. Such initial localisation usually comes from linkage and association studies as well as studies looking at chromosomal abnormalities [70]. Positional cloning is done by constructing detailed genetic and physical

maps which should narrow down the chromosomal region of interest followed by selection of genes likely to be involved in the trait.

### 1.5.3.2 Increasing the odds for successful mapping

In searching for the genes for complex traits, the main objective is to identify the genetic variation which truly influences the phenotype, or more often, a genetic marker that is correlated with the variant which influences the phenotype. When setting up association and linkage studies one is usually trying to minimise environmental effects by selecting ethnically and culturally homogenous groups. Such populations are not going to be exceptionally homogenous, but the selection is likely to reduce residual genetic effects that are unrelated to the disease and sometimes even genetic heterogeneity which is a common characteristic of complex traits [4].

Another approach that can be considered is the selection of samples from the extremes of phenotypic distribution (quantitative traits), or examination of very specific phenotypes. This should result in enrichment of the sample with trait-increasing alleles, thereby increasing the power for detection of the locus [101], [102].

Additionally, increasing the number of marker loci studied increases the probability that one of them will be in significant association with a trait and therefore in strong linkage with a nearby susceptibility gene [4]. Genetic markers and choices regarding them are discussed later.

### 1.5.3.3 Linkage studies

In linkage studies related individuals, either siblings or extended pedigrees, are studied in an attempt to localise genes influencing a trait. This is done by examining the co-segregation of the phenotype with genetic markers [62].

If two markers are at loci that are close together they may exhibit linkage. This is because it is highly unlikely that a crossover will occur between them at meiosis resulting in a particular haplotype being passed on intact. The proportion of meioses in which recombination occurs between two markers is called the recombination fraction ($\theta$). If the two loci are distant, or on different chromosomes, $\theta$ will equal 0.5 [103]. In other words, their alleles will assort independently, according to Mendel's second law. Linked loci will show departures from independent assortment and will have recombination fraction of less than 50%.

To find out how strong the evidence for linkage is, statistical analysis is performed (Figure No. 15). It is based on a likelihood ratio, the ratio between the likelihood assuming linkage at a given value of $\theta$ ($L_\theta$) and the likelihood assuming non-linkage ($L_{0.5}$). The logarithm to the base 10 of that ratio is the LOD (logarithm of odds) [104].

$$LOD(\theta) = \log_{10} (L_\theta / L_{0.5})$$

**Figure No. 15:** Formula for calculating LOD score

The traditional criterion for accepting evidence for linkage is the LOD >3. This value was chosen on the basis of calculations for single gene disorders that if such a threshold is used, 95% of LOD >3 would reflect true linkage [72]. If multiple genetic

models are tested, the weight of a given LOD score should accordingly be reduced. For this reason, LOD scores between 3 and 6 are considered nowadays as suggestive rather than a definite evidence of linkage for complex traits [105].

One situation that may be difficult to deal with is a case when a different gene influences a trait in different families - a phenomenon called locus heterogeneity. For this reason, linkage for very heterogeneous traits is usually established in single large families [104].

Classical methods of linkage analysis require detailed knowledge of the underlying genetic model. For complex traits, specification of parameters essential for analysis can be problematic, leading to reduced power and spurious results [106]. Several non-parametric methods have been proposed as an alternative to classical linkage studies to avoid the problem of mis-specifying parameters. The most commonly used non-parametric methods examine sibling or other relative pairs for increased allele-sharing [105]. Although more appropriate and convenient for studying complex characters, all non-parametric methods are less powerful than the classic parametric ones and require larger sample sizes dependent on the number of genes involved, trait/disease prevalence and heritability of the trait [106].

Since linkage studies rely on the analysis of related subjects they are only capable of locating a relatively large chromosomal region. With complex traits, linkage peaks may span regions much greater than 10-20 cM, with an inconveniently large number of genes underneath [107]. This means that this type of study may not provide the necessary resolution for locating genes underlying phenotypic variation, but it can be used for initial localisation, especially in cases where there is no prior information on what kind of gene/genes might be involved in influencing a particular trait.

### 1.5.3.4 Association studies

Association is a phenomenon whereby two features occur together with a frequency greater than expected on the basis of chance alone [108]. Association studies focus on population allele frequencies and test whether a trait and an allele show correlated occurrence in a population.

Genetic loci implicated in complex traits are likely to be of small/modest effect size [63]. The linkage method, although successful in detecting major genes, has limited power to detect genes of modest effect. For loci conferring genotypic relative risks (GRR) of 2 or less, the sample sizes required are generally beyond reach (> ~2500 families) [109]. On the other hand, the sample sizes required for association studies are much smaller, even for detecting smaller signals with GRR of 1.5, and are generally less than 1000 affected individuals [109]. So, association is a powerful technique and is increasingly seen as the method of choice for mapping complex trait loci.

The measure of association can be qualitative (presence or absence of a significant association) or quantitative (strength of association). The latter one can be evaluated by several statistical parameters such as: risk difference, relative risk, odds ratio and attributable risk [108]. The $\chi^2$ test detects the existence of a significant association.

The basic approach of the method is to identify marker loci at which some alleles are more frequent (or deficient) among individuals with a particular phenotype than among controls. A statistically significant difference (positive association) can arise for several reasons [62]:

A) if the marker itself is contributing to overall phenotypic variance (pure or direct association),
B) if there is a tight physical linkage between a marker and a trait locus (linkage disequilibrium or indirect association), or

C) if the association observed results from population admixture and an inadequately matched control group (spurious association).

Spurious associations are a big problem in population-based studies. To significantly reduce the number of such, it is crucial to select suitable controls. It is of utmost importance that the two groups compared are matched for ethnicity to avoid any false associations that have nothing to do with the examined trait but rather reflect ethnic differences in marker allele frequencies.

Family-based association methods help overcome this problem by using 'internal' controls [108]. The two most frequently used are the *Haplotype Relative Risk (HRR)* [110] and the *Transmission Disequilibrium Test (TDT)* [111]. In the HRR allele frequencies in affected offspring are compared with frequencies in a "control group" consisting of non-transmitted parental alleles. In the TDT, the frequency of transmission of alleles from heterozygous parents to affected offspring is examined.

In addition to being more costly than case-control studies, family-based studies have to face problems in relation to sample collection. Diseases with a later age of onset are especially unfavourable to the family based design, because the parents are often dead and/or other members of the nuclear family are unreachable.

A further argument to be considered in favour of case-control studies is that the effect of population stratification is actually negligible within an ethnic group, even if defined vaguely as Caucasian, Black, Asian or Hispanic [112]. So, unless an investigator is very careless and mixes different ethnic groups with different disease prevalence and marker frequencies, spurious associations due to population stratification should be extremely rare [112].

One easy and cost effective way to control for stratification in population based studies was proposed by Pritchard [6]. As allele frequencies at random marker loci may differ among ethnic groups [113], one should expect to find allele frequency differences between such groups throughout the genome. If cases and controls are well matched,

then significant allele frequency differences should be observed only in close proximity to a disease susceptibility allele. In other words, if stratification due to different ethnic representation in case and control groups is present, then markers unlinked to a disease locus should also show associations with the phenotype. It has therefore been suggested that case-control based studies should include in their design 15-20 randomly chosen microsatellite markers, or 30 biallelic markers in order to keep the overall Type I error rate in reporting associations to ≤5% [6].

Another common problem which may lead to false positive results is the application of insufficiently rigorous statistical analysis. This is often due to inadequate correction for the number of questions asked/markers tested. Since association studies of complex traits are likely to involve testing a large number of markers, correcting the final p value using the Bonferroni correction method [114] will require a statistical significance several orders of magnitude greater in order to remain significant. Such levels of significance are almost impossible to reach, and applying such a conservative multiple-test correction could result in real signals being missed. Approaches such as consistent replication of positive findings on an independent sample have been suggested as an alternative yet powerful way of circumventing the problem [115].

An important limitation to be considered in association studies of complex traits is heterogeneity. Many different environmental and genetic factors involved may result in a difficulty to find a single genetic marker associated with the trait in question within the practical scale of a study [116].

Even with all these limitations, association studies are still one of the easiest ways of filtering out genes which may influence a trait of interest, assuming all precautionary measures are taken. Also, they are frequently the inevitable choice if the character studied is complex in nature and is likely to be influenced by a number of genes in combination with environment.

### 1.5.4 Linkage disequilibrium (LD) mapping

Association studies can either be direct or indirect [117]. Direct studies rely on studying variants in coding and regulatory regions of genes in the hope that these will include variants contributing directly to a particular trait. Indirect studies rely on association between a trait and a marker allele located close to and in *linkage disequilibrium (LD)* with an allele contributing to phenotypic variance. In other words, even variants that are not included in the screen would be assayed indirectly through LD with nearby markers. This approach is known as *LD mapping* [118].

LD mapping relies on the assumption that a single ancestral mutation is responsible for a large proportion of disease carriers (individuals with a certain trait) in a present-day population [119]. The mutation arises on a particular ancestral haplotype which slowly degrades by recombination with passing generations. However, this haplotype should largely be preserved within the region around the mutation because loci in close proximity to each other will rarely be separated by recombination events. The approach makes use of the large number of generations since the first appearance of the mutation and allows substantial narrowing of the interval in which a disease gene might lie.

#### 1.5.4.1 Measures of LD

Two alleles are said to be in linkage disequilibrium, also called allelic association, if they occur together at a frequency greater than expected if the alleles were segregating at random. In other words, the frequency at which these alleles occur together will depart from the simple product of their allele frequencies, $p_A \times p_B$, where $p_A$ is the frequency of allele A at the first locus and $p_B$ the frequency of allele B at the second

locus [120]. This means that unrelated people who have a certain allele at one locus will have statistically greater chance of having a particular allele at a second locus if these two loci are in LD.

One of the earliest measures of disequilibrium, symbolised by $D$, was proposed nearly forty years ago [121]. It quantified LD as a difference between expected and observed haplotype frequencies for two loci in question. Because its numerical value is dependent on allele frequencies, $D$ cannot be used as a measure of LD strength or for comparing levels of LD [118].

In addition to $D$, there are two other commonly used measures of LD: $D'$ and $r^2$. $D'$ is determined by dividing $D$ by its maximum possible value. $D' = 1$ when LD is complete. Intermediate values, $D' < 1$, tend to be biased upward inversely with sample size [122] and should not be used for comparisons of the strength of LD between studies [118].

$$D = p_{AB} - p_A p_B$$
$$D' = D / D_{max}$$
$$r^2 = D^2 / p_A p_a p_B p_b$$

**Figure No. 16:** Calculating different measures of LD.

$r^2$, the correlation of alleles at the two sites, is formed by dividing $D^2$ by the product of the four allele frequencies at the two sites. Although it is allele frequency and sample size dependent, $r^2$ is much less inflated in small samples than $D'$ [122], and has been proposed as the measure of choice for quantifying and comparing LD levels [123]. While

$r^2$ = 1 is known as perfect LD, values of $r^2$ above 1/3 might indicate sufficiently strong LD to be useful for mapping [118].

In addition to pairwise measures of LD (*i.e.* $D$, $D'$, $r^2$), the population recombination parameter, $4N_e r$, facilitates comparisons between regions [118]. However, estimation of this parameter from genotyping data is computationally very challenging and is currently a subject of intense scientific research [124].

### 1.5.4.2 What is the extent of LD?

Linkage disequilibrium is not a strict monotonic function of distance. Both strong LD between very distant markers and virtually non-existent LD between physically close markers are readily observed in the human genome. Mutation and recombination might have the most evident impact on linkage disequilibrium, but other factors are likely to further shape the extent of LD and change the simple relationship between the strength of LD and the physical distance. Some of these are:

A) *Genetic drift*: This causes changes in haplotype frequencies due to random sampling of gametes. The effect is especially strong in small and stable populations, where low frequency haplotypes tend to be lost, resulting in increased LD [125].

B) *Admixture*: This is the phenomenon whereby two or more genetically distinct populations admix and create a so-called "hybrid" population. In such populations that have undergone recent admixture very strong LD is observed over large distances. The speed of LD decay will depend on the size of the hybrid population,

and whether it experiences a constant external influx from one of its "originator" populations (slow LD decay), or has become isolated (steep LD decay) [126].

C) *Population growth*: Rapid population growth causes faster LD decay through reducing the effect of genetic drift.

D) *Population bottlenecks*: Bottlenecks result in substantial reduction in population size and haplotype diversity leading to increase in the extent of LD [127].

E) *Sub-populations and inbreeding*: Tend to increase LD.

F) *Natural selection*: The entire haplotype flanking the favoured variant becomes highly frequent in the population, a phenomenon called "hitchhiking effect". Negative selection, although milder in effect, can also strengthen LD by removing haplotypes with deleterious variants.

G) *Variable recombination and mutation rates*: There will be little or no LD between loci within regions with high recombination rate, so-called "hot spots", and within those characterised by higher mutation rate, like CpG islands [128].

H) *Gene conversion*: This is the phenomenon whereby a short stretch of one copy of chromosome is transferred to the other copy during meiosis. The process is equivalent to two very closely spaced recombination events [118], and results in rapid LD breakdown [129, 130].

So, depending on the size and history of the population, LD is seen over varying distances from a trait locus. Haplotype signatures tend to be shorter in older outbred populations [127] than in younger population isolates [131]. Recent computer based

simulations predicted that sufficient LD for indirect association studies is unlikely to extend beyond an average distance of 3 kb in the general population, [117]. The 3 kb figure was met with a widespread scepticism [132, 133, 127, 128].

Collins (1999) predicted that in large outbred populations disequilibrium should be detectable within 100-300 kb (0.1 – 0.3cM) of a trait locus after analysing data from 1000 pairs of loci [132]. In a similar attempt to throw some light on extent and distribution of LD, Abecasis (2001) examined 127 polymorphisms in a sample of 575 chromosomes from unrelated individuals of British ancestry and found out that for half of the markers LD extends to 50 kb [133].

LD may extend over much larger distances (>5 cM). This is particularly true for small population isolates [134, 131], making such populations attractive for whole genome scans for association with complex traits. When talking population isolates, one usually thinks of populations coming from Finland, Iceland, Sardinia and Japan. However, these populations show levels of LD (<1 cM) similar to other populations in Europe. This is likely to be due to combination of two factors: the relatively large number of founders and the growth these populations experienced since their foundation [135]. By contrast, small populations that have remained stable in size over long periods, like Scandinavian Saami [136], are more likely to have experienced much stronger effect of genetic drift which resulted in LD extending over larger regions (>10 cM) [125, 137]. Shifman and Darvasi (2001) argue that apparent similarity in LD between outbred populations and population isolates (*i.e.* Finland and Iceland) is true only for SNPs separated by up to 200 kb, while those separated by more than 200 kb show significant differences in LD level between these populations [138].

Excess of LD is also observed in admixed populations. Lemba, a hybrid population generated by Bantu-Semitic admixture, is a fine example of one. Significant LD is present between markers ≤21 cM apart [139]. Pfaff *et al.* reported similar findings in two African American populations where LD extended over 10 cM [126].

The attractiveness of populations that show extensive linkage disequilibrium lies mainly in the fact that genome-wide mapping would be achievable with a smaller set of already available markers from public databases, but one has to bear in mind that the resolution of such mapping will be relatively low [127].

### 1.5.4.3 Block structure of LD

Several recent studies [140, 141, 142, 143, 144] indicate that linkage disequilibrium in the human genome is highly structured into blocks of low haplotype diversity separated by recombination hot-spots (Figure No. 17) [1].

Haplotype blocks, characterised by extensive LD, can span from only a few to >100 kb. Even though most ( ~94%) of all crossovers were found to lie within recombination hot spots, resulting in a rapid LD breakdown, they are not exclusive to these short stretches of chromosomes ( 1 to 2 kb) [142]. This may explain a within-block decay of LD which, if present, is usually much more gradual than that in hot spot regions.

If there is very little recombination between the blocks, a high inter-block correlation called long-range LD, will be present. The phenomenon is characterised by lower haplotype diversity where a small number of haplotypes observed over longer stretches of chromosomes account for a large proportion of chromosomes in a given sample [141].

Since haplotype structure is highly dependent on population history, different patterns of LD are likely to be observed in different populations and different chromosomal regions. In some extreme cases of large and stable populations, a sufficient number of crossover events may have occurred within blocks for the clear block-like structure of LD to be completely absent [1]. The lower LD seen in the African population [127] could therefore be a result of sharper decay of disequilibrium within and/or between blocks.

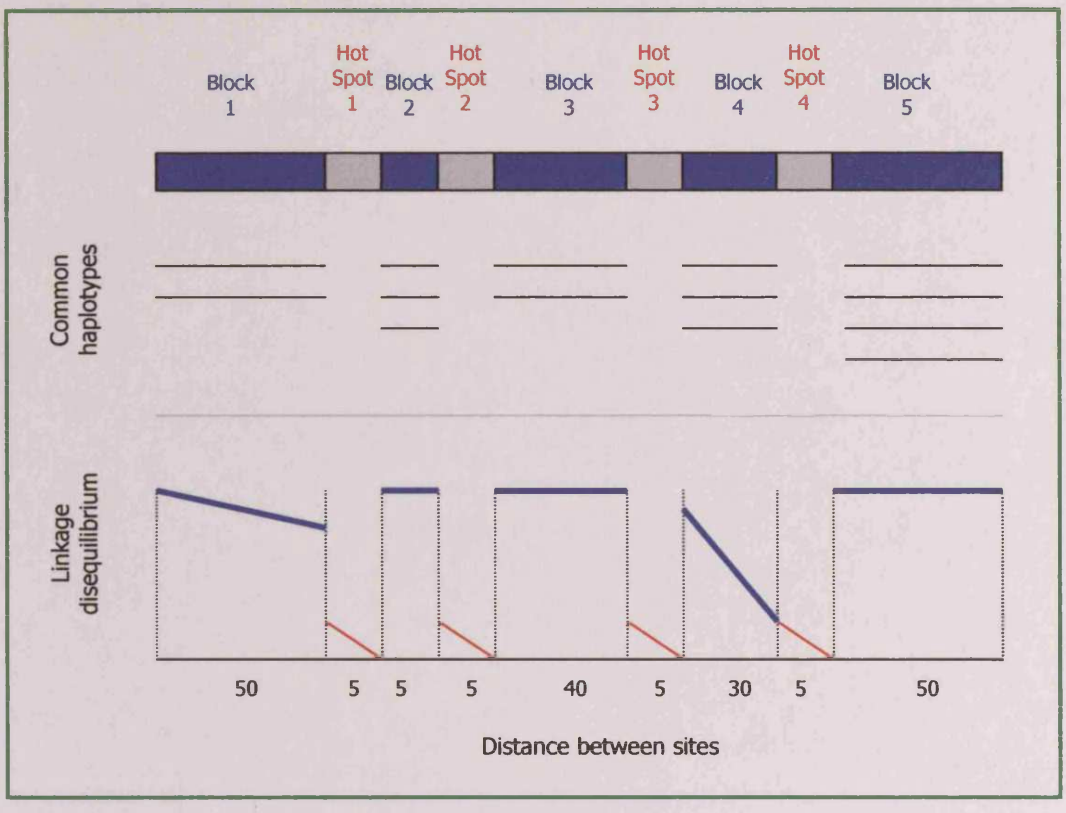



**Figure No. 17:** Islands of LD [1]

If the block-like structure of the human genome proves to be a rule rather than exception, it is likely to have a profound impact on association studies. The number of SNPs one will have to genotype could drastically be reduced to only 2-5 per domain, bringing down the total number of SNPs necessary for a genome-wide scan to 80,000-200,000 [142]. For variants located within hot spot clusters this number will have to be higher.

More recently, a term linkage disequilibrium unit (LDU) was introduced [145]. One LDU corresponds to one swept radius which is defined as the average extent of useful LD [146]. Regions of extensive LD have fewer LD units while regions with high levels of recombination have greater number of LDUs. Based on chromosome 22 data, Tapper

*et al.* estimate a total of ~50,000 LDUs per human genome and anticipate that a number of SNPs required for an efficient genome screen will be a multiple of this number (200,000 – 300,000).

Although attractive in terms of resources necessary for carrying out an association study, the block structure of LD might mean that the maximum resolution that can be achieved with LD mapping cannot go much further than ~100kb [127, 146]. If finer resolution is desired, populations with less extensive LD may have to be considered.

#### 1.5.4.4 Power to detect LD

The power to detect LD is affected by the method used, the number of available samples, the recombination and mutation pattern of the region, the degree of locus and allele heterogeneity, age of the mutation, the number and type of markers used, and population history [137].

Long and Langley (1999) estimate that sample sizes of ~500 individuals typed with ~20 SNPs spaced throughout a candidate gene region (~250 kb, if one assumes that one 4Nc in humans is equal to 5 kb and that there are 50 such units in a candidate region) should provide sufficient power to detect QTLs that account for ~5% of the variation in a complex trait [124]. Results from simulation studies they carried out also suggest that a greater increase in power will be achieved by increasing sample sizes than by increasing the number of typed SNPs. That is particularly true for smaller regions. Increasing the number of SNPs will be helpful only if recombinationally larger regions are considered.

Allelic and locus heterogeneity can result in marked decrease in power [147, 4]. Increased allelic heterogeneity gives rise to a greater number of possible haplotype backgrounds each harbouring a particular disease locus variant. This dilutes the signal

coming from any one of these haplotypes, resulting in less LD and lower power to detect it [4]. The chances of detecting LD can also be reduced if the allele conferring susceptibility is common relative to the population size [125, 108].

The length of time since the disease causing mutation occurred is also important. Mutations responsible for complex diseases tend to be older because they are typically not subject to strong natural selection. Older mutations allow greater resolution with LD mapping studies, but being too old might result in equilibrium already being reached and/or many new mutations accumulated [147].

Because of its short range in large outbred populations, LD mapping should preferably be used as a method for fine mapping once a region of interest has been found by linkage analysis - the strategy successfully applied for mapping complex disease loci such as a Crohn's disease susceptibility gene [148], susceptibility genes for asthma [149, 150], ulcerative colitis [151], diabetes mellitus type 2 [152], autism [153] and schizophrenia [154].

### 1.5.5 Choice of markers

Genetic markers are heritable characters (loci) with multiple states (alleles) at each character [155].

Microsatellites, or Short Tandem Repeats (STRs) became the markers of choice for genetic mapping studies in 1989 [156]. Their main feature is the presence of many alleles and high levels of heterozygosity which maximises the number of informative meiotic events essential for resolving the segregation of alleles through a family. Such markers provide an ideal tool for pedigree-based linkage analysis [106]. Three types commonly

used are recognised on the basis of the number of nucleotides comprising each repeat unit. Those are di-nucleotides, the most abundant, tri- and tetra-nucleotides.

Recent developments in mutation detection technology led to emergence of a new generation of markers called Single Nucleotide Polymorphisms (SNP). These markers are usually biallelic and less polymorphic than microsatellites, but they have the advantage of being highly abundant in the genome and, because of their binary nature, favourable for high-throughput automated genotyping technologies.

The power to detect linkage between a marker and a disease locus and the accuracy of the estimation of the map location of the disease gene greatly, but not exclusively, depends on the right choice of markers. In general, if the frequency of the associated allele is unknown and the distance between the marker and the disease locus is relatively large, a microsatellite marker would be a better choice for population-based studies [156]. Sham *et al.* have shown that highly polymorphic markers also have greater power in cases where there are multiple ancestral disease mutations and suggested ten alleles to be a sufficient allele number for detection of LD [157]. Nevertheless, it has been suggested that when choosing between the markers with the same heterozygosity one should opt for the one with fewer alleles [158].

However, SNPs are likely to outperform microsatellites in LD mapping studies where the distance between the marker and the disease locus is small, so that the microsatellite mutation rate is not negligible compared with the recombination rate [156]. Typing a large number of SNPs overcomes the problem of the low informativeness that an individual SNP suffers from and offers the opportunity to further increase power by generating haplotype information in a small genomic region of interest [156, 159]. On the other hand, if SNPs are in strong linkage disequilibrium with one another less information and less power will be contributed [137].

The use of biallelic markers rather than multiallelic microsatellites for LD testing in genome screens requires much denser distribution of markers to achieve the same power [158]. Apart from the feasibility question one has to consider the increase in Type I

error related to a larger number of markers tested. Even more problematic is the fact that linkage disequilibrium does not necessarily have to exist even for SNPs within the disease gene [106]. An increasing number of studies are trying to deal with these issues through investigating the application of haplotype analysis to LD mapping. So far, promising results have been reported for single-gene disease models [159], whereas complex traits and diseases are still to be considered.

Linkage analysis with microsatellites used to identify regions of interest followed by association-based analysis employing SNPs to narrow down the region to a size manageable for positional cloning is a strategy recommended by Almasy and Blangero [106]. Many research groups have adopted such approach and are focusing on SNP genotyping of linkage peaks previously detected with microsatellites [149-151, 154].

However, more recent reports suggest that microsatellite linkage mapping may soon become a strategy of the past. Direct comparisons of linkage peaks obtained through the use of microsatellite markers and densely distributed SNP indicated a high degree of correspondence between the two approaches, but there were instances where more traditional microsatellite approaches had insufficient power to detect highly significant linkage peaks [160, 161]. In addition, high-density SNP scans offer considerable savings in downstream mapping owing to better definition of the linkage peaks [160].

## *1.6 Mapping studies on intelligence*

Although a number of studies offer support for a relatively large genetic contribution to variance in *g*, there have been a few attempts to identify the specific genes.

The first mapping study on intelligence focused on available DNA markers, mostly RFLPs (restriction fragment length polymorphisms), in or near genes likely to be of neurological relevance [162, 163]. 100 markers were examined for allelic association by looking at allele frequency differences between high-IQ (IQ > 130) and low-IQ groups (IQ < 82). Despite a two-stage study design (original and replication stage) and selection of individuals from extremes of the phenotypic distribution, small sample sizes offered limited statistical power to detect genes of small effect size. Several significant associations ($p < 0.05$) were found in the first stage of the study. Two of them, alcohol dehydrogenase 5 (ADH5) and nerve growth factor beta polypeptide (NGFB) showed allele frequency differences in the same direction in the replication sample but failed to reach statistical significance.

A more systematic approach was taken by Chorney *et al.* (1998) who, instead of looking at candidate genes, used a dense map of microsatellite markers on chromosome 6p to look for allelic association with general cognitive ability [164]. The authors reported an association with a marker situated within the gene for insulin-like growth factor 2 receptor (IGF2R), a finding later replicated on a larger sample and different polymorphism in IGF2R [165].

With fast development of new molecular techniques and increased public availability of markers, both microsatellites and SNPs, it would be reasonable to expect further mapping efforts aimed at detecting some of the QTLs responsible for phenotypic variance in *g*.

## *1.7 Animal research*

Cognitive ability, especially learning, memory and problem solving, is also studied in other species such as rat, mouse, marine snail *Aplysia* and fruit fly. In one of the first

such experiments during 1920s, rats were selectively bred for their performance in learning the maze in order to find food [166]. During the 1950s and 1960s studies of inbred strains of mice showed that genes are important contributors to learning and memory. These maze-dull and maze-bright rats were also used in a study of genotype-environment interaction [167], the only research on learning, so far, that managed to show differential response of genotypes to environment [168]. Late 1970s were characterised by introduction of *Aplysia* for the study of learning and memory. This simple organism allowed relatively easy mapping of learned behaviour to synaptic change [169]. Even though mammalian brain is much more complex, *Aplysia* provided an important example of the relationship between learning and plasticity, as well as elucidation of various mechanisms of plasticity that are conserved across the species [170].

Candidate gene studies using knock-out strains of mammals have generated a substantial amount of data on the involvement of various genes and their products in particular aspects of mental functioning. An important contribution to knock-out studies was made by Tsien *et al.* who developed a method to create mice in which the deletion of almost any gene will be restricted to specific brain region such as the hippocampal CA1 region [171]. Creation of such knockouts allows more precise analysis of the impact of specific gene mutations on animal behaviour.

A substantial proportion of current animal research focuses on investigation of molecules involved in neuronal migration, differentiation, axon outgrowth and synapse formation, with particular focus on their role in activity-dependent synaptic plasticity. Integrin-deficient [172], neural cell adhesion molecule (NCAM)-deficient [173] and LIM kinase 1-deficient [174] mice are just some of the many strains that provided strong evidence for involvement of such molecules not only in synapse architecture and development, but also in mediation and regulation of synaptic plasticity.

Another vigorously explored avenue is directed towards examining receptors that may have substantial impact on processes leading to changes in synaptic transmission and, by doing so, on learning and memory. The importance of AMPA and NMDA

receptors has been acknowledged for a long time, but it was the production of mutant mice expressing different receptor subinuts that provided more insight into their roles in learning and memory. GluR2 subunit of AMPA receptor is correlated with lower calcium ion permeability [175], and not surprisingly GluR2 mutant mice were found to exhibit enhanced levels of synaptic potentiation [176]. A few years later, mice with enhanced learning ability and over-expressing the 2B subunit of NMDA receptors were generated [177].

In addition to these two most abundant receptor groups, other receptors were also found to have an impact on behavioural processes, particularly spatial learning. Opioid [178], interleukin-1 [179] and oestrogen receptor knockouts [180] all show disturbed synaptic plasticity and profound deficits in spatial ability.

Such studies are throwing more light on complex processes of learning-related synaptic plasticity and are bringing our attention to genes likely to be implicated. Results reported by many research groups have confirmed some pre-existing hypotheses regarding molecular mechanisms that take place during synaptic activity. Equally, if not more, important is the contribution made by groups that brought our attention to novel mechanisms that appear not to be alternative, but rather complementary to already known ones.

# 2 AIMS OF THE STUDY

General cognitive ability is one of the most heritable human behavioural traits [88, 89], and a good candidate for molecular genetic research.

This study will attempt to identify some of the genetic loci responsible for overall variance in *g*. Two approaches will be used: 1) a QTL association study with a dense map of microsatellite markers (aiming for a 1cM grid), and 2) an investigation of functional candidate genes.

# 3 GENERAL MATERIALS AND METHODS

## *3.1 Sample*

The project includes three samples: 1) an original sample of 101 high *g* cases and 101 average *g* controls, 2) a replication sample of 100 very high *g* cases and 100 average *g* controls, and 3) a within-family replication sample of 196 parent-offspring trios in which children have high g scores. The third sample was available only to our project collaborators in London.

The samples were obtained from areas around Cleveland (Ohio, USA). Subjects were restricted to non-Hispanic, white children in order to reduce Error 1 rate due to population stratification. Subjects were excluded if one of the maternal or paternal grand-parents was not white. Informed consent was obtained from all participants. For all subjects in the original case-control groups, permanent cell lines were established from blood samples. DNA for the remaining subjects was extracted from cheek swabs.

### 3.1.1 Original sample

The original controls were selected for IQ scores between 90 and 110 (mean IQ 102.2; 6.3 SD) from a larger group of children assessed using the Wechsler Intelligence Scale for Children Revised (WISC-R) [181]. Average age was 13.

First 51 children of the original cases group were selected for IQs above 130 from the same group of children (mean IQ 136.0; 9.3 SD). Average age was 10. The remaining 50 original high *g* individuals were selected from the highest-scoring

individuals from the Study of Mathematically Precocious Youth (SMPY) [182] that sat the Scholastic Aptitude Test (SAT, college entrance exam) four years early. These children obtained SAT scores 4 standard deviations above the mean (equivalent to IQ scores > 160).

### 3.1.2 Replication sample

Replication samples were obtained from the same Cleveland area and in a similar fashion as the original group. The controls had an average IQ of 102.4 (7.6 SD), but unlike the original group, the subjects were adults (33 years average age). They were assessed by the Wechsler Adult Intelligence Scale-III [183].

The replication high *g* group was selected from the remaining highest-scoring individuals in the SMPY sample. IQs estimated from their SAT results were > 160.

### 3.1.3 Parent-offspring trios

The sample of 196 parent-offspring trios was obtained from the SMPY study. High scoring individuals were selected on a basis of availability of their parents. About one third of the children forming this group were also used as replication cases for the study.

## 3.2 Creation of DNA pools

DNA pools were constructed from individual DNA samples that were brought down to 6ng/µl prior to pooling, making sure that every individual contribute the same amount of DNA. The volumes of the pools were determined from the number of markers planned to be tested and from the amount of DNA that is used per standard PCR reaction. All pools were constructed in triplicate.

DNA quantification prior to pooling was performed at three concentration levels (~80ng/µl, ~20ng/µl and at 6ng/µl) using the PicoGreen™ fluorescent assay. PicoGreen dye was added to genomic DNA samples that were diluted in 1xTE buffer (1:20 or 1:50 dilution factor depending on the expected DNA concentration) and pre-aliquoted into White Cliniplates™ (Labsystems). After incubation of 4min at room temperature samples were analysed by a Flouroscan Ascent fluorometer. The method gives more reliable readings than the simple spectrophotometry because it relies on the dsDNA intercalating property of the dye. This ensures that only double stranded DNA molecules generate signal that is translated by the Ascent software into a DNA concentration of the tested sample. However, presence of some compounds that commonly contaminate nucleic acid preparations, like phenol, ethanol and chloroform, can increase signal intensity by 13%, 12% and 14% respectively (figures adopted from Molecular Probes product information sheet).

## 3.3 Polymerase Chain Reaction (PCR)

The polymerase chain reaction (PCR) is a technique for rapid amplification of a target DNA sequence *in vitro*.

The DNA fragment to be amplified is determined by selecting primers. Primers are short oligonucleotide strands (usually 18-25 bp) that are complementary to the beginning (forward primer complementary to the "minus" strand) and end (reverse primer complementary to the "plus" strand) of the DNA fragment to be amplified. The PCR reaction is carried out in a thermal cycler which heats and cools the reaction mixture to the precise temperature and length of time required for each step of the reaction.

PCR primers for microsatellite markers used in the Genome Scan study were designed by their original submitters (http://cedar.genetics.soton.ac.uk/), and were purchased from MWG-Biotech (Germany). Primers for candidate gene studies were designed using the Whitehead Institute for Biomedical Research software *Primer 3* (http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi) and purchased from Sigma-Genosys.

### 3.3.1 Standard PCR

Standard PCR reaction mixture and thermal-cycling protocol were used for candidate gene studies. Annealing temperatures and optimal $Mg^{2+}$ concentrations were deduced empirically from optimisation reactions. Main features of a standard PCR are presented in Tables 3 and 4.

**Table No. 3:** PCR reaction mixture preparation. All values apply for **ONE** reaction.

| PCR mixture component | Volume (µl) |
|---|---|
| DNA (6 ng/µl) | 4.00 |
| $H_2O$ | 4.96 |
| $MgCl_2$ (25 mM) | 0.24 |
| Reaction buffer (x10) | 1.20 |
| dNTPs mix (5 mM each) | 0.96 |
| Forward primer (5 pmpl/µl) | 0.28 |
| Reverse primer (5 pmpl/µl) | 0.28 |
| *Taq* polymerase | 0.06 |

**Table No. 4:** Thermal cycling protocol for a **STANDARD** PCR reaction.

| Cycling Step | Temperature | Time | No of cycles |
|---|---|---|---|
| Denaturation | 94° | 5 min | 1 |
| Denaturation | 94° | 30 sec | 35 |
| Annealing | T° deduced from optimisation experiments | 30 sec | |
| Extension | 72° | 45 sec | |
| Final Extension | 72° | 10 min | 1 |

### 3.3.2 Touchdown PCR

In addition to standard, a Touchdown protocol was used. Here, a template DNA is amplified using a range of annealing temperatures in a single reaction. Touchdown PCR simplifies the process of determining optimal primer annealing temperatures and is particularly useful for studies investigating a large number of genetic markers. It is also helpful in cases where a non-specific primer binding is suspected [184].

During the initial cycles of touchdown PCR, annealing takes place at temperatures 5° to 10°C above the calculated Tm. In subsequent cycles, the annealing temperature is gradually reduced by 1-2°C each cycle until it has reached approximately 5°C below the calculated Tm. The touchdown thermal cycling protocol successfully employed for

genotyping of more than a thousand microsatellite markers in the Genome wide screen for association with IQ (Chapter 4) is presented in Table 5.

| Cycling Step | Temperature | Time | No of cycles |
|---|---|---|---|
| Denaturation | 94° | 5 min | 1 |
| Denaturation | 94° | 30 sec | 10 |
| Annealing | 60° -1° with each additional cycle | 30 sec | |
| Extension | 72° | 45 sec | |
| Denaturation | 94° | 30 sec | 25 |
| Annealing | 50° | 30 sec | |
| Extension | 72° | 45 sec | |
| Final Extension | 72° | 10 min | 1 |

**Table No. 5:** Thermal cycling protocol for a **TOUCHDOWN** (60°-50°C) PCR.

### 3.3.3 Post-PCR product treatment

#### 3.3.3.1 Klenow fragment treatment

Post-PCR treatment of amplified products with *E. coli* DNA polymerase I Klenow fragment (Amersham) to repair 3' overhanging ends and generate blunt-ended products was used for the pooled genotyping stage of the genome-wide screen with microsatellites for the purpose of reducing the stutter band effect (for detailed discussion see Chapter 4).

| Mixture component | Volume (µl) |
|---|---|
| Pooled PCR products | 12.5 |
| $H_2O$ | 0.6 |
| Reaction buffer (x10) | 1.5 |
| Klenow fragment (5u/µl) | 0.4 |

Table No. 6:

Components of a "Klenow mixture" for post-PCR treatment of amplified products. Reaction mixture is incubated at 30° for 90 minutes.

3.3.3.2 *Exonuclease I* and Shrimp alkaline phosphatase treatment

Certain techniques, such as DNA sequencing and SNaPshot genotyping (described further down), require PCR products to be free from unincorporated primers and dNTPs. Purification was achieved by adding 2U of *Exonuclease I* (*ExoI*) (Amersham) and 2U of Shrimp alkaline phosphatase (SAP) (Amersham) to a 12µl PCR product. The mixture was incubated at 37°C for 75 min, after which the reaction was stopped by heating the products at 80°C for 15 min.

## *3.4 Agarose Gel Electrophoresis*

Agarose gels provide the most commonly-used means of isolating, purifying and size separating fragments of DNA. The equipment and supplies necessary for conducting agarose gel electrophoresis are relatively simple and include:

1) Agarose MS (Boehringer Mannheim) for separation of 50bp – 1500bp fragments
2) An electrophoresis chamber and power supply

3) Gel casting trays and well-forming combs

4) Electrophoresis buffer, *i.e.* Tris-borate-EDTA (0.5X TBE, National Diagnostics)

5) Loading buffer which contains a dense substance to allow the sample to "fall" into the sample wells and one or two tracking dyes, which migrate in the gel and allow monitoring or how far the electrophoresis has proceeded. It can be prepared by adding 3.0g Ficoll 400 (BDH), 0.22g EDTA (Sigma), 0.02g Bromophenol Blue (Sigma) and 0.02g Xylenecyanol (Sigma) to 20ml $H_2O$. The dye mixture is kept over-night in a warm water bath.

6) Ethidium bromide, intercalating fluorescent dye used for staining nucleic acids

7) 1kb DNA ladder

8) Transilluminator (an ultraviolet lightbox) fitted with Polaroid camera

To pour a gel, agarose powder is mixed with electrophoresis buffer to the desired concentration (1.5 – 3.0%), then heated in a microwave oven until completely melted. By using gels with different concentrations of agarose, one can resolve different sizes of DNA fragments. Higher concentrations of agarose facilitate separation of small DNAs, while low agarose concentrations allow resolution of larger DNAs. Ethidium bromide (1µl of 10mg/ml is sufficient for a 100ml gel) is added to the gel at this point to facilitate visualization of DNA after electrophoresis. After cooling the solution to about 60°C, it is poured into a casting tray containing a sample comb and allowed to solidify at room temperature or in a refrigerator. After the gel has solidified, the comb is removed.

DNA samples are mixed with 4µl loading buffer per 12µl sample. 10µl of this mixture are then loaded into the sample wells and a current is applied (~100 volts for 30-90 min depending on the fragment size). DNA migrates towards the anode. The distance DNA has migrated in the gel can be judged by visually monitoring migration of the tracking dyes. Bromophenol blue and xylene cyanol dyes migrate through agarose gels at roughly the same rate as double-stranded DNA fragments of 300 and 4000 bp,

respectively. To visualize DNA, the gel is placed on a UV transilluminator and a photograph is taken.

## *3.5 Mutation detection*

Analysis of candidate genes often involves examination of sequence polymorphisms and their possible associations with a trait investigated. Denaturing High Performance Liquid Chromatography (DHPLC) is a mutation detection method that was used in this study and was accompanied by subsequent sequencing for ultimate confirmation of polymorphisms detected by DHPLC.

### 3.5.1 Denaturing High Performance Liquid Chromatography (DHPLC)

Denaturing high performance liquid chromatography (DHPLC) analysis is a reverse-phase ion-pair chromatography that allows the identification of single-nucleotide mismatches and small insertion-deletion polymorphisms within an amplified DNA fragment several hundred base pairs in length [185]. It is a rapid and semi-automated method based on differential separation and detection of mismatched DNA strands (heteroduplexes). Formation of heteroduplexes is a crucial step. It is performed on crude PCR products which are rapidly heated at 96°C for 2 min to allow melting followed by gradual decreasing sample temperature from 96°C to 56°C over a period of 40 min (temperature decrement of 1°C per minute) to allow re-annealing and enable the formation of heteroduplexes. At the appropriate column temperature, partially melted DNA heteroduplexes, formed after re-annealing of normal and mutant DNA

strands, have relatively reduced affinity to the hydrophobic matrix and are therefore eluted earlier from the column, enabling DNA variants to be detected with high sensitivity.

The method was performed using the Transgenomic WAVE system. Aliquots of 5-7 µl PCR samples (that underwent melting procedure described above) were loaded onto a preheated reverse-phase column based on non-porous polystyrene-divinyl-benzene particles (DNASep column, Transgenomic). DNA (homoduplexes with or without heteroduplexes) was eluted from the column by a linear acetonitrile gradient in 0.1 mM triethylamine acetate buffer (TEAA, Transgenomic), at a constant flow rate of 0.9 ml/min. DNA was detected at 260 nm. The gradient was formed by mixing buffer A (0.1 mM TEAA) and buffer B (0.1 mM TEAA, 25% v/v acetonitrile). Optimal combination of buffers and the column temperature, to allow fragment retention between 3 and 5 min, depends on a nucleotide composition of each of the fragments analysed. A simple and quick publicly available software based on simulated annealing algorithms was used to obtain optimal parameters for each of the fragments analysed (http://insertion.stanford.edu/meltdoc.html).

Each fragment was analysed on 15 high *g* individuals – mutation screening set. The size of the screening set gives sufficient power (95%) to detect polymorphisms with a minor allele frequency of 0.1. For each of the fragments analysed chromatograms were obtained for all the samples in the screening set. Elution curves were compared and a shift in a trace pattern was treated as indicative of the presence of a heteroduplex.

### 3.5.2 DNA Sequencing

Fragments showing elution curve shifts on DHPLC were subjected to sequencing for confirmation and characterisation of polymorphisms.

DNA sequencing is the determination of the precise sequence of nucleotides in a sample of DNA. The most popular method for doing this is called the dideoxy method. The dideoxy method gets its name from the critical role played by synthetic nucleotides that lack the -OH at the 3' carbon atom. Chain elongation stops once a dideoxynucleotide (ddNTP) is incorporated because there is no 3' -OH for the next nucleotide to be attached to. For this reason, the dideoxy method is also called the chain termination method [186].

Because all four deoxynucleotides are present in a standard sequencing reaction, chain elongation proceeds normally until, by chance, DNA polymerase inserts a dideoxynucleotide instead of the normal deoxynucleotide. The fragments are size separated on ABI sequencers with a maximum resolution of 1 bp. Each of the four dideoxynucleotides is labelled with a different fluorochrome tag which fluoresces at a characteristic wavelength when illuminated by a laser beam. Emitted fluorescence is collected by a camera and stored in a form of digitally encoded signals. Sequencing analysis software is used to visualise the sequence and allow base calling at each data point.

| ddNTP | Floorochrome | Colour (as seen in a chromatogram) |
|---|---|---|
| A | dR6G | Green |
| C | dROX | Blue |
| G | dR110 | Black |
| T | dTAMRA | Red |

**Table No. 7:** Fluorescently labelled ddNTPs used in sequencing reaction and their corresponding colours.

Sequencing reactions using the Big Dye Terminator (v.2.0) Cycle Sequencing kit (Applied Biosystems) were performed on PCR products that had previously been cleaned from unincorporated primers and dNTPs. For each fragment two reactions

were performed: with sense and anti-sense PCR primer. Reaction mix details are presented in Table No. 8, and thermal cycling conditions in Table No. 9.

**Table No. 8:** Sequencing reaction mixture preparation. All values apply for ONE reaction.

| Mixture component | Volume (µl) |
|---|---|
| Cleaned PCR product | 5.00 |
| Big Dye reaction mix | 4.00 |
| Primer (10 pmol/µl) | 0.35 |
| $H_2O$ | 0.65 |

**Table No. 9:** Sequencing reaction thermal cycling protocol.

| Cycling Step | T° | Time | No of cycles |
|---|---|---|---|
| Denaturation | 96° | 2 min | 1 |
| Denaturation | 96° | 30 sec | 25 |
| Annealing | 55° | 15 sec | |
| Extension | 60° | 4 min | |

Before sequencing reactions could be loaded onto a sequencer they had to be purified from any unincorporated nucleotides, salts and other impurities. That was done by gel filtering of sequencing products using MultiScreen HV filtration plates (Millipore) loaded with Sephadex G-50. The filtration plate had to be prepared few hours in advance (330µl $ddH_2O$ added to each of the wells containing dry Sephadex) to allow the resin to swell at an ambient temperature. Excess water was spun out of the filtration plates in a centrifuge at 2500 rpm for 5 min, after which individual sequencing products could be added to the centre of individual columns. The filtration plate was then placed on top of a clean 96-well collection plate and spun at 2500 rpm for 5 min. Eluted products were dried in a speed vacuum at 65°C for 45min.

Final step involved re-suspension of dried sequencing products in 10 µl HiDi formamide (Applied Biosystems) and their electrophoresis on ABI Prism 3100 Genetic Analyzer using a 36cm capillary array and POP-6 polymer.

DNA sequence was determined and polymorphisms characterised using a data analysis software Sequencher.

## *3.6 Genotyping*

### 3.6.1 Restriction Fragment Length Polymorphism (RFLP)

Restriction fragment length analyses uses restriction endonuclease (RE) to digest DNA at specific 4-6 bp recognition sites. Sample DNA is cut (digested) with one or more REs and resulting fragments are separated according to molecular size using gel electrophoresis. Molecular size standards (*i.e.* 1kb ladder) are used to estimate fragment sizes of digestion products. Ethidium bromide (intercalating dye that fluoresces under UV light) is used for staining and revealing the fragments under UV (260 nm) light. Polymorphisms like base substitutions, additions and deletions result in sequence rearrangements within RE recognition sequences and in a subsequent loss/gain of a restriction enzyme cut site. This, in turn, results in digestion products of different sizes that are readily observed on gel electrophoresis.

Details on RFLP genotyping assays used in the study are described in Chapter 5.

### 3.6.2 Genotyping with microsatellites

Microsatellite DNA is defined as a small array of tandem repeats that are simple in sequence and occur regularly throughout the genome (approximately every 6 kb) [187].

The high heterozygosity of microsatellite markers makes them suitable for population based association studies.

Many earlier studies used radioactive labelling of the PCR primers for microsatellite genotyping. The PCR products were run on a long polyacrylamide gel and alleles visualised by autoradiography. Between three to five markers of non-overlapping size could be run simultaneously. The method was effective for small studies, but it was time consuming and allele calling was often problematic.

Today, a fluorescent-based high-throughput technology provides a more efficient method for microsatellite genotyping. It can successfully be applied to both individual and pooled DNA samples. Up to six PCR product of non-overlapping size can be run concurrently. In addition, as one primer of each pair is labelled with one of the three available fluorescent labels, even products of overlapping size can be run together (see Figure No. 18).

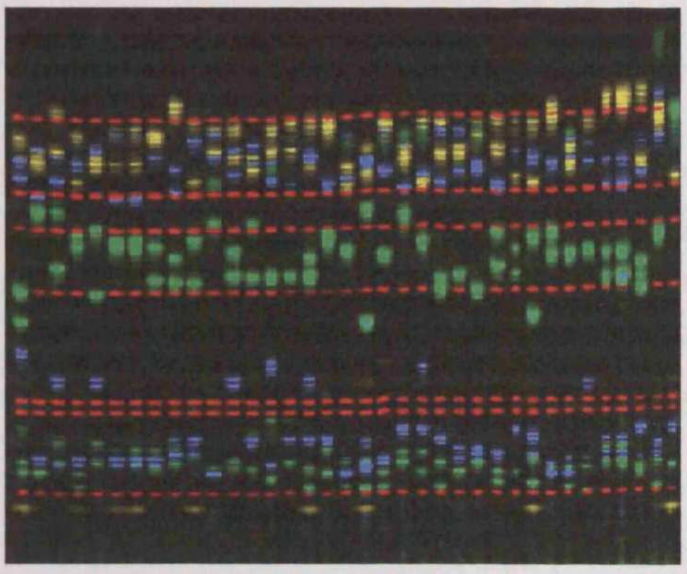


**Figure No. 18:** Image of a 96-well gel generated using GeneScan software. The size standard (TAMRA) appears red, the fluorescent microsatellite labels FAM, TET and HEX appear blue, green and yellow, respectively (with a kind permission of Dr N. Williams, Psychological Medicine, UHW, Cardiff)

More details on selection of microsatellite markers and genotyping using fluorescently labelled primers are given in Chapter 4 (individual and pooled DNA samples).

### 3.6.3 The SNaPshot™ Multiplex System assay genotyping

The chemistry is based on the dideoxy single nucleotide extension of an unlabeled oligonucleotide primer. The primer binds to a complementary template in the presence of fluorescently labelled ddNTPs and AmpliTaq DNA Polymerase. The polymerase extends the primer one nucleotide, adding a single fluorophore-labeled ddNTP to its 3′ end. Because the primer is designed to anneal directly adjacent to the variant base of interest and the reaction does not include dNTPs, incorporation occurs only at a single site. Each of the four possible dye labelled terminators in a SNaPshot ddNTP Primer Extension reaction is labelled with a different fluorescent dye. The labelled primer extension products are detected and analyzed by the ABI 3100 DNA Analyzer.

| ddNTP | Dye | Label Color of Analyzed Data |
|---|---|---|
| A | dR6G | Green |
| C | dTAMRA | Black |
| G | dR110 | Blue |
| T (U) | dROX | Red |

**Table No. 10:** Individual ddNTPs and their assigned fluorescent dyes.

A disadvantage of the method is that it is prone to yielding false positives. When running SNaPshot, negative control reactions always must be included, and if they do

not give the expected negative result, the rest of the data should be considered flawed. If even a small proportion of primers falsely extend on a "hairpin" or duplex the result will be a false positive. Extension primers with a complex secondary structure may also result in a SNaPshot assay failure. To reduce a probability of a primer forming a "hairpin" a software *FP Primer 1.0.1b* - by Dobril Ivanov (http://m034.pc.uwcm.ac.uk/FP_Primer.html) was used to asses sequence flanking a SNP in question. A flanking stretch of nucleotides with the least amount of a secondary structure would be selected and used as an extension primer. Detailed list of extension primers used in the CaMKII-α candidate gene study for SNaPshot genotyping is presented in Chapter 6.

| Stage | Protocol |
|---|---|
| PCR reaction | See above |
| Post-PCR cleaning with SAP and *ExoI* | See above |
| SNaPshot reaction | See below |
| Post-SNaPshot cleaning with SAP | 2.0 μl $H_2O$ and 2 μl SAP (1U/μl) are added to 10 μl SNaPshot reaction product and incubated at 37°C for 75 min |
| Preparation for analysis on ABI sequencers | 2.0 μl cleaned SNaPshot product with 8 μl HiDi Formamide |
| Store frozen (-20°C) | |

**Table No. 11:** Essential procedures involved in SNaPshot genotyping. Some of the stages were described in more detail earlier in the Chapter.

**Table No. 12:** SNaPshot reaction preparation

| SNaPshot reaction component | Volume (μl) |
|---|---|
| PCR product (cleaned with *ExoI* and SAP) | 2.0 |
| $H_2O$ | 2.0 |
| Reaction buffer | 2.5 |
| SNaPshot Ready Reaction Mix | 2.5 |
| Extension primer (0.5 pmol/μl) | 1.0 |

**Table No. 13:** SNaPshot reaction thermal cycling protocol

| Cycling Step | T° | Time | No of cycles |
|---|---|---|---|
| Denaturation | 94° | 2 min | 1 |
| Denaturation | 94° | 5 sec | 25 |
| Annealing | 43° | 5 sec | |
| Extension | 60° | 5 sec | |

The SNaPshot system can successfully be used for individual and pooled DNA genotyping. However, great attention has to be paid to interpreting pooled samples genotyping results as primer extension products are not equally represented for all alleles of a given genetic marker. Differential PCR amplification of alleles [188] and differential efficiency in ddNTP incorporation [189] are well known reasons for unequal representation of alleles. Fluorescent dyes used in the assay have unequal emission energies and result in different peak heights even when an individual heterozygous DNA sample is assayed. Therefore, estimated pool allele frequencies have to be corrected.

In order to allow for unequal allele representation, a simple and universal method [20] can be used to correct estimated allele frequencies from pools (see Fig. No. 19).

$$A_{corrected} = A / (A + kB)$$

**Figure No. 19:** Correction of estimated allele frequencies from pools. Frequency of allele A is calculated by applying the above formula, where A and B are the peak heights of the primer extension products (representing alleles A and B in pools) and $k$ is the mean of replicates of A/B ratios observed in a heterozygote [20].

## *3.7 Statistical Analysis*

Statistical tools employed for analysis of data collected from different studies are described in relevant chapters.

# 4 GENOME-WIDE SCREEN FOR ASSOCIATION WITH *g*

## *4.1 Introduction*

### 4.1.1 Background of the study

General cognitive ability, *g*, is one of the most heritable human behavioural traits [33]. A model-fitting analysis of cumulative data from more than a hundred world-wide studies on intelligence [87] suggests that about half of the variance in IQ scores can be attributed to genetic factors, most of which are additive in nature [33]. The next natural step in studying general cognitive ability would be to identify some of the genes responsible for its high heritability.

The complex nature of inheritance and a roughly normally distributed IQ scores suggest that universally observed variation in intelligence is governed by a mixture of genetic and environmental factors. The exact number of genes, quantitative trait loci, involved in shaping *g* is not known, and even the most accurate estimate is likely to be conservative (reasons for this are discussed in paragraph 1.4.2 of the General Introduction). The total genetic effect on the trait is not evenly distributed between genetic loci. The most plausible scenario will probably involve few modest effect loci and a large number of small effect loci [63, 67].

A small effect QTL is usually described as a locus that accounts for only about 1% or less of the variance [162]. Finding genes of such a small effect requires the power of an association rather than a linkage study design [109]. In the first attempt to identify some of the genetic loci contributing to variance in *g*, Plomin *et al.* (1994) genotyped

100 DNA markers in or near genes involved in brain functioning, but no replicated associations with *g* were found [163].

Instead of a candidate gene approach, an allelic association study can be made more systematic if a large number of densely distributed markers are genotyped along a chromosome. Chorney *et al.* (1998) made a first step in that direction by typing 47 microsatellite markers on the long arm of chromosome 6 [164]. A significant association for a marker positioned within the gene for insulin-like growth factor-2 receptor (IGF2R) was successfully replicated by the same group using larger samples [165]. Considering how difficult it is to find a true association, this was an encouraging result, which offered justification for further development of the strategy and continuation of the genome screen in search of association with *g*.

### 4.1.2 DNA pooling

#### 4.1.2.1 DNA pooling makes large genome-wide scans feasible

The problem with using a dense map of markers for whole-genome screens is the amount of genotyping required. Linkage disequilibrium with microsatellite markers is typically found within a distance of 500 kb (~0.5 cM) [190]. To cover the human genome with markers spaced at no more than 1 cM, a density desirable for a systematic genome screen for allelic association studies, one would need to genotype about 3500 microsatellite markers. This would ensure that, on average, any trait locus is less than 0.5 cM (~500 kb) from the marker. For LD studies using bi-allelic markers the situation is much worse since the suggested number of markers varies between 35,000 [132], >100,000 [133], 250,000 [146] and most pessimistic 1,000,000 [117]. For power to be large enough to detect small effect loci, one also has to aim for larger sample sizes (several

hundred). This would require the creation of several million genotypes, which is beyond the capabilities of most research laboratories.

However, the problem can be overcome by adopting a DNA pooling strategy [191, 192]. The need for genotyping is reduced by pooling DNA from all individuals within a group (both case group and control group) and then genotyping the pooled groups. So, for example, instead of genotyping 202 individuals in the first stage of the IQ genome scan (101 high *g* and 101 average *g* controls), we need to generate only six pool amplifications per microsatellite marker if the pools are amplified in triplicates (for description, see Chapter 2: General Materials and Methods). This method allows a rapid screening of a large number of markers and an efficient identification of those that are most likely to show allelic association. The method has been applied successfully in QTL association studies of both humans and animals [193, 194].

The strategy also benefits from the possibility of selective pooling of individuals. It has been shown that greater power can be achieved by selecting individuals at the phenotypic extremes of the population [195], and that fewer individuals need to be genotyped in order to retain the same level of power as with an unselected sample [194]. In comparison with standard selective genotyping, the selective pooling means doing even fewer genotypings per marker while only slightly decreasing the power (assuming the allele frequencies are estimated without error) [195].

It is worth noting that the collection of sufficiently large phenotypically extreme samples would require that initially a relatively large number of individuals be phenotyped. Nevertheless, the benefits associated with selective pooling justify greater initial efforts in phenotyping. With quantitative traits such as *g*, different lower thresholds can be chosen for cases (and controls). Table No. 14 gives examples of sample sizes that would be required for the IQ QTL study if different lower thresholds were to be used for selecting individuals for a high *g* group while maintaining the same level of power (in this example set to 95%).

| Total QTL variance | Distance of the selected High *g* group from the mean | | | |
|---|---|---|---|---|
| | +1 SD | +1.5 SD | +2 SD | + 2.5 SD |
| 0.05 | 194 | 118 | 78 | 54 |
| 0.025 | 404 | 246 | 162 | 113 |
| 0.01 | 1047 | 641 | 423 | 295 |

**Table No. 14:**

Estimates of sample size necessary for detection of QTLs that account for 5%, 2.5% and 1% of the total variance, assuming complete linkage disequilibrium (D'=1). For the original set, 101 cases (high *g* individuals) and 101 controls (average *g* individuals) were considered. Lower threshold for cases was taken to vary between 1 SD and 2.5 SD above the mean. Controls were selected to be ±0.67 SD from the mean; QTL and marker allele frequencies were set to 0.25.

#### 4.1.2.2 Problems associated with DNA pooling

Unlike individual genotyping, DNA pooling cannot provide us with real allele frequencies for a marker typed on pools. To be able to analyse data from pooled DNA amplifications and successfully employ them in allelic association studies, we must estimate marker allele frequencies.

The power of selective DNA pooling, as strong as it could be for detecting genes with small effects, can decrease steeply with the increase of error in estimating allele frequencies [195]. The errors mainly stem from three PCR artefacts: stutter and differential amplifications, both commonly associated with microsatellite repeats, and non-templated nucleotide addition.

**Stutter bands** result from the amplification of products one or two, and rarely more, repeat units shorter than the correct sized amplicon. This is thought to be due to slippage of *Taq* DNA polymerase on the repeated sequence [196], and/or out-of-register

annealing of truncated PCR products [197]. Stutter is marker specific and occurs more frequently in di-nucleotide repeats, although it can be seen, to a lesser extent, in tri- and tetra-nucleotides. The stutter bands overlap and add to the total signal coming from real alleles, making it difficult for us to determine accurately the frequencies of pooled alleles.

**Differential amplification** is a PCR artefact observed in heterozygotes, and it is a consequence of preferential amplification of the smaller allele. In some extreme circumstances, differential amplification may result in complete disappearance of the larger allelic product [198]. This is thought to result from longer alleles re-annealing at a faster rate, which reduces the efficiency of PCR amplification [198].

**Non-templated addition of a nucleotide to the 3' end** of amplification products, catalysed by *Taq* DNA polymerase, is a marker-specific activity and presents another potential source of error in genotyping studies [199]. Which nucleotide will be added depends on the type of polymerase used as well as the identity of the last residue of the DNA template [200]. Various DNA polymerases can add nucleotides with varying efficiencies [200]. Smith *et al.* estimate that between 1% and 3% genotyping error rate is attributable to this phenomenon [201].

#### 4.1.2.3 Solving problems

Several groups have attempted to solve some of the problems associated with previously described PCR artefacts. Two main approaches can be recognised: practical and mathematical.

I) ***Practical solutions***: Practical solutions have been offered for planning stages, when a researcher is advised to choose for his/her study only stutter-free microsatellite

markers. These can be tetra-nucleotide [202] and/or carefully selected di-nucleotide repeat markers [203]. However, solving the stutter problem by using tetra-nucleotides may result in an exaggerated problem of differential amplification [202].

It has been suggested that the problem of differential amplification can be avoided by raising the molar ratio of the *Taq* DNA polymerase to the template DNA during the PCR [204], or simply by decreasing the amount of DNA template subjected to the PCR [202].

The tendency to preferentially amplify lower molecular weight alleles can also be reduced by partial replacement of the physiologic dGTP by 7-deaza-2'-dGTP during amplification, which weakens the stability of inter-molecular G-C base pairing and ensures that DNA polymerase successfully completes primer extension [205, 206].

Demers *et al.* (1995) suggested addition of peptide nucleic acid (PNA) to PCR reaction. PNA is a DNA mimic with a unique ability to hybridise to DNA template and make it unavailable for premature re-annealing with a complementary DNA strand while allowing polymerase to displace it and extend the primer to completion [198].

There are also a few options for treatment of non-templated addition of a nucleotide at the 3' end: **1)** employment of a high fidelity (*Pyrococcus furiosus*) *Pfu* DNA polymerase, which is highly efficient in making blunt-ended PCR products [200], **2)** "pig-tail" primers, where a non-labelled primer of each pair has a consensus sequence tail (GTTTCTT) appended to its 5' end and produces nearly 100% adenylation of the 3' end of the forward strand [207], **3)** post-PCR treatment of amplified products with *E. coli* DNA polymerase I Klenow fragment to repair 3' overhanging ends and generate blunt-ended products, **4)** T4 DNA polymerase treatment of PCR products [208], or **5)** prolongation of the final extension step at 72°C following the amplification to 90 min, the method being probably most amenable to automated high throughput genotyping [201].

II) ***Mathematical solutions***: Le Duc *et al.* (1995) improved the accuracy in the estimation of allele frequencies by arithmetic removal of stutter bands [209]. They measured peak heights of all alleles, but before converting them into allele frequencies the stutter ratio was applied as a correction factor. The method was tested on two

microsatellites, D11S903 and D14S80, and estimated allele frequencies reflected real allele frequencies quite accurately [209]. However, the stutter ratio was calculated on the basis of the amount of stutter from the smallest allele only and not adjusted for the fact that the amount of stutter is allele-dependent.

More accurate mathematical stutter-correction was developed by Perlin *et al.* (1995). This method is based on the finding that each microsatellite marker has its own very reproducible stutter pattern. Also, rather than assuming an invariant stutter pattern for all alleles, Perlin *et al.* advise employment of algorithms that permit allele-dependent variation. Such algorithms showed, on average, less error by a factor of 10 in estimating allele frequencies than those assuming a constant stutter pattern [210].

Barcellos *et al.* (1997) have combined Perlin's method with mathematical correction for differential amplification. A preferential-amplification factor is calculated as a ratio of peak heights (X/Y), where X denotes the smaller peak. In the pooled samples, in addition to Perlin's correction for stutter, observed values for peak heights are further modified and divided by the preferential-amplification factor prior to conversion to allele frequencies [191].

Even though Perlin and Barcelloes created methods capable of producing quite accurate allele frequency estimates from pooled data, each individual marker required a lengthy work-up, which would be a substantial slowing-down factor in large-scale studies requiring many markers to be tested.

#### 4.1.2.4 The ΔAIP method

Dealing with differential amplification and stutter problems, at either practical or mathematical level, can be very challenging for studies planning to test thousands of microsatellite markers. To by-pass these problems altogether, Daniels *et al.* (1998)

developed a simple method for analysis of pooled genotypes that relies on statistical comparison of Allele Image Patterns (AIPs) for pooled cases and pooled controls [192]. For more detailed description of the method see Materials and Methods section.

The main difference between this and other methods is that the ΔAIP (Allele Image Pattern Difference) method is used as a rapid initial screening step to identify markers for individual genotyping. Its goal is not to estimate absolute allele frequencies for a single group (cases or controls) which requires adjustments for technical problems discussed, but simply to compare the uncorrected products of pooled microsatellite amplification.

The method does not require lengthy analysis of each marker prior to genotyping, which allows the application of DNA pooling to large-scale association studies.

## *4.2 Aim*

The aim of this study was to perform a genome-wide screen for the association with general cognitive ability, *g*, using a large number of densely (1 cM grid) spaced microsatellite markers. To make the study feasible, completed in time and within available resources, selective DNA pooling and the ΔAIP method were employed. A multi-stage study design was used to obtain a balance between false positives and false negatives in the search for QTLs of small effect size.

## *4.3 Materials and Methods*

### 4.3.1 Sample

See General Materials and Methods

### 4.3.2 Microsatellite marker selection

This computer-intensive initial stage of the study was carried out mainly by the project's post-doctoral research fellow, Paul J. Fisher, with some input from Linzy Hill (collaborating group in London IOP) and myself.

DNA markers were selected from the Genetic Location Database (LDB) summary maps (http://cedar.genetics.soton.ac.uk/public_html/summaryml.html) and ordered according to sex-averaged genetic distances available at the Marshfield Centre for Medical Genetics web site (http://research.marshfieldclinic.org/genetics/Map_Markers/maps/IndexMapFrames.html). When no genetic distance information was available, markers were positioned according to the physical map. Ideally, markers were 1 cM apart and had between 5 and 9 alleles and high heterozygosity scores (0.5 and 0.9). However, due to scarcity of markers in some regions, some less than ideal markers were chosen.

In total, 2871 microsatellite markers were selected for a screening set using the above conditions. Primer sets with fluorescently labelled reverse primers were purchased from MWG-Biotech in Germany. Additional 117 primer pairs (ABI PRISM® Linkage Mapping Set v2.5 ) were used from marker stocks already available at the laboratory. This brought the total marker number to 2988 with average inter-marker genetic distance of 1.5 cM, and with ~80% of the genome within 1 cM of a marker.

Coverage success varied between and within chromosomes, with somewhat poorer coverage toward telomeric regions.

Details on the 1847 microsatellite markers (allele numbers, heterozygosity scores, amplicon sizes, primer sequences) that were successfully genotyped on pools using standard PCR conditions (described in section 4.3) can be found on the web site developed by the London project team (http:\\www.sgdp.org.uk).

### 4.3.3 ΔAIP statistic calculation

Allele image patterns are produced for cases and controls by a standard ABI Genescan software, overlaid in Genotyper (see Figure No. 20) and imported into commercially available graphics software Debabelizer (Equilibrium).

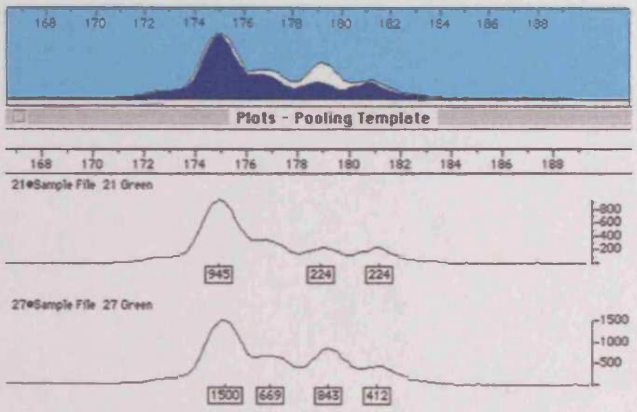



**Figure No. 20:**

Overlaid Allele Image Patterns (AIPs) with differently coloured shared (dark blue) and non-shared (white) areas; area sizes are determined from the pixel number for each of the colours representing shared and non-shared areas.

The ΔAIP statistic is calculated as a function of the total shared and non-shared area, whose sizes are determined from the pixel number of colours used to distinguish the areas:

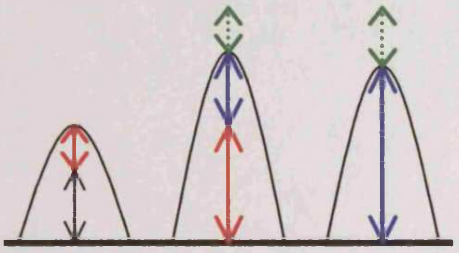
$$\Delta AIP = \frac{\text{Non-shared area}}{\text{Non-shared + Shared area}}$$

The statistic is an estimate of relative difference in allele frequencies between the groups. The significance of the ΔAIP is then determined by a computer simulation program [192].

### 4.3.4 Allele Specific Test (AST) statistic calculation

The test incorporates a simple algorithm for "blanket" correction of stutter and differential amplification. Algorithms for di-, tri- and tetra-nucleotide repeat markers were derived empirically by Paul J. Fisher (unpublished). Pearson $\chi^2$ with a 2x2 contingency table is used to compare estimated frequency of the target allele versus all other alleles.

**Figure No. 21:**



Simplified explanation of the algorithm used for stutter and differential amplification correction prior to AST analysis. Three peaks from amplified pooled DNA are shown. Differently coloured arrows denote various sources contributing to peak height generation. So, for example, middle peak height will have to be reduced for the amount of the stutter coming form the right peak (blue arrow), increased for the amount of its own stutter present in the left peak (red arrow) and finally increased for the amount of differential amplification expected for the particular allele (green dashed arrow).

### 4.3.5 Multi-stage study design

As previously suggested by Risch and Teng (1998), studies using pooled DNA should ideally have at least two stages: first, screening DNA pools by pooled genotyping and second, individual genotyping of positive markers identified in the initial screen [211]. In this study, a four stage design with two independent samples was used in order to provide balance between false positive and false negative results in search for QTLs of small effect size (See Figure No. 22).

Stage 1: Original case/control sample – DNA pools

This is the initial screening stage in which all markers were tested on pooled samples of original 101 high *g* individuals and 101 average *g* controls. ΔAIP and simulated p-values are calculated for each of the successfully amplified microsatellites.

Markers yielding $p<0.05$ were subjected to an Allele Specific Test (AST) in order to establish which alleles show the largest difference between the high *g* and control group. Only markers with $p<0.05$ for both ΔAIP and AST qualified for the Stage 2.

Stage 2: Replication case/control sample – DNA pools

The allele-specific directional hypothesis from Stage 1 was tested in the replication case-control sample using pooled DNA. A one-tailed test of significance was used because the target allele was required to show allele frequency difference between the groups in the same direction as in Stage 1. For a marker to qualify for the next stage, a p-value had to be $<0.05$.

Stage 3: Individual genotyping of the original sample set

Markers surviving Stages 1 and 2 were individually genotyped on the original case/control sample. The allele-specific directional hypothesis was tested for significance using one-tailed Pearson $\chi^2$ test comparing that allele against all others.

Stage 4: Individual genotyping of the replication sample set

Finally, Stage 3 markers with $p<0.05$ were individually genotyped on the replication case/control set. Pearson $\chi^2$ test (one-tailed) was used to asses significance of allele-specific directional hypothesis.

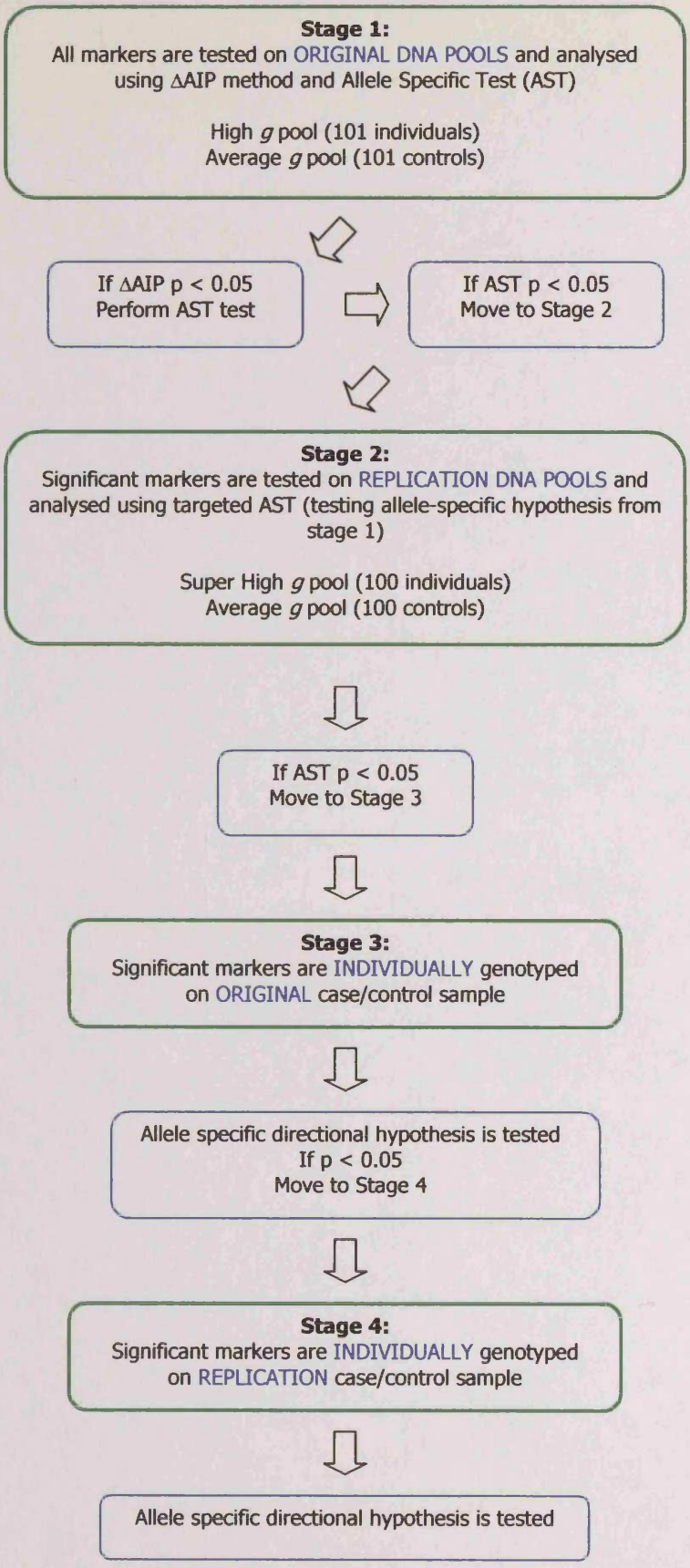



**Figure No. 22:** Graphical presentation of a multi-stage study design adopted for the genome-wide screen for association with general cognitive ability.

### 4.3.6 DNA pooling

See General Materials and Methods

### 4.3.7 Pool genotyping

Each of the three replica pools for "high" *g* and "average" *g* was amplified at the same time, using a touchdown PCR protocol. For each microsatellite marker a separate reaction was performed. Amplification products were pooled while care was taken to combine different fluorescent dyes (Hex, Fam and Tet) and product sizes. Up to four different markers were loaded into one gel lane. Because of the difference in fluorescing intensity, PCR products had to be diluted prior to pooling: Fam labelled 8x, Tet 6x and Hex 5x. Diluted and pooled PCR products (1.5μl) were mixed with loading dye (1.5μl) and GeneScan®-500 size ladder (0.5μl) and size separated on 12.5% denaturing gels using ABI373A sequencer.

Gels were analysed and allele image patterns (AIPs) generated for each of the products with the GENESCAN® software. Visualisation and overlaying of AIPs was carried out using GENOTYPER®1.1.1. Only one of the three AIP images generated by each of the two groups was selected as a consensus image.

In order to compare the results of the pooled genotyping between high *g* and control group, consensus images were overlaid and imported into a graphics software DeBabelizer™. Different colours were used to fill in shared and non-shared areas (see paragraph No. 4.1.2.4), whose sizes were then expressed by the software as a number of pixels and used in the formula for ΔAIP calculation. Markers were tested for significant ΔAIP using a computer simulation program (for description see paragraph No. 4.3.3).

Allele counts and frequencies were estimated form pooled data using the algorithm described in section No. 4.3.4. Target allele significance was tested using Pearson $\chi^2$.

### 4.3.8 Simulated p-value

Absolute ΔAIP value does not realistically reflect the significance of difference in allele frequencies observed between two pools because the significance of ΔAIP will depend on the number of parameters, such as sample size, number of marker alleles and allele frequency. The test for significance was performed by a statistical method based on a computer program, which, with 2000 simulations per test, simulates case and control samples from a population with allele frequencies estimated from the peak heights of the control sample [192].

### 4.3.9 Individual genotyping

For individual genotyping, PCRs and gel runs were performed using the same protocol, equipment and software as described for pooled genotyping, except for the last step, which involves image overlaying and ΔAIP calculation. The significance of the target allele was estimated using Pearson $\chi^2$.

### 4.3.10 Power

Power analysis was carried out using publicly available software "Genetic Power Calculator" (http://statgen.iop.kcl.ac.uk/gpc/index.html#qcc_ins). All calculations in the program are based upon formula derived by Sham *et al.* (2000) [212]. Detailed instructions on entering values into required fields are available on the web site.

Calculations were performed separately for original and replication sample sets. Nine different scenarios were considered: QTL that accounts for 5%, 2.5% and 1% of the phenotypic variance, assuming three different levels of LD (D'=1, 0.75 and 0.5). The results are presented in Tables No. 15 and No. 16.

| | Power to detect QTL ($\alpha$=0.05) | | |
|---|---|---|---|
| Total QTL variance | D'=1 | D'=0.75 | D'=0.50 |
| 0.05 | 100% | 96% | 73% |
| 0.025 | 93% | 74% | 42% |
| 0.01 | 56% | 36% | 19% |

**Table No. 15:** Power estimates used to detect QTLs that account for 5%, 2.5% and 1% of the total variance under different levels of linkage disequilibrium (D'); For the original set, 101 cases (high *g* individuals) and 101 controls (average *g* individuals) were considered; Lower threshold for cases was taken to be 2.5 SD above the mean; Controls were selected to be ±0.67 SD from the mean; QTL and marker allele frequencies were set to 0.25.

| | Power to detect QTL ($\alpha$=0.05) | | |
|---|---|---|---|
| Total QTL variance | D'=1 | D'=0.75 | D'=0.50 |
| 0.05 | 100% | 100% | 99% |
| 0.025 | 100% | 99% | 91% |
| 0.01 | 97% | 85% | 53% |

**Table No. 16:** Power estimates used to detect QTLs that account for 5%, 2.5% and 1% of the total variance under different levels of linkage disequilibrium (D'); For the replication set 100 cases (super-high *g* individuals) and 96 controls (average *g* individuals) were considered; Lower threshold for cases was taken to be 5 SD above the mean; Controls were selected to be ±0.67 SD from the mean; QTL and marker allele frequencies were set to 0.25.

## 4.4 *Results*

### 4.4.1 General notes

Genome-wide screen workload was shared between Cardiff and London groups collaborating on the IQ QTL project. All DNA pools were created in Cardiff: the original by myself, and replication by the project's post-doctoral research fellow Dr Paul Fisher. Both groups were assigned a set of chromosomes to work on. Each of the chromosomes was expected to be carried through all the stages by the same group it initially had been assigned to. The exception were the ABI linkage panel microsatellite markers, which were tested in Cardiff regardless of their chromosomal position. I was responsible for genotyping most of the microsatellite markers assigned to Cardiff group (~1600 markers). All ΔAIPs, a proportion of the simulated p-value calculations and some AST analysis for Cardiff markers were calculated by myself, while the rest of the simulated p values, AST and other statistical tests were carried out by Dr Paul Fisher. The analysis of markers tested in London was carried out by the London project group.

| Chr. No. | Average inter-marker distance (cM) | Largest inter-marker distance (cM) | Chr. No. | Average inter-marker distance (cM) | Largest inter-marker distance (cM) |
|---|---|---|---|---|---|
| **1** | 2.18 | 10.01 | **12** | 1.65 | 5.76 |
| **2** | 1.77 | 9.04 | **13** | 1.54 | 7.82 |
| **3** | 1.56 | 5.59 | **14** | 2.38 | 8.58 |
| **4** | 2.01 | 13.55 | **15** | 2.66 | 11.63 |
| **5** | 1.83 | 6.47 | **16** | 2.05 | 10.33 |
| **6** | 1.78 | 9.18 | **17** | 2.68 | 9.60 |
| **7** | 1.78 | 8.94 | **18** | 2.91 | 19.37 |
| **8** | 2.14 | 7.52 | **19** | 2.87 | 10.97 |
| **9** | 2.31 | 11.70 | **20** | 2.35 | 8.72 |
| **10** | 2.86 | 10.32 | **21** | 2.23 | 8.05 |
| **11** | 2.71 | 9.59 | **22** | 2.57 | 10.38 |

**Table No. 17:** Average inter-marker distances (cM) for each of the chromosomes analysed. Limited microsatellite availability resulted in a number of gaps in our final map. Largest gaps (cM) for each of the chromosomes are presented.

### 4.4.2 Stage 1

Of the 2988 microsatellite markers tested, ΔAIP and simulated p-values were obtained for 1847 (62%). Using standard touchdown PCR conditions (see General Material and Methods), 1588 markers (53%) yielded amplification products with highly replicable allele image patterns, where at least two out of three AIPs generated for each of the groups gave nearly complete overlays. The remaining 1400 markers went through a second amplification protocol, using *Taq* Gold polymerase (see General Materials and Methods). This generated replicable AIPs for another 259 markers (9%).

| | **Total No. of markers tested** | **No. of $\Delta AIP_{ori}$ results (%)** | **No. of markers with simulated $p_{ori}<0.05$ (%)** | **No. of markers with AST $p<0.05$ (%)** |
|---|---|---|---|---|
| **Cardiff** | 1641 | 1021 (62%) | 51 (5%) | 38 (75%) |
| **London** | 1347 | 826 (61%) | 186 (23%) | 70 (38%) |
| **TOTAL** | 2988 | 1847 (62%) | 237 (13%) | 108 (46%) |

**Table No. 18:** Total number of markers considered for the IQ QTL study and their progress through different levels of the Stage 1 are presented for each of the two collaborative centres (Cardiff and London). Cardiff markers come from chromosomes 2, 3, 4, 5, 6, 7, 12, 14, 15, and 17. Cardiff centre tested only a very small proportion of microsatellite markers belonging to other chromosomes (those that were part of the ABI linkage mapping panel).

For a large number of markers (~1/3 of all markers purchased), ΔAIP could not be calculated for various reasons. First, for a significant proportion of the markers no amplification product could be recovered using standard PCR conditions with either of the *Taq* polymerases. Second, for a number of microsatellites amplicons were of poor

quality, which meant that images could not be overlaid and could not be used for ΔAIP analysis. Third, in some instances non-templated addition of a nucleotide at the 3' end was very pronounced and could not be resolved by Klenow fragment treatment of the product. Some markers had to be discarded because AIPs showed significant variation within the group and no consensus image could be selected as a representative of that group.

Out of 1847 generated ΔAIPs, 237 had significant simulated $p<0.05$. Only 92 markers were expected to be significant by chance alone with $p<0.05$. More than three quarters of positive markers came from the London group (see Table 18). The most likely explanation was generation of false positive results through accepting less than perfect AIP overlays. A large proportion of these (62%) were filtered out after the AST analysis was performed. In contrast, only 5% of Cardiff markers had significant p-values before the AST analysis, with most of them (75%) staying significant after further analysis, possibly due to more rigorous criteria for accepting AIP overlays.

Of the total of 237 markers, 108 had significant AST results ($p<0.05$) and qualified for the second stage.

### 4.4.3 Stage 2

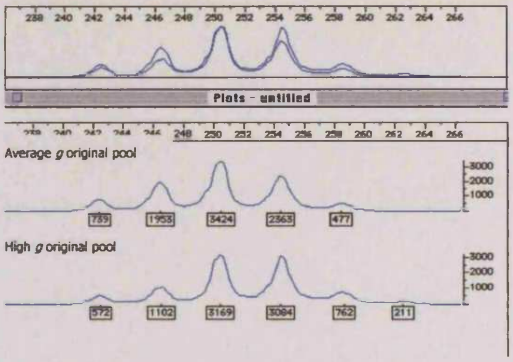
108 markers reached Stage 2, where they were tested on independent replication sets of pooled DNA samples (Stage 2 markers that were done in our laboratories are listed in Table No. 19). Of these, 7 (D2S427, D3S2425, D4S2460, D4S175, D6S277, D16S687 and D18S1110) showed significant simulated p-values ($p<0.05$) when their AIP images were analysed, while only 5 were expected to be significant by chance alone. Four markers (D3S2425, D4S175, D16S687 and D18S1110) did not have a significant AST result, or they had a significant AST, but their association was in the

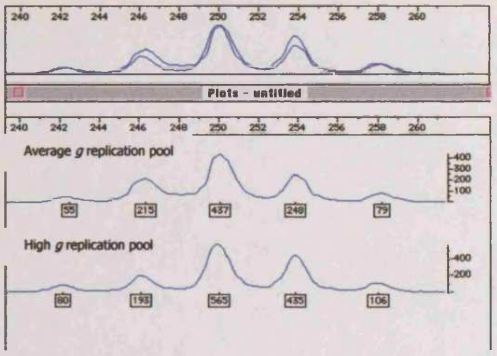
opposite direction (*i.e.* allele was more frequent in high *g* individuals than in controls in the original sample set, but was significantly less frequent in high *g* individuals than in controls in the replication set, and *vice versa*; which was observed for D3S2425), or they had a different allele showing association in the replication set (observed for D6S277).

| No. | Marker ID | Simulated $p_{ori}$ | AST $p_{ori}$ | No. | Marker ID | Simulated $p_{ori}$ | AST $p_{ori}$ |
|---|---|---|---|---|---|---|---|
| 1 | D2S285 | 0.023 | 0.019 | 20 | D5S1726 | 0.040 | 0.010 |
| 2 | D2S2264 | 0.000 | 0.000 | 21 | **D6S277** | 0.020 | 0.039 |
| 3 | D2S2277 | 0.004 | 0.038 | 22 | D6S506 | 0.020 | 0.010 |
| 4 | D2S1776 | 0.013 | 0.019 | 23 | D6S1597 | 0.010 | 0.000 |
| 5 | D2S2978 | 0.021 | 0.049 | 24 | D6S1564 | 0.010 | 0.038 |
| 6 | **D2S427** | 0.026 | 0.007 | 25 | D6S446 | 0.000 | 0.008 |
| 7 | D3S1217 | 0.003 | 0.011 | 26 | D7S1508 | 0.001 | 0.000 |
| 8 | D3S2433 | 0.000 | 0.003 | 27 | D7S1874 | 0.003 | 0.011 |
| 9 | **D3S2425** | 0.020 | 0.001 | 28 | D7S2501 | 0.010 | 0.004 |
| 10 | D3S3699 | 0.005 | 0.009 | 29 | D12S1632 | 0.018 | 0.046 |
| 11 | D3S3578 | 0.001 | 0.005 | 30 | D12S90 | 0.003 | 0.005 |
| 12 | MSX1 | 0.002 | 0.005 | 31 | D12S1727 | 0.010 | 0.000 |
| 13 | D4S404 | 0.046 | 0.015 | 32 | D14S749 | 0.002 | 0.000 |
| 14 | D4S3001 | 0.020 | 0.001 | 33 | D14S65 | 0.040 | 0.028 |
| 15 | **D4S2460** | 0.009 | 0.027 | 34 | D17S831 | 0.049 | 0.013 |
| 16 | **D4S175** | 0.026 | 0.004 | 35 | D17S1863 | 0.004 | 0.000 |
| 17 | D4S1607 | 0.019 | 0.003 | 36 | D17S1794 | 0.009 | 0.048 |
| 18 | D4S2943 | 0.038 | 0.015 | 37 | D17S1299 | 0.023 | 0.021 |
| 19 | D5S1980 | 0.000 | 0.000 | 38 | D17S1350 | 0.040 | 0.000 |

**Table No. 19:** The microsatellite markers tested in Cardiff as part of Stage 2. Those typed in **blue** and bold had simulated $p<0.05$ in the replication sample set.

Figure No. 23: D2S427 allele image patterns




Left

D2S427 allele image patterns generated by the ORIGINAL high *g* and average *g* DNA pools and the overlaid image used for ΔAIP calculation.

ΔAIP=0.16

sim p=0.03

AST p=0.01 (peak 2)

Right

D2S427 allele image patterns generated by the REPLICATION high *g* and average *g* DNA pools and the overlaid image used for ΔAIP calculation.

ΔAIP=0.15

sim p=0.04

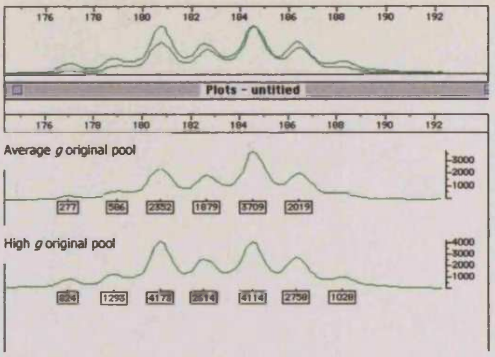
AST p=0.03 (peak 2)

| D2S427 | Description | Peak 1 | Peak 2 | Peak 3 | Peak 4 | Peak 5 |
|---|---|---|---|---|---|---|
| Average *g* original pool | Uncorrected peak height | 739 | 1953 | 3424 | 2363 | 477 |
| | Corrected peak height | 678 | 1869 | 3577 | 2997 | 572 |
| | Estimated allele frequency (%) | 7.0 | 19.3 | 36.9 | 30.9 | 5.9 |
| **High *g* original pool** | Uncorrected peak height | 572 | 1102 | 3169 | 3084 | 762 |
| | Corrected peak height | 550 | 968 | 3247 | 3479 | 914 |
| | Estimated allele frequency (%) | 6.0 | 10.6 | 35.5 | 38.0 | 10.0 |
| **Allele frequency difference (%)** | | **1.0** | **8.7** | **1.4** | **-7.1** | **-4.1** |

| D2S427 | Description | Peak 1 | Peak 2 | Peak 3 | Peak 4 | Peak 5 |
|---|---|---|---|---|---|---|
| Average *g* replication pool | Uncorrected peak height | 55 | 215 | 437 | 248 | 79 |
| | Corrected peak height | 47 | 201 | 466 | 277 | 95 |
| | Estimated allele frequency (%) | 4.3 | 18.5 | 42.9 | 25.5 | 8.7 |
| **High *g* replication pool** | Uncorrected peak height | 80 | 193 | 565 | 435 | 106 |
| | Corrected peak height | 75 | 168 | 590 | 491 | 127 |
| | Estimated allele frequency (%) | 5.2 | 11.6 | 40.7 | 33.8 | 8.8 |
| **Allele frequency difference (%)** | | **-0.8** | **6.9** | **2.2** | **-8.3** | **0.0** |

**Tables No. 20 and 21:** Calculation of estimated allele frequencies for D2S427 in original and replication DNA pools. Selection of a target allele (Peak 2) was based on the ORIGINAL pools' results. Target allele is printed in RED.

**Figure No. 24:** D4S2460 allele image patterns



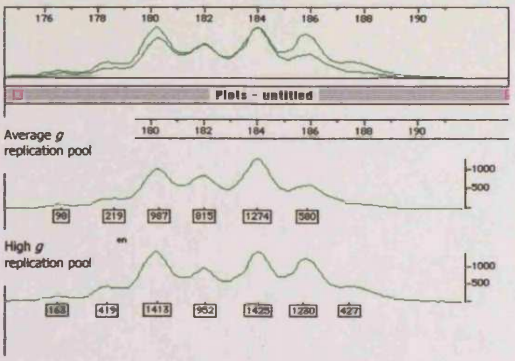

Left

**D4S2460** allele image patterns generated by the ORIGINAL high *g* and average *g* DNA pools and the overlaid image used for ΔAIP calculation.

ΔAIP=0.22

sim p=0.01

AST p=0.03 (peak 5)




Right

**D4S2460** allele image patterns generated by the REPLICATION high *g* and average *g* DNA pools and the overlaid image used for ΔAIP calculation.

ΔAIP=0.21

sim p=0.00

AST p=0.01 (peak 5)

| D4S2460 | Description | Peak 1 | Peak 2 | Peak 3 | Peak 4 | Peak 5 | Peak 6 |
|---|---|---|---|---|---|---|---|
| Average *g* original pool | Uncorrected peak height | 277 | 586 | 2352 | 1879 | 3709 | 2019 |
| | Corrected peak height | 288 | 179 | 2554 | 1304 | 3965 | 2650 |
| | Estimated allele frequency (%) | 2.6 | 1.6 | 23.3 | 11.9 | 36.2 | 24.2 |
| **High *g* original pool** | Uncorrected peak height | 824 | 1293 | 4173 | 2614 | 4114 | 2758 |
| | Corrected peak height | 842 | 610 | 4567 | 2159 | 4198 | 3620 |
| | Estimated allele frequency (%) | 5.3 | 3.8 | 28.6 | 13.5 | 26.2 | 22.6 |
| **Allele frequency difference (%)** | | **-2.6** | **-2.2** | **-5.2** | **-1.6** | **10.0** | **1.6** |

| D4S2460 | Description | Peak 1 | Peak 2 | Peak 3 | Peak 4 | Peak 5 | Peak 6 |
|---|---|---|---|---|---|---|---|
| Average *g* original pool | Uncorrected peak height | 98 | 219 | 987 | 815 | 1274 | 580 |
| | Corrected peak height | 104 | 48 | 1049 | 648 | 1406 | 761 |
| | Estimated allele frequency (%) | 2.6 | 1.2 | 26.1 | 16.1 | 35.0 | 19.0 |
| **High *g* original pool** | Uncorrected peak height | 163 | 419 | 1413 | 952 | 1425 | 1230 |
| | Corrected peak height | 155 | 189 | 1523 | 834 | 1348 | 1614 |
| | Estimated allele frequency (%) | 2.7 | 3.3 | 26.9 | 14.7 | 23.8 | 28.5 |
| **Allele frequency difference (%)** | | **-0.1** | **-2.1** | **-0.8** | **1.4** | **11.2** | **-9.6** |

**Tables No. 22 and 23:** Calculation of estimated allele frequencies for D4S2460 in original and replication DNA pools. Selection of a target allele (Peak 5) was based on the ORIGINAL pools' results. Target allele is printed in RED.

### 4.4.4 Stages 3 and 4

The two markers reaching Stage 3 (D2S427 and D4S2460) were individually genotyped on the original sample set (Table No. 24). The individual genotyping confirmed the results of DNA pooling in that both markers showed differences between the groups for the same allele and in the same direction as those detected in Stages 1 and 2 (D2S427, $\chi^2$=2.76, p=0.048 and D4S2460, $\chi^2$=5.40, p=0.010).

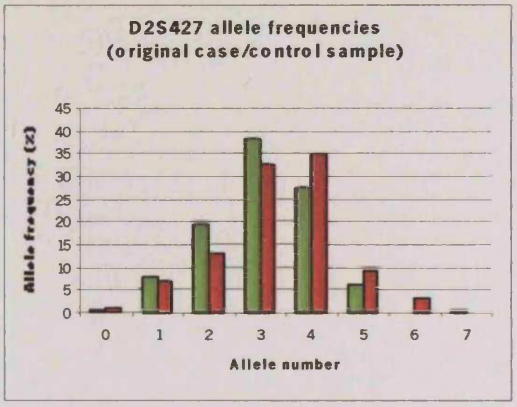
Allele specific results were also in the expected direction in the case-control replication sample. Stage 4 results were significant for one of the two markers, D4S2460 ($\chi^2$=2.98, p=0.042). Detailed genotyping results for the two markers are shown in Table No. 24. Graphical illustration of markers' allele frequencies can be seen in figures No. 25 and 26.

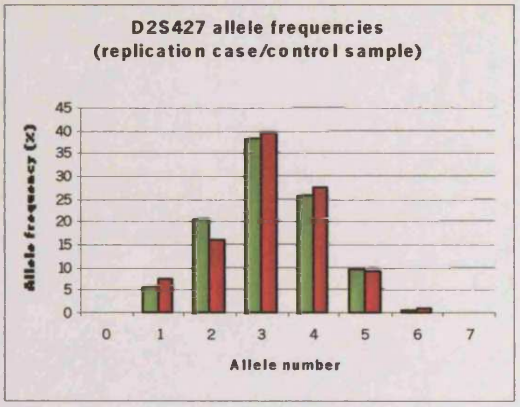
**Table No. 24:** Genotyping results for 7 markers qualifying for Stage 2.

| | Stage 1 | | Stage 2 | | Stage 3 | | | | Stage 4 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Original case-control sample | | Replication case-control sample | | Original case-control sample | | | | Replication case-control sample | | | |
| Marker (target allele) | ΔAIP-sim p-value | AST p-value | ΔAIP-sim p-value | AST p-value | High *g* | Control *g* | $X^2$ | p | High *g* | Control *g* | $X^2$ | p |
| D2S427 (2) | 0.026 | 0.007 | 0.04 | **0.030** | 27 (13%) | 39 (19%) | 2.76 | **0.048** | 25 (16%) | 38 (21%) | 1.08 | 0.149 |
| D3S2425 (4) | 0.020 | 0.001 | 0.00 | 0.059[1] | | | | | | | | |
| D4S2460 (5) | 0.009 | 0.027 | 0.00 | **0.013** | 56 (29%) | 77 (40%) | 5.40 | **0.010** | 50 (29%) | 73 (38%) | 2.98 | **0.042** |
| D4S175 | 0.026 | 0.004 | 0.01 | 0.292 | | | | | | | | |
| D6S277 (3) | 0.020 | 0.039 | 0.00 | 0.371 | | | | | | | | |
| D16S687 (4) | 0.000 | 0.001 | * | 0.073 | | | | | | | | |
| D18S1110 (5) | 0.000 | 0.035 | * | 0.160 | | | | | | | | |

[1]Target allele in opposite direction to the original case-control sample set

*Data not available



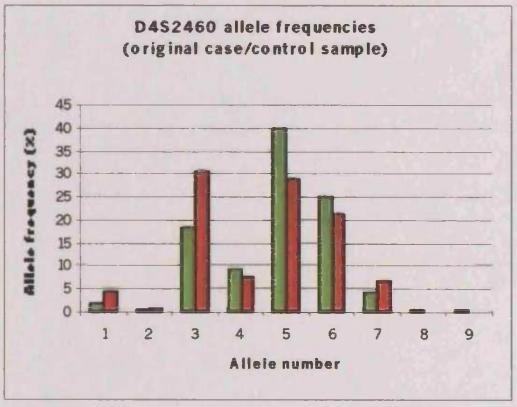

AST (allele 2): $\chi^2 = 3.13$
$p = 0.038$

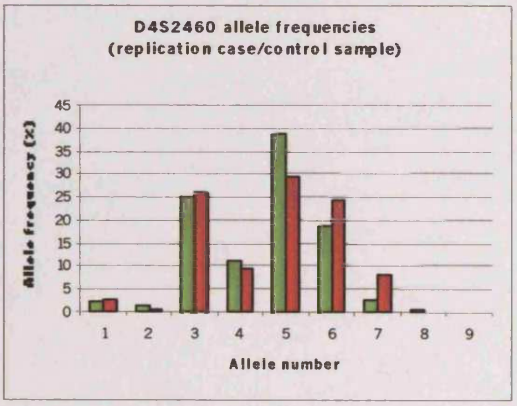
AST (allele 2): $\chi^2 = 1.25$
$p = 0.132$

Average *g* individuals

High *g* individuals

**Figure No. 25:** Illustration of allele frequencies for microsatellite marker **D2S427** obtained by individual genotyping in the original sample set (left) and the replication sample set (right). Also shown are AST results with one-tailed p-values.




AST (allele 5): $\chi^2 = 5.40$
$p = 0.010$

AST (allele 5): $\chi^2 = 3.51$
$p = 0.030$

Average *g* individuals

High *g* individuals

**Figure No. 26:** Illustration of allele frequencies for microsatellite marker **D4S2460** obtained by individual genotyping in the original sample set (left) and the replication sample set (right). Also shown are AST results with one-tailed p-values.

### 4.4.5 Population stratification control

To check whether significant results generated during the study arise from population stratification, a method suggested by Pritchard and Rosenberg (1999) was applied [6]. One modification to the method was that marker allele frequencies were estimated and did not come from individual genotyping of the sample.

A set of 37 randomly chosen microsatellite markers on chromosome 6 was used for this purpose. For each marker allele frequencies were estimated on pooled DNA samples from Stage 1, using the algorithm described in section 4.3.4. The most frequent allele for each of the 37 markers was then compared to all other alleles. This yielded a non-significant result ($\chi^2$=22.914, p=0.966), suggesting that population stratification does not contribute significantly to observed differences between high *g* and average *g* individuals (See Table 25).

| Marker ID | Group | Allele 1 Counts | Allele 2 Counts | $\chi^2$ | | Marker ID | Group | Allele 1 Counts | Allele 2 Counts | $\chi^2$ |
|---|---|---|---|---|---|---|---|---|---|---|
| D6S1697 | Mid | 90 | 112 | | | D6S1644 | Mid | 83 | 119 | |
| | High | 97 | 105 | 0.488 | | | High | 78 | 124 | 0.258 |
| D6S1598 | Mid | 82 | 120 | | | D6S1717 | Mid | 75 | 127 | |
| | High | 91 | 111 | 0.819 | | | High | 82 | 120 | 0.510 |
| D6S1955 | Mid | 70 | 132 | | | D6S408 | Mid | 96 | 106 | |
| | High | 79 | 123 | 0.861 | | | High | 84 | 118 | 1.443 |
| D6S1279 | Mid | 82 | 120 | | | D6S1620 | Mid | 64 | 138 | |
| | High | 62 | 140 | 4.316 | | | High | 67 | 135 | 0.102 |
| D6S1034 | Mid | 66 | 136 | | | D6S413 | Mid | 130 | 72 | |
| | High | 67 | 135 | 0.011 | | | High | 130 | 72 | 0.000 |
| D6S1721 | Mid | 82 | 120 | | | D6S270 | Mid | 89 | 113 | |
| | High | 85 | 117 | 0.092 | | | High | 94 | 108 | 0.250 |
| D6S1593 | Mid | 109 | 93 | | | D6S1699 | Mid | 83 | 119 | |
| | High | 104 | 98 | 0.248 | | | High | 82 | 120 | 0.010 |
| D6S285 | Mid | 83 | 119 | | | D6S314 | Mid | 48 | 154 | |
| | High | 90 | 112 | 0.495 | | | High | 47 | 155 | 0.014 |
| D6S1266 | Mid | 68 | 134 | | | D6S1704 | Mid | 62 | 140 | |
| | High | 60 | 142 | 0.732 | | | High | 68 | 134 | 0.408 |
| D6S1029 | Mid | 133 | 69 | | | D6S1637 | Mid | 50 | 152 | |
| | High | 127 | 75 | 0.388 | | | High | 53 | 149 | 0.117 |
| D6S1629 | Mid | 62 | 140 | | | D6S1064 | Mid | 111 | 91 | |
| | High | 65 | 137 | 0.103 | | | High | 110 | 92 | 0.009 |
| D6S1680 | Mid | 76 | 126 | | | D6S425 | Mid | 73 | 129 | |
| | High | 67 | 135 | 0.877 | | | High | 60 | 142 | 1.894 |
| D6S1641 | Mid | 119 | 83 | | | D6S1612 | Mid | 82 | 120 | |
| | High | 134 | 68 | 2.379 | | | High | 91 | 111 | 0.819 |
| D6S282 | Mid | 37 | 165 | | | D6S411 | Mid | 113 | 89 | |
| | High | 31 | 171 | 0.636 | | | High | 108 | 94 | 0.250 |
| D6S1650 | Mid | 64 | 138 | | | D6S1719 | Mid | 68 | 134 | |
| | High | 65 | 137 | 0.011 | | | High | 64 | 138 | 0.180 |
| D6S1628 | Mid | 63 | 139 | | | D6S1027 | Mid | 97 | 105 | |
| | High | 57 | 145 | 0.427 | | | High | 93 | 109 | 0.159 |
| D6S1625 | Mid | 90 | 112 | | | D6S281 | Mid | 85 | 117 | |
| | High | 75 | 127 | 2.305 | | | High | 79 | 123 | 0.369 |
| D6S1634 | Mid | 92 | 110 | | | D6S1626 | Mid | 67 | 135 | |
| | High | 97 | 105 | 0.248 | | | High | 72 | 130 | 0.274 |
| D6S1601 | Mid | 67 | 135 | | | | | | | |
| | High | 61 | 141 | 0.412 | | | | | | |
| | | | | | | **Total** | | **df=37** | $\Sigma X^2$ | **22.914** |
| | | | | | | | | | **p** | **0.966** |

**Table No. 25:** Population stratification analysis [6]. Analysis was performed according to suggestion by Pritchard and Rosenberg (1999). Sum of individual $\chi^2$ values ($\Sigma X^2$) and the overall p-value of randomly selected 37 chromosome 6 microsatellite markers are presented at the bottom of the table.

## 4.5 Discussion

A systematic genome wide screen with close to 2000 microsatellite markers was carried out in an attempt to locate some of the genetic loci contributing to the total variance in *g,* and it is probably the first association study to employ a DNA pooling technique on such a large scale.

The objective of the study was to test ~3500 microsatellite markers with average inter-marker distance of 1 cM. To be selected for the screen, markers had to be highly polymorphic and to have high heterozygosity in order to increase the power to detect linkage disequilibrium with a trait locus [157], and preferably be 1 cM away from the first neighbouring selected microsatellite. Less then 3000 markers fitted those requirements.

Initially, the main focus was on finding optimal PCR conditions at which most markers would successfully amplify at first attempt. Once a successful protocol was established, the focus shifted to a non-templated addition of a nucleotide at the 3' end, which lead to serious problems of misinterpreted AIPs, falsely inflated simulated p-values and non-reproducible results. This was mainly resolved by introducing a universal and amenable to high throughput screening post-PCR Klenow fragment treatment of all markers, regardless of their predisposition to such PCR artefacts. In addition to increasing the quality of obtained AIPs, standardised protocols insured less variation in results when tests were carried out by different researchers, and it greatly contributed to a considerable speed at which the screening process was performed.

Although such measures ensured speed and reproducibility, they resulted in a relatively large number of markers which failed to amplify (38%) since optimisation of individual markers was not carried out. In consequence, only a 2 cM average marker spacing was achieved instead of the aimed tighter 1 cM grid.

Unlike markers with uncertain PCR behaviour selected by our laboratory, commercially designed genome screening panels of microsatellites (Applied Biosystems),

with incorporated "pig-tails" at their 5' ends, proved to be much more successful, with only 13% failing to generate good quality AIPs. For this reason, research groups planning systematic genome screens may wish to consider readily available marker panels that have been designed and optimised to produce accurate and reliable genotyping results under standard conditions. Greater initial cost may soon prove to be a good investment since, by having fewer dropouts, better genome coverage is ensured and fewer marker replacements required.

Centralised DNA pooling had an important role in reducing between-research groups variation by providing DNA templates of the same composition to both centres involved in the study. The quality of DNA comprising the pool proved to be of great importance. Old and already fragmented DNA can lead to rapid deterioration of the pool composition and make such pools unrepresentative of the pooled sample. The IQ QTL association project was fortunate in that its almost unlimited DNA supplies came from permanent cell lines established for each individual participating in the first stage of the study. High quality DNA derived from the cell lines resulted in DNA pools still highly replicable and truly representative of the sample even five years after their construction. In addition, equal contribution of DNA from all individuals comprising the pool is of utmost importance for studies based on estimated allele frequencies using pooled DNA [195]. In this project, Pico-Green DNA quantification method was used.

Out of 1847 markers for which AIP images were generated in the first stage, only 1 (D4S2460) survived the final stage. The London team had access to sample of 196 parent-offspring trios, for which children were selected for high *g* [213]. The surviving marker was genotyped individually in the trios sample for Transmission Disequilibrium Test (TDT) analysis [111]. The target allele 5 for D4S2460 was less frequent in the high *g* group in both original and replication case/control sets, while the other alleles were expected to be preferentially transmitted in high *g* individuals. Not only that the TDT result was not significant, but it also was in the wrong direction. That is, the target

allele was more frequently transmitted to high *g* offspring than not (55% *vs.* 43% respectively).

One reason for the TDT's failure to replicate the positive finding from the case/control sample may be population stratification. However, all subjects used in the study were classified as Caucasian, which means that population stratification is an unlikely contributor to observed allele frequency differences between average *g* and high *g* individuals [112]. In addition, the genomic control analysis using pooled DNA data showed no evidence of population stratification ($\chi 2=22.914$, $p=0.966$).

It may seem that 1847 markers is a large number, especially when compared to 300-400 markers usually considered as sufficient for a good genome coverage in linkage based studies. Nevertheless, this number is far from the ideal of 35,000 – 250,000 bi-allelic markers thought to be necessary for a genome screen for association studies based on linkage disequilibrium [132, 133, 146], not to mention a daunting 1 million figure if Kruglyak's predictions prove to be correct [117].

The power to detect QTL contributing to high *g* drops steeply with distance from the nearest marker. When *D'* is 1.0, the power to detect a QTL that accounts for 5%, 2.5% and 1% of the total variance is, respectively, 100%, 93% and 56%. When *D'* drops to 0.50, the power is reduced to 73%, 42% and 19% respectively. The average inter-marker distance for markers used in this study is 1.5 Mb, which means that no QTL should be further than 750 kb from the nearest marker. However, the average distance over which an LD between a microsatellite and a trait locus can be detected in outbred population is suggested to be between 300 and 500 kb [135,214,215]. That leaves a relatively large proportion of the genome unaccounted for in this study. It would not come as a great surprise then if true associations with high *g* were missed. Low marker density problem may not be solved simply by testing additional several thousand microsatellite markers, because the power and accuracy of LD mapping with microsatellites becomes small if genetic distance between a microsatellite and a trait locus is very short [156]. A more reasonable following step would be to genotype densely

distributed SNPs, in regions around microsatellite markers that gave positive signals in the study and in candidate genes/gene systems and their regulatory regions. Assuming that a coding SNP is the QTL, focusing efforts on them may further increase chances for detection of loci contributing to *g*.

Another possible reason for so few markers surviving to the final stage could be the over-stringent criteria imposed by the multiple stage study design. This is particularly true for the Stage 1, where the power to detect a QTL of small effect size (accounting for 1% of the total variance), even in the presence of the complete LD, is only 56%.

One way to approach this problem could be to relax requirements for the passage through to the next stage by allowing all markers with $p<0.1$ to qualify for the Stage 2 which should halve Type 2 error. Even though lower stringency at the initial stage would result in doubling the Type I error (0.00025), the study design would still provide adequate protection against false positive results in a scan of 2000 markers. Reduction in the number of false negative results can also be achieved by using a more powerful sample. Larger sample sizes would have provided more power, but size changes were not a feasible option since all DNA collection aimed for the Stage 1 ended well before the molecular genetic study started. Choosing individuals with phenotypes further away from the mean ("super high" *g*) was another possibility, which actually did take place, but only in time for the replication stages. It has to be said, though, that permanent cell lines were not created from "super high *g*" individuals and that the very limited amount of DNA available rendered the sample set unsuitable for the initial screen of ~2000 markers. In addition, a more powerful sample set for replication stages has its advantage since replication of original findings requires more power.

Selection of the most extreme individuals in a phenotypic distribution is known to increase the power of a study [195], but in some instances the reverse may happen if a phenotypically different population subset is used for replication stages. When a phenotype-like general cognitive ability is influenced by a large number of genes and environment, individuals may show extreme phenotypes due to different sets of QTLs

and/or due to environmental factors [216]. This could be one of the reasons why out of 108 microsatellite markers found associated with high *g* in the original sample set, only 4 generated significant p-values in the more powerful replication set.

General cognitive ability is a complex trait with undetermined number of loci contributing to phenotypic variation. There is a real possibility that this number is large and that every individual QTL accounts for less than 1% of the total variance. Sample sizes used in the study do not provide sufficient power to detect such loci. Significantly larger sample sizes (~1000 or more) would increase the power, a benefit which does not necessarily have to come at a greater cost if DNA pooling is employed.

# 5 INVESTIGATING CANDIDATE GENES FOR *g*

High cognitive ability has been associated with various phenomena ranging from decreased glucose intake caused by more efficient synapses [52], to faster nerve conduction [49, 48], to quicker reaction times [49, 50, 51], to better working memory and more effective executive functioning [10]. The correlation between psychometric *g* and these physical variables means that in some way *g* is connected with underlying biological systems. We now face a complex task of discovering the nature of these diverse biological systems and the way each of them shapes our cognitive ability.

## *5.1 Apolipoprotein E*

### 5.1.1 Introduction

There are two processes thought to be essential for cognitive performance: learning and memory [33]. A poor cognitive performance, often associated with a loss of synapses, is readily observable as part of the normal ageing processes [217]. While most healthy elderly people will show some degree of forgetfulness and will perform less well than younger persons, some of them will demonstrate a significant cognitive decline. By the age of 85, about one third of the population will be affected by major impairments in intellectual abilities, and Alzheimer's disease (AD) is the commonest cause [218].

The ε4 variant of the apolipoprotein E (apoE) gene is one of the widely accepted factors which increase the risk and decrease the age of onset of Alzheimer's disease in direct relation to the number of ε4 alleles expressed [219, 218]. In addition, more than

20% of the overall risk of impaired cognitive function in the general population is attributable to the ε4 variant [220]. A three-year longitudinal study of elderly men (mean age=75.1, SD=4.6) conducted in the Netherlands has shown that ApoE4 is also responsible for a more rapid cognitive decline in men with two copies of the allele, while in those with only one copy of the allele, the decline, although slightly slower, is still faster than for non-ApoE4 carriers [220]. Even though some studies failed to show such relationship [221], many others confirmed it and strengthened the hypotheses that poor cognitive performance is associated with the apoE ε4 allele [222, 223, 224].

### 5.1.2 Molecular biology of apolipoprotein E – general features

Apolipoprotein E is a plasma protein involved in metabolism of cholesterol and triglycerides. It is present in very low-density lipoprotein (VLDL) and high-density lipoprotein (HDL), and serves as a ligand for uptake of these molecules in the liver and other organs. Apart from mediating extracellular cholesterol transport, apolipoprotein E is known to regulate multiple additional metabolic pathways, mainly through its interactions with different receptors and through binding other proteins [11]. Much research is now directed at identifying these metabolic pathways in the hope that some of these could explain the development of neurologic and vascular disorders.

The apolipoprotein E gene is positioned on chromosome 19q13.32. Genomic sequence containing four transcribed exons is 3590 bp long (Figure No. 1). The expression of the gene gives rise to a 34-kDa protein made up of 229 amino acids. The protein is present in three common isoforms: apoE2, apoE3 and apoE4 [225]. The isoforms have identical amino acid sequence (Figure No. 27) except for amino acids 112 and 158, where Cys112 and Arg158 (apoE3) are replaced by arginine at 112 (apoE4), or cystein at 158 (apoE2). The amino acid changes determine the binding properties of

the molecule to receptors and other proteins, and are ultimately responsible for isotype-specific roles in pathology of neurologic [219, 226, 227], and common vascular diseases [228, 229].

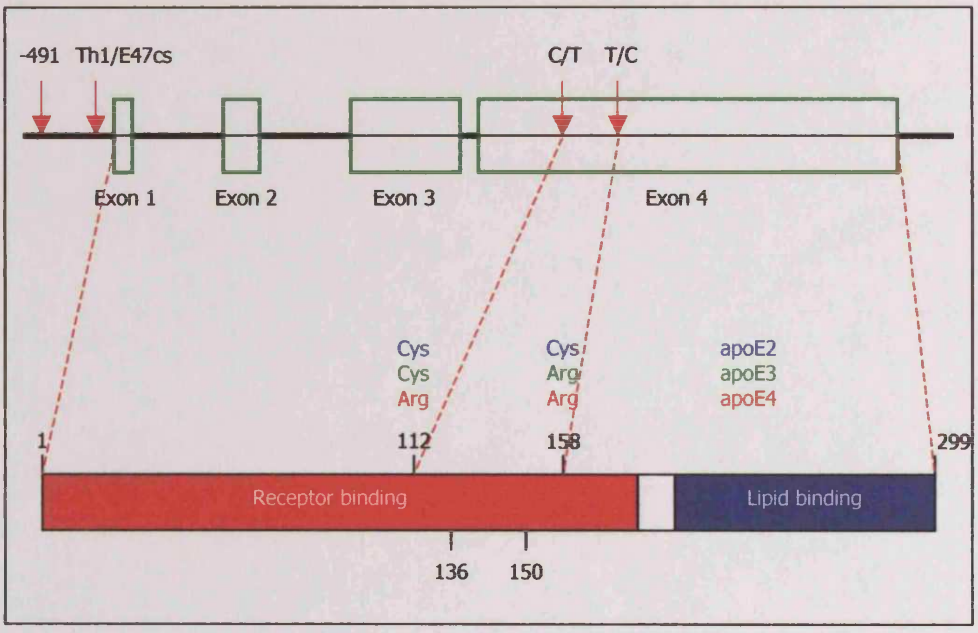



**Figure No. 27:** Molecular structure of the ApoE gene (top); three common isoforms of the apoE protein and its functional domains (bottom); binding of apoE to the LDL family of receptors requires conserved αα136 – αα150 region of the receptor binding domain [11].

Functionally, there are two distinct domains of the protein: a receptor-binding and lipid-binding domain (Figure No. 27). The polymorphic sites determining apoE isoforms are located in the receptor-binding domain and, although not directly within the conserved region (αα136-150), they influence the binding affinity for LDL receptors. Apart from a well known role in mediating cellular uptake of lipoproteins, the LDL receptors have recently been shown to be involved in cellular signalling and neuronal migration during brain development [230, 231]. The variation in affinity with which the

apolipoprotein E isoforms bind to the LDL receptors is, therefore, likely to cause both disturbances in lipoprotein metabolism, leading to atherosclerosis [11] and to affect the efficiency of cellular signalling pathways [230].

### 5.1.3 ApoE in neuronal maintenance and repair

Apolipoprotein E is the predominant type of apolipoprotein in the brain, and is involved in cholesterol metabolism in this organ. It is expressed by astrocytes, microglia and, to a lesser extent, neurons [232] in both normal and AD brain. Ignatius *et al.* (1986) were among the first to demonstrate that apoE levels are up-regulated in response to brain injury [233]. More recent studies further suggest that apolipoprotein E helps protect the brain against injury [234], but the effect seems to be isoform specific. Possession of apoE4 variant increases the risk of an unfavourable outcome after a traumatic brain injury [235, 236], whereas apoE3 shows a neuroprotective effect [237].

The exact mechanism underlying these processes is still unclear, but it could involve the influence of apoE on neuronal repair through its effects on neurite outgrowth and cytoskeletal stability [225]. It seems that apoE might also be necessary for maintenance of neuronal integrity during normal ageing of the CNS [238, 239].

*In vitro* studies have demonstrated that apoE4 and apoE3 have differential effects on neuronal sprouting. Addition of apoE4 to the culture medium inhibits neuronal outgrowth, while apoE3 variant enhances neurite extension [240, 241]. These findings were later confirmed by a number of *in vivo* studies using transgenic mice expressing human apoE4 and apoE3 variants [242, 239, 243]. The adverse effect of apoE4 on neurite outgrowth was attributed to its ability to destabilise the neuronal cytoskeleton, especially the micro-tubular system, which is known to be essential for neuronal development and remodelling [244].

Buttini *et al.* (2000) showed that detrimental apoE4 effects are dose-dependent and that defects in neuronal sprouting are further exacerbated with increased levels of apoE4 [245]. Not only is apoE4 defective in supporting neuronal extension, but it also acts as a dominant negative factor interfering with the beneficial effects of apoE3 [245, 232].

In experiments using human ApoE transgenic mice, Cambon *et al.* (2000) have shown that apoE4 is associated with a significant decline in a synapse per neuron ratio [239]. The effect becomes more apparent later in life when neuronal repair is required. It can be argued that the ε4 isoform is less efficient in reparation of damage incurred by natural ageing processes such as synaptic loss. The loss of synapses and/or replacement by aberrant ones are likely to disrupt hippocampal circuitry, which could affect the storage and retrieval of information in the hippocampus and, quite possibly, result in cognitive impairment [246].

### 5.1.4 Transcriptional activity of the ApoE gene

Even though studies described so far show that the ApoE ε4 allele increases the risk of developing Alzheimer's disease and that it has a deleterious effect on neuronal maintenance and repair, the studies do not explain why a relatively large proportion (~ 50%) of aged homozygotes for ε4 allele [17] do not develop cognitive impairments.

It is possible that polymorphisms within the transcriptional regulatory region of the ApoE gene influence these phenomena. To date, five polymorphisms have been described in the ApoE promoter region (see Figure No. 1 and 2) [16, 17]. Two of those seem to play a role in AD. Lambert and colleagues characterised a G>T mutation at –186 bp, named Th1/E47cs, located in a consensus sequence of the basic helix-loop-helix transcription factor, E47 [17], while Bullido *et al.* described an A>T polymorphism at –491 bp [247].

Both groups found the A allele of the –491 bp site to be an independent risk factor for the disease. *In vitro* studies have shown that the –491A allele is associated with higher levels of ApoE promoter activity than the T allele [247]. The authors hypothesised that the homozygous AA genotype, but not the heterozygous AT, was responsible for the increase in risk for Alzheimer's type dementia. This hypothesis was supported in North American and Caucasian French samples [247, 248].


**Figure No. 28:**
1kb region upstream from the ApoE transcriptional start site; five characterised polymorphisms are presented in red colour; two of the polymorphisms found to be associated with AD, -491A/T and -186G/T (Th1/E47cs) are boxed. [16,17].

The Th1/E47cs mutation has a more complex theory behind it. Epidemiological studies have shown that its adverse effects are most pronounced in ε3/ε4

heterozygotes, who are at the same time heterozygous for the Th1/E47cs polymorphism [17]. Lambert *et al.* pointed that individuals heterozygous for both loci will have one of two haplotypes with the ε4 allele combined with either Th1/E47cs T or G allele. In their sample of 152 ε3/ε4 cases and 91 ε3/ε4 control subjects, AD cases were estimated to have ε4 more frequently combined with the Th1/E47cs T allele than controls (70% *vs.* 58%) resulting in an estimated overall AD risk of 1.7 [17]. This was in accordance with their previous data which suggest that the ε3/ε4 individuals will have increased risk of developing AD because they expressed ε4 allele at a greater level [249]. Thus, the Th1/E47cs mutation in the APOE promoter may lead to preferential expression of the ε4 allele in ε3/ε4 individuals [248], resulting in overall favouring of potentially deleterious action of the E4 isoform.

### 5.1.5 Aims

The findings reviewed above suggest that genetic variation in APOE modifies risk for cognitive decline with age. However, another hypothesis is that variation in APOE might influence cognitive function rather than influencing decline. Therefore, we investigated the influence of apolipoprotein E variations on general cognitive ability, *g*, in a young sample. A possible role of the differential expression of apoE alleles in cognitive performance was examined by observing two polymorphisms in the ApoE promoter region, Th1/E47cs and –491AT, which were repeatedly found to be associated with AD.

## 5.1.6 Materials and Methods

### 5.1.6.1 Sample

See General Materials and Methods

### 5.1.6.2 Genotyping of ApoE, Th1/E47cs and –491 AT polymorphisms

The ApoE genotypes were produced by PCR amplification of 8ng/μl DNA samples. 5'tccaaggactgcaggcggcga3' was used as a sense primer, and 5'acagaattcgccccggcctggtacactgcca3' as an antisense primer in the following conditions: 3min at 94°C, (30s at 94°C, 30s at 66°C, 1min at 72°C) x 35, 10min at 72°C. Restriction fragment length polymorphism (RFLP) analysis was performed by restriction digestion of the PCR products with *Cfo*I (2U per reaction, at 37°C for 16 h) [250].

Th1/E47cs polymorphism was typed by nested PCR followed by double restriction digestion with *Hpa*II and *Eco*M. The primers used for PCR reactions were as follows: forward - 5'caaggtcacacagctggcaac3'; reverse - 5'tccaatcgacggctagctacc3'. First, they were used to generate a 1423 bp fragment spanning -1017 to +406 relative to the transcriptional start site [16]. Then, the product was used as a template to amplify smaller 234 bp fragment containing the polymorphic site. Because the polymorphism itself did not cause a change in restriction sites, the forward primer was designed to create a restriction site (forward: 5'agaatggaggagggtgtcc**g**3'; reverse: 5'actcaaggatcccagacttg3'; the underlined letters show positions of mismatched bases).

The -491 AT polymorphism was genotyped using the same nested PCR protocol as for the Th1/E47cs. With the larger fragment serving as a template, the smaller fragment with a forced restriction site was amplified using the primers: forward, 5'tgttggccaggctggt**t**t**t**aa3'; reverse 5'cctcctttcctgaccctgtcc3' (the underlined letters show

positions of mismatched bases). RFLP was analysed by double restriction digestion with *DraI* and *AluI* restriction enzymes.

#### 5.1.6.3 Statistical analysis

Genotype frequencies for ApoE were analysed by 2x5 $\chi^2$ using the CLUMP program [251]. Genotype and allele frequencies for Th1/E47cs and -491 polymorphisms were analysed by 2x3 $\chi^2$ using MINITAB 10.5 (1995). The haplotype analysis was performed using the EH+ program [252].

### 5.1.7 Results

Restriction isotyping of apolipoprotein E was performed by polymerase chain reaction (PCR) amplification of a discrete genomic region followed by *CfoI* restriction digestion. The size of the uncut amplified product was 227 bp. Polymorphic sites determining apoE isoforms were easily recognised by either the presence or the absence of cutting sites for *CfoI* restriction enzyme (see Figure No. 29 for graphical presentation of ApoE *CfoI* restriction map) [250].

Genotypes were obtained for 97 high *g* individuals++ and 98 controls (see Picture No. 1). The genotype and allele frequencies are shown in Table No. 26. No association of any of the apoE isoforms with *g* was found ($\chi^2$=2.877, p=0.237), although the differences in allelic frequencies were in the expected direction for the high *g vs.* average *g* groups (82% *vs.* 77%, respectively for ε3, and 13% *vs.* 15% for ε4). Genotypes also showed no significant difference ($\chi^2$=3.175, p=0.211), although the

genotypic frequencies were also in the expected direction for the high and average groups (65% *vs.* 58% for ε3ε3 and 25% *vs.* 22% for ε3ε4).

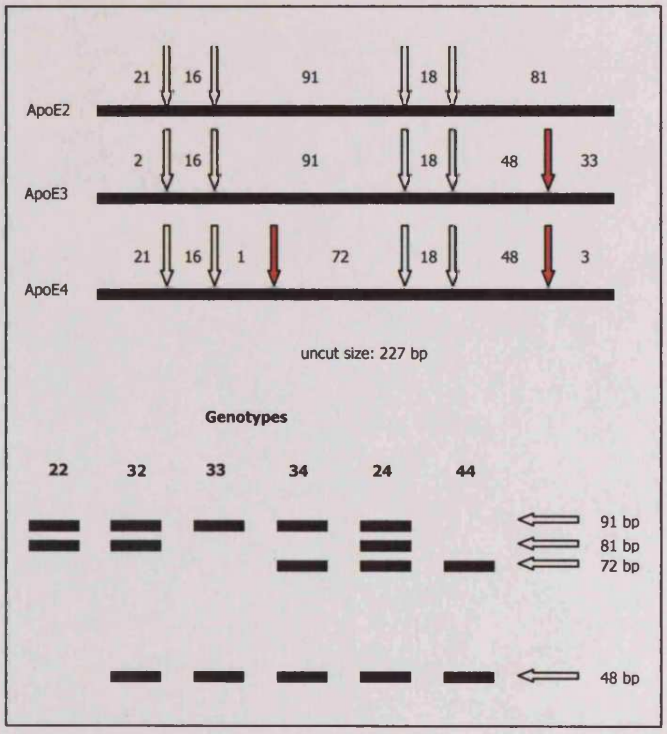



**Figure No. 29:**

ApoE *Cfo*I restriction map; red arrows denote polymorphic restriction sites; the lower half of the figure shows restriction digests for different allele combinations and the way they would appear on agarose gel.

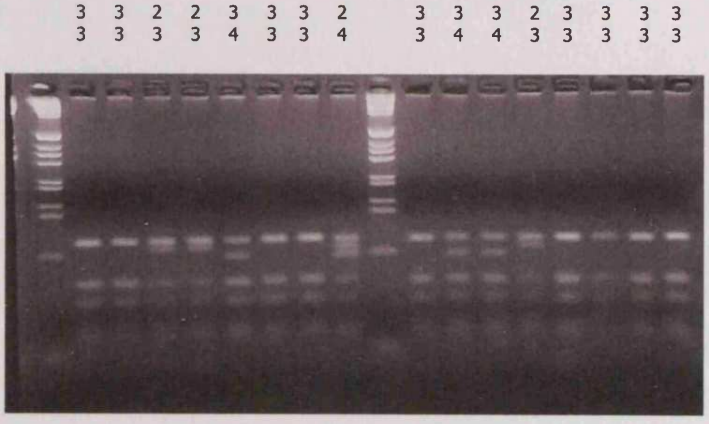



**Figure No. 30:**

ApoE *Cfo*I restriction digestion products size separated on an agarose gel.

| ApoE | Genotype n (*f*%)[a] | | | | | | Allele n (*f*%)[b] | | |
|---|---|---|---|---|---|---|---|---|---|
| | **22** | **23** | **24** | **33** | **34** | **44** | **2** | **3** | **4** |
| High *g* | 0 *(0.00)* | 9 *(9.27)* | 0 *(0.00)* | 63 *(64.95)* | 24 *(24.74)* | 1 *(1.03)* | 9 *(4.64)* | 159 *(81.96)* | 26 *(13.40)* |
| Average *g* | 0 *(0.00)* | 14 *(14.28)* | 3 *(3.09)* | 57 *(58.16)* | 22 *(22.45)* | 2 *(2.04)* | 17 *(8.67)* | 150 *(76.53)* | 29 *(14.79)* |

**Table No. 26:** ApoE genotype and allele counts (n) and frequencies (*f*%)

a $\chi^2$=3.17; p=0.211

b $\chi^2$=2.88; p=0.237

To examine the possibility of the association between the apoE expression levels and cognitive ability, two polymorphisms in the transcriptional region of the ApoE gene (Th1/E47cs G/T and –491 A/T) were investigated.

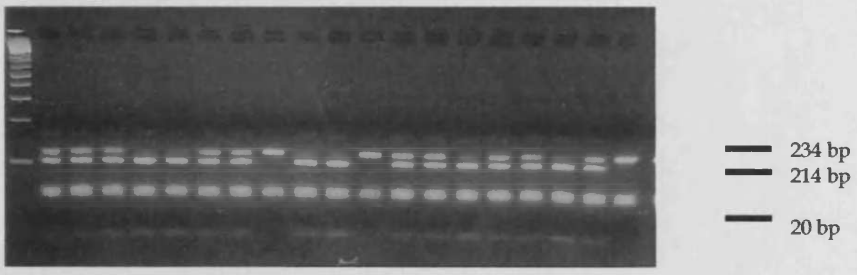



**Figure No. 31:** Th1/E47cs G/T polymorphism genotyping. Presented are *Hpa*II and *EcoN*I restriction digestion products that are size separated on an agarose gel.

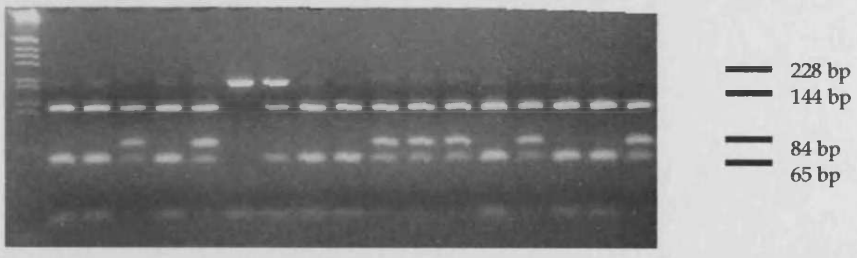
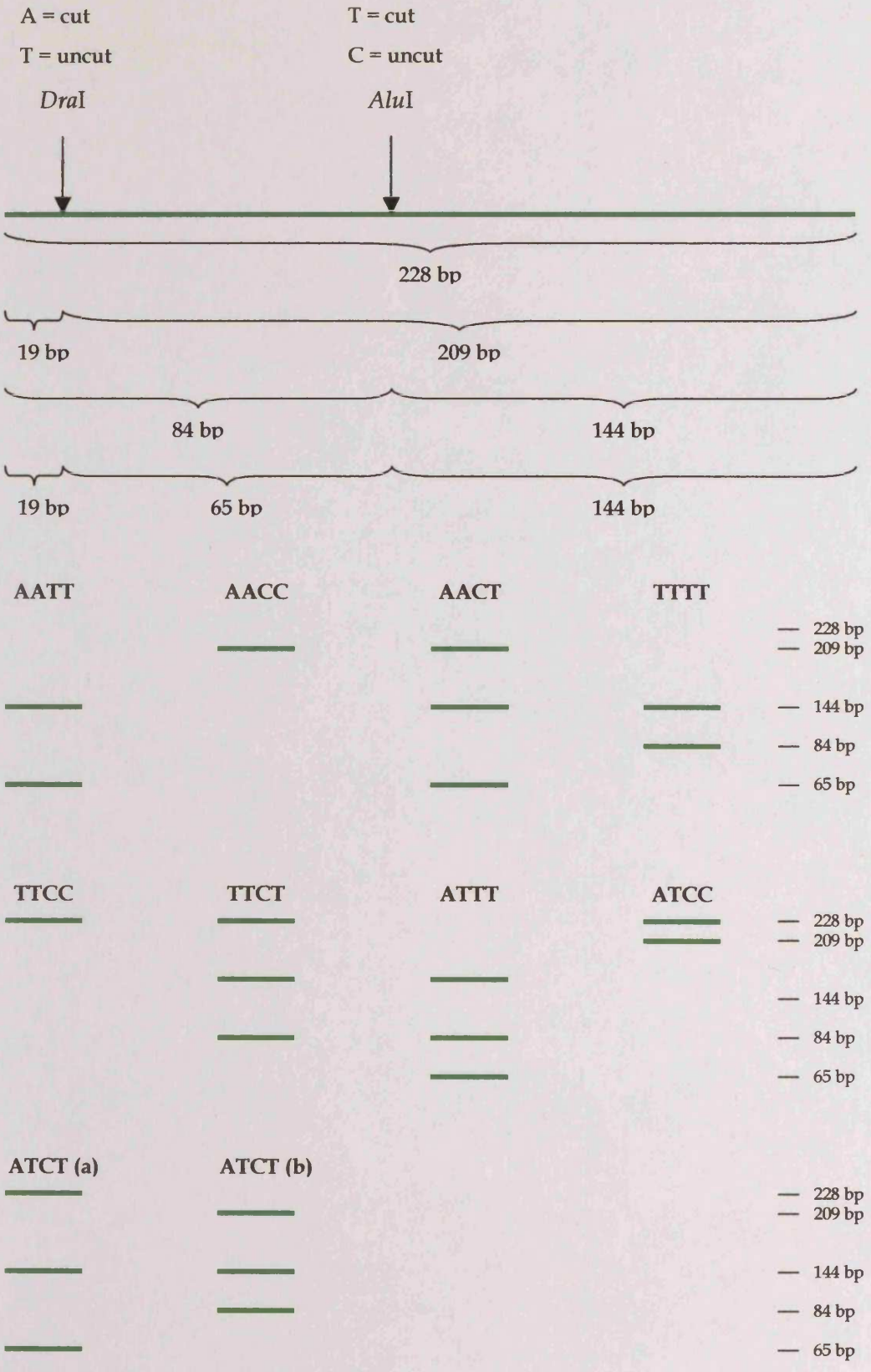



**Figure No. 32:** -491 A/T polymorphism genotyping. Presented are *Dra*I and *Alu*I restriction digestion products that are size separated on an agarose gel. An additional rare polymorphic site within the *Alu*I recognition sequence was noticed. As a result, a variety of restriction products could be visualised on a gel. All possible combinations are described in a diagram below (Figure No. 33).

**Figure No. 33:** Diagram of the -491 A/T polymorphism RFLP analysis. -491 A/T genotype was determined by observing presence/absence of a DraI restriction enzyme cut site. Different length restriction products are presented schematically.



The allelic and genotypic frequencies of the two polymorphisms investigated (Table No. 27 and 28) were not different in the high *g* group when compared with the control group (Th1/E47cs genotypic frequency, $\chi^2$=1.757, p=0.416; Th1/E47cs allelic frequency, $\chi^2$=0.389, p=0.533; -491 AT genotypic frequency, $\chi^2$=0.616, p=0.735; ; -491 AT allelic frequency, $\chi^2$=0.081, p=0.775).

| -491 A/T | Genotype n (*f*%)[a] | | | Allele n (*f*%)[b] | |
|---|---|---|---|---|---|
| | **AA** | **AT** | **TT** | **A** | **T** |
| High *g* | 61 *(64.89)* | 27 *(28.72)* | 6 *(6.38)* | 149 *(79.26)* | 39 *(20.74)* |
| Average *g* | 60 *(61.22)* | 33 *(33.67)* | 5 *(5.10)* | 153 *(78.06)* | 43 *(21.94)* |

**Table No. 27:** -491 polymorphism genotype and allele counts (n) and frequencies (*f*%)

a $\chi^2$=0.62; p=0.735; df=2
b $\chi^2$=0.08; p=0.775; df=1

| Th1/E47cs | Genotype n (*f*%)[a] | | | Allele n (*f*%)[b] | |
|---|---|---|---|---|---|
| | **GG** | **GT** | **TT** | **G** | **T** |
| High *g* | 24 *(25.53)* | 54 *(57.45)* | 16 *(17.02)* | 102 *(54.26)* | 86 *(45.74)* |
| Average *g* | 25 *(26.32)* | 47 *(49.47)* | 23 *(24.21)* | 97 *(51.05)* | 93 *(48.95)* |

**Table No. 28:** Th1/E47cs polymorphism genotype and allele counts (n) and frequencies (*f*%)

a $\chi^2$=1.76; p=0.415; df=2
b $\chi^2$=0.39; p=0.533; df=1

The possibility that *g* might be influenced by differential expression of ApoE alleles was examined by analysis of the effect of promoter region variants in combination with ApoE isoforms. Differences in haplotype frequencies (Table No. 29 and 30) were not statistically significant for either of the promoter variants (Th1/E47cs – ApoE, $\chi^2$=3.38, p=0.495; -491 – ApoE, $\chi^2$=5.18, p=0.300; -491-Th1/E47cs-ApoE, $\chi^2$=8.16, p=0.363).

| **ApoE genotype** | **ε2ε3** | **ε3ε3** | **ε2ε4** | **ε3ε4** | **ε4ε4** | **n (*f*%)** |
|---|---|---|---|---|---|---|
| **High *g*** | | | | | | |
| **AA** | 3 | 37 | 0 | 19 | 2 | 61 *(64.89)* |
| **AT** | 2 | 21 | 0 | 4 | 0 | 27 *(28.72)* |
| **TT** | 3 | 3 | 0 | 0 | 0 | 6 *(6.38)* |
| **n (*f*%)** | 8 *(8.51)* | 61 *(64.89)* | 0 *(0.00)* | 23 *(24.47)* | 2 *(2.13)* | 94 |
| **Average *g*** | | | | | | |
| **AA** | 8 | 34 | 2 | 14 | 2 | 60 *(61.22)* |
| **AT** | 5 | 20 | 1 | 7 | 0 | 33 *(33.67)* |
| **TT** | 1 | 4 | 0 | 0 | 0 | 5 *(5.10)* |
| **n (*f*%)** | 14 *(14.29)* | 58 *(59.18)* | 3 *(3.06)* | 21 *(21.43)* | 2 *(2.04)* | 98 |

**Table No. 29:** ApoE and –491 A/T polymorphism combination

$\chi^2$=5.18; p=0.300; df=7

| **ApoE genotype** | **ε2ε3** | **ε3ε3** | **ε2ε4** | **ε3ε4** | **ε4ε4** | **n (*f*%)** |
|---|---|---|---|---|---|---|
| **High *g*** | | | | | | |
| **GG** | 6 | 16 | 0 | 2 | 0 | 24 *(25.53)* |
| **GT** | 3 | 34 | 0 | 17 | 0 | 54 *(57.45)* |
| **TT** | 0 | 11 | 0 | 4 | 1 | 16 *(17.02)* |
| **n (*f*%)** | 9 *(9.57)* | 61 *(64.89)* | 0 *(0.00)* | 23 *(24.47)* | 1 *(1.06)* | 94 |
| **Average *g*** | | | | | | |
| **GG** | 7 | 16 | 0 | 2 | 0 | 25 (26.32) |
| **GT** | 6 | 27 | 3 | 11 | 0 | 47 (49.47) |
| **TT** | 0 | 12 | 0 | 9 | 2 | 23 (24.21) |
| **n (*f*%)** | 13 *(13.68)* | 55 *(57.89)* | 3 *(3.16)* | 22 *(23.16)* | 2 *(2.11)* | 95 |

**Table No. 30:** ApoE and Th1/E47cs G/T polymorphism combination

$\chi^2$=3.38; p=0.495; df=7

### 5.1.8 Discussion

Alzheimer's disease is a complex disease likely to be caused by a number of genes working together with environmental factors [218]. The ε4 allele of the apolipoprotein E gene is accepted to be one of the risk factors for the disease. Several recent studies have offered evidence that not only the presence of ApoE ε4 allele, but its differential expression are strongly associated with cognitive decline [253], [248]. It can be argued that the ε4 isoform is less efficient in processes requiring reparation of damage incurred by brain injury and natural ageing processes such as synaptic loss. In these instances fewer new dendrites are formed and fewer lost synapses replaced, quite often with aberrant ones [246]. Therefore, it will not be surprising if these changes disrupted the finely tuned and complex signal transduction network, leading to less efficient cognitive functioning. This lead us to hypothesise that apolipoprotein E variations may also influence efficiency of cognitive functioning and general cognitive ability, *g*, in a younger sample. We investigated whether the different apoE isoforms and their differential expression plays a role in cognitive performance by observing polymorphisms in the ApoE coding sequence and its promoter region (–491AT and Th1/E47cs).

No association between the ApoE isoforms ($\chi^2$=2.88; p=0.237), its promoter polymorphisms (–491AT: $\chi^2$=0.08; p=0.775 and Th1/E47cs: $\chi^2$=0.39; p=0.533) and general cognitive ability has been observed in the studied sample. The ε4 allele of the ApoE was not significantly less frequent in the high *g* group than in the control group and likewise the ε3 allele was not significantly more frequent in the high *g* group. The selected extremes design used in the present analysis has 100% power to detect QTLs that account for 5% of the variance and 93% power to detect QTLs that account for 2.5% of the variance (p = 0.05). Thus, the possibility that ApoE accounts for more than 2.5% of the variance could be excluded. However, it is possible that ApoE has a much weaker effect, as suggested by the allelic frequency differences for ε3 between

the two groups, although much larger samples would be needed to detect such a small effect.

Although there is evidence that the two ApoE promoter variants confer risk for developing AD in association with and independently of ε4 allele [17,253], the results that obtained in this study do not show that either the –491 A allele or the Th1/E47cs T allele is associated with *g* in the young healthy population.

In conclusion, apolipoprotein E is a major apolipoprotein constituent of the CNS where it plays an important role in injured and ageing brain by mediating neuronal remodelling. Its three isoforms have been shown not to be equally efficient in such processes. ε4, the least efficient of all is a well-established risk factor for AD which is the commonest cause of cognitive decline in elderly. However, this study suggests that apoE is not a factor with a measurable impact on general cognitive ability in the young healthy population.

## 5.2 *Investigating Calcium/calmodulin dependent protein kinase II alpha (CaMKII-α) as a candidate gene for g*

### 5.2.1 Introduction

#### 5.2.1.1 Involvement of the hippocampus in learning and memory

Learning is manifested as a change in behaviour or in a probability of a particular response in the presence of a particular stimulus as the result of prior experience [10]. Memory is the retention of the altered behaviour over a period of time.

Definitions of intelligence often include the ability to learn and to remember. The kind of learning and memory that is referred to as "spatial" is the ability to navigate effectively through an environment, integrating self-motion with prominent features of the external space [254]. This ability has repeatedly been found to be one of the most heritable components of general cognitive ability, *g* (See Table No. 31) [7, 15, 19].

| Study | Heritability estimate | |
|---|---|---|
| | Spatial ability | Verbal ability |
| McGue & Bouchard (1989) | 71% | 57% |
| Pedersen (1992) | 46% | 58% |
| Ando (2001) | 65% | 65% |

**Table No. 31:** Heritability estimates for specific cognitive abilities from three independent twin studies [7, 15, 19].

The hippocampal formation is closely related to learning spatial relations. In fact, spatial memory is extremely hippocampus-dependent in rats [255, 256]. The hippocampus

has been implicated in the normal learning and memory in humans since the 1950s. Surgical removal of both medial temporal lobes in patient H. M. [257] resulted in profound memory deficits, inability to acquire new information and to navigate in new environments [258].

Impairments in spatial ability have also been recognised in patients with damage restricted to the right temporal lobe [259] as well as in those with ischemic lesions limited to field CA1 of the hippocampus (See Figure No. 34) [260].

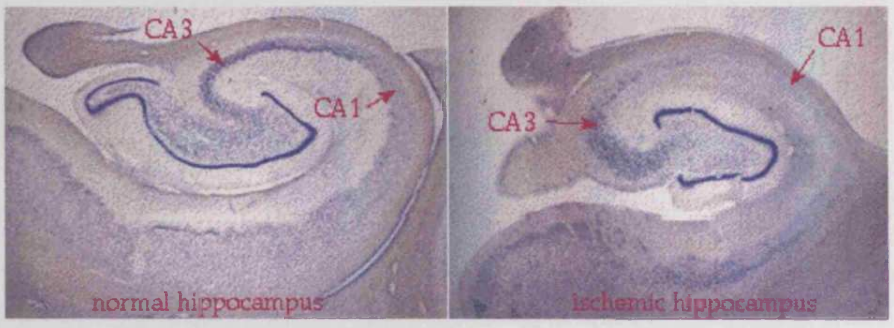



**Figure No. 34:**
A photomicrograph of a frontal section of the human hippocampal formation. The image on the left represents the normal hippocampus. The image on the right shows the ischemic hippocampus. Degeneration of CA1 and the absence of cell bodies (stained purple) is readily observable in the affected hippocampus [12].

Progressive learning and memory ability impairments and general cognitive decline are the main features of Alzheimer's disease, which is characterised by pronounced histopathological changes (extra-cellular plaques, intra-cellular neurofibrillary tangles and neuronal loss) in the CA1 region of the hippocampus (See Figure No. 35) [261, 262, 263].

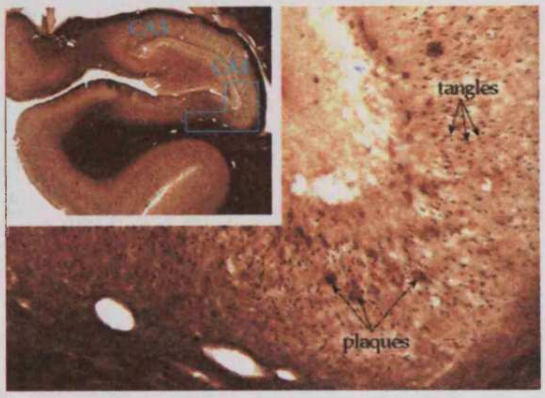



**Figure No. 35:**

A photomicrograph of a frontal section of the hippocampal formation in the brain affected by Alzheimer's disease. Pronounced histopathological changes in the CA1 region with visible neurofibrillary tangles and plaques (stained dark brown) [12].

Taken together, these findings strongly indicate involvement of the hippocampus and its CA1 region in normal processes of human learning and memory.

### 5.2.2 Hippocampal anatomy and circuitry

The hippocampal formation is the folded archaecortex, which consists of two main features: the *dentate gyrus* and the *cornu ammonis* (CA), plus the subiculum (See Figure No. 36). The dentate gyrus and CA have the cortex organised into three layers. The subiculum is a transitional formation linking the hippocampus proper with the six layered entorhinal region of the neocortex [264].

The *perforant pathway* is a major afferent pathway bringing signals from the entorhinal neocortex to the hippocampus. Here it forms synapses with *granule cells* of the dentate gyrus and pyramidal cells of the CA3 region of the CA. CA3 cells also receive input form granule cells via *mossy fibers* [264].

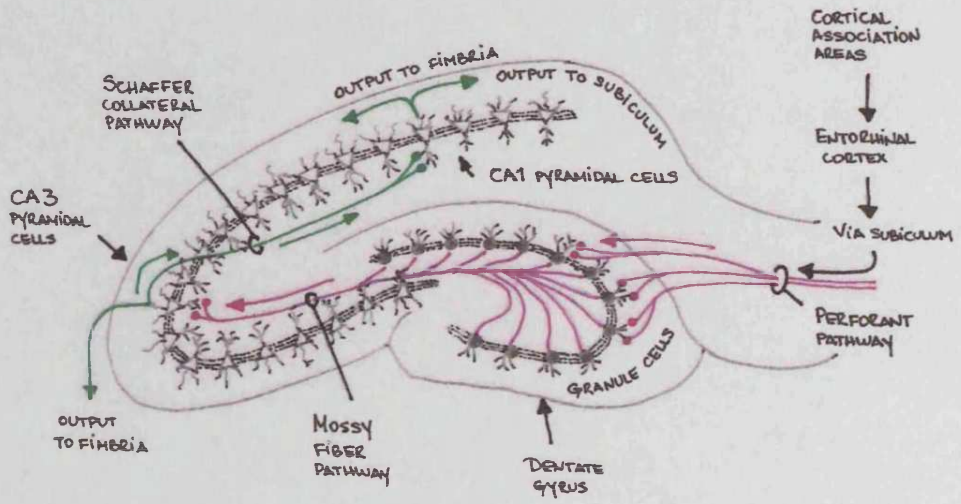



**Figure No. 36:** Diagram of the synaptic organization of the rat hippocampus (figure taken from http://psych.fullerton.edu/swillis/amnesia.html).

Axons of CA3 pyramidal cells form several efferent branches. Those that pass to the opposite side of the brain and make contact with the contralateral CA3 region are called *commissural fibers*. A proportion of the branches leaving the hippocampus will terminate in the hypothalamus and thalamus. Finally, some of the CA3 axons make synapses with neighbouring CA3 cells (*recurrent collaterals*) or with pyramidal cells in the CA1 region (*Schaffer collaterals*). CA1 cell axons make synapses in the subiculum and entorhinal cortex [264].

The three sets of excitatory synapses that are formed by the fibers of the perforant pathway terminating on granule cells and by the mossy fibers and Schaeffer collaterals making connections with the pyramidal cells are known as the *trisynaptic loop*, a formation that has been often referred to in studies on learning and memory mechanisms.

### 5.2.3 Neural basis of learning and memory

During development neuronal pathways are formed. Some of these connections remain unaltered and it is said of such pathways that they are "hard-wired". Pathways that are continuously rewired as a result of learning are referred to as "plastic" and the process itself as "plasticity" [264].

Plasticity is observed across the animal kingdom, including animals with very rudimentary nervous systems. However, the less complex nervous system is less capable of making plastic changes in response to the environment because it has fewer neurons and fewer connections that are available for such modifications. In animals with more complex nervous systems, including humans, billions of neurons interconnect to form large networks *via* vast number of synapses, which continuously undergo changes in their strengths in response to various stimuli.

The idea of changes in synaptic strength as the basis of learning and memory is the core idea of the modern age neuroscience [265]. The first experimental demonstration of such changes took place in the early 1970s, when it was shown that repetitive activation of excitatory synapses in the hippocampus resulted in a long lasting increase in synaptic strength [266]. The observed phenomenon is termed "long-term potentiation" (LTP) and is usually defined as a persistent (hours) enhancement of an excitatory postsynaptic potential (EPSP) following brief high-frequency stimulation of afferent pathways [266]. A counterpart of LTP, long-term depression (LTD), is another event thought to be a component of a neural basis of learning and memory [267]. LTD can be induced by low frequency stimulation protocols, but its role in learning and memory is less well understood since only a small proportion of studies has focused on hippocampal LTD [268, 269].

All three types of excitatory synapses in the trisynaptic loop use glutamate as their neurotransmitter which can bind to metabotropic and ionotropic receptors. Ionotropic receptors can be further subdivided into two major types termed *α-amino-3-hydroxy-5-*

*methyl-4-isoxazolepropionate* (AMPA) and *N-methyl-D-aspartate* (NMDA) receptors. Depending on whether the activation of NMDA receptors is necessary for the inclusion of prolonged changes in synaptic strength, two distinct forms of hippocampal LTP can be recognised. While one of these, observed in the CA1 region, requires activation of NMDA receptors [270], the other, occurring at mossy fiber synapses in the CA3 region, is independent of NMDAR status and post-synaptic activity [271, 272]. Because the CA1 region of the hippocampus is regarded as a crucial neuro-anatomical formation for learning and memory, its NDMAR-dependent form of LTP has been more extensively studied in relation to these two behavioural processes.

The main feature of the CA1 type of LTP induction is its dependence on $Ca^{2+}$ influx through activated postsynaptic NMDA receptors. The presynaptically released neurotransmitter glutamate binds to both types of postsynaptic ionotropic receptors, AMPAR and NMDAR. When the neuronal cell is close to its resting state, AMPA receptors are mainly responsible for influx of $Na^{+}$ and $K^{+}$ necessary for maintenance of the resting membrane potential. NMDARs are largely non-responsive at low frequency synaptic transmission, owing to voltage-dependent block by $Mg^{2+}$. Only strong afferent stimulation can produce sufficient postsynaptic depolarisation to remove the $Mg^{2+}$ block and allow $Ca^{2+}$ and $Na^{+}$ influx.

As described, for successful CA1 LTP induction two events have to occur simultaneously, namely presynaptic release of sufficient amounts of glutamate and sufficient depolarisation of the postsynaptic membrane to remove the $Mg^{2+}$ block and open NMDAR channels. The mechanism involved in LTP induction was proposed more than 50 years ago by Donald Hebb, who postulated that a synaptic connection would be strengthened if the presynaptic and postsynaptic neurons were active simultaneously [273]. Therefore, the coincidence detection property exhibited by NMDA receptors makes them crucial for memory formation under the basic principle of Hebb's rule.

### 5.2.4 Role of α-CaMKII in learning and memory

The sequence of events following LTP induction *via* NMDA receptor-gated $Ca^{2+}$ influx is called LTP expression [264]. One mechanism by which $Ca^{2+}$ signals are converted into physiological changes in the neuron is through the activation of protein kinases. A family of serine/threonine protein kinases that is directly activated by $Ca^{2+}$ is the $Ca^{2+}$/calmodulin-dependent protein kinase II (CaMKII) family.

#### 5.2.4.1 Biochemical properties of CaMKII

CaMKII comprises a family of 28 similar isoforms derived from four genes (α, β, γ and δ) [274]. CaMKII is a multi-unit holoenzyme (460 – 654 kDa) with a ring-like structure composed of 8 to 12 subunits [275]. Most of the CaMKII activity in the forebrain comes from the hetero-multimers with an α:β isoform ratio of 3:1 [275]. In addition, homomers composed of the α isoform have also been identified in the rat forebrain [276].

A typical CaMKII subunit possesses three domains: catalytic, autoregulatory and association (See figure No. 37). Of these, the catalytic domain is a substrate binding domain, but it is also capable of binding the autoinhibitory domain, which has amino-acid residues that mimic a protein substrate [277]. The association domain is necessary for formation and assembly of a multimeric CaMKII holoenzyme.

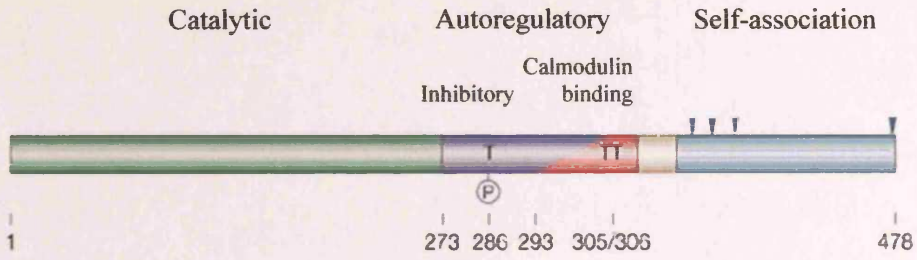



**Figure No. 37:** A typical CaMKII subunit possesses three domains: catalytic, autoregulatory and association domain. Denoted are phosphorylation sites (P) and amino-acid positions (1 - 478).

In the absence of $Ca^{2+}$/calmodulin ($Ca^{2+}$/CaM), the kinase is inactive owing to the autoinhibitory domain, which blocks the catalytic domain. $Ca^{2+}$/CaM binding alters the conformation of the autoinhibitory domain and permits substrate access to the catalytic site. This activates a particular subunit, but will not lead to autophosphorylation of the enzyme unless another $Ca^{2+}$/CaM is bound to a neighbouring subunit. This conformational change, triggered by $Ca^{2+}$/CaM binding, exposes a site for phosphorylation at $Thr^{286}$ in the adjacent subunit. Autophosphorylation of this residue further disrupts the autoinhibitory interaction with the catalytic domain and continues to produce such disinhibition in the absence of its activator $Ca^{2+}$/CaM.

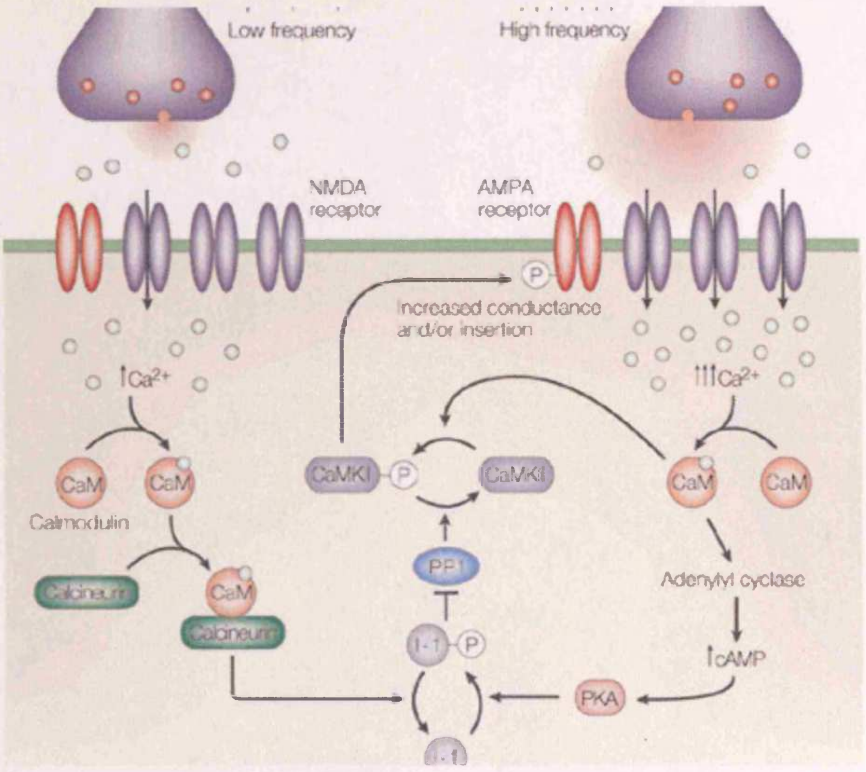



**Figure No. 38:** The Lisman model of kinase-phosphatase signalling in LTP [5]. Figure taken from [13].

Autophosphorylation of Thr[286] increases the affinity of CaMKII for CaM over 1000-fold, called "*CaM trapping*", by exposing residues Phe[293] and Asn[294] , which are critical for a switch from low to high-affinity CaM binding [278]. This is followed by another wave of autophosphorylation, or "burst phosphorylation", of a number of residues within the $Ca^{2+}$/CaM binding domain, which ultimately results in CaM dissociation and $Ca^{2+}$/CaM independent enzyme activity. The inability of the phosphorylated and activated kinase to bind $Ca^{2+}$/CaM is called "*CaM capping*" [279]. It has been suggested that the potential function of the burst phosphorylation might be to mark autonomously active subunits of CaMKII, so that these no longer compete for CaM binding and activation [277].

The activity of phosphorylated CaMKII is also regulated through the action of protein phosphatase 1 (PP1), which is very effective in dephosphorylating $Thr^{286}$ residue [280]. The activity of PP1 is suppressed during LTP in order to prevent rapid dephosphorylation and inactivation of CaMKII, whereby the suppression results from interaction of PP1 with inhibitor proteins, one of which is known as inhibitor 1 (I1) [281]. When I1 is phosphorylated by protein kinase A (PKA), it inhibits PP1 and thus prolongs autonomous CaMKII activity [282].

#### 5.2.4.2 CaMKII is sensitive to frequency of $Ca^{2+}$ oscillations

In the brain, synaptic activity is marked by oscillations in intracellular $Ca^{2+}$. Information on the stimulus is encoded by size, frequency and duration of these oscillations [283]. CaMKII is believed to be a central molecular mechanism involved in the decoding of such information.

Scientific evidence for such a proposal came from an *in vitro* study which showed that generation of autonomous CaMKII activity is sensitive to the frequency and size of $Ca^{2+}$ spikes [283]. At lower $Ca^{2+}$-spiking frequencies, kinase autonomous activity was low even though the autophosphorylation was evident. Stimulation with high frequency pulses (4.0 Hz) and/or with those of longer duration (500 to 1000 ms) resulted in a marked increase in enzyme autonomy. The observed CaMKII frequency-dependent autonomy occurs possibly because spike intervals are shorter at higher frequencies than the time required for complete dissociation of CaM [283], which increases the probability of CaM molecules binding to two neighbouring subunits and subsequently reaches the "CaM trapped" state and full autonomous kinase activity.

The initial state of CaMKII activity also affects the way the enzyme responds to subsequent $Ca^{2+}$ fluctuations. If several subunits are already autophosphorylated, the

enzyme sensitivity will shift to lower stimulation frequencies. In this way, CaMKII sets the stimulus frequency threshold at which the enzyme responds and determines the frequency threshold required to induce LTP or LTD [284]. It was postulated that this enzyme could function as a molecular switch of learning and memory, whereby past $Ca^{2+}$ influx is "memorised" in a form of various levels of autonomous enzyme activity and by regulated responsiveness to subsequent stimuli [284].

### 5.2.4.3 CaMKII is crucial for hippocampal LTP induction

The ability of the kinase to convert a transient $Ca^{2+}$ signal into a long-lasting increase in enzyme activity has lead to the suggestion that the kinase is directly involved in induction and expression of long-term potentiation [285]. Evidence in support of the suggestion is overwhelming.

Research groups using various synthetic peptide analogues corresponding to the autoregulatory domain of CaMKII showed that extracellular application of such inhibitors can block LTP [286, 287]. However, the activity of these inhibitors was not specific [288] and, although a role of CaMKII in LTP and learning and memory could not be ruled out, it needed further scientific support.

In 1992, Silva and colleagues generated genetically modified mice that do not express the α-CaMKII [289], the isoform which is highly enriched in postsynaptic formations of hippocampus. These mice had no obvious neuroanatomical deficits, and despite intact postsynaptic NMDA receptor function, they had almost complete loss of normal LTP and displayed a severe deficit in spatial memory tasks [289].

By using regulated and targeted expression of the α-CaMKII gene, Mayford and colleagues (1996) addressed the question of whether the changes observed in genetically modified animals were due to an indirect effect on neuronal circuit

development. Using a forebrain specific promoter, they were able not only to time-control the gene expression, but also to restrict its expression to specific regions of the brain, particularly the hippocampus. Results of the study led to conclusion that the CaMKII signalling pathway is critical for memory processes in a manner that is independent of its potential role in brain development [290].

The importance of autophosphorylation of $Thr^{286}$ of the αCaMKII in LTP and learning was demonstrated by introducing a point mutation into the αCaMKII gene, which blocked the autophosphorylation of the $Thr^{286}$ residue without affecting its CaM-dependent activity. The mutant mice had no NMDAR-dependent LTP in the hippocampal CA1 area and showed massive reduction of spatial learning ability [291].

#### 5.2.4.4 CaMKII and LTP expression

The postsynaptic density (PSD) is a tiny, amorphous structure located beneath the postsynaptic membrane of synapses in the CNS [292]. It is usually seen as a thickening in the membrane caused by the accumulation of proteins, receptors and enzymes involved in the postsynaptic response to transmitter [264] and is believed to be the site of LTP expression [265].

It has been known for some time that CaMKII is localised in post-synaptic densities, where together with NMDARs it makes up a large proportion of the total PSD protein [293]. Leonard *et al.* (1999) have shown that CaMKII recruitment to PSD is increased after the NMDAR-gated $Ca^{2+}$ influx [294]. Here the kinase becomes associated with NR1 [294] and NR2A/B subunits of NMDA receptors [295], which places CaMKII at an ideal position for activation by $Ca^{2+}$ influx through NMDARs.

The anchoring of the CaMKII *via* NMDARs brings the kinase in close proximity to AMPA receptors, whose phosphorylation plays a central role in NMDAR-dependent LTP

expression [296]. The induction of LTP specifically increases the phosphorylation of the GluR1 subunit of the AMPA receptor [296], the subunit demonstrated to be crucial for generation of LTP in hippocampal CA1 synapses [297]. This phosphorylation event mediated by CaMKII enhances AMPA receptor-mediated currents due to an increase in single-channel conductance which causes prolonged changes in synaptic strength in response to synaptic activity [297].

In addition to modifying the conductance of existing AMPA receptors, CaMKII induces the insertion of new GluR1 containing AMPARs into the synapse, which causes further increase in AMPAR-mediated transmission [298]. The insertion of AMPA receptors leads to increase in PSD size and to eventual formation of multi-spine synapses and synapse multiplication [299]. Such structural changes in synapses triggered by CaMKII activity may well be a mechanism by which experiences are translated into memories.

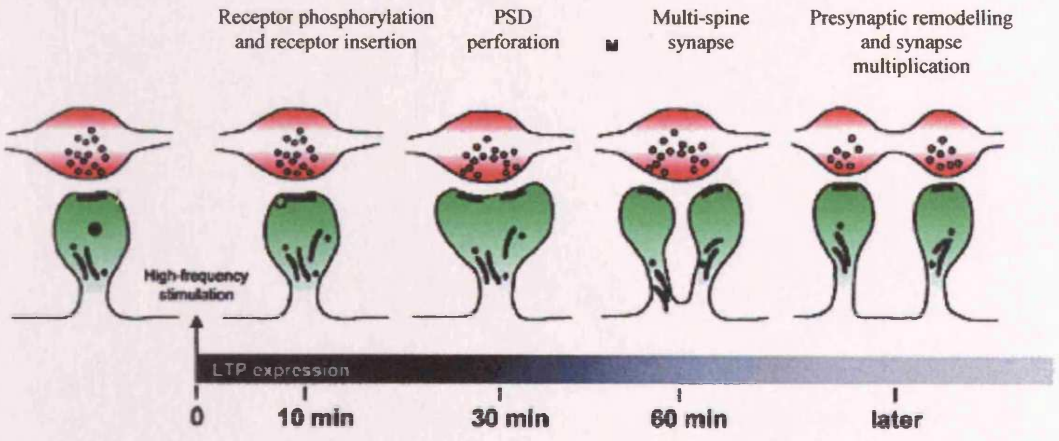



**Figure No. 39:** LTP expression mechanism (picture adopted from [299]): Activation of CaMKII results in phosphorylation of AMPA receptors and an increase in their single-channel conductance. In addition, AMPA receptors are inserted into the postsynaptic membrane, which leads to an increase in the size of the PSD. Through an unknown mechanism, some synapses use the expanded membrane area to generate multi-spine synapses. Retrograde communication triggers presynaptic structural changes, eventually increasing the total number of synapses [299].

### 5.2.5 Organisation and expression of the α-CaMKII gene

The gene coding for α-CaMKII maps to human chromosome 5q33.1 [300]. The gene is organised into 18 exons and 17 introns with an overall genomic size of 66641 bp. The gene is transcribed into a 4836 bp mRNA, which is alternatively spliced to produce two forms of the α-CaMKII. The longer transcript, variant A, is expressed in the brain where it is involved in LTP induction and expression, while the shorter transcript, variant B, lacks the characteristic kinase activity and is preferentially expressed in skeletal muscle [301].

The α-CaMKII mRNA is one of a limited number of mRNAs, including glutamate receptor family transcripts, shown to be present within the dendrites of neurons [302]. Protein synthetic machinery has also been identified in dendrites, and it was clearly demonstrated that dendritic translation occurs independently of neuronal cell bodies [303], [304]. Local protein translation in dendrites could be a means for delivering synaptic proteins to their sites of action in a fashion that could contribute to plasticity.

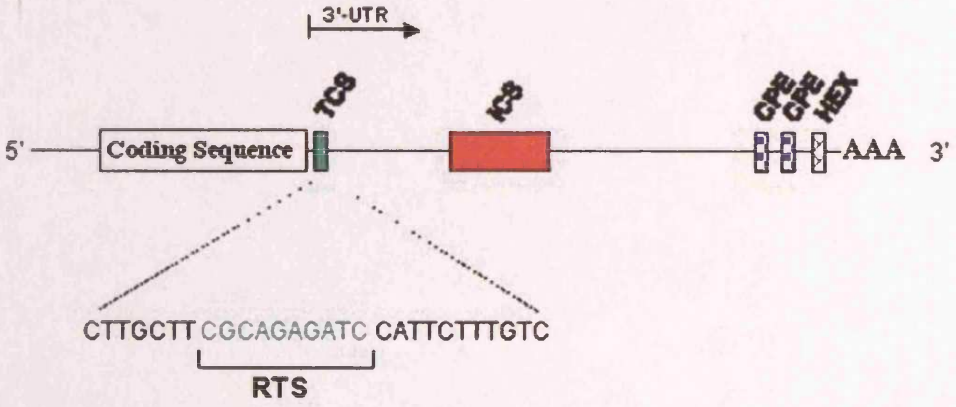



**Figure No. 40:** Functional map of the CaMKII 3'UTR, showing the cis-element regulation sites in the 3'UTR thought to be involved in dendritic targeting and translational control. Within the targeting cis-sequence (TCS) is an oligonucleotide sequence homologous to the RNA transport sequence (RTS) necessary for mRNA transport in oligodendrocytes. ICS denotes inhibitory cis-sequence that can prevent dendritic mRNA transport [14].

While the forebrain specific expression of α-CaMKII is driven by the gene promoter region, the targeting of α-CaMKII mRNA to dendritic processes requires the presence of cis-acting elements in the 3′ untranslated region (3′UTR) of the mRNA [305, 306, 307]. Two particularly interesting cis-acting elements present in the 3′UTR of the α-CaMKII mRNA were described for the first time by Mori and colleagues (2000). α-CaMKII/Neurogranin dendritic localisation element (CNDLE) is a 27bp long domain located at the most 5′ end of the 3′UTR, whose presence seems to be sufficient for dendritic targeting (See Figure No. 40). CNDLE was also identified in the 3′UTR of neurogranin mRNA, which is another transcript known to be targeted to dendrites [307]. The second regulatory element described is an inhibitory element that prevents mRNA translocation to dendrites. The authors have shown that synaptic activity marked by prolonged depolarisation can override this inhibition [307]. In addition to these two novel elements, two well defined sequences, the cytoplasmic polyadenylation element (CPE) and the poly-(A) addition site (HEX) are likely to be involved in activity-dependent translation of α-CaMKII [308].

The importance of dendritic localisation of α-CaMKII mRNA and its local translation for hippocampal LTP was established by Miller and colleagues (2002) [309]. The group generated a mutant mouse whose protein-coding region of α-CaMKII was intact, but in which the dendritic localization signal in the mRNA (CNDLE) was disrupted and mRNA restricted to the soma. The removal of α-CaMKII mRNA from dendrites resulted in a remarkable reduction of the kinase in postsynaptic densities (PSDs), a reduction in LTP, and impairment in spatial memory [309]. These results demonstrate that local translation is important for synaptic delivery of the kinase and that local translation contributes to synaptic plasticity.

### 5.2.6 Aim of the study

Neuronal plasticity, as detected through a phenomenon of LTP, is a basis of learning and memory. CA1 region of the hippocampus is regarded as a crucial neuro-anatomical formation for spatial learning and memory. For successful CA1 LTP induction, activation of NMDAR *via* $Ca^{2+}$ influx is necessary. One mechanism by which $Ca^{2+}$ signals are converted into physiological changes in the neuron is through the activation of protein kinases such as α-CaMKII. The kinase is hypothesised to be directly involved in induction and expression of LTP through conversion of a transient $Ca^{2+}$ signal into a long-lasting increase in enzyme activity and through mediating structural changes in synapses *via* insertion of new AMPARs. Dendritic localisation of α-CaMKII mRNA, mediated through *cis*-acting regulatory elements, and its local translation are crucial for hippocampal LTP and memory formation.

Therefore, we hypothesised that α–CaMKII isoform may influence *g* by affecting two of its major components, spatial learning and memory. We screened the gene for polymorphisms, including promoter, all exons, intron/exon boundaries, 3′ and 5′ UTRs. Detected polymorphisms were tested for association with general cognitive ability on high *g* and average *g* DNA pools using a primer extension method.

## 5.2.7 Materials and methods

### 5.2.7.1 Subjects

See General materials and methods section.

### 5.2.7.2 PCR amplification of the CaMKII-α genomic sequence

A complete genomic sequence for the CaMKII-α gene (Reference sequence ID: NM_015981), together with defined intron/exon boundaries, was obtained on the UCSC Genome Browser web site using a November 2002 sequence assembly version (NCBI Build 31) (http://genome.ucsc.edu/).

PCR primers (Table No. 32) were designed using the publicly available program Primer3 [310].

PCR conditions were determined using the protocol described in General Materials and Methods section.

### 5.2.7.3 Denaturing High Performance Liquid Chromatography

Detailed description of the method is provided in the General Material and Methods chapter. PCR primers used for amplification of DNA fragments to be analysed by the method are listed in Table No. 32.

| No. | Fragment ID | Forward primer sequence | Reverse primer sequence |
|---|---|---|---|
| | | | |
| 1 | CaMKII-5' flank | acggggcaatacctgtctct | agccttcgtcagcatcca |
| 2 | CaMKII-exo1 | cgttttggaggtggttgc | tacaaggaatgctccccaag |
| 3 | CaMKII-exo2 | gtccctgagttgggtcacag | cccctgccataagtatcacc |
| 4 | CaMKII-exo3 | aggcctcttcctctttccac | ctccatccccagagaagaag |
| 5 | CaMKII-exo4 | agcccagccctctttcagta | aagagggaattcagcccaag |
| 6 | CaMKII-exo5 | cccatccttagtttcctggag | gctggatacagtgactagggaga |
| 7 | CaMKII-exo6 | caggtgagtgtcaggtgctg | ccttccagcctaacacagaga |
| 8 | CaMKII-exo7 | gactgtctcttggcgctttc | acagcatctccagccatacc |
| 9 | CaMKII-exo8 | gctgttttcctcagctccag | tacaggatgacccctggaaa |
| 10 | CaMKII-exo9 | gtacgggaagcctgtggac | cccctgtgccagaactagag |
| 11 | CaMKII-exo10 | gagtggatgggcagtaggac | ttacccctgagaaccacagc |
| 12 | CaMKII-exo11 | gatgtcaaccacacctccattaag | gtctgtcctctccactagagcaa |
| 13 | CaMKII-exo12 | gctttggcttcttctctgaatg | ccatctagaactcaatcccctg |
| 14 | CaMKII-exo13 | gtggatgctgctcgtcct | aaagcctaggtcgtgggtct |
| 15 | CaMKII-exo14 | cagggttttcaggacagcat | aatgatgcctcccagtgtg |
| 16 | CaMKII-exo15 | tctgcatgaggtcttcaggtc | tacccccctgaacaacaatcc |
| 17 | CaMKII-exo16 | tccaagggtccaatgagaag | gtaccaaccccagtccacac |
| 18 | CaMKII-exo17 | gctgggagtggagaaagtga | ggcaacactcccatccttta |
| 19 | CaMKII-3'UTR-1 | gactccatctaaagcggcag | gagagggacggacggatg |
| 20 | CaMKII-3'UTR-2 | ggctgtaggctagaatgccc | cagaaatttggctataaaaaggc |
| 21 | CaMKII-3'UTR 3 | tcttcctcctgtgctgacct | tccacagtcccaaaaggaac |
| 22 | CaMKII-3'UTR 4 | gtcacctgtggttattgggg | cctagtggatgtgccaaggt |
| 23 | CaMKII-3'UTR 5 | tctgatgaccacagcaaagc | cagttctgctccaaccaaca |
| 24 | CaMKII-3'UTR 6 | gaagagccacgtccccag | tagccagacgtttgctcctt |
| 25 | CaMKII-3'UTR-7 | agggttctttgttcccatcc | ggcacttggagagtgacaca |
| 26 | CaMKII-3'UTR-8 | aagagcagtgatttggggtg | tttgtctggtgagggtaggc |
| 27 | CaMKII-3'UTR 9 | tgtcactctccaagtgccag | ctttgaagggagacaggagg |
| 28 | CaMKII-3'UTR 10 | taaataaaaatccaaacccaagtca | tctggaactggaaactaacaaaaac |

**Table No. 32:** Oligonucleotide sequences used for PCR amplification of CaMKII-α fragments.

#### 5.2.7.4 Fluorescent automated DNA sequencing using ABI PRISM® BigDye™ Terminators v 3.0

Detailed description of the method is provided in the General Material and Methods chapter.

#### 5.2.7.5 Genotyping DNA pools using The SNaPshot™ Multiplex System

Detailed description of the method is provided in the General Material and Methods chapter.

#### 5.2.7.6 Statistical analysis

Allele frequencies for identified SNPs were estimated from pooled DNA genotyping applying a correction formula, $A = A / (A + kB)$, described by Hoogendoorn *et al.*, where A and B are the peak heights of the primer extension products (representing alleles A and B in pools) and $k$ is the mean of replicates of A/B ratios observed in a heterozygote [20].

Polymorphisms for which estimated allele frequencies were calculated were tested for allelic association with a general cognitive ability by a 2x2 $\chi^2$ using MINITAB 10.5 (1995).

## 5.2.8 Results

### 5.2.8.1 Locating CaMKII-α promoter region

Transcription initiation in eukaryotes requires the presence of various combinations of short sequence elements positioned in the immediate vicinity of a gene, *cis*-acting, and at relatively constant distances from the transcription start site (TSS). These act as recognition signals for transcription factors and collectively constitute a promoter. A minimal DNA element necessary to produce an accurate initiation of transcription lies typically from -60 to +40 bp relative to TSS and is called a core promoter. Different types of genes will have different promoter compositions. Promoters of genes that are expressed at either a specific developmental stage or in a specific tissue and cell type will always include a TATA box. The TATA box (consensus sequence: TATA[A/T]A[A/T]) [311] is usually positioned ~25-30 base pairs upstream from the TSS. Many other genes, like housekeeping genes, are TATA-less. That is, they have CAAT box (CCAAT) at around -80 bp or a GC box (consensus sequence: GGGCGG) instead [70].

Even though the CaMKII-α promoter activity and its importance in localised forebrain expression was extensively studied in rats, no such reports on the human gene variant or any information on its core promoter sequence and exact location have been published so far.

Comprehensive screening of a gene for sequence variants would usually include detailed analysis of its promoter region. In attempt to localise one in the human CaMKII-α a number of publicly available databases and prediction programs were exploited. Results from the following three programs are presented:

**1. PROSCAN** - developed by BioInformatics & Molecular Analysis Section (BIMAS). The software predicts promoter regions based on scoring homologies with putative eukaryotic *Pol II* promoter sequences. No promoter regions were predicted in the first 500bp upstream from the TSS (see Figure No. 41).

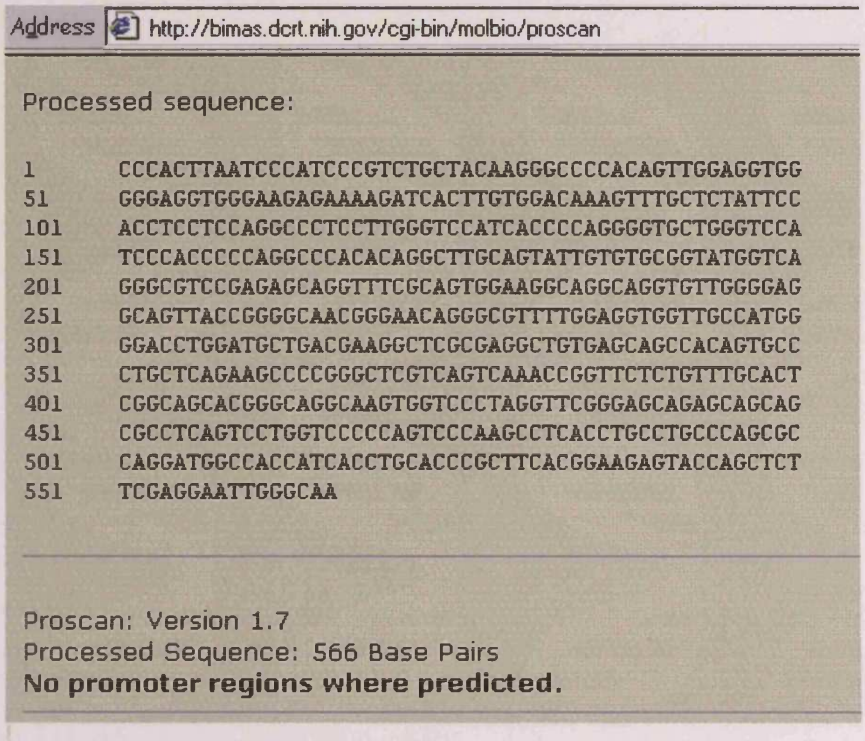



**Figure No. 41:** PROSCAN output file for the first 500 bp upstream from the CaMKII-α transcription start site.

**2. CISTER** - developed by ZLAB (Boston University, USA). Cister predicts regulatory regions in DNA sequences by searching for clusters of cis-elements [312]. No promoter or any other transcriptional regulatory elements were predicted in the first 500bp upstream from the TSS of the human CaMKII-α gene.

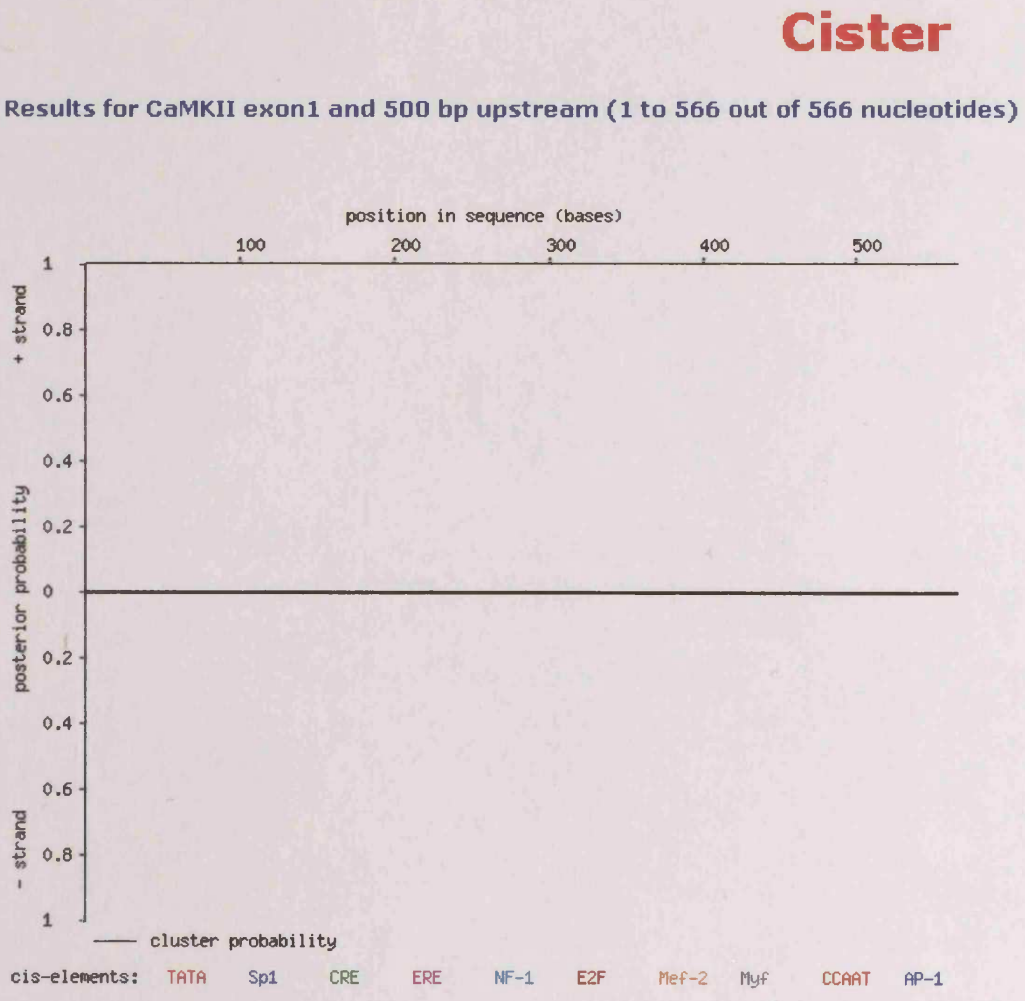



**Figure No. 4:** CISTER output file for Exon 1 and the first 500 bp upstream from the CaMKII-α transcription start site.

**3. NIX** - distributed by UK HGMP Resource Centre. NIX is a WWW tool to simultaneously view the results of running many DNA analysis programs on a particular DNA sequence. The analysis programs run include: GRAIL, Fex, Hexon, GENESCAN, Genemark, Genefinder, FGene, BLAST, RepeatMasker, tRNAscan *etc.* No promoter elements were predicted to exist in the region immediately upstream from the Exon 1 of the CaMKII-α gene. The only promoter element predicted was located downstream from the Exon 1 and was of a marginal prediction quality (displayed as a dark green triangle).

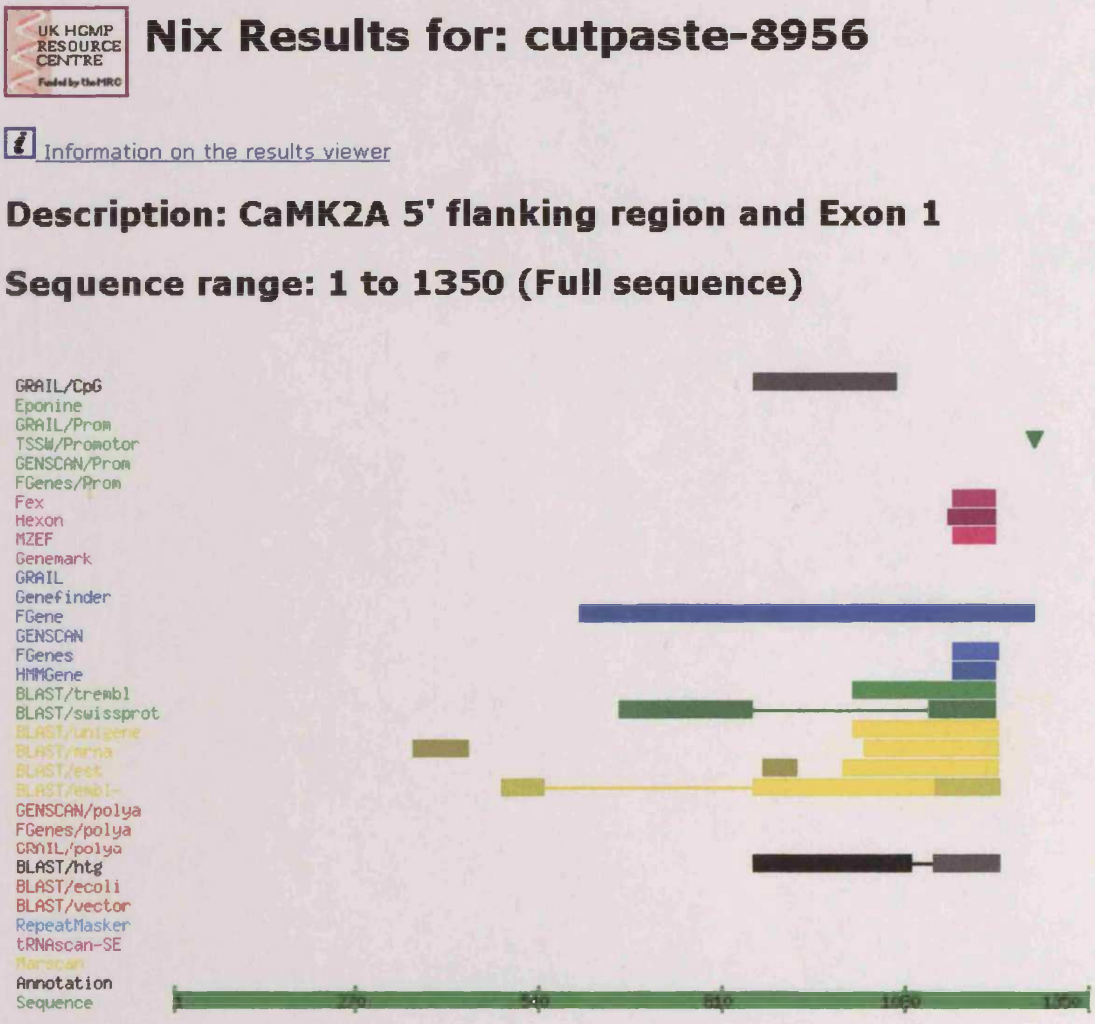



**Figure No. 43:** NIX output file for 1350bp including Exon 1 and the 5′ flanking region of the CaMKII-α.

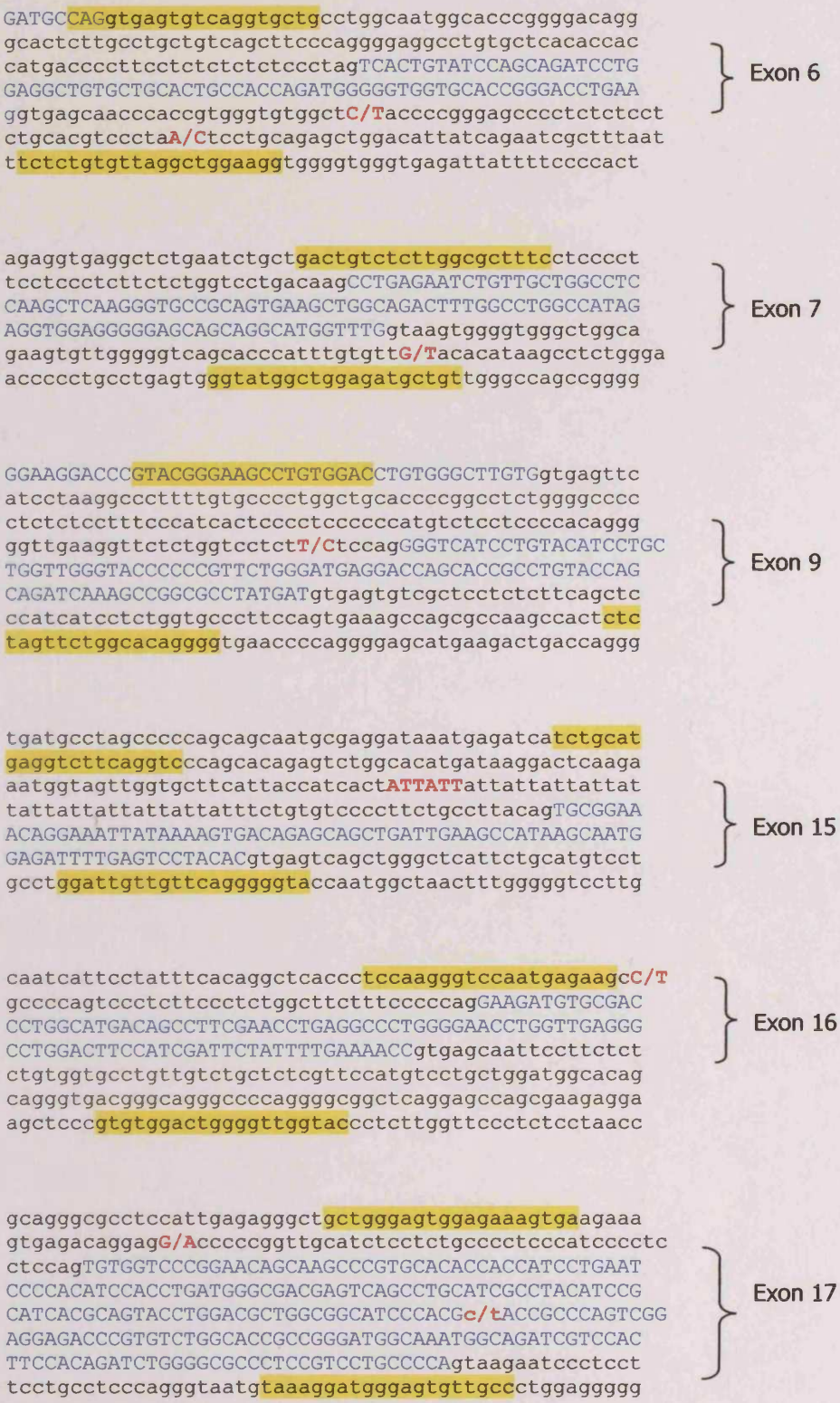
5.2.8.2 PCR, DHPLC and sequencing

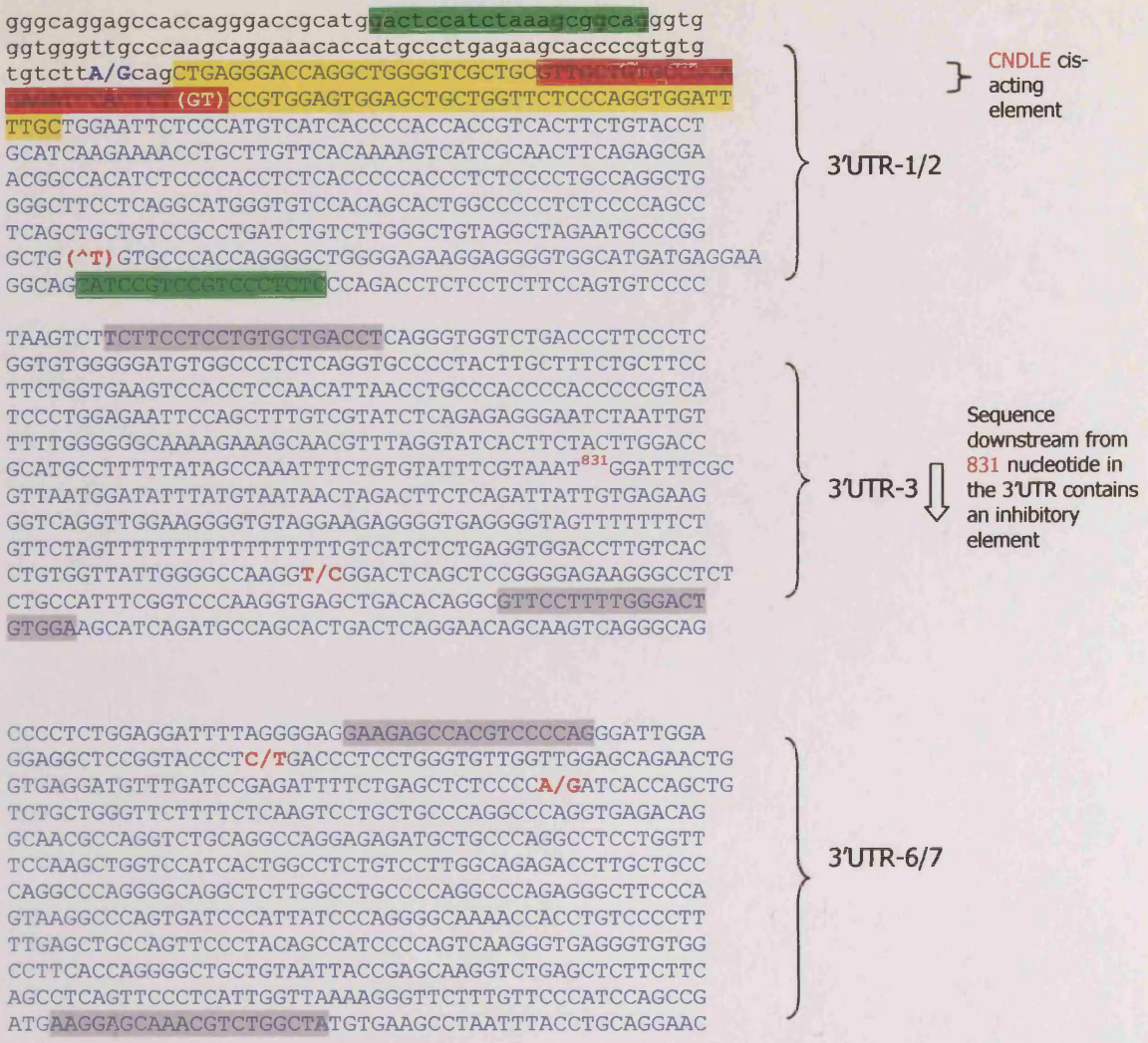
A successful PCR amplification was achieved for most exons, putative promoter region and the whole length of the 3'UTR. Exon 12 is the only fragment for which the amplification assay was not successfully optimised even though different PCR conditions and two sets of primers were attempted, so that it had to be excluded from further assays and analyses.

For the purpose of mutation detection using DHPLC, each fragment was amplified on fifteen individuals from the high *g* group (95% power to detect polymorphisms with a minor allele frequency of 0.1). PCR products were denatured using the melt program (See General Material and Methods) and were subsequently run on a pre-programmed DHPLC machine. Amplified sequencies containing exons 6, 7, 9, 15, 16, 17 and five amplicons from the 3'UTR exhibited changes in elution profiles. Profiles for all individual samples that subsequently underwent sequencing are presented in the Appendix section.

The nature of the polymorphisms and their respective positions were determined by sequencing (See Figure No. 44). In total, fourteen sequence changes were detected, three of which were insertion/deletions. Only one change was exonic (C/T change in exon 17, not causing an amino acid change), 5 were located in the 3'UTR while the remaining 8 were intronic.

**Figure No. 44:** Polymorphisms detected by dHPLC method and characterised by sequencing for 11 PCR fragments.




CNDLE cis-acting element
3'UTR-1/2
3'UTR-3
Sequence downstream from 831 nucleotide in the 3'UTR contains an inhibitory element
3'UTR-6/7

| SNP No. | Polymorphism ID | Alternative ID | Gene position |
|---|---|---|---|
| SNP1 | g.33228C>T | CaM-6-SNP1 | 26bp downstream Exon 6 |
| SNP2 | g.33264A>C | CaM-6-SNP2 | 62bp downstream Exon 6 |
| SNP3 | g.36382G>T | CaM-7-SNP1 | 51bp downstream Exon 7 |
| SNP4 | g.37924T>C | CaM-9-SNP1 | 5bp upstream Exon 9 |
| SNP5 | g.58331_58336delATTATT | CaM-15-SNP1 | 63bp upstream Exon 15 |
| SNP6 | g.61450C>T | CaM-16-SNP1 | 37bp upstream Exon 16 |
| SNP7 | g.66513G>A | CaM-17-SNP1 | 43bp upstream Exon 17 |
| SNP8 | g.66687C>T | CaM-17-SNP2 | Exon 17 |
| SNP9 | g.67057A>G | CaM-3'UTR-SNP1 | 3bp upstream 3'UTR |
| SNP10 | g.67113_67114delGT | CaM-3'UTR-SNP2 | 3'UTR |
| SNP11 | g.67404_67405insT | CaM-3'UTR-SNP3 | 3'UTR |
| SNP12 | g.680727T>C | CaM-3'UTR-SNP4 | 3'UTR |
| SNP13 | g.68818C>T | CaM-3'UTR-SNP5 | 3'UTR |
| SNP14 | g.68889A>G | CaM-3'UTR-SNP6 | 3'UTR |

**Table No. 33:** Polymorphisms detected by DHPLC and characterised by DNA sequencing.

### 5.2.8.3 Genotyping

Genotyping was performed on DNA pools, using the SNaPshot assay (see General Materials and Methods). Extension primers for the SNaPshot assay were designed for all but the Exon 15 insertion/deletion change (g.58331_58336delATTATT) since the nature of the change and its surrounding sequence did not allow for it.

| SNP No | Extension primer ID | Extension primer sequence | dG (kCal/mol) |
|---|---|---|---|
| SNP1 | CaM-6-SNP1-er | ggagagaggggctcccggggt | -2.1 |
| SNP2 | CaM-6-SNP2-ef | cctctctcctctgcacgtcccta | 0 |
| SNP3 | CaM-7-SNP1-er | ggttcccagaggcttatgtgt | 0 |
| SNP4 | CaM-9-SNP1-ef | gggttgaaggttctctggtcctct | -0.4 |
| SNP6 | CaM-16-SNP1-er | cagagggaagagggactggggc | -0.7 |
| SNP7 | CaM-17-SNP1-ef | gtgaagaaagtgagacaggag | 0 |
| SNP8 | CaM-17-SNP2-er | acgggtctcctccgactgggcggt | -1.9 |
| SNP9 | CaM-3'UTR-SNP1-ef | aagcaccccgtgtgtgtctt | -1.1 |
| SNP10 | CaM-3'UTR-SNP2-ef | ttgctgtgccgcagagatccactct | -1.7 |
| SNP11 | CaM-3'UTR-SNP3-ef | tgtaggctagaatgcccgggctg | -1.9 |
| SNP12 | CaM-3'UTR-SNP4-ef | ctgtggttattggggccaagg | -0.9 |
| SNP13 | CaM-3'UTR-SNP5-er | caaccaacacccaggagggtc | -3.0 |
| SNP14 | CaM-3'UTR-SNP6-ef | agattttctgagctctcccc | 0 |

**Table No. 34:** Extension primer IDs, sequences and corresponding dG values

SNaPshot assays for characterised SNPs were optimised using heterozygous individuals previously identified by sequencing. The results of pool genotyping using the assay are presented in Table No. 35.

Estimated allele counts and p-values for high and average *g* pools were obtained for four SNPs (SNP2, SNP3, SNP4 and SNP7). For SNP6, SNP8 and SNP10 only one allele could be visualised in the pools, probably due to inability of the pooling method to detect low frequency alleles. For the remaining SNPs (SNP1, SNP9, SNP11, SNP12, SNP13 and SNP14), the assay failed in the pools. Since individuals heterozygous for those SNPs generated optimum height peaks for both expected alleles, it is possible that the failures in pools were due to DNA degradation as these were rather old (4 years) at

the time of the experimental analysis. Unfortunately, no further access to individual DNA samples meant that fresh pools could not be constructed and experiments repeated.

| SNP No. | SNP ID | Pool ID | Estimated Allele1 count (%) | Estimated Allele2 count (%) | Allelic p-value |
|---|---|---|---|---|---|
| SNP 1 | g.33228C>T | | **Failed** | | |
| SNP 2 | g.33264A>C | mid *g*<br>high *g* | 97 (*0.48*)<br>92 (*0.45*) | 105 (*0.52*)<br>110 (*0.55*) | $\chi 2$=0.2485<br>p=0.618 |
| SNP 3 | g.36382G>T | mid *g*<br>high *g* | 40 (*0.20*)<br>37 (*0.18*) | 162 (*0.80*)<br>165 (*0.82*) | $\chi 2$=0.1444<br>p=0.704 |
| SNP 4 | g.37924T>C | mid *g*<br>high *g* | 96 (*0.47*)<br>100 (*0.49*) | 106 (*0.53*)<br>102 (*0.51*) | $\chi 2$=0.1585<br>p=0.691 |
| SNP 6 | g.61450C>T | mid *g*<br>high *g* | **Low frequency polymorphism** | | |
| SNP 7 | g.66513G>A | mid *g*<br>high *g* | 94 (*0.46*)<br>96 (*0.47*) | 108 (*0.54*)<br>106 (*0.53*) | $\chi 2$=0.0397<br>p=0.842 |
| SNP 8 | g.66687C>T | mid *g*<br>high *g* | **Low frequency polymorphism** | | |
| SNP 9 | g.67057A>G | | **Failed** | | |
| SNP 10 | g.67113_67114delGT | mid *g*<br>high *g* | **Low frequency polymorphism** | | |
| SNP 11 | g.67404-67405insT | | **Failed** | | |
| SNP 12 | g.680727T>C | | **Failed** | | |
| SNP 13 | g.68818C>T | | **Failed** | | |
| SNP 14 | g.68889A>G | | **Failed** | | |

**Table No. 35:** SNaPshot assay pool genotyping results with estimated allele counts, frequencies and corresponding allelic p-values. For SNPs 1, 9, 11, 12, 13 and 14 assays failed with pooled DNA, while for SNPs 6, 8 and 10 rare allele frequency was too low for allele counts to be accurately estimated with the pooling method.

None of the successfully genotyped SNPs generated a significant p-value (g.33264A>C, p=0.618; g.36382G>T, p=0.704; g.37924T>C, p=0.691 and g.66513G>A, p=0.842), as calculated from the estimated allele counts, and entered the final stage of individual genotyping.

### 5.2.9 Discussion

Learning and memory are elementary constituents of intelligence. These cognitive abilities and their spatial component in particular, are among the most heritable human behavioural traits. Sustained changes in synaptic strength, such as long term potentiation (LTP), are believed to be the neural basis of learning and memory [265]. Genetic and biological investigations of the calcium/calmodulin dependent protein kinase II alpha (CaMKII-α) support the hypothesis that CaMKII-α is crucial for induction of LTP and formation of spatial memories in the hippocampus [291]. This study investigated CaMKII-α gene as candidate gene for *g*.

Using DHPLC for mutation detection fourteen intra-genic sequence changes were detected, only one of which was previously reported and deposited into the dbSNP database (g.68889A>G; rs2163766). Polymorphisms were genotyped on DNA pools by a highly accurate protocol for SNP allele frequency estimation based upon the SNaPshot primer extension chemistry [313]. For six polymorphisms assays optimised on individual DNA samples repeatedly failed to generate good quality electropherograms in pooled DNA (g.33228C>T, g.67057A>G, g.67404_67405insT, g.680727T>C, g.68818C>T and g.68889A>G). For those changes allele frequencies could be calculated. For g.61450C>T, g.66687C>T and g.67113_67114delGT allele frequency estimates were not obtained as only single dye electropherograms could be visualised on pooled DNA

samples. This is probably due to presence of low frequency alleles ($f$<0.05) [313], rather than failed assays, as individuals heterozygous for these SNPs generated electropherograms with peaks of expected dye and height representing both alleles. Estimated allele frequencies for four SNPs at positions g.33264A>C, g.36382G>T, g.37924T>C and g.66513G>A were not significantly different between the high $g$ and average $g$ groups ($p_{g.33264A>C}$=0.693, $p_{g.36382G>T}$=0.806, $p_{g.37924T>C}$=0.754 and $p_{g.66513G>A}$=0.910).

Particularly interesting from the point of sequence polymorphism detection was exon11. This exon harbours the codons for the $Thr^{286}$, $Phe^{293}$ and $Asn^{294}$ residues showed to be critical for the autophosphorylation and high-affinity calmodulin binding of the CaMKII-α protein [285] changes in which result in LTP and learning and memory deficits [290]. The high functional importance of the protein domain this CaMKII-α exon is coding for is well reflected by a 100% amino-acid sequence match between human, rat and mouse as well as by this study's failure to detect any sequence changes in it. Such finding is within the study's expectations, as changes in the protein sequence leading to major alteration of its function are not likely to explain normal variation in human cognitive ability in the general population.

The 3' untranslated region of the CaMKII-α gene with its cis-acting regulatory elements is known to play important role in dendritic targeting of the transcript [305,306] disruption of which results in profound spatial learning and memory impairments [309]. Thorough screening of the 3'UTR resulted in characterisation of six polymorphisms, one of which (g.67113_67114delGT) appeared to be within the putative cis-regulatory element (CNDLE). The results of pooled genotyping suggested that the deletion exists at a very low frequency in a general population (only one allele visualised in DNA pools). Bearing in mind the importance of this regulatory element, it would be of great interest to individually genotype all the subjects included in the study and to extend the research into analysis of a functional importance of the detected deletion. It is not beyond the possibility that subtle changes in the CNDLE element sequence, rather than

its complete exclusion, may result in lowered, but not abolished, efficacy of the transcript targeting to dendritic processes and subsequent reduced representation of the kinase in postsynaptic densities.

The CaMKII-α promoter region is known to be responsible for forebrain-specific expression of the gene. Mutation screening of the first 500 bp upstream from the transcription start site failed to detect any sequence changes. However, the study cannot firmly conclude that the human CaMKII-α transcription regulatory region is free from sequence polymorphisms that may influence its expression. The main reason for that is the lack of evidence for the presence of any known transcription regulatory elements in the first 500bp 5′ from the transcription initiation site. The study explored two avenues in attempt to localise a putative promoter, literature and bio-informatics research. Even though there is a plethora of publications on CaMKII-α, its structure and function, a very limited amount of scientific work has been dedicated to analysis of CaMKII-α promoter. In fact, only one group reported their work on functional identification of the rat core promoter for the CaMKII-α (consensus TATA sequence at -162) [314] which influenced the decision to analyse the first few hundred base pairs in the 5′ flanking region. As yet, no such work has been carried out/published on the human gene variant.

Bio-informatics findings were even more disappointing as a number of different transcription element and gene prediction programs (NIX-Nucleotide Identification of unknown sequences, http://www.hgmp.mrc.ac.uk/; CISTER [312], http://zlab.bu.edu/~mfrith/cister.shtml) failed to recognise transcription regulatory elements of any kind not only in the first 500bp, but in the 10kb long stretch of genomic DNA upstream from the transcription start site. Given the importance of the CaMKII-α promoter it would be of great scientific interest to thoroughly examine the region immediately upstream from the TSS and identify a core sequence that is necessary and sufficient to drive the human CaMKII-α forebrain expression.

A number of polymorphisms detected in the gene were not successfully analysed on the study sample mainly for the lack of time and funding for the study to be brought to completion. These should become a priority if the study of the CaMKII-α is to be continued. One interesting avenue to explore could be stratification of the sample into those that performed particularly well in tests that examine spatial ability component of the general intelligence and those that performed less well on them, not forgetting that increasing the total sample size to at least several hundred subjects may further increase a chance of detecting smaller signals contributing to variance in *g*.

To summarise, sequence variants detected and analysed in the human variant of the CaMKII-α gene are not showing evidence of association with *g* in the studied sample and cannot explain the difference in general intelligence scores observed between the high *g* and average *g* groups.

# 6 GENERAL DISCUSSION

For a very long time the human mind was considered a perfect, immaterial and universal attribute of humans [10]. The idea of evolution, brought about by Darwin in the 19th century, resulted in formal recognition that mental ability is a product of evolutionary processes involving natural selection and hereditary variation. However, it was Sir Francis Galton's pioneering work on the heritability of human intelligence that instigated a series of studies that resulted in important advances in data analysis, mental ability test development and the invention of factor analysis by Spearman, all of which ultimately led to the discovery of general cognitive ability, *g*.

At present, it is widely accepted that variation in general cognitive ability is significantly associated with a difference in genetic composition. The belief is firmly supported by a large amount of empirical evidence from a number of family studies carried so far. Unfortunately, suggested heritability levels range from 0.30 to 0.75, a window difficult to narrow any further with currently available data and mathematical models. The pattern of correlation in scores on mental ability tests between various relative pairs suggests complex inheritance and involvement of a number of genetic and environmental factors in shaping heritability. The most likely scenario involves a large number of genetic loci with varied but very small effect on a trait. Statistical models predict that only a small fraction of them are likely to have a modest effect and explain much of the phenotypic variance observed.

Although the heritable nature of intelligence was established a long time ago, the magnitude of this effect is still not fully recognized. Moreover, the difficulties of undertaking large, genome-wide association studies, which are currently the only way to systematically detect multiple small-effect genetic loci [109], has meant that there have been relatively few molecular genetic studies of *g*. Candidate gene studies are easier to perform and indeed several have been reported in recent years [315-319]. Unfortunately, the number of potential candidates is large and unlikely to be much reduced by animal

research that opened new avenues of thinking and uncovered complex interacting biological networks underlying different cognitive processes like learning and memory.

Recent advances in laboratory techniques (DNA pooling), equipment (ABI sequencers, dHPLC machines) and publicly available resources (large databases of multi-allelic and bi-allelic markers) have eliminated most of the problems encountered in the past in genome-wide studies. With these advances, it has been possible to carry out a genome-wide screen for association with IQ – the first study of this kind.

Almost 2000 micro-satellite markers with average spacing of 2 cM were tested on a case/control set, using a multiple-stage study design combined with pooled genotyping. Such a design allowed time- and cost-effective screening of a large number of markers, reduced generation of false positives to a minimum and, by incorporating a replication stage, excluded the need for p-value correction for multiple testing [115]. However, a single marker (D4S2460), found to be associated with IQ in our case/control study, failed to be replicated by TDT testing on trios. Even though the study failed to generate markers convincingly associated with human intelligence, it showed that the DNA pooling is a reliable method and a tool of choice for future studies of this kind. It has also highlighted the need for relaxation of what appeared to be the over-stringent criteria imposed by the multiple-stage design in order to reduce the level of false negatives. By shifting the focus of the study from children to adults, known to have higher heritability levels for *g*, it may be possible to further improve the prospects for detecting a genuine association.

A new generation of markers, bi-allelic SNPs, is increasingly used in genetic studies. Their generous availability, dense distribution and amenability for high-throughput projects have made them the markers of choice for genome-wide studies. The question of optimal number of SNPs to be genotyped in such studies is still to be answered and suggestions are currently ranging from ~250,000 (few SNPs per 1 LDU) [146] to more pessimistic ~1,000,000 SNPs (1 SNP every 3kb) [117]. Even so, such strategy is certainly to be considered in future attempts to identify genes contributing to variation in *g*.

Candidate gene studies using knockout strains of mice have pointed to a number of genes having a possible role in different aspects of mental functioning. Molecular genetic analysis of two such genes, ApoE and CaMKII-α, was an attempt to throw some light on potential biological systems underlying general cognitive ability. Neither Apolipoprotein E, a well-known protein with established role in neuronal maintenance and repair [225, 239], nor CaMKII-α, believed to be a central molecular mechanism involved in learning and memory [285], have shown significant association with general cognitive ability. The negative reports on these two genes do not exclude them from involvement in *g*; they rather emphasise the need for much larger study samples if small effect genes are to be detected.

Similar to combing human genome with a dense battery of genetic markers, a more systematic approach to candidate gene studies could potentially lead to significant discoveries. Rather than analysing isolated genes, investigation of whole pathways should take place. Genetic studies of such kind should run in parallel to, and not isolated from, animal research which creates an inexhaustible pool of biological hypotheses through creation of transgenic animals. CaMKII-α is a good candidate gene, but it is only one of the many components of a pathway leading to synaptic plasticity. Not only that, but its function is dependent upon other molecular interactions preceding activation of the pathway it belongs to. In addition, this pathway's activity initiates a set of cell signalling cascades crucial for memory consolidation. Clearly, investigating whole pathways and their networks is a future of candidate gene studies, but expertise in different scientific fields and joint effort of many research groups is going to be necessary to carry out and interpret results of such large studies.

The identification of genes for human intelligence is still in the initial stage, but soon the first genuine associations will start emerging. It is reasonable to believe that reports on such genes will give a new impetus to ethical criticism of the intelligence research and instigate general public concern about the wider application of the scientific knowledge. However, the criticism thus far has largely been based on

misconceptions about behavioural genetics. Therefore, it must be stressed that a large number of genes is likely to be involved in shaping human intelligence in concert with environmental factors, which makes development of genetic tests of intelligence improbable. What the critics of intelligence research often fail to appreciate is the significant contribution it provides to better understanding of molecular and cellular mechanisms of human behaviour and the mind.

As scientists, we have a responsibility to strive to further the knowledge and understanding of our environment and ourselves. In doing so, we have to make sure that the power of knowledge be not abused but that it be used solely for the enrichment of human life and the advancement of well-being of our species as a whole.

# 7 APPENDIX

## *7.1 Genome-wide screen for association with g:*

Table with listed microsatellite markers, their positions on the Marshfield Genetic map and inter-marker distances (cM). For each of the chromosomes investigated **average** and **maximum** inter-marker distances are presented (printed in **RED** at the bottom of each chromosome's marker list).

| chr1 | cM | inter-marker distance (cM) | chr2 | cM | inter-marker distance (cM) | chr3 | cM | inter-marker distance (cM) |
|---|---|---|---|---|---|---|---|---|
| D1S468 | **4.22** | | D2S2584 | **1.95** | | D3S1270 | **6.96** | |
| D1S2795 | **11.87** | 7.65 | D2S2268 | **1.95** | 0 | D3S1307 | **7.12** | 0.16 |
| D1S214 | **14.04** | 2.17 | D2S2245 | **7.05** | 5.1 | D3S2426 | **7.38** | 0.26 |
| D1S2642 | **14.59** | 0.55 | D2S319 | **7.6** | 0.55 | D3S1297 | **8.31** | 0.93 |
| D1S253 | **14.59** | 0 | D2S304 | **10.04** | 2.44 | D3S3630 | **10.7** | 2.39 |
| D1S1646 | **14.59** | 0 | D2S2211 | **15.61** | 5.57 | D3S1620 | **14.46** | 3.76 |
| D1S548 | **15.13** | 0.54 | D2S2164 | **18.42** | 2.81 | D3S1515 | **18.38** | 3.92 |
| D1S2694 | **15.13** | 0 | D2S162 | **20.03** | 1.61 | D3S3706 | **18.97** | 0.59 |
| D1S1160 | **20.61** | 5.48 | D2S2243 | **20.03** | 0 | D3S1304 | **22.33** | 3.36 |
| D1S503 | **20.61** | 0 | D2S423 | **22.1** | 2.07 | D3S3728 | **24.29** | 1.96 |
| D1S450 | **20.61** | 0 | D2S398 | **23.57** | 1.47 | D3S3591 | **24.89** | 0.6 |
| D1S2736 | **20.61** | 0 | D2S2278 | **26.52** | 2.95 | D3S1537 | **27.72** | 2.83 |
| D1S1597 | **29.93** | 9.32 | D2S2199 | **31.05** | 4.53 | D3S1597 | **29.92** | 2.2 |
| D1S489 | **29.93** | 0 | D2S219 | **31.2** | 0.15 | D3S3601 | **31.13** | 1.21 |
| D1S2718 | **29.93** | 0 | D2S2267 | **33.31** | 2.11 | D3S1263 | **36.1** | 4.97 |
| D1S2834 | **31.02** | 1.09 | D2S149 | **34.04** | 0.73 | D3S3680 | **36.1** | 0 |
| D1S2728 | **33.75** | 2.73 | D2S312 | **34.78** | 0.74 | D3S3610 | **37.2** | 1.1 |
| D1S407 | **33.75** | 0 | D2S2346 | **36.72** | 1.94 | D3S2403 | **37.2** | 0 |
| D1S507 | **33.75** | 0 | D2S305 | **38.87** | 2.15 | D3S1585 | **38.83** | 1.63 |
| D1S2697 | **37.05** | 3.3 | D2S2342 | **40.47** | 1.6 | D3S2338 | **42.1** | 3.27 |
| D1S1592 | **38.51** | 1.46 | D2S2373 | **42.65** | 2.18 | D3S3726 | **42.64** | 0.54 |
| D1S2826 | **41.92** | 3.41 | D2S218 | **44.09** | 1.44 | D3S3038 | **44.81** | 2.17 |
| D1S199 | **45.33** | 3.41 | D2S2226 | **45.3** | 1.21 | D3S3659 | **47.44** | 2.63 |
| D1S552 | **45.33** | 0 | D2S2350 | **46.37** | 1.07 | D3S2336 | **49.18** | 1.74 |
| D1S2647 | **45.33** | 0 | D2S165 | **47.43** | 1.06 | D3S2337 | **50.25** | 1.07 |
| D1S478 | **48.53** | 3.2 | D2S2283 | **50.65** | 3.22 | D3S1266 | **52.6** | 2.35 |
| D1S458 | **52.7** | 4.17 | D2S265 | **54.96** | 4.31 | D3S1283 | **52.6** | 0 |
| D1S2698 | **52.7** | 0 | D2S367 | **54.96** | 0 | D3S3547 | **55.11** | 2.51 |
| D1S2838 | **52.7** | 0 | D2S1325 | **54.96** | 0 | D3S1609 | **55.11** | 0 |
| D1S234 | **55.1** | 2.4 | D2S1788 | **55.51** | 0.55 | D3S3567 | **56.69** | 1.58 |
| D1S2787 | **56.48** | 1.38 | D2S2230 | **56.15** | 0.64 | D3S1619 | **60.98** | 4.29 |
| D1S2639 | **56.74** | 0.26 | D2S2186 | **57.22** | 1.07 | D3S1612 | **61.52** | 0.54 |
| D1S2854 | **56.74** | 0 | D2S2220 | **59.36** | 2.14 | D3S3718 | **61.52** | 0 |
| D1S441 | **64.38** | 7.64 | D2S2163 | **59.36** | 0 | D3S1260 | **63.12** | 1.6 |
| D1S255 | **65.47** | 1.09 | D2S2305 | **61.66** | 2.3 | D3S3658 | **65.26** | 2.14 |
| D1S2723 | **65.47** | 0 | D2S2294 | **64.84** | 3.18 | D3S2304 | **67.94** | 2.68 |
| D1S380 | **67.22** | 1.75 | D2S2174 | **67.58** | 2.74 | D3S3678 | **68.47** | 0.53 |
| D1S1598 | **70.41** | 3.19 | D2S2291 | **69.77** | 2.19 | D3S3647 | **68.47** | 0 |
| D1S447 | **73.81** | 3.4 | D2S2182 | **70.31** | 0.54 | D3S3582 | **69.19** | 0.72 |
| D1S451 | **75.66** | 1.85 | D2S288 | **70.63** | 0.32 | D3S1767 | **69.9** | 0.71 |
| D1S2797 | **75.66** | 0 | D2S123 | **73.61** | 2.98 | D3S1621 | **70.61** | 0.71 |
| D1S232 | **76.27** | 0.61 | D2S2251 | **76.34** | 2.73 | D3S3660 | **72.21** | 1.6 |
| D1S2661 | **78.25** | 1.98 | D2S2292 | **76.34** | 0 | D3S3717 | **74.35** | 2.14 |
| D1S475 | **82.41** | 4.16 | D2S2279 | **77.97** | 1.63 | D3S1592 | **77.01** | 2.66 |
| D1S2690 | **83.07** | 0.66 | D2S2165 | **80.69** | 2.72 | D3S1547 | **78.64** | 1.63 |
| D1S2867 | **85.68** | 2.61 | D2S2225 | **82.82** | 2.13 | D3S1313 | **78.64** | 0 |
| D1S2890 | **85.68** | 0 | D2S296 | **83.88** | 1.06 | D3S1300 | **80.32** | 1.68 |
| D1S405 | **86.77** | 1.09 | D2S290 | **84.42** | 0.54 | D3S3631 | **82.24** | 1.92 |
| D1S2869 | **86.77** | 0 | D2S285 | **86.02** | 1.6 | D3S1312 | **82.24** | 0 |
| D1S2741 | **87.86** | 1.09 | D2S2152 | **87.62** | 1.6 | D3S1600 | **85.97** | 3.73 |
| D1S3728 | **89.49** | 1.63 | D2S303 | **88.15** | 0.53 | D3S1285 | **91.18** | 5.21 |
| D1S2788 | **93.86** | 4.37 | D2S2115 | **88.15** | 0 | D3S3545 | **96.12** | 4.94 |
| D1S2638 | **96.04** | 2.18 | D2S2111 | **90.29** | 2.14 | D3S1296 | **97.75** | 1.63 |
| D1S515 | **97.49** | 1.45 | D2S2112 | **90.29** | 0 | D3S3614 | **101.55** | 3.8 |
| D1S368 | **100.39** | 2.9 | D2S2110 | **90.82** | 0.53 | D3S2406 | **102.64** | 1.09 |
| D1S1665 | **102.02** | 1.63 | D2S1394 | **90.82** | 0 | D3S3581 | **104.83** | 2.19 |
| D1S2614 | **104.79** | 2.77 | D2S286 | **94.05** | 3.23 | D3S3653 | **107.19** | 2.36 |
| D1S500 | **107.56** | 2.77 | D2S2116 | **95.67** | 1.62 | D3S3681 | **109.22** | 2.03 |
| D1S430 | **109.04** | 1.48 | D2S169 | **99.41** | 3.74 | D3S1604 | **109.22** | 0 |
| D1S2856 | **113.69** | 4.65 | D2S139 | **101.56** | 2.15 | D3S3049 | **109.22** | 0 |

| chr1 | cM | inter-marker distance (cM) | chr2 | cM | inter-marker distance (cM) | chr3 | cM | inter-marker distance (cM) |
|---|---|---|---|---|---|---|---|---|
| D1S454 | **113.69** | 0 | D2S394 | **103.16** | 1.6 | D3S1577 | **109.22** | 0 |
| D1S551 | **113.69** | 0 | D2S2161 | **105** | 1.84 | D3S1276 | **111.89** | 2.67 |
| D1S488 | **114.24** | 0.55 | D2S2232 | **107.46** | 2.46 | D3S3671 | **112.96** | 1.07 |
| D1S2807 | **114.24** | 0 | D2S2181 | **110** | 2.54 | D3S1552 | **114.02** | 1.06 |
| D1S2766 | **118.14** | 3.9 | D2S2216 | **111.21** | 1.21 | D3S1271 | **117.76** | 3.74 |
| D1S1588 | **125.51** | 7.37 | D2S2175 | **111.21** | 0 | D3S3655 | **117.76** | 0 |
| D1S435 | **125.51** | 0 | D2S2209 | **112.28** | 1.07 | D3S2459 | **119.09** | 1.33 |
| D1S2868 | **126.16** | 0.65 | D2S2264 | **114.42** | 2.14 | D3S3632 | **120.43** | 1.34 |
| D1S424 | **126.16** | 0 | D2S2972 | **114.42** | 0 | D3S1291 | **121.67** | 1.24 |
| D1S236 | **128.73** | 2.57 | D2S2356 | **115.49** | 1.07 | D3S1616 | **124.16** | 2.49 |
| D1S223 | **134.2** | 5.47 | D2S135 | **116.02** | 0.53 | D3S3045 | **124.16** | 0 |
| D1S2896 | **134.2** | 0 | D2S2386 | **118.16** | 2.14 | D3S1281 | **124.16** | 0 |
| D1S206 | **134.2** | 0 | D2S1890 | **119.22** | 1.06 | D3S3044 | **126.83** | 2.67 |
| D1S1671 | **134.2** | 0 | D2S1889 | **120.29** | 1.07 | D3S2496 | **126.83** | 0 |
| D1S2626 | **136.34** | 2.14 | D2S160 | **122.96** | 2.67 | D3S3683 | **127.89** | 1.06 |
| D1S1631 | **136.88** | 0.54 | D2S347 | **131.51** | 8.55 | D3S3670 | **129.73** | 1.84 |
| D1S2759 | **136.88** | 0 | D2S2339 | **132.58** | 1.07 | D3S1579 | **131.83** | 2.1 |
| D1S2726 | **144.38** | 7.5 | D2S1273 | **132.58** | 0 | D3S1558 | **133.93** | 2.1 |
| D1S2809 | **144.38** | 0 | D2S112 | **141.62** | 9.04 | D3S3649 | **134.64** | 0.71 |
| D1S2837 | **146.53** | 2.15 | D2S2219 | **142.83** | 1.21 | D3S3515 | **136.32** | 1.68 |
| D1S2744 | **149.2** | 2.67 | D2S1334 | **145.08** | 2.25 | D3S3576 | **138** | 1.68 |
| D1S440 | **150.81** | 1.61 | D2S314 | **146.15** | 1.07 | D3S3620 | **138** | 0 |
| D1S514 | **152.45** | 1.64 | D2S2376 | **146.15** | 0 | D3S3552 | **139.12** | 1.12 |
| D1S442 | **154.74** | 2.29 | D2S2288 | **147.4** | 1.25 | D3S1267 | **139.12** | 0 |
| D1S498 | **155.89** | 1.15 | D2S2362 | **150.96** | 3.56 | D3S3569 | **140.19** | 1.07 |
| D1S2858 | **159.32** | 3.43 | D2S381 | **151.5** | 0.54 | D3S3584 | **143.94** | 3.75 |
| D1S2624 | **162.57** | 3.25 | D2S2184 | **152.04** | 0.54 | D3S3607 | **143.94** | 0 |
| D1S2125 | **163.34** | 0.77 | D2S2277 | **154.48** | 2.44 | D3S1292 | **146.6** | 2.66 |
| D1S1653 | **164.09** | 0.75 | D2S2299 | **157.55** | 3.07 | D3S3657 | **148.2** | 1.6 |
| D1S484 | **169.68** | 5.59 | D2S2360 | **161.26** | 3.71 | D3S3641 | **149.8** | 1.6 |
| D1S104 | **175.62** | 5.94 | D2S142 | **161.26** | 0 | D3S1590 | **150.64** | 0.84 |
| D1S1677 | **175.62** | 0 | D2S284 | **161.81** | 0.55 | D3S3528 | **151.49** | 0.85 |
| D1S2628 | **177.86** | 2.24 | D2S306 | **164.81** | 3 | D3S1615 | **151.49** | 0 |
| D1S1158 | **179.1** | 1.24 | D2S2370 | **165.06** | 0.25 | D3S1549 | **151.49** | 0 |
| D1S196 | **181.49** | 2.39 | D2S2380 | **166.51** | 1.45 | D3S3586 | **152.62** | 1.13 |
| D1S431 | **182.35** | 0.86 | D2S399 | **169.41** | 2.9 | D3S1309 | **153.74** | 1.12 |
| D1S433 | **184.21** | 1.86 | D2S124 | **169.41** | 0 | D3S3546 | **154.48** | 0.74 |
| D1S210 | **188.85** | 4.64 | D2S1379 | **169.41** | 0 | D3S1569 | **158.38** | 3.9 |
| D1S218 | **191.52** | 2.67 | D2S2345 | **171.04** | 1.63 | D3S1550 | **159.8** | 1.42 |
| D1S2791 | **192.9** | 1.38 | D2S1776 | **173** | 1.96 | D3S3704 | **161.04** | 1.24 |
| D1S2818 | **198.3** | 5.4 | D2S2284 | **174.3** | 1.3 | D3S1306 | **164.25** | 3.21 |
| D1S444 | **200.96** | 2.66 | D2S2381 | **175.91** | 1.61 | D3S3705 | **165.32** | 1.07 |
| D1S191 | **200.96** | 0 | D2S335 | **175.91** | 0 | D3S1555 | **165.32** | 0 |
| D1S238 | **202.73** | 1.77 | D2S2307 | **180.79** | 4.88 | D3S1299 | **166.93** | 1.61 |
| D1S428 | **204.51** | 1.78 | D2S300 | **182.96** | 2.17 | D3S1315 | **168.94** | 2.01 |
| D1S413 | **212.44** | 7.93 | D2S385 | **182.96** | 0 | D3S1584 | **170.14** | 1.2 |
| D1S2840 | **212.44** | 0 | D2S2281 | **186.21** | 3.25 | D3S1570 | **171.47** | 1.33 |
| D1S2717 | **218.46** | 6.02 | D2S364 | **186.21** | 0 | D3S3708 | **173.87** | 2.4 |
| D1S1725 | **219.92** | 1.46 | D2S2366 | **186.84** | 0.63 | D3S2415 | **174.94** | 1.07 |
| D1S249 | **220.65** | 0.73 | D2S426 | **190** | 3.16 | D3S3643 | **176.54** | 1.6 |
| D1S205 | **229.13** | 8.48 | D2S2262 | **190** | 0 | D3S3622 | **177.75** | 1.21 |
| D1S217 | **231.11** | 1.98 | D2S161 | **191.08** | 1.08 | D3S1614 | **177.75** | 0 |
| D1S425 | **231.11** | 0 | D2S315 | **193.26** | 2.18 | D3S1243 | **180.8** | 3.05 |
| D1S237 | **232.81** | 1.7 | D2S309 | **198.65** | 5.39 | D3S1282 | **180.8** | 0 |
| D1S2827 | **234.52** | 1.71 | D2S2189 | **199.18** | 0.53 | D3S2433 | **181.87** | 1.07 |
| D1S2621 | **238.52** | 4 | D2S2237 | **200.43** | 1.25 | D3S1574 | **181.87** | 0 |
| D1S2871 | **241.26** | 2.74 | D2S369 | **202.92** | 2.49 | D3S3520 | **183.47** | 1.6 |
| D1S2763 | **242.34** | 1.08 | D2S2358 | **203.46** | 0.54 | D3S2425 | **185.4** | 1.93 |
| D1S225 | **245.05** | 2.71 | D2S322 | **206.13** | 2.67 | D3S2421 | **186.68** | 1.28 |
| D1S2709 | **247.23** | 2.18 | D2S317 | **208.58** | 2.45 | D3S2412 | **189** | 2.32 |

| chr1 | cM | inter-marker distance (cM) | chr2 | cM | inter-marker distance (cM) | chr3 | cM | inter-marker distance (cM) |
|---|---|---|---|---|---|---|---|---|
| D1S2800 | **252.12** | 4.89 | D2S137 | **210.43** | 1.85 | D3S3715 | **190.43** | 1.43 |
| D1S2850 | **256.26** | 4.14 | D2S2394 | **210.43** | 0 | D3S3699 | **191.79** | 1.36 |
| D1S2785 | **266.27** | 10.01 | D2S2382 | **213.49** | 3.06 | D3S1618 | **194.68** | 2.89 |
| D1S184 | **267.51** | 1.24 | D2S301 | **214.71** | 1.22 | D3S3578 | **195.6** | 0.92 |
| D1S304 | **267.51** | 0 | D2S2210 | **215.78** | 1.07 | D3S3592 | **198.68** | 3.08 |
| D1S321 | **267.51** | 0 | D2S173 | **215.78** | 0 | D3S1617 | **198.68** | 0 |
| D1S2842 | **273.46** | 5.95 | D2S295 | **215.78** | 0 | D3S3570 | **201.14** | 2.46 |
| D1S1609 | **274.53** | 1.07 | D2S120 | **218.45** | 2.67 | D3S1602 | **201.14** | 0 |
| D1S423 | **277.8** | 3.27 | D2S377 | **220.59** | 2.14 | D3S3600 | **203.81** | 2.67 |
| D1S2836 | **285.75** | 7.95 | D2S2197 | **222.2** | 1.61 | D3S1580 | **207.73** | 3.92 |
| D1S2132 | **?** | | D2S130 | **222.73** | 0.53 | D3S3530 | **209.41** | 1.68 |
| D1S168 | **?** | | D2S2228 | **224.33** | 1.6 | D3S1294 | **210.09** | 0.68 |
| D1S3471 | **?** | | D2S2308 | **227** | 2.67 | D3S2747 | **212.61** | 2.52 |
| D1S1634 | **?** | | D2S401 | **229.14** | 2.14 | D3S3557 | **214.45** | 1.84 |
| D1S2145 | **?** | | D2S2158 | **229.14** | 0 | D3S1662 | **214.45** | 0 |
| D1S519 | **?** | | D2S2185 | **231.27** | 2.13 | D3S1523 | **217.24** | 2.79 |
| D1S2129 | **?** | | D2S2213 | **231.81** | 0.54 | D3S3590 | **217.24** | 0 |
| D1S2144 | **?** | | D2S396 | **232.9** | 1.09 | D3S1305 | **222.83** | 5.59 |
| D1S3473 | **?** | | D2S427 | **236.7** | 3.8 | D3S1272 | **224.88** | 2.05 |
| D1S3468 | **?** | | D2S2344 | **238.33** | 1.63 | D3S1311 | **224.88** | 0 |
| D1S2131 | **?** | | D2S206 | **240.79** | 2.46 | D3S3707 | **224.88** | 0 |
| D1S550 | **?** | | D2S2176 | **242.17** | 1.38 | D3S3550 | **227.57** | 2.69 |
| D1S3469 | **?** | | D2S2205 | **243.01** | 0.84 | D3S1448 | **?** | |
| | max | **10.01** | D2S2202 | **249.22** | 6.21 | D3S1210 | **?** | |
| | ave | **2.18** | D2S1397 | **249.22** | 0 | D3S4550 | **?** | |
| | | | D2S2968 | **251.94** | 2.72 | D3S1449 | **?** | |
| | | | D2S2285 | **258.49** | 6.55 | D3S659 | **?** | |
| | | | D2S125 | **260.63** | 2.14 | D3S1352 | **?** | |
| | | | D2S395 | **261.34** | 0.71 | D3S1217 | **?** | |
| | | | D2S140 | **263.56** | 2.22 | D3S2389 | **?** | |
| | | | D2S2585 | **263.56** | 0 | D3S1246 | **?** | |
| | | | D2S2338 | **269.07** | 5.51 | D3S1233 | **?** | |
| | | | D2S2736 | **?** | | D3S4544 | **?** | |
| | | | D2S2966 | **?** | | | max | **5.59** |
| | | | D2S1335 | **?** | | | ave | **1.56** |
| | | | D2S1387 | **?** | | | | |
| | | | D2S1774 | **?** | | | | |
| | | | D2S2969 | **?** | | | | |
| | | | D2S1392 | **?** | | | | |
| | | | D2S2970 | **?** | | | | |
| | | | D2S2978 | **?** | | | | |
| | | | D2S2973 | **?** | | | | |
| | | | D2S100 | **?** | | | | |
| | | | D2S266 | **?** | | | | |
| | | | D2S207 | **?** | | | | |
| | | | | max | **9.04** | | | |
| | | | | ave | **1.77** | | | |

| chr4 | cM | inter-marker distance (cM) | chr5 | cM | inter-marker distance (cM) | chr6 | cM | inter-marker distance (cM) |
|---|---|---|---|---|---|---|---|---|
| D4S412 | **4.74** | | D5S678 | **1.72** | | D6S942 | **0** | |
| D4S3023 | **8.24** | 3.5 | D5S1981 | **1.72** | 0 | D6S1685 | **9.18** | 9.18 |
| D4S431 | **12.35** | 4.11 | D5S2005 | **1.72** | 0 | D6S1598 | **11.89** | 2.71 |
| D4S403 | **25.9** | 13.55 | D5S2849 | **7.77** | 6.05 | D6S309 | **14.07** | 2.18 |
| D4S1602 | **25.9** | 0 | D5S1980 | **7.77** | 0 | D6S277 | **14.61** | 0.54 |
| D4S3048 | **29.14** | 3.24 | D5S675 | **9.41** | 1.64 | D6S410 | **15.43** | 0.82 |
| D4S1567 | **29.68** | 0.54 | D5S405 | **9.41** | 0 | D6S940 | **16.84** | 1.41 |
| D4S419 | **33.42** | 3.74 | D5S406 | **11.85** | 2.44 | D6S470 | **18.22** | 1.38 |
| D4S2926 | **33.96** | 0.54 | D5S464 | **14.3** | 2.45 | D6S1034 | **23.23** | 5.01 |
| D4S3020 | **35.03** | 1.07 | D5S635 | **14.91** | 0.61 | D6S1593 | **24.51** | 1.28 |
| D4S2953 | **35.56** | 0.53 | D5S676 | **16.72** | 1.81 | D6S2434 | **25.08** | 0.57 |
| D4S3017 | **35.56** | 0 | D5S1957 | **19.02** | 2.3 | D6S1721 | **25.08** | 0 |
| D4S404 | **37.16** | 1.6 | D5S1976 | **19.67** | 0.65 | D6S1578 | **26.71** | 1.63 |
| D4S1551 | **38.77** | 1.61 | D5S2004 | **21.81** | 2.14 | D6S259 | **27.78** | 1.07 |
| D4S3044 | **38.77** | 0 | D5S432 | **22.88** | 1.07 | D6S1676 | **29.93** | 2.15 |
| D4S3052 | **38.77** | 0 | D5S2081 | **24.48** | 1.6 | D6S289 | **29.93** | 0 |
| D4S1091 | **42.74** | 3.97 | D5S1991 | **26.73** | 2.25 | D6S1584 | **32.62** | 2.69 |
| D4S418 | **44.66** | 1.92 | D5S1954 | **28.22** | 1.49 | D6S285 | **34.23** | 1.61 |
| D4S2912 | **47.58** | 2.92 | D5S416 | **28.76** | 0.54 | D6S1597 | **37.79** | 3.56 |
| D4S3027 | **49.47** | 1.89 | D5S1997 | **30.51** | 1.75 | D6S1029 | **39.2** | 1.41 |
| D4S3001 | **49.47** | 0 | D5S268 | **33.04** | 2.53 | D6S1660 | **40.14** | 0.94 |
| D4S2955 | **50.53** | 1.06 | D5S2096 | **33.04** | 0 | D6S1545 | **42.98** | 2.84 |
| D4S2995 | **50.53** | 0 | D5S2031 | **36.25** | 3.21 | D6S258 | **44.41** | 1.43 |
| D4S1581 | **51.6** | 1.07 | D5S502 | **39.46** | 3.21 | D6S1281 | **44.41** | 0 |
| D4S405 | **56.95** | 5.35 | D5S819 | **41.06** | 1.6 | D6S1666 | **45.5** | 1.09 |
| D4S2382 | **56.95** | 0 | D5S477 | **45.34** | 4.28 | D6S1629 | **47.71** | 2.21 |
| D4S2919 | **57.75** | 0.8 | D5S2062 | **45.34** | 0 | D6S1583 | **47.71** | 0 |
| D4S1547 | **60.16** | 2.41 | D5S651 | **47.09** | 1.75 | D6S291 | **49.5** | 1.79 |
| D4S1536 | **60.79** | 0.63 | D5S426 | **51.99** | 4.9 | D6S1576 | **50.75** | 1.25 |
| D4S3002 | **61.42** | 0.63 | D5S1994 | **51.99** | 0 | D6S1548 | **52** | 1.25 |
| D4S2971 | **61.42** | 0 | D5S2023 | **53.67** | 1.68 | D6S1680 | **53.81** | 1.81 |
| D4S2996 | **63.58** | 2.16 | D5S1490 | **57.3** | 3.63 | D6S1641 | **53.81** | 0 |
| D4S428 | **64.24** | 0.66 | D5S2022 | **58.55** | 1.25 | D6S2427 | **53.81** | 0 |
| D4S1518 | **68.94** | 4.7 | D5S2092 | **60.92** | 2.37 | D6S1549 | **61.86** | 8.05 |
| D4S3000 | **68.94** | 0 | D5S474 | **63.6** | 2.68 | D6S1017 | **63.28** | 1.42 |
| D4S1569 | **71.77** | 2.83 | D5S407 | **64.67** | 1.07 | D6S400 | **63.28** | 0 |
| D4S3004 | **75.2** | 3.43 | D5S491 | **66.81** | 2.14 | D6S1582 | **65.14** | 1.86 |
| D4S1541 | **75.2** | 0 | D5S2102 | **66.81** | 0 | D6S282 | **66.37** | 1.23 |
| D4S2931 | **78.97** | 3.77 | D5S398 | **68.03** | 1.22 | D6S1650 | **68.65** | 2.28 |
| D4S2969 | **78.97** | 0 | D5S1956 | **70.44** | 2.41 | D6S1669 | **73.13** | 4.48 |
| D4S1517 | **82.09** | 3.12 | D5S2048 | **73.35** | 2.91 | D6S269 | **73.13** | 0 |
| D4S2990 | **83.29** | 1.2 | D5S2072 | **73.35** | 0 | D6S1960 | **76.62** | 3.49 |
| D4S2363 | **87.06** | 3.77 | D5S647 | **74.07** | 0.72 | D6S466 | **77.78** | 1.16 |
| D4S2964 | **88.35** | 1.29 | D5S2019 | **75.89** | 1.82 | D6S1573 | **77.78** | 0 |
| D4S2932 | **90.28** | 1.93 | D5S1351 | **78.31** | 2.42 | D6S428 | **77.78** | 0 |
| D4S2922 | **90.28** | 0 | D5S2122 | **78.31** | 0 | D6S1628 | **80.99** | 3.21 |
| D4S2361 | **93.48** | 3.2 | D5S424 | **81.95** | 3.64 | D6S421 | **84.15** | 3.16 |
| D4S2460 | **97.3** | 3.82 | D5S672 | **86.26** | 4.31 | D6S1625 | **89.23** | 5.08 |
| D4S1544 | **100.06** | 2.76 | D5S2067 | **92.38** | 6.12 | D6S460 | **89.83** | 0.6 |
| D4S3037 | **100.75** | 0.69 | D5S2094 | **93.59** | 1.21 | D6S1707 | **90.43** | 0.6 |
| D4S2909 | **101.55** | 0.8 | D5S1726 | **94.8** | 1.21 | D6S1634 | **92.25** | 1.82 |
| D4S1559 | **104.75** | 3.2 | D5S428 | **95.4** | 0.6 | D6S1601 | **92.85** | 0.6 |
| D4S2986 | **105.29** | 0.54 | D5S459 | **97.21** | 1.81 | D6S1644 | **96.05** | 3.2 |
| D4S2961 | **107.95** | 2.66 | D5S401 | **97.82** | 0.61 | D6S462 | **99.01** | 2.96 |
| D4S1570 | **109.02** | 1.07 | D5S1452 | **99.42** | 1.6 | D6S1570 | **99.01** | 0 |
| D4S3026 | **109.02** | 0 | D5S815 | **101.02** | 1.6 | D6S450 | **100.91** | 1.9 |
| D4S411 | **109.02** | 0 | D5S484 | **105.29** | 4.27 | D6S458 | **100.91** | 0 |
| D4S1571 | **114.67** | 5.65 | D5S652 | **105.29** | 0 | D6S424 | **104.08** | 3.17 |
| D4S2945 | **116.37** | 1.7 | D5S2079 | **108.07** | 2.78 | D6S1717 | **107.25** | 3.17 |
| D4S406 | **117.06** | 0.69 | D5S433 | **111.97** | 3.9 | D6S468 | **107.88** | 0.63 |

| chr4 | cM | inter-marker distance (cM) |
|---|---|---|
| D4S2392 | **121.61** | 4.55 |
| D4S191 | **121.61** | 0 |
| D4S1524 | **126.15** | 4.54 |
| D4S430 | **126.15** | 0 |
| D4S1615 | **128.31** | 2.16 |
| D4S2938 | **129.38** | 1.07 |
| D4S1527 | **132.05** | 2.67 |
| D4S175 | **134.74** | 2.69 |
| D4S397 | **140.64** | 5.9 |
| D4S1565 | **143.84** | 3.2 |
| D4S1610 | **145.98** | 2.14 |
| D4S3008 | **152.98** | 7 |
| D4S1588 | **154.63** | 1.65 |
| D4S1549 | **154.63** | 0 |
| D4S3016 | **157.99** | 3.36 |
| D4S2980 | **159.3** | 1.31 |
| D4S3033 | **161.04** | 1.74 |
| D4S3046 | **162.47** | 1.43 |
| D4S2952 | **166.85** | 4.38 |
| D4S349 | **167.55** | 0.7 |
| D4S2910 | **171.39** | 3.84 |
| D4S1545 | **173.59** | 2.2 |
| D4S2977 | **176.19** | 2.6 |
| D4S3030 | **177.33** | 1.14 |
| D4S1529 | **181.64** | 4.31 |
| D4S2417 | **181.93** | 0.29 |
| D4S1537 | **183.63** | 1.7 |
| D4S2956 | **183.63** | 0 |
| D4S1607 | **183.63** | 0 |
| D4S1584 | **184.2** | 0.57 |
| D4S3015 | **185.28** | 1.08 |
| D4S1530 | **185.28** | 0 |
| D4S2951 | **185.28** | 0 |
| D4S3041 | **187.64** | 2.36 |
| D4S2920 | **190.02** | 2.38 |
| D4S2943 | **190.02** | 0 |
| D4S2954 | **193.74** | 3.72 |
| D4S171 | **199.93** | 6.19 |
| D4S1540 | **199.93** | 0 |
| D4S3051 | **203.77** | 3.84 |
| MSX1 | **?** | |
| D1S503 | **?** | |
| DRD5 | **?** | |
| D4S2614 | **?** | |
| 259 | **?** | |
| D4S622 | **?** | |
| | max | **13.55** |
| | ave | **2.01** |

| chr5 | cM | inter-marker distance (cM) |
|---|---|---|
| D5S505 | **111.97** | 0 |
| D5S475 | **114.75** | 2.78 |
| D5S2051 | **116.98** | 2.23 |
| D5S489 | **123.45** | 6.47 |
| D5S404 | **127.93** | 4.48 |
| D5S471 | **129.83** | 1.9 |
| D5S622 | **130.38** | 0.55 |
| D5S467 | **131.48** | 1.1 |
| D5S809 | **133.65** | 2.17 |
| D5S490 | **133.65** | 0 |
| D5S808 | **137.46** | 3.81 |
| D5S1995 | **137.95** | 0.49 |
| D5S458 | **139.33** | 1.38 |
| D5S479 | **141.27** | 1.94 |
| D5S1983 | **141.82** | 0.55 |
| D5S2116 | **142.92** | 1.1 |
| D5S2119 | **144.06** | 1.14 |
| D5S402 | **147.49** | 3.43 |
| D5S638 | **148.63** | 1.14 |
| D5S413 | **150.34** | 1.71 |
| D5S434 | **150.34** | 0 |
| D5S640 | **152.62** | 2.28 |
| D5S2014 | **153.17** | 0.55 |
| D5S2077 | **155.92** | 2.75 |
| D5S670 | **156.47** | 0.55 |
| D5S497 | **157.02** | 0.55 |
| D5S1499 | **159.22** | 2.2 |
| D5S2049 | **160.87** | 1.65 |
| D5S403 | **162.47** | 1.6 |
| D5S422 | **164.19** | 1.72 |
| D5S2066 | **165.13** | 0.94 |
| D5S621 | **167.69** | 2.56 |
| D5S2050 | **171.06** | 3.37 |
| D5S1458 | **172.67** | 1.61 |
| D5S619 | **172.67** | 0 |
| D5S504 | **175.34** | 2.67 |
| D5S625 | **177.06** | 1.72 |
| D5S614 | **177.92** | 0.86 |
| D5S462 | **178.57** | 0.65 |
| D5S2069 | **182.35** | 3.78 |
| D5S2108 | **182.35** | 0 |
| D5S498 | **184.66** | 2.31 |
| D5S2034 | **187.81** | 3.15 |
| D5S469 | **189.23** | 1.42 |
| D5S2008 | **190.18** | 0.95 |
| D5S2073 | **194.88** | 4.7 |
| D5S408 | **195.49** | 0.61 |
| D5S2846 | **?** | |
| D5S1468 | **?** | |
| D5S435 | **?** | |
| D5S204 | **?** | |
| D5S253 | **?** | |
| | max | **6.47** |
| | ave | **1.83** |

| chr6 | cM | inter-marker distance (cM) |
|---|---|---|
| D6S1555 | **107.88** | 0 |
| D6S434 | **109.19** | 1.31 |
| D6S1580 | **112.2** | 3.01 |
| D6S1546 | **113.03** | 0.83 |
| D6S1664 | **113.03** | 0 |
| D6S447 | **113.61** | 0.58 |
| D6S278 | **116.26** | 2.65 |
| D6S1698 | **118.08** | 1.82 |
| D6S404 | **118.64** | 0.56 |
| D6S287 | **121.97** | 3.33 |
| D6S1608 | **122.51** | 0.54 |
| D6S1712 | **122.51** | 0 |
| D6S408 | **125.71** | 3.2 |
| D6S1690 | **126.52** | 0.81 |
| D6S1620 | **128.93** | 2.41 |
| D6S1572 | **129.46** | 0.53 |
| D6S457 | **130** | 0.54 |
| D6S262 | **130** | 0 |
| D6S413 | **131.07** | 1.07 |
| D6S1656 | **131.07** | 0 |
| D6S975 | **133.18** | 2.11 |
| D6S87 | **133.18** | 0 |
| D6S270 | **135.47** | 2.29 |
| D6S1626 | **136.97** | 1.5 |
| D6S1699 | **141.15** | 4.18 |
| D6S314 | **143.4** | 2.25 |
| D6S1704 | **144.46** | 1.06 |
| D6S1703 | **146.06** | 1.6 |
| D6S1637 | **147.13** | 1.07 |
| D6S1564 | **149.13** | 2 |
| D6S1687 | **153.04** | 3.91 |
| D6S290 | **154.1** | 1.06 |
| D6S425 | **155.17** | 1.07 |
| D6S442 | **157.31** | 2.14 |
| D6S1708 | **157.84** | 0.53 |
| D6S1612 | **159.98** | 2.14 |
| D6S1581 | **164.78** | 4.8 |
| D6S305 | **166.39** | 1.61 |
| D6S411 | **166.39** | 0 |
| D6S1579 | **166.39** | 0 |
| D6S1277 | **173.31** | 6.92 |
| D6S1719 | **177.88** | 4.57 |
| D6S264 | **179.07** | 1.19 |
| D6S297 | **182.11** | 3.04 |
| D6S1697 | **182.11** | 0 |
| D6S1027 | **187.23** | 5.12 |
| D6S446 | **189** | 1.77 |
| D6S281 | **190.14** | 1.14 |
| D6S1279 | **?** | |
| D6S2417 | **?** | |
| D6S506 | **?** | |
| D6S106 | **?** | |
| D6S1955 | **?** | |
| D6S1266 | **?** | |
| | max | **9.18** |
| | ave | **1.78** |

| chr7 | cM | inter-marker distance (cM) | chr8 | cM | inter-marker distance (cM) | chr9 | cM | inter-marker distance (cM) |
|---|---|---|---|---|---|---|---|---|
| D7S2477 | **0** | | D8S264 | **0.73** | | D9S1779 | **0** | |
| D7S2521 | **5.28** | 5.28 | D8S1824 | **2.92** | 2.19 | D9S1873 | **9.83** | 9.83 |
| D7S511 | **7.44** | 2.16 | D8S1806 | **3.83** | 0.91 | D9S1852 | **14.78** | 4.95 |
| D7S3056 | **7.44** | 0 | D8S518 | **5.63** | 1.8 | D9S1686 | **14.78** | 0 |
| D7S2553 | **13.37** | 5.93 | D8S561 | **8.34** | 2.71 | D9S168 | **21.88** | 7.1 |
| D7S513 | **17.74** | 4.37 | D8S277 | **8.34** | 0 | D9S269 | **24.07** | 2.19 |
| D7S2464 | **17.74** | 0 | D8S1819 | **9.96** | 1.62 | D9S268 | **26.24** | 2.17 |
| D7S2557 | **23.29** | 5.55 | D8S516 | **17** | 7.04 | D9S267 | **26.24** | 0 |
| D7S507 | **28.74** | 5.45 | D8S1721 | **17** | 0 | D9S235 | **27.32** | 1.08 |
| D7S2495 | **29.28** | 0.54 | D8S520 | **20.61** | 3.61 | D9S1782 | **30.61** | 3.29 |
| D7S2532 | **29.28** | 0 | D8S1695 | **21.87** | 1.26 | D9S1839 | **30.61** | 0 |
| D7S2551 | **30.9** | 1.62 | D8S1754 | **27.4** | 5.53 | GATA124D09 | **32.24** | 1.63 |
| D7S2535 | **33.05** | 2.15 | D8S549 | **31.73** | 4.33 | D9S162 | **34.42** | 2.18 |
| D7S493 | **34.69** | 1.64 | LPL | **39.25** | 7.52 | D9S171 | **42.73** | 8.31 |
| D7S2458 | **36.03** | 1.34 | D8S258 | **41.55** | 2.3 | D9S1679 | **44.28** | 1.55 |
| D7S2463 | **38.48** | 2.45 | D8S1116 | **42.85** | 1.3 | D9S265 | **44.28** | 0 |
| D7S2534 | **40.62** | 2.14 | D8S560 | **43.41** | 0.56 | D9S761 | **51.21** | 6.93 |
| D7S2440 | **41.69** | 1.07 | D8S298 | **43.96** | 0.55 | D9S161 | **51.81** | 0.6 |
| D7S516 | **41.69** | 0 | D8S1733 | **45.41** | 1.45 | D9S263 | **51.81** | 0 |
| D7S2515 | **43.84** | 2.15 | D8S1739 | **48.79** | 3.38 | D9S248 | **54.5** | 2.69 |
| D7S2496 | **47.08** | 3.24 | D8S1725 | **50.05** | 1.26 | D9S165 | **58.26** | 3.76 |
| D7S2492 | **48.15** | 1.07 | D8S1048 | **54.28** | 4.23 | D9S1817 | **59.34** | 1.08 |
| D7S2491 | **48.69** | 0.54 | D8S585 | **54.98** | 0.7 | D9S1859 | **59.87** | 0.53 |
| D7S2252 | **50.29** | 1.6 | D8S1477 | **60.34** | 5.36 | D9S1800 | **64.72** | 4.85 |
| D7S656 | **52.7** | 2.41 | D8S535 | **60.87** | 0.53 | D9S1806 | **66.32** | 1.6 |
| D7S484 | **53.5** | 0.8 | D8S540 | **60.87** | 0 | D9S226 | **67.93** | 1.61 |
| D7S497 | **53.5** | 0 | D8S255 | **64.6** | 3.73 | D9S284 | **70.33** | 2.4 |
| D7S528 | **57.79** | 4.29 | D8S1817 | **65.13** | 0.53 | D9S175 | **70.33** | 0 |
| D7S2507 | **59.39** | 1.6 | D8S1831 | **67.27** | 2.14 | D9S1860 | **73.03** | 2.7 |
| D7S485 | **59.93** | 0.54 | D8S165 | **69.4** | 2.13 | D9S1807 | **73.03** | 0 |
| D7S2524 | **59.93** | 0 | D8S285 | **71** | 1.6 | D9S1123 | **77.56** | 4.53 |
| D7S2541 | **61.53** | 1.6 | D8S1113 | **77.89** | 6.89 | D9S1843 | **80.31** | 2.75 |
| D7S521 | **62.07** | 0.54 | D8S260 | **79.36** | 1.47 | D9S781 | **80.31** | 0 |
| D7S671 | **62.87** | 0.8 | D8S510 | **79.94** | 0.58 | D9S245 | **80.31** | 0 |
| D7S2428 | **65.55** | 2.68 | D8S1748 | **81.68** | 1.74 | D9S167 | **83.41** | 3.1 |
| D7S2528 | **67.43** | 1.88 | D8S512 | **81.68** | 0 | D9S776 | **87.29** | 3.88 |
| D7S2436 | **67.96** | 0.53 | D8S1840 | **82.84** | 1.16 | D9S1865 | **87.29** | 0 |
| D7S1508 | **69.56** | 1.6 | D8S1767 | **83.51** | 0.67 | D9S1680 | **90.4** | 3.11 |
| D7S2467 | **73.31** | 3.75 | D8S543 | **87.54** | 4.03 | D9S1812 | **90.4** | 0 |
| D7S2542 | **74.91** | 1.6 | D8S1807 | **90.33** | 2.79 | D9S283 | **94.85** | 4.45 |
| D7S659 | **75.98** | 1.07 | D8S1776 | **92.58** | 2.25 | D9S1820 | **95.65** | 0.8 |
| D7S2549 | **77.91** | 1.93 | D8S1829 | **94.61** | 2.03 | D9S1836 | **96.99** | 1.34 |
| D7S2530 | **77.91** | 0 | D8S501 | **95.15** | 0.54 | D9S1796 | **97.53** | 0.54 |
| D7S2489 | **78.65** | 0.74 | D8S1736 | **97.28** | 2.13 | D9S1781 | **99.4** | 1.87 |
| D7S2483 | **78.65** | 0 | D8S1707 | **101.69** | 4.41 | D9S196 | **101.27** | 1.87 |
| D7S502 | **78.65** | 0 | D8S273 | **102.62** | 0.93 | D9S1851 | **103.42** | 2.15 |
| D7S645 | **80.42** | 1.77 | D8S1724 | **103.69** | 1.07 | D9S1809 | **103.42** | 0 |
| D7S2435 | **80.42** | 0 | D8S270 | **103.69** | 0 | D9S910 | **104.48** | 1.06 |
| D7S2476 | **84.52** | 4.1 | D8S1772 | **107.97** | 4.28 | D9S1690 | **106.63** | 2.15 |
| D7S1870 | **86.12** | 1.6 | D8S546 | **113.16** | 5.19 | D9S1866 | **110.91** | 4.28 |
| D7S2490 | **87.38** | 1.26 | D8S521 | **116.14** | 2.98 | D9S172 | **111.99** | 1.08 |
| D7S2470 | **89.88** | 2.5 | D8S1834 | **117.62** | 1.48 | D9S2026 | **117.37** | 5.38 |
| D7S2499 | **91.67** | 1.79 | D8S1714 | **117.62** | 0 | D9S261 | **117.37** | 0 |
| D7S634 | **93.1** | 1.43 | D8S1784 | **118.15** | 0.53 | D9S105 | **120.04** | 2.67 |
| D7S660 | **93.63** | 0.53 | D8S200 | **120.29** | 2.14 | D9S930 | **120.04** | 0 |
| D7S2212 | **95.43** | 1.8 | D8S539 | **123.54** | 3.25 | D9S1683 | **120.04** | 0 |
| D7S2417 | **97.38** | 1.95 | D8S1015 | **123.54** | 0 | D9S51 | **123.33** | 3.29 |
| D7S524 | **97.38** | 0 | D8S555 | **123.54** | 0 | D9S1776 | **123.33** | 0 |
| D7S630 | **98.44** | 1.06 | D8S1694 | **125.27** | 1.73 | D9S1811 | **126.41** | 3.08 |
| D7S492 | **100.05** | 1.61 | D8S1726 | **130** | 4.73 | D9S754 | **126.41** | 0 |

| chr7 | cM | inter-marker distance (cM) | chr8 | cM | inter-marker distance (cM) | chr9 | cM | inter-marker distance (cM) |
|---|---|---|---|---|---|---|---|---|
| D7S657 | **104.86** | 4.81 | D8S1804 | **131.69** | 1.69 | D9S275 | **127.98** | 1.57 |
| D7S2430 | **105.39** | 0.53 | D8S1832 | **132.49** | 0.8 | D9S258 | **130.52** | 2.54 |
| COL1A2 | **106.46** | 1.07 | D8S266 | **136.5** | 4.01 | D9S1840 | **136.47** | 5.95 |
| D7S2482 | **108.59** | 2.13 | D8S508 | **137.92** | 1.42 | D9S904 | **136.47** | 0 |
| D7S491 | **109.12** | 0.53 | D8S1128 | **139.53** | 1.61 | D9S290 | **140.86** | 4.39 |
| D7S1522 | **109.66** | 0.54 | D8S263 | **141.68** | 2.15 | D9S260 | **140.86** | 0 |
| D7S2498 | **111.26** | 1.6 | D8S1765 | **144.36** | 2.68 | D9S1795 | **142.51** | 1.65 |
| D7S651 | **111.26** | 0 | D8S557 | **145.97** | 1.61 | D9S1830 | **145.21** | 2.7 |
| D7S477 | **111.79** | 0.53 | D8S1835 | **145.97** | 0 | D9S164 | **147.91** | 2.7 |
| D7S515 | **112.32** | 0.53 | D8S1740 | **147.04** | 1.07 | D9S1826 | **159.61** | 11.7 |
| D7S658 | **113.92** | 1.6 | D8S1746 | **149.46** | 2.42 | D9S2156 | **?** | |
| D7S2453 | **115.96** | 2.04 | D8S1710 | **150.8** | 1.34 | D9S924 | **?** | |
| D7S2459 | **119.81** | 3.85 | D8S272 | **154.02** | 3.22 | D9S1112 | **?** | |
| D7S2456 | **120.61** | 0.8 | D8S1837 | **156.59** | 2.57 | D9S2154 | **?** | |
| D7S2418 | **122.48** | 1.87 | D8S1741 | **162.94** | 6.35 | D9S147E | **?** | |
| D7S2543 | **124.08** | 1.6 | D8S1704 | **164.25** | 1.31 | D9S906 | **?** | |
| D7S643 | **125.15** | 1.07 | D8S1713 | **164.38** | 0.13 | D9S298 | **?** | |
| D7S648 | **128.41** | 3.26 | D8S1729 | **165.93** | 1.55 | D9S55 | **?** | |
| D7S1874 | **129.58** | 1.17 | D8S1926 | **167.9** | 1.97 | D9S748 | **?** | |
| D7S514 | **130.17** | 0.59 | D8S1989 | **?** | | D9S199 | **?** | |
| D7S2501 | **130.76** | 0.59 | D8S1462 | **?** | | D9S967 | **?** | |
| D7S530 | **134.55** | 3.79 | D8S2332 | **?** | | D9S123 | **?** | |
| D7S2531 | **136.06** | 1.51 | D8S386 | **?** | | D9S935 | **?** | |
| D7S640 | **137.83** | 1.77 | CRH | **?** | | | max | **11.70** |
| D7S2533 | **138.42** | 0.59 | D8S2330 | **?** | | | ave | **2.31** |
| D7S2452 | **138.42** | 0 | | max | **7.52** | | | |
| D7S800 | **140.63** | 2.21 | | ave | **2.14** | | | |
| D7S509 | **143.33** | 2.7 | | | | | | |
| D7S495 | **144.72** | 1.39 | | | | | | |
| D7S2505 | **148.11** | 3.39 | | | | | | |
| D7S1824 | **149.9** | 1.79 | | | | | | |
| D7S661 | **155.1** | 5.2 | | | | | | |
| D7S1486 | **155.79** | 0.69 | | | | | | |
| D7S498 | **155.79** | 0 | | | | | | |
| D7S2511 | **156.33** | 0.54 | | | | | | |
| D7S2442 | **159.53** | 3.2 | | | | | | |
| D7S636 | **162.33** | 2.8 | | | | | | |
| D7S642 | **162.33** | 0 | | | | | | |
| D7S2439 | **163.74** | 1.41 | | | | | | |
| D7S1523 | **165.18** | 1.44 | | | | | | |
| D7S798 | **168.98** | 3.8 | | | | | | |
| D7S2546 | **173.03** | 4.05 | | | | | | |
| D7S2423 | **181.97** | 8.94 | | | | | | |
| D7S1492 | **?** | | | | | | | |
| D7S3050 | **?** | | | | | | | |
| D7S1491 | **?** | | | | | | | |
| D7S3048 | **?** | | | | | | | |
| D7S794 | **?** | | | | | | | |
| D7S3084 | **?** | | | | | | | |
| D7S3069 | **?** | | | | | | | |
| D7S3068 | **?** | | | | | | | |
| D7S804 | **?** | | | | | | | |
| D7S1815 | **?** | | | | | | | |
| | max | **8.94** | | | | | | |
| | ave | **1.78** | | | | | | |

| chr10 | cM | inter-marker distance (cM) | chr11 | cM | inter-marker distance (cM) | chr12 | cM | inter-marker distance (cM) |
|---|---|---|---|---|---|---|---|---|
| D10S249 | **2.13** | | D11S1758 | **7.03** | | D12S352 | **0** | |
| D10S1706 | **5.21** | 3.08 | D11S2362 | **8.9** | 1.87 | D12S91 | **1.71** | 1.71 |
| D10S591 | **13.49** | 8.28 | D11S1760 | **11.67** | 2.77 | D12S1656 | **2.79** | 1.08 |
| D10S552 | **14.33** | 0.84 | D11S1331 | **12.92** | 1.25 | D12S100 | **3.33** | 0.54 |
| D10S189 | **19** | 4.67 | D11S1338 | **12.92** | 0 | D12S1689 | **3.33** | 0 |
| D10S547 | **29.15** | 10.15 | D11S926 | **20.4** | 7.48 | D12S1615 | **4.64** | 1.31 |
| D10S585 | **30** | 0.85 | D11S2368 | **22.56** | 2.16 | D12S372 | **6.42** | 1.78 |
| D10S527 | **33.48** | 3.48 | D11S928 | **26.34** | 3.78 | D12S1626 | **7.12** | 0.7 |
| D10S1721 | **33.48** | 0 | D11S1359 | **29.18** | 2.84 | D12S1725 | **9.52** | 2.4 |
| D10S1725 | **36.06** | 2.58 | D11S1750 | **33.02** | 3.84 | D12S314 | **11.37** | 1.85 |
| D10S1653 | **40.36** | 4.3 | D11S904 | **33.57** | 0.55 | D12S99 | **12.6** | 1.23 |
| D10S1476 | **42.5** | 2.14 | D11S1392 | **43.16** | 9.59 | D12S356 | **14.23** | 1.63 |
| D10S211 | **48.36** | 5.86 | D11S935 | **45.94** | 2.78 | D12S336 | **19.68** | 5.45 |
| D10S197 | **52.1** | 3.74 | D11S4185 | **45.94** | 0 | D12S1695 | **19.68** | 0 |
| D10S572 | **52.1** | 0 | D11S1279 | **50.88** | 4.94 | D12S1690 | **20.27** | 0.59 |
| D10S1641 | **54.23** | 2.13 | D11S1763 | **53.56** | 2.68 | D12S89 | **23.41** | 3.14 |
| D10S588 | **54.23** | 0 | D11S903 | **54.75** | 1.19 | D12S98 | **24.45** | 1.04 |
| D10S1684 | **57.42** | 3.19 | D11S1777 | **58.4** | 3.65 | D12S1581 | **29.73** | 5.28 |
| D10S193 | **59.03** | 1.61 | D11S1335 | **58.4** | 0 | D12S364 | **30.6** | 0.87 |
| D10S208 | **60.64** | 1.61 | D11S1765 | **61.78** | 3.38 | D12S308 | **31.22** | 0.62 |
| D10S1654 | **62.23** | 1.59 | D11S913 | **67.48** | 5.7 | D12S1303 | **32.48** | 1.26 |
| D10S1791 | **65.97** | 3.74 | D11S987 | **67.48** | 0 | D12S1728 | **34.85** | 2.37 |
| D10S578 | **65.97** | 0 | D11S4184 | **75.3** | 7.82 | D12S310 | **36.06** | 1.21 |
| D10S1787 | **68.63** | 2.66 | D11S916 | **76.13** | 0.83 | D12S1650 | **38.5** | 2.44 |
| D10S225 | **70.06** | 1.43 | D11S1789 | **79.98** | 3.85 | D12S1682 | **38.5** | 0 |
| D10S196 | **70.23** | 0.17 | D11S937 | **79.98** | 0 | D12S1654 | **41.55** | 3.05 |
| D10S538 | **70.23** | 0 | D11S901 | **85.48** | 5.5 | D12S1688 | **42.1** | 0.55 |
| D10S539 | **72.9** | 2.67 | D11S4135 | **86.98** | 1.5 | D12S1591 | **43.38** | 1.28 |
| D10S546 | **75.57** | 2.67 | D11S1354 | **87.89** | 0.91 | D12S1617 | **44.03** | 0.65 |
| D10S1762 | **75.57** | 0 | D11S4082 | **89.69** | 1.8 | D12S1596 | **45.95** | 1.92 |
| D10S464 | **79.03** | 3.46 | D11S1780 | **90.29** | 0.6 | D12S1643 | **50.9** | 4.95 |
| D10S549 | **79.03** | 0 | D11S4175 | **91.47** | 1.18 | D12S1584 | **51.99** | 1.09 |
| D10S1225 | **80.77** | 1.74 | D11S1332 | **93.12** | 1.65 | D12S331 | **54.46** | 2.47 |
| D10S1652 | **80.77** | 0 | D11S898 | **98.98** | 5.86 | D12S1592 | **56.38** | 1.92 |
| D10S1640 | **80.77** | 0 | D11S1886 | **100.62** | 1.64 | D12S85 | **61.34** | 4.96 |
| D10S1743 | **82.5** | 1.73 | D11S2017 | **101.75** | 1.13 | D12S1713 | **61.34** | 0 |
| D10S1646 | **86.2** | 3.7 | D11S1391 | **104.6** | 2.85 | D12S1661 | **63.89** | 2.55 |
| D10S1647 | **88.41** | 2.21 | DRD2 | **105.17** | 0.57 | D12S1627 | **64.43** | 0.54 |
| D10S537 | **91.13** | 2.72 | D11S4206 | **105.17** | 0 | D12S1629 | **65.49** | 1.06 |
| D10S188 | **93.92** | 2.79 | D11S908 | **108.59** | 3.42 | D12S368 | **66.03** | 0.54 |
| D10S1136 | **97.29** | 3.37 | D11S4129 | **115.53** | 6.94 | D12S1622 | **69.82** | 3.79 |
| D10S202 | **98.41** | 1.12 | D11S925 | **118.47** | 2.94 | D12S90 | **71.61** | 1.79 |
| D10S573 | **106.11** | 7.7 | D11S1353 | **122.47** | 4 | D12S1632 | **71.61** | 0 |
| D10S1735 | **109.33** | 3.22 | D11S4094 | **123** | 0.53 | D12S1644 | **72.2** | 0.59 |
| D10S1753 | **112.58** | 3.25 | D11S933 | **124.07** | 1.07 | D12S305 | **74.58** | 2.38 |
| D10S1755 | **114.19** | 1.61 | D11S990 | **126.21** | 2.14 | D12S83 | **75.17** | 0.59 |
| D10S185 | **116.34** | 2.15 | D11S4158 | **126.21** | 0 | D12S1655 | **75.17** | 0 |
| D10S1758 | **118.94** | 2.6 | D11S4151 | **127.33** | 1.12 | D12S1601 | **75.17** | 0 |
| D10S1726 | **121.98** | 3.04 | D11S4110 | **129.02** | 1.69 | D12S1585 | **75.76** | 0.59 |
| D10S597 | **128.73** | 6.75 | D11S2367 | **138.56** | 9.54 | D12S1676 | **78.74** | 2.98 |
| D10S1693 | **137.39** | 8.66 | D11S969 | **146.6** | 8.04 | D12S313 | **79.93** | 1.19 |
| D10S542 | **138.47** | 1.08 | D11S968 | **147.77** | 1.17 | D12S1680 | **80.52** | 0.59 |
| D10S587 | **147.57** | 9.1 | D11S4125 | **147.77** | 0 | D12S92 | **83.19** | 2.67 |
| D10S217 | **157.89** | 10.32 | HBB | **?** | | D12S80 | **83.19** | 0 |
| D10S1676 | **160.04** | 2.15 | D11S4967 | **?** | | D12S326 | **86.4** | 3.21 |
| D10S1655 | **162.38** | 2.34 | D11S2018 | **?** | | D12S1709 | **86.4** | 0 |
| D10S1651 | **168.77** | 6.39 | FCER1B | **?** | | D12S106 | **90.77** | 4.37 |
| D10S1675 | **170.4** | 1.63 | D11S2008 | **?** | | D12S1719 | **94.49** | 3.72 |
| D10S590 | **170.4** | 0 | D11S490 | **?** | | D12S316 | **94.49** | 0 |
| D10S212 | **170.94** | 0.54 | D11S4966 | **?** | | D12S1345 | **96.09** | 1.6 |

| chr10 | cM | inter-marker distance (cM) |
|---|---|---|
| D10S179 | ? | |
| D10S1257 | ? | |
| D10S1187 | ? | |
| D10S1555 | ? | |
| D10S71 | ? | |
| | max | **10.32** |
| | ave | **2.86** |

| chr11 | cM | inter-marker distance (cM) |
|---|---|---|
| D11S863 | ? | |
| D11S2011 | ? | |
| | max | **9.59** |
| | ave | **2.71** |

| chr12 | cM | inter-marker distance (cM) |
|---|---|---|
| D12S1346 | **97.16** | 1.07 |
| D12S348 | **100.92** | 3.76 |
| D12S101 | **100.92** | 0 |
| D12S302 | **101.98** | 1.06 |
| D12S332 | **105.18** | 3.2 |
| D12S1727 | **107.19** | 2.01 |
| D12S1607 | **107.86** | 0.67 |
| D12S360 | **111.27** | 3.41 |
| D12S338 | **111.87** | 0.6 |
| D12S317 | **114.28** | 2.41 |
| D12S353 | **115.18** | 0.9 |
| D12S1613 | **116.08** | 0.9 |
| D12S1605 | **116.66** | 0.58 |
| D12S1339 | **118.68** | 2.02 |
| D12S105 | **118.68** | 0 |
| D12S1616 | **119.55** | 0.87 |
| D12S1665 | **125.31** | 5.76 |
| D12S1602 | **125.31** | 0 |
| D12S79 | **125.31** | 0 |
| D12S1718 | **128.05** | 2.74 |
| D12S366 | **133.33** | 5.28 |
| D12S349 | **134.54** | 1.21 |
| D12S76 | **136.82** | 2.28 |
| D12S1666 | **136.82** | 0 |
| D12S395 | **136.82** | 0 |
| D12S321 | **136.82** | 0 |
| D12S1603 | **140.17** | 3.35 |
| D12S1614 | **144.83** | 4.66 |
| D12S340 | **146.39** | 1.56 |
| D12S324 | **147.17** | 0.78 |
| D12S1658 | **148.24** | 1.07 |
| D12S1675 | **150.7** | 2.46 |
| D12S1679 | **153.19** | 2.49 |
| D12S1609 | **153.33** | 0.14 |
| D12S1659 | **155.94** | 2.61 |
| D12S60 | **159.59** | 3.65 |
| D12S1714 | **159.59** | 0 |
| D12S1045 | **160.68** | 1.09 |
| D12S97 | **160.68** | 0 |
| D12S343 | **163.55** | 2.87 |
| D12S1599 | **164.63** | 1.08 |
| D12S1628 | **165.69** | 1.06 |
| D12S1638 | **168.79** | 3.1 |
| D12S70 | ? | |
| D12S2194 | ? | |
| D12S833 | ? | |
| D12S2196 | ? | |
| | max | **5.76** |
| | ave | **1.65** |

| chr13 | cM | inter-marker distance (cM) | chr14 | cM | inter-marker distance (cM) | chr15 | cM | inter-marker distance (cM) |
|---|---|---|---|---|---|---|---|---|
| D13S1236 | **2.77** | | D14S132 | **0** | | D15S128 | **6.11** | |
| D13S175 | **6.03** | 3.26 | D14S261 | **6.46** | 6.46 | D15S97 | **9.85** | 3.74 |
| D13S250 | **6.03** | 0 | D14S50 | **12.46** | 6 | D15S156 | **14.58** | 4.73 |
| D13S232 | **6.99** | 0.96 | D14S283 | **13.89** | 1.43 | D15S1002 | **14.58** | 0 |
| D13S292 | **8.87** | 1.88 | D14S972 | **21.51** | 7.62 | D15S1019 | **19.12** | 4.54 |
| D13S1243 | **9.79** | 0.92 | D14S64 | **22.66** | 1.15 | D15S165 | **20.24** | 1.12 |
| D13S283 | **11.45** | 1.66 | D14S1032 | **23.2** | 0.54 | D15S231 | **24.06** | 3.82 |
| D13S1285 | **12.91** | 1.46 | D14S597 | **28.01** | 4.81 | D15S144 | **25.3** | 1.24 |
| D13S1294 | **12.91** | 0 | D14S1042 | **28.01** | 0 | D15S1007 | **25.86** | 0.56 |
| D13S1244 | **15.19** | 2.28 | D14S615 | **28.01** | 0 | D15S1040 | **28.35** | 2.49 |
| D13S625 | **16.2** | 1.01 | D14S1040 | **31.75** | 3.74 | D15S971 | **31.46** | 3.11 |
| D13S217 | **17.21** | 1.01 | D14S1060 | **34.43** | 2.68 | D15S118 | **32.58** | 1.12 |
| D13S1250 | **17.21** | 0 | D14S741 | **36.76** | 2.33 | D15S1012 | **35.95** | 3.37 |
| D13S1242 | **17.21** | 0 | D14S70 | **40.11** | 3.35 | D15S978 | **45.62** | 9.67 |
| D13S629 | **19.36** | 2.15 | D14S1049 | **40.94** | 0.83 | D15S1003 | **47.85** | 2.23 |
| D13S1299 | **19.36** | 0 | D14S253 | **42.81** | 1.87 | D15S962 | **47.85** | 0 |
| D13S1246 | **20.44** | 1.08 | D14S69 | **45.12** | 2.31 | D15S117 | **51.21** | 3.36 |
| D13S1287 | **21.51** | 1.07 | D14S278 | **45.12** | 0 | D15S987 | **59.05** | 7.84 |
| D13S289 | **21.51** | 0 | D14S266 | **47.51** | 2.39 | D15S993 | **59.61** | 0.56 |
| D13S260 | **23.65** | 2.14 | D14S288 | **47.51** | 0 | D15S1009 | **61.28** | 1.67 |
| D13S171 | **25.08** | 1.43 | D14S976 | **50.5** | 2.99 | D15S651 | **61.84** | 0.56 |
| D13S267 | **26.87** | 1.79 | D14S1068 | **50.5** | 0 | D15S213 | **62.4** | 0.56 |
| D13S1293 | **26.87** | 0 | D14S1031 | **52.49** | 1.99 | D15S153 | **62.4** | 0 |
| D13S305 | **28.04** | 1.17 | D14S991 | **55.82** | 3.33 | D15S1025 | **69.46** | 7.06 |
| D13S219 | **28.87** | 0.83 | D14S276 | **56.36** | 0.54 | D15S650 | **70.73** | 1.27 |
| D13S218 | **32.9** | 4.03 | D14S1064 | **57.43** | 1.07 | D15S215 | **72.94** | 2.21 |
| D13S263 | **38.32** | 5.42 | D14S66 | **59.43** | 2 | D15S1027 | **74.69** | 1.75 |
| D13S1247 | **38.96** | 0.64 | D14S994 | **66.81** | 7.38 | D15S1037 | **75.85** | 1.16 |
| D13S1272 | **41.71** | 2.75 | D14S997 | **67.99** | 1.18 | D15S1041 | **77.69** | 1.84 |
| D13S291 | **41.71** | 0 | D14S63 | **69.18** | 1.19 | D15S205 | **78.92** | 1.23 |
| D13S326 | **42.8** | 1.09 | D14S1069 | **73.03** | 3.85 | D15S1045 | **85.64** | 6.72 |
| D13S287 | **45** | 2.2 | D14S1065 | **74.01** | 0.98 | D15S127 | **86.81** | 1.17 |
| D13S155 | **45.55** | 0.55 | D14S540 | **74.01** | 0 | D15S649 | **98.44** | 11.63 |
| D13S1301 | **47.19** | 1.64 | D14S1011 | **74.96** | 0.95 | D15S1038 | **100.59** | 2.15 |
| D13S233 | **47.19** | 0 | D14S258 | **76.28** | 1.32 | D15S130 | **100.59** | 0 |
| D13S1320 | **49.38** | 2.19 | D14S268 | **79.94** | 3.66 | D15S657 | **104.86** | 4.27 |
| D13S1231 | **51.57** | 2.19 | D14S74 | **87.36** | 7.42 | D15S1014 | **107.71** | 2.85 |
| D13S275 | **52.1** | 0.53 | D14S1022 | **92.69** | 5.33 | D15S212 | **109.29** | 1.58 |
| D13S1324 | **53.17** | 1.07 | D14S140 | **95.89** | 3.2 | D15S985 | **110.94** | 1.65 |
| D13S1257 | **55.31** | 2.14 | D14S256 | **96.42** | 0.53 | D15S966 | **112.58** | 1.64 |
| D13S152 | **55.31** | 0 | D14S1015 | **105** | 8.58 | D15S120 | **112.58** | 0 |
| D13S269 | **56.38** | 1.07 | D14S280 | **105** | 0 | D15S514 | **?** | |
| D13S792 | **57.46** | 1.08 | D14S617 | **105.53** | 0.53 | D15S112 | **?** | |
| D13S1281 | **59.79** | 2.33 | D14S1050 | **107.13** | 1.6 | D15S222 | **?** | |
| D13S160 | **61.17** | 1.38 | D14S749 | **108.22** | 1.09 | D15S172 | **?** | |
| D13S921 | **62.83** | 1.66 | D14S265 | **113.17** | 4.95 | D15S111 | **?** | |
| D13S264 | **64.97** | 2.14 | D14S987 | **114.81** | 1.64 | | max | **11.63** |
| D13S251 | **64.97** | 0 | D14S51 | **115.6** | 0.79 | | ave | **2.66** |
| D13S271 | **64.97** | 0 | D14S979 | **115.91** | 0.31 | | | |
| D13S253 | **66.59** | 1.62 | D14S65 | **117.3** | 1.39 | | | |
| D13S266 | **67.65** | 1.06 | D14S998 | **117.75** | 0.45 | | | |
| D13S265 | **68.73** | 1.08 | D14S78 | **125.88** | 8.13 | | | |
| D13S795 | **71.06** | 2.33 | D14S1426 | **125.88** | 0 | | | |
| D13S1300 | **71.06** | 0 | D14S985 | **126.61** | 0.73 | | | |
| D13S167 | **73.04** | 1.98 | D14S293 | **133.03** | 6.42 | | | |
| D13S1241 | **76.26** | 3.22 | D14S118 | **133.03** | 0 | | | |
| D13S1252 | **77.47** | 1.21 | D14S260 | **134.3** | 1.27 | | | |
| D13S1284 | **79.49** | 2.02 | D14S292 | **134.3** | 0 | | | |
| D13S779 | **82.93** | 3.44 | D14S1007 | **138.18** | 3.88 | | | |
| D13S225 | **83.57** | 0.64 | D14S582 | **?** | | | | |

| chr13 | cM | inter-marker distance (cM) | chr14 | cM | inter-marker distance (cM) |
|---|---|---|---|---|---|
| D13S158 | **84.87** | 1.3 | D14S1278 | **?** | |
| D13S1256 | **84.87** | 0 | D14S750 | **?** | |
| D13S1323 | **84.87** | 0 | D14S1433 | **?** | |
| D13S1311 | **90.27** | 5.4 | | max | **8.58** |
| GATA135E01 | **90.27** | 0 | | ave | **2.38** |
| D13S286 | **93.52** | 3.25 | | | |
| D13S278 | **93.52** | 0 | | | |
| D13S796 | **93.52** | 0 | | | |
| D13S778 | **96.7** | 3.18 | | | |
| D13S1265 | **98.82** | 2.12 | | | |
| D13S1315 | **102.73** | 3.91 | | | |
| D13S1295 | **110.55** | 7.82 | | | |
| D13S285 | **110.55** | 0 | | | |
| D13S293 | **114.98** | 4.43 | | | |
| D13S141 | **?** | | | | |
| D13S127 | **?** | | | | |
| D13S789 | **?** | | | | |
| D13S115 | **?** | | | | |
| D13S138 | **?** | | | | |
| D13S231 | **?** | | | | |
| D13S120 | **?** | | | | |
| D13S894 | **?** | | | | |
| D13S1818 | **?** | | | | |
| D13S802 | **?** | | | | |
| D13S1491 | **?** | | | | |
| | max | **7.82** | | | |
| | ave | **1.54** | | | |

| chr16 | cM | inter-marker distance (cM) |
|---|---|---|
| D16S521 | **1.08** | |
| D16S3124 | **7.05** | 5.97 |
| D16S3065 | **8.16** | 1.11 |
| D16S423 | **10.36** | 2.2 |
| D16S3042 | **13.12** | 2.76 |
| D16S3092 | **13.67** | 0.55 |
| D16S3030 | **13.67** | 0 |
| D16S3108 | **14.77** | 1.1 |
| D16S3087 | **16.97** | 2.2 |
| D16S687 | **18.07** | 1.1 |
| D16S414 | **20.77** | 2.7 |
| D16S519 | **20.77** | 0 |
| D16S2613 | **22.02** | 1.25 |
| D16S748 | **22.65** | 0.63 |
| D16S3060 | **28.3** | 5.65 |
| D16S764 | **29.97** | 1.67 |
| D16S3017 | **32.07** | 2.1 |
| D16S3056 | **35.44** | 3.37 |
| D16S749 | **39.04** | 3.6 |
| D16S3045 | **40.65** | 1.61 |
| D16S417 | **43.89** | 3.24 |
| D16S420 | **44.45** | 0.56 |
| D16S3068 | **48.53** | 4.08 |
| D16S3116 | **50.6** | 2.07 |
| D16S769 | **50.6** | 0 |
| D16S3145 | **52.26** | 1.66 |
| D16S540 | **57.79** | 5.53 |
| D16S536 | **58.46** | 0.67 |
| D16S409 | **58.46** | 0 |
| D16S517 | **58.46** | 0 |
| D16S3080 | **59.68** | 1.22 |
| D16S3136 | **62.11** | 2.43 |
| D16S3396 | **63.78** | 1.67 |
| D16S757 | **65.1** | 1.32 |
| D16S2623 | **66.1** | 1 |
| D16S419 | **67.4** | 1.3 |
| D16S415 | **67.62** | 0.22 |
| D16S3034 | **69.05** | 1.43 |
| D16S771 | **70.69** | 1.64 |
| D16S3253 | **71.77** | 1.08 |
| D16S3039 | **74.44** | 2.67 |
| D16S494 | **77.15** | 2.71 |
| D16S3143 | **81.15** | 4 |
| D16S3129 | **81.95** | 0.8 |
| D16S3111 | **83.55** | 1.6 |
| D16S400 | **83.55** | 0 |
| D16S3095 | **87.06** | 3.51 |
| D16S512 | **89.63** | 2.57 |
| D16S3115 | **92.1** | 2.47 |
| D16S3101 | **93.78** | 1.68 |
| D16S3049 | **97.03** | 3.25 |
| D16S3096 | **99.44** | 2.41 |
| D16S516 | **100.39** | 0.95 |
| D16S3104 | **105.17** | 4.78 |
| D16S3119 | **105.17** | 0 |
| D16S3098 | **108.32** | 3.15 |
| D16S505 | **108.96** | 0.64 |
| D16S3091 | **111.12** | 2.16 |
| D16S422 | **111.12** | 0 |
| D16S3061 | **121.45** | 10.33 |

| chr17 | cM | inter-marker distance (cM) |
|---|---|---|
| D17S1308 | **0.63** | |
| D17S831 | **6.60** | 5.97 |
| D17S829 | **10.02** | 3.42 |
| D17S1828 | **10.02** | 0 |
| D17S1298 | **10.72** | 0.7 |
| D17S796 | **14.69** | 3.97 |
| D17S960 | **15.77** | 1.08 |
| D17S1805 | **16.84** | 1.07 |
| D17S954 | **22.24** | 5.4 |
| D17S1303 | **23.56** | 1.32 |
| D17S799 | **31.96** | 8.4 |
| D17S936 | **31.96** | 0 |
| D17S1856 | **35.55** | 3.59 |
| D17S921 | **36.14** | 0.59 |
| D17S1857 | **43.01** | 6.87 |
| D17S1794 | **47** | 3.99 |
| D17S935 | **50.74** | 3.74 |
| D17S1863 | **50.74** | 0 |
| D17S1800 | **51.63** | 0.89 |
| D17S1293 | **56.48** | 4.85 |
| D17S1842 | **57.71** | 1.23 |
| D17S1846 | **57.71** | 0 |
| D17S1299 | **62.01** | 4.3 |
| D17S1834 | **64.16** | 2.15 |
| D17S1868 | **64.16** | 0 |
| D17S1795 | **68.44** | 4.28 |
| D17S1877 | **74.45** | 6.01 |
| D17S787 | **74.99** | 0.54 |
| D17S1799 | **74.99** | 0 |
| D17S1853 | **80.38** | 5.39 |
| D17S944 | **82.56** | 2.18 |
| D17S942 | **85.94** | 3.38 |
| D17S795 | **89.32** | 3.38 |
| D17S2059 | **93.27** | 3.95 |
| D17S1350 | **93.98** | 0.71 |
| D17S1351 | **95.99** | 2.01 |
| D17S1823 | **97.6** | 1.61 |
| D17S1602 | **99.21** | 1.61 |
| D17S1839 | **102.46** | 3.25 |
| D17S1817 | **103.53** | 1.07 |
| D17S785 | **103.53** | 0 |
| UT952 | **105.68** | 2.15 |
| D17S937 | **105.68** | 0 |
| D17S1847 | **111.22** | 5.54 |
| D17S836 | **112.92** | 1.7 |
| D17S784 | **116.86** | 3.94 |
| D17S1822 | **116.86** | 0 |
| D17S928 | **126.46** | 9.6 |
| D17S516 | **?** | |
| D17S71 | **?** | |
| D17S2195 | **?** | |
| | max | **9.60** |
| | ave | **2.68** |

| chr18 | cM | inter-marker distance (cM) |
|---|---|---|
| D18S59 | **0** | |
| D18S1146 | **2.08** | 2.08 |
| D18S476 | **2.84** | 0.76 |
| D18S481 | **6.94** | 4.1 |
| D18S1138 | **8.3** | 1.36 |
| D18S63 | **8.3** | 0 |
| D18S471 | **18.7** | 10.4 |
| D18S452 | **18.7** | 0 |
| D18S458 | **18.7** | 0 |
| D18S1163 | **24.08** | 5.38 |
| D18S464 | **31.17** | 7.09 |
| D18S1153 | **35.46** | 4.29 |
| D18S1150 | **37.15** | 1.69 |
| D18S1158 | **38.92** | 1.77 |
| D18S1116 | **41.24** | 2.32 |
| D18S53 | **41.24** | 0 |
| D18S1149 | **49.55** | 8.31 |
| D18S1107 | **51.21** | 1.66 |
| D18S478 | **52.86** | 1.65 |
| D18S819 | **52.86** | 0 |
| D18S463 | **56.71** | 3.85 |
| D18S456 | **61.73** | 5.02 |
| D18S1128 | **62.84** | 1.11 |
| D18S1102 | **62.84** | 0 |
| D18S1130 | **64.48** | 1.64 |
| D18S475 | **64.48** | 0 |
| D18S454 | **66.66** | 2.18 |
| D18S1118 | **68.91** | 2.25 |
| D18S1110 | **71.32** | 2.41 |
| D18S539 | **74.93** | 3.61 |
| D18S484 | **74.93** | 0 |
| D18S69 | **77.36** | 2.43 |
| D18S1152 | **80.41** | 3.05 |
| D18S1155 | **83.46** | 3.05 |
| D18S64 | **84.8** | 1.34 |
| D18S466 | **104.17** | 19.37 |
| D18S848 | **106.81** | 2.64 |
| D18S1106 | **106.81** | 0 |
| D18S469 | **109.18** | 2.37 |
| D18S58 | **111.2** | 2.02 |
| D18S1009 | **116.44** | 5.24 |
| D18S498 | **?** | |
| D18S977 | **?** | |
| D18S64 | **?** | |
| D18S1376 | **?** | |
| D18S812 | **?** | |
| D18S378 | **?** | |
| MBP | **?** | |
| D18S37 | **?** | |
| D18S862 | **?** | |
| | max | **19.37** |
| | ave | **2.91** |

| chr16 | cM | inter-marker distance (cM) |
|---|---|---|
| D16S539 | **124.73** | 3.28 |
| D16S520 | **125.82** | 1.09 |
| D16S3063 | **127.99** | 2.17 |
| D16S3077 | **127.99** | 0 |
| D16S3023 | **132.6** | 4.61 |
| PKD1 | **?** | |
| D16S310 | **?** | |
| D16S292 | **?** | |
| D16S301 | **?** | |
| D16S291 | **?** | |
| D16S3098 | **?** | |
| D16S79 | **?** | |
| D16S488 | **?** | |
| D16S319 | **?** | |
| D16S3023 | **?** | |
| D16S481 | **?** | |
| D16S541 | **?** | |
| D16S2618 | **?** | |
| D16S3255 | **?** | |
| | max | **10.33** |
| | ave | **2.05** |

| chr19 | cM | inter-marker distance (cM) |
|---|---|---|
| D19S886 | **0** | |
| D19S424 | **10.97** | 10.97 |
| D19S209 | **10.97** | 0 |
| D19S216 | **20.01** | 9.04 |
| D19S427 | **20.75** | 0.74 |
| D19S177 | **20.75** | 0 |
| D19S406 | **25.17** | 4.42 |
| D19S583 | **32.94** | 7.77 |
| D19S914 | **36.22** | 3.28 |
| D19S415 | **42.28** | 6.06 |
| D19S385 | **42.28** | 0 |
| D19S432 | **42.28** | 0 |
| D19S593 | **45.48** | 3.2 |
| D19S925 | **48.52** | 3.04 |
| D19S434 | **49.75** | 1.23 |
| D19S414 | **54.01** | 4.26 |
| D19S874 | **58.69** | 4.68 |
| D19S245 | **58.69** | 0 |
| D19S587 | **59.36** | 0.67 |
| D19S220 | **62.03** | 2.67 |
| D19S417 | **63.1** | 1.07 |
| D19S420 | **66.3** | 3.2 |
| D19S913 | **67.37** | 1.07 |
| D19S902 | **72.72** | 5.35 |
| D19S879 | **75.41** | 2.69 |
| D19S867 | **77.54** | 2.13 |
| D19S585 | **79.48** | 1.94 |
| D19S601 | **83.19** | 3.71 |
| D19S571 | **84.08** | 0.89 |
| D19S924 | **88.85** | 4.77 |
| D19S418 | **92.56** | 3.71 |
| D19S887 | **100.01** | 7.45 |
| D19S210 | **100.01** | 0 |
| D19S544 | **100.01** | 0 |
| D19S214 | **100.61** | 0.6 |
| D19S254 | **100.61** | 0 |
| D19S711 | **?** | |
| D19S418 | **?** | |
| D19S1171 | **?** | |
| DMD | **?** | |
| D19S414 | **?** | |
| | max | **10.97** |
| | ave | **2.87** |

| chr20 | cM | inter-marker distance (cM) |
|---|---|---|
| D20S864 | **0** | |
| D20S117 | **2.83** | 2.83 |
| D20S113 | **8.97** | 6.14 |
| D20S181 | **9.53** | 0.56 |
| D20S889 | **11.2** | 1.67 |
| D20S116 | **11.2** | 0 |
| D20S895 | **13.98** | 2.78 |
| D20S192 | **18.79** | 4.81 |
| D20S448 | **18.79** | 0 |
| D20S900 | **21.15** | 2.36 |
| D20S851 | **24.7** | 3.55 |
| D20S901 | **26.13** | 1.43 |
| D20S188 | **29.09** | 2.96 |
| D20S894 | **30.56** | 1.47 |
| D20S163 | **31.43** | 0.87 |
| D20S186 | **32.3** | 0.87 |
| D20S172 | **35.51** | 3.21 |
| D20S904 | **37.65** | 2.14 |
| D20S98 | **37.65** | 0 |
| D20S112 | **39.25** | 1.6 |
| D20S182 | **39.9** | 0.65 |
| D20S471 | **42.28** | 2.38 |
| D20S54 | **45.77** | 3.49 |
| D20S912 | **46.71** | 0.94 |
| D20S837 | **50.81** | 4.1 |
| D20S174 | **54.09** | 3.28 |
| D20S855 | **56.83** | 2.74 |
| D20S99 | **56.83** | 0 |
| D20S899 | **58.48** | 1.65 |
| D20S911 | **61.77** | 3.29 |
| D20S481 | **62.32** | 0.55 |
| D20S836 | **64.88** | 2.56 |
| D20S197 | **66.16** | 1.28 |
| D20S178 | **66.16** | 0 |
| D20S159 | **69.5** | 3.34 |
| D20S176 | **72.27** | 2.77 |
| D20S185 | **77.75** | 5.48 |
| D20S854 | **80.63** | 2.88 |
| D20S120 | **83.51** | 2.88 |
| D20S100 | **84.78** | 1.27 |
| D20S102 | **86.98** | 2.2 |
| D20S171 | **95.7** | 8.72 |
| D20S173 | **98.09** | 2.39 |
| D20S164 | **101.22** | 3.13 |
| D20S57 | **?** | |
| D20S1143 | **?** | |
| D20S1149 | **?** | |
| D20S477 | **?** | |
| D20S151 | **?** | |
| D20S110 | **?** | |
| D20S56 | **?** | |
| D20S43 | **?** | |
| | max | **8.72** |
| | ave | **2.35** |

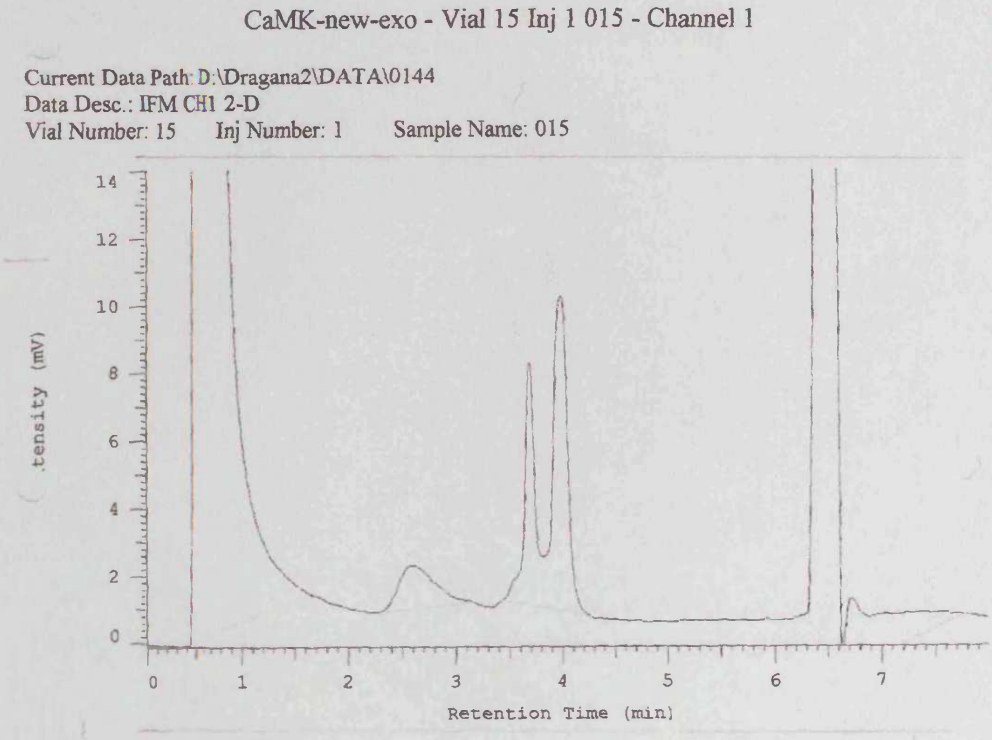
| chr21 | cM | inter-marker distance (cM) |
|---|---|---|
| D21S1904 | **2.13** | |
| D21S1431 | **2.13** | 0 |
| D21S120 | **2.99** | 0.86 |
| D21S1256 | **9.72** | 6.73 |
| D21S1899 | **9.72** | 0 |
| D21S1905 | **11.3** | 1.58 |
| D21S1918 | **16.22** | 4.92 |
| D21S210 | **20.45** | 4.23 |
| D21S1907 | **20.45** | 0 |
| D21S1896 | **21.88** | 1.43 |
| D21S260 | **22.59** | 0.71 |
| D21S1916 | **25.26** | 2.67 |
| D21S1913 | **25.8** | 0.54 |
| D21S263 | **27.4** | 1.6 |
| D21S1252 | **35.45** | 8.05 |
| D21S1894 | **35.45** | 0 |
| D21S1917 | **38.65** | 3.2 |
| D21S1893 | **43.67** | 5.02 |
| D21S1887 | **45.87** | 2.2 |
| D21S266 | **45.87** | 0 |
| D21S1260 | **46.71** | 0.84 |
| D21S1839 | **?** | |
| D21S368 | **?** | |
| D21S1235 | **?** | |
| D21S1830 | **?** | |
| D21S1264 | **?** | |
| D21S1261 | **?** | |
| D21S409 | **?** | |
| D21S82 | **?** | |
| pfkl | **?** | |
| D21S366 | **?** | |
| D21S416 | **?** | |
| | max | **8.05** |
| | ave | **2.23** |

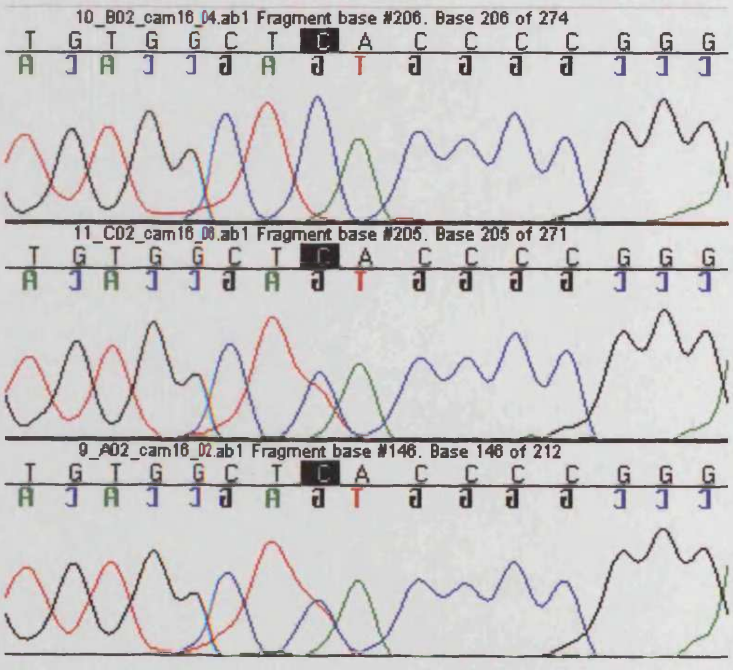
| chr22 | cM | inter-marker distance (cM) |
|---|---|---|
| D22S420 | **4.06** | |
| D22S539 | **14.44** | 10.38 |
| D22S257 | **17.71** | 3.27 |
| D22S926 | **21.47** | 3.76 |
| D22S315 | **21.47** | 0 |
| D22S429 | **22.01** | 0.54 |
| D22S533 | **22.01** | 0 |
| D22S1150 | **28.57** | 6.56 |
| D22S1176 | **29.66** | 1.09 |
| D22S280 | **31.3** | 1.64 |
| D22S685 | **32.39** | 1.09 |
| D22S422 | **32.93** | 0.54 |
| D22S683 | **36.22** | 3.29 |
| D22S1173 | **37.82** | 1.6 |
| D22S1156 | **44.32** | 6.5 |
| D22S284 | **46.42** | 2.1 |
| D22S1171 | **48.19** | 1.77 |
| D22S279 | **48.19** | 0 |
| D22S1153 | **51.54** | 3.35 |
| D22S444 | **51.54** | 0 |
| D22S274 | **51.54** | 0 |
| D22S1161 | **59.5** | 7.96 |
| D22S1169 | **60.61** | 1.11 |
| D22S431 | **?** | |
| D22S304 | **?** | |
| CRYBB2 | **?** | |
| D22S572 | **?** | |
| | max | **10.38** |
| | ave | **2.57** |

## *7.2 Investigating Calcium/calmodulin dependant protein kinase II alpha (CaMKII-$\alpha$) as a candidate gene for g*

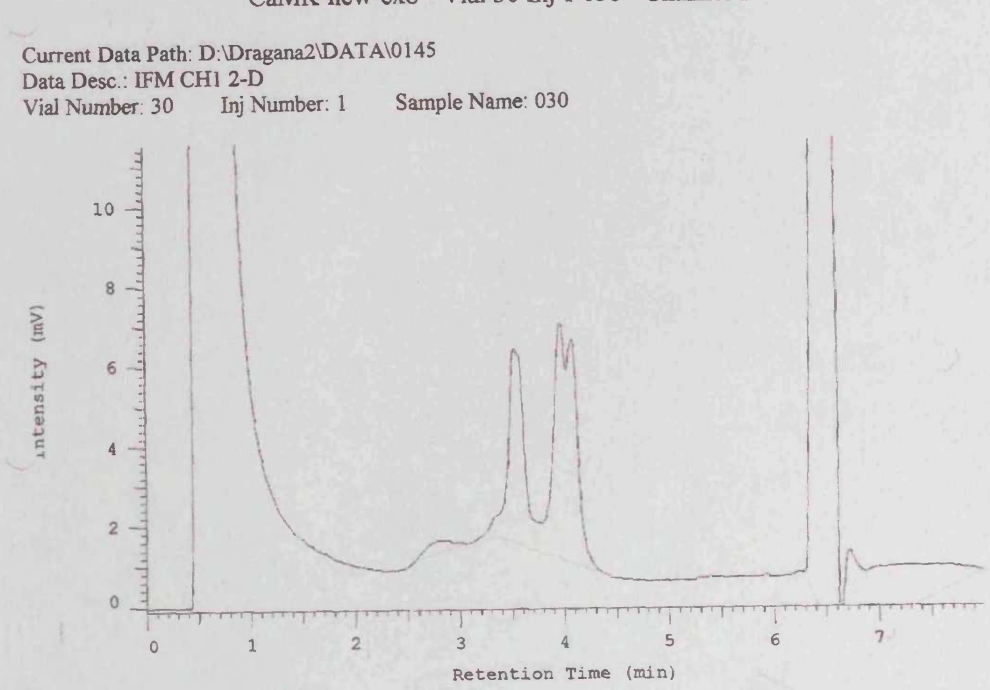
Selected dHPLC elution profile images were generated by amplified genomic DNA fragments speculated to harbour sequence polymorphisms. Profiles are accompanied by their respective sequencing electropherograms.

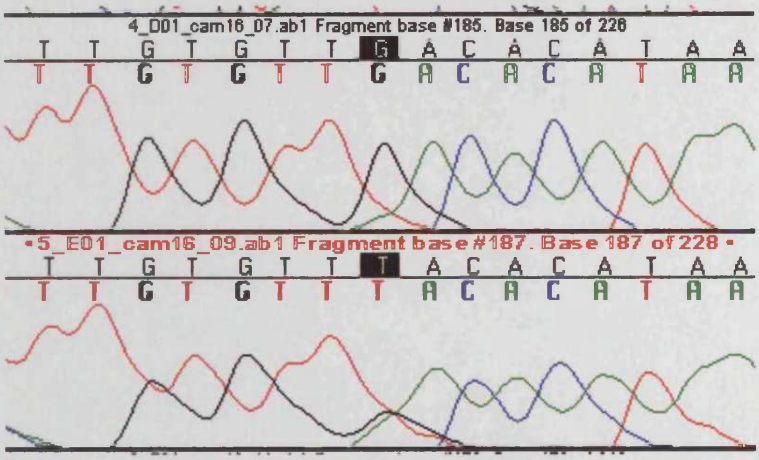
CaMKII-exo 6 - SNP 1 (C/T)
Heterozygous individuals: 026R, 029R



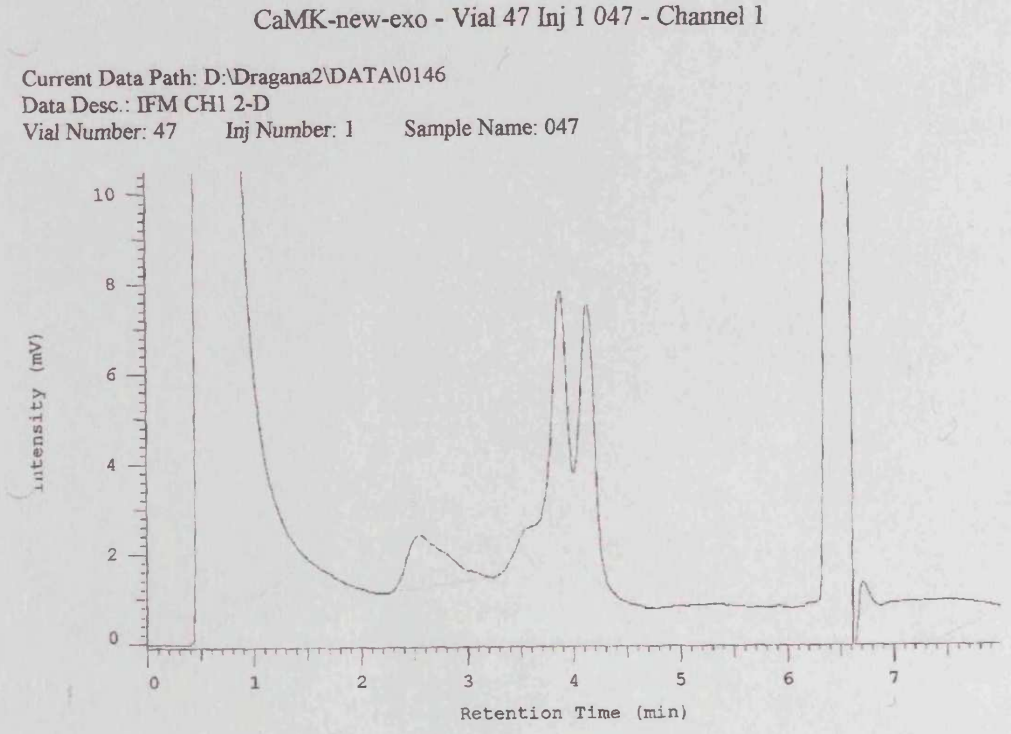



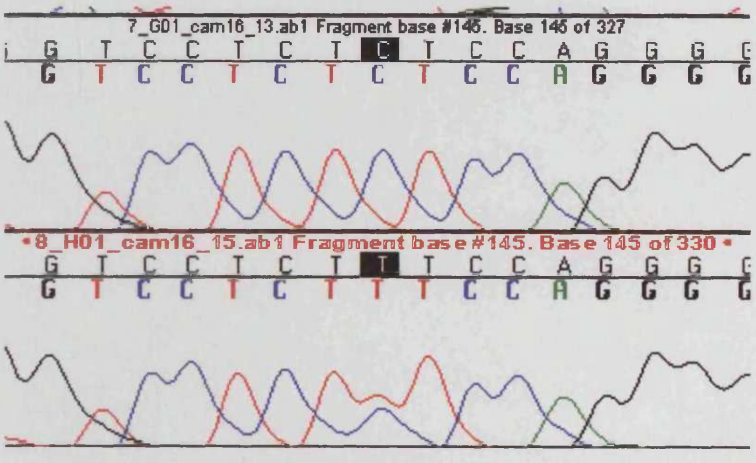
CaMKII-exo7 – SNP 1 (G/A)
Heterozygous individuals: 028R



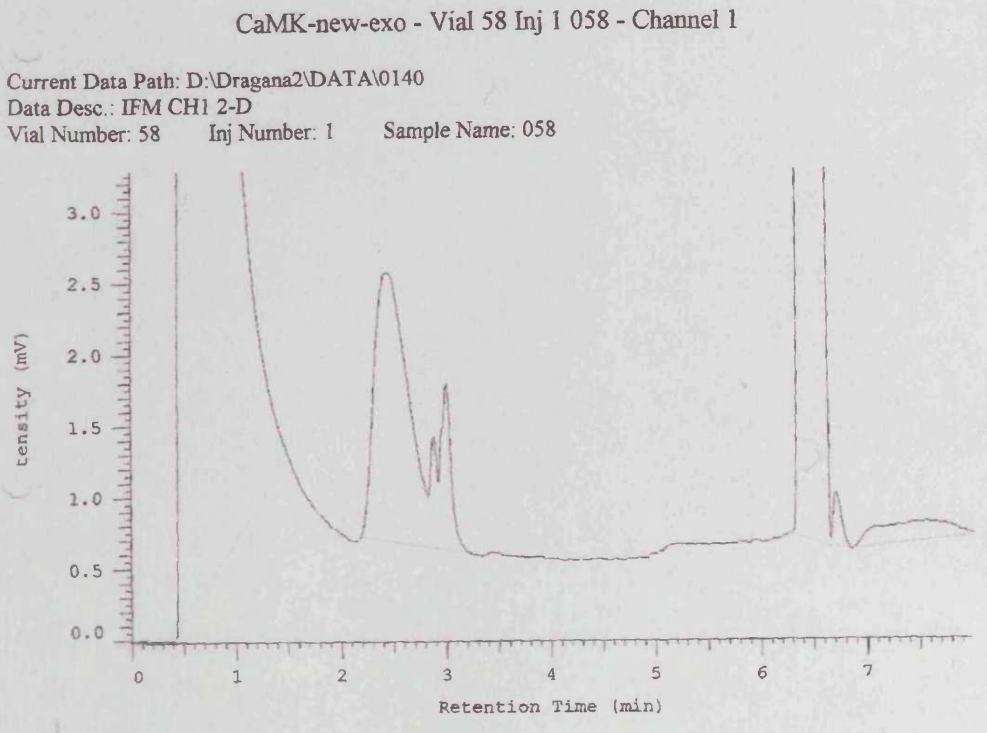



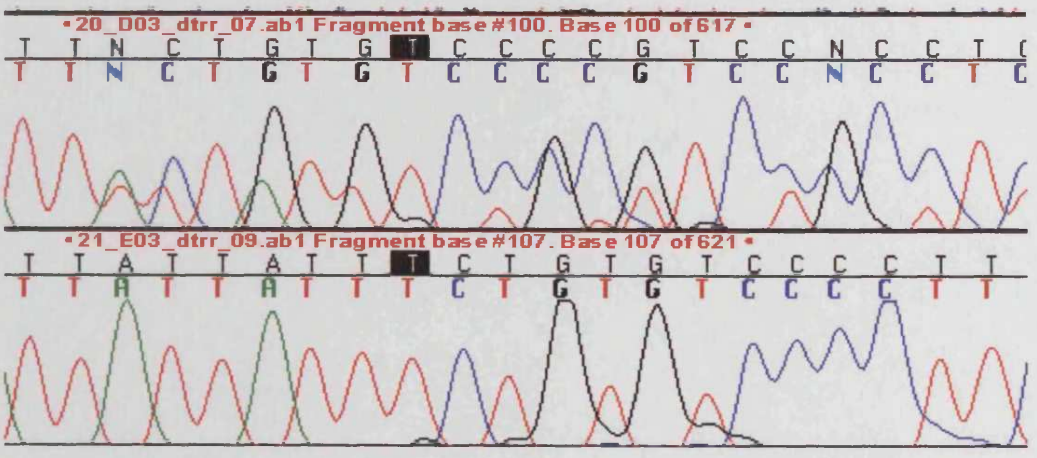
CaMKII-exo9 – SNP 1 (C/T)
Heterozygous individuals: 028R



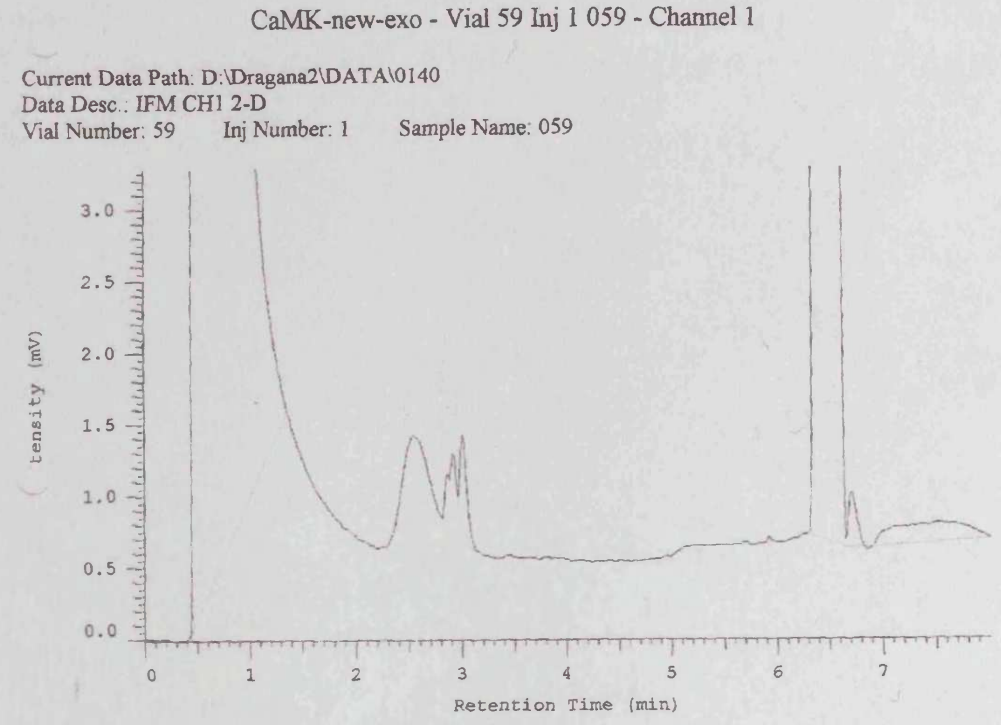



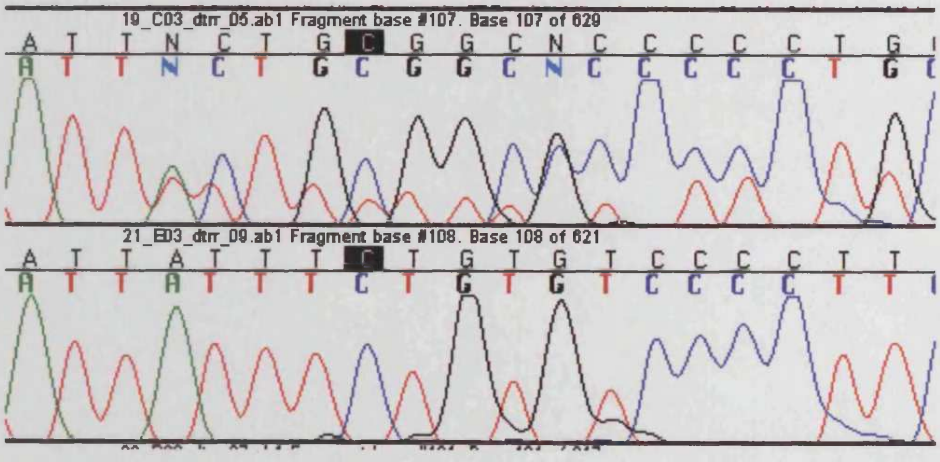
CaMKII-exo15 – SNP 1 (ATTATT deletion)
Heterozygous individuals: 020R



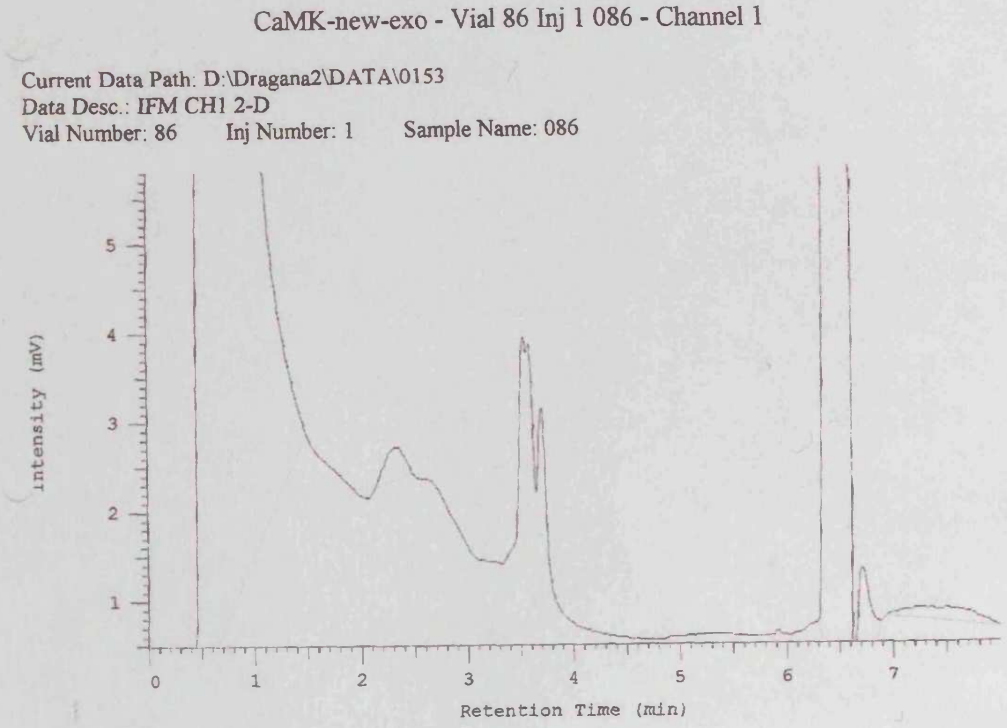



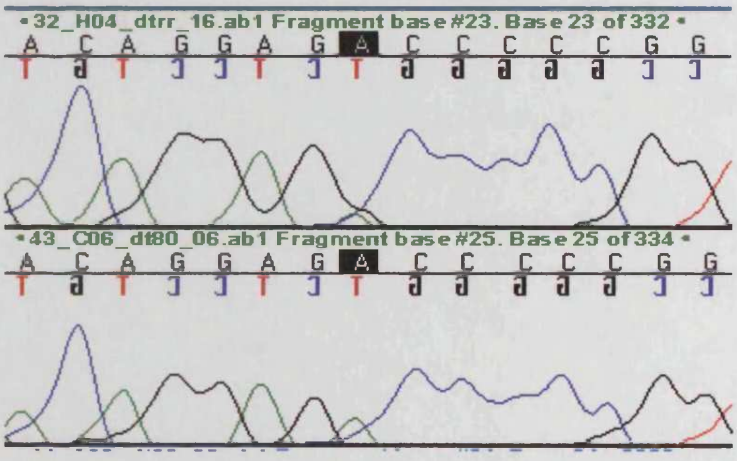
CaMKII-exo15 – SNP 2 (ATT deletion)
Heterozygous individuals: 010R



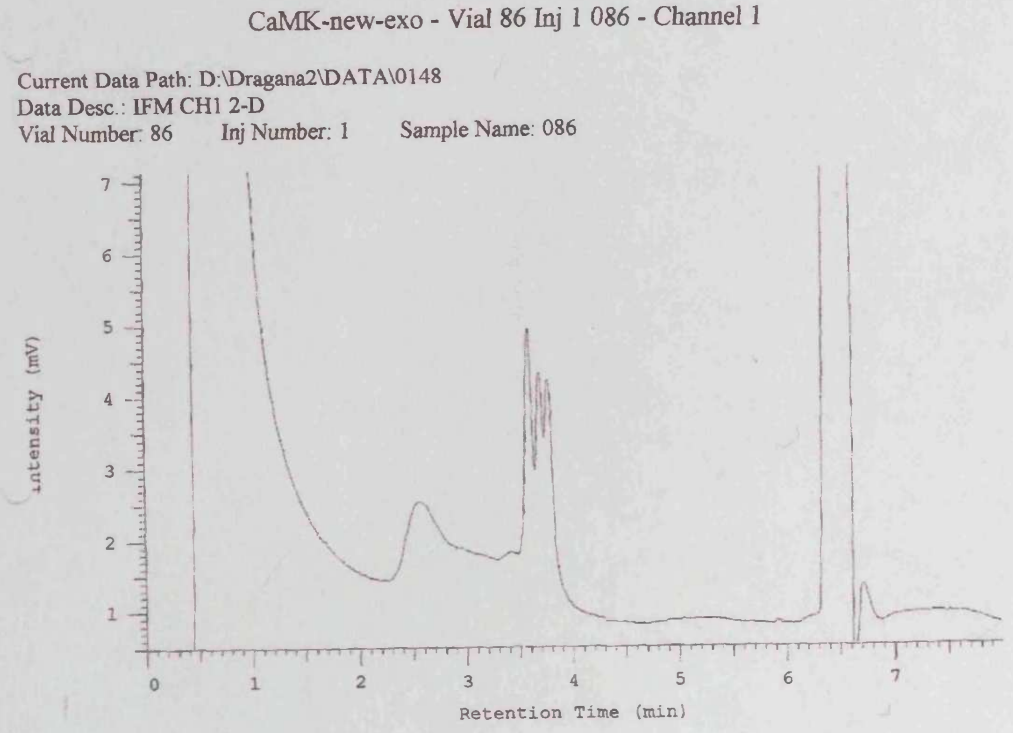



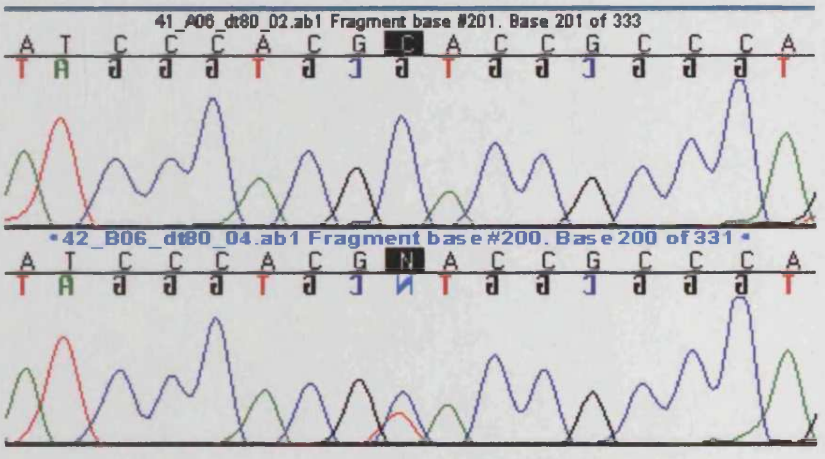
CaMKII-exo 17 – SNP 1 (G/A)
Heterozygous individuals: 001R



CaMKII-exo 17 – SNP 2 (C/T)
Heterozygous individuals: 015R

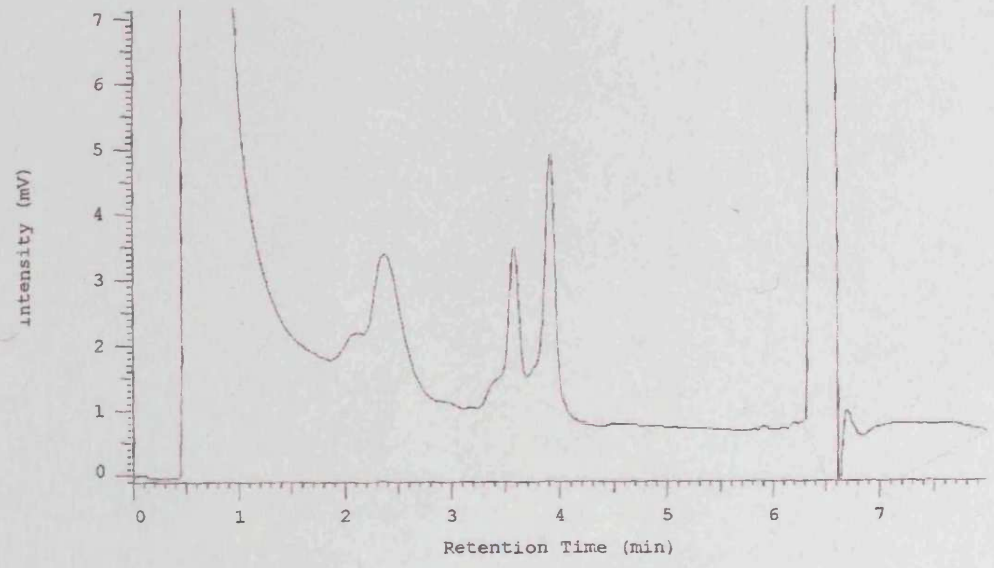


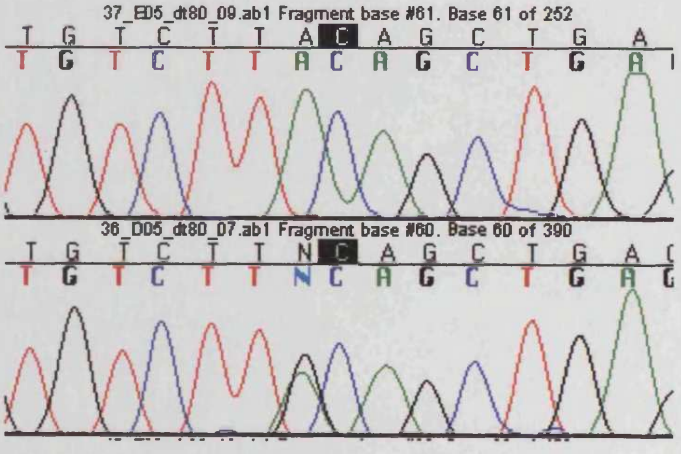



CaMKII-3'UTR-1- SNP 1 (A/G)
Heterozygous individuals: 004R

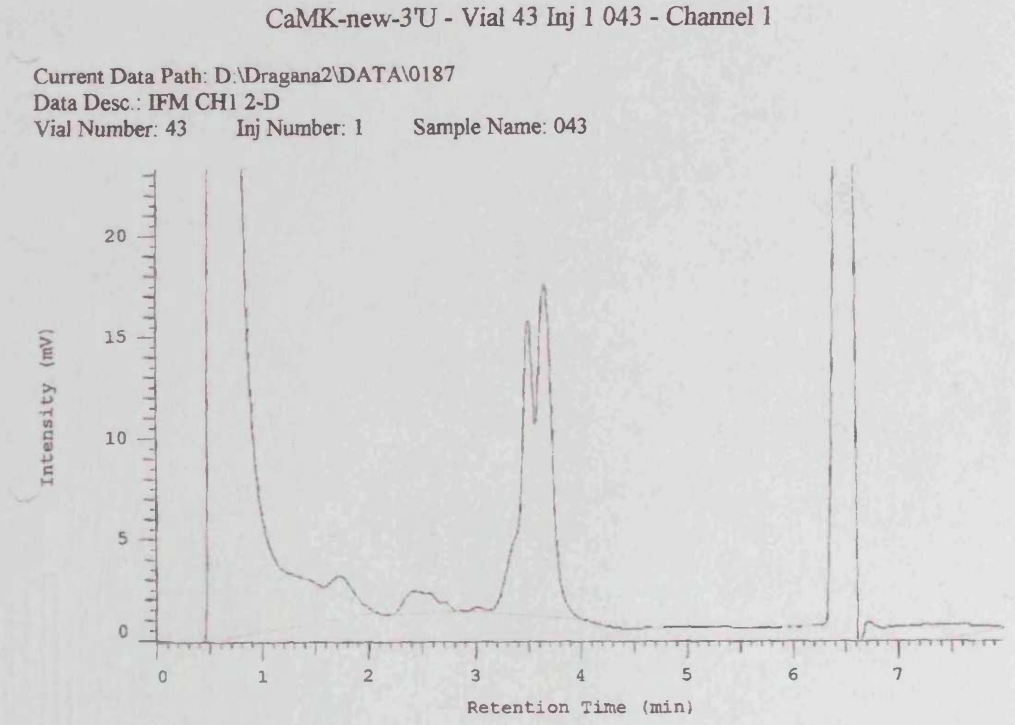
CaMK-new-3'- - Vial 6 Inj 1 006 - Channel 1

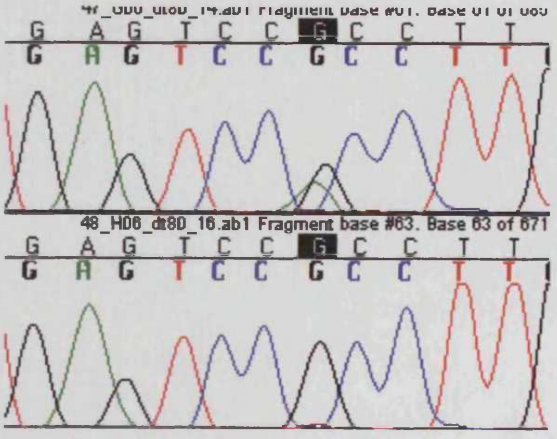
Current Data Path: D:\Dragana2\DATA\0160
Data Desc.: IFM CH1 2-D
Vial Number: 6 Inj Number: 1 Sample Name: 006



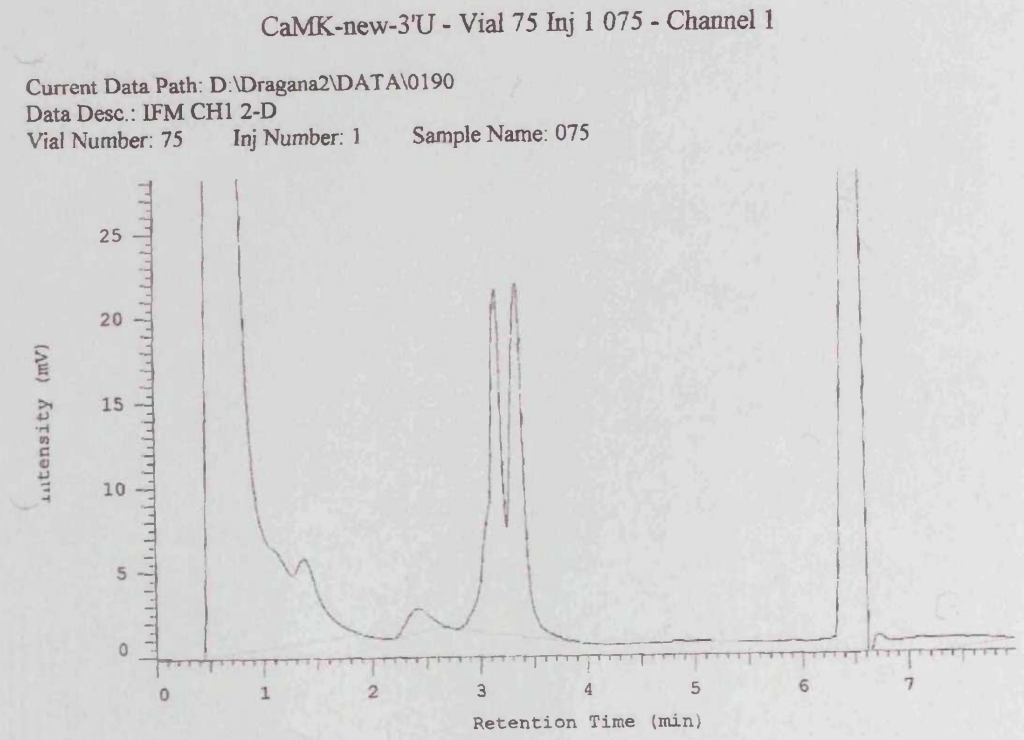



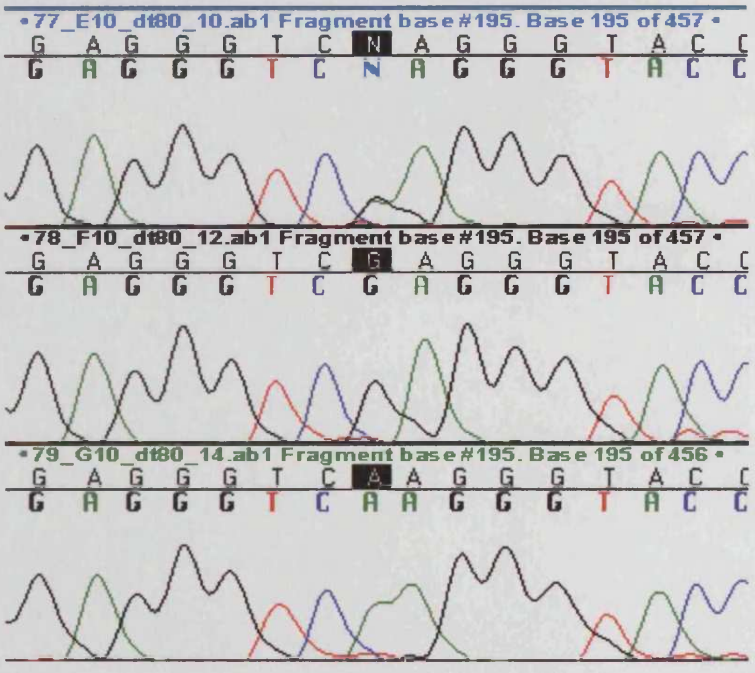
CaMKII-3'UTR-new 3 – SNP 1 (G/A)
Heterozygous individuals: 080R



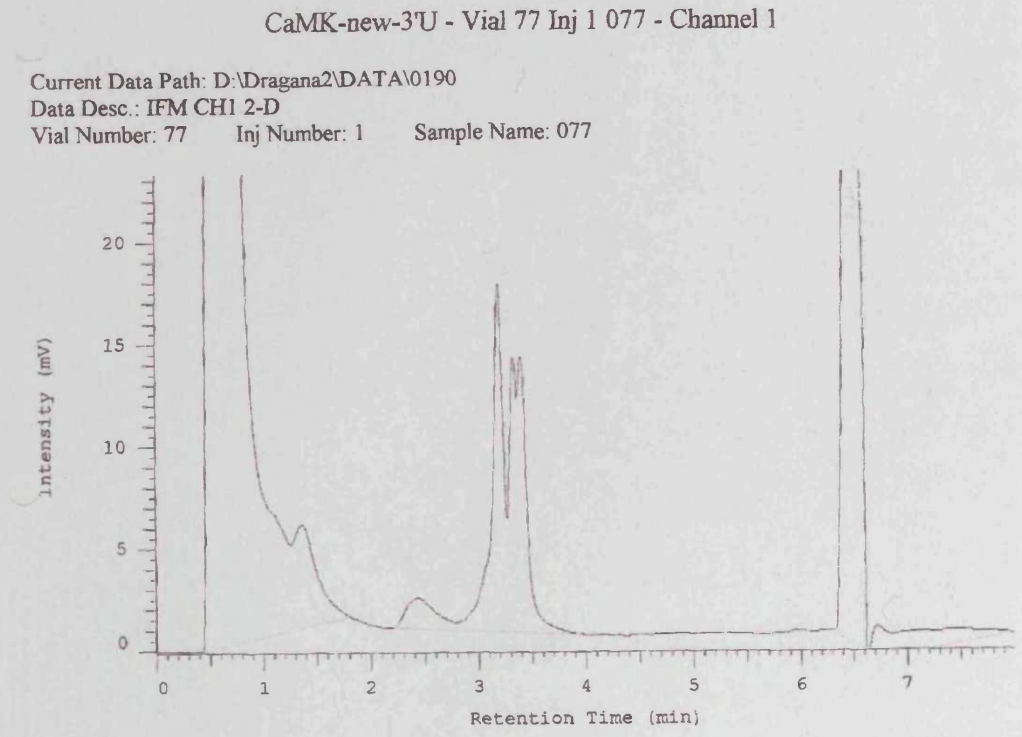



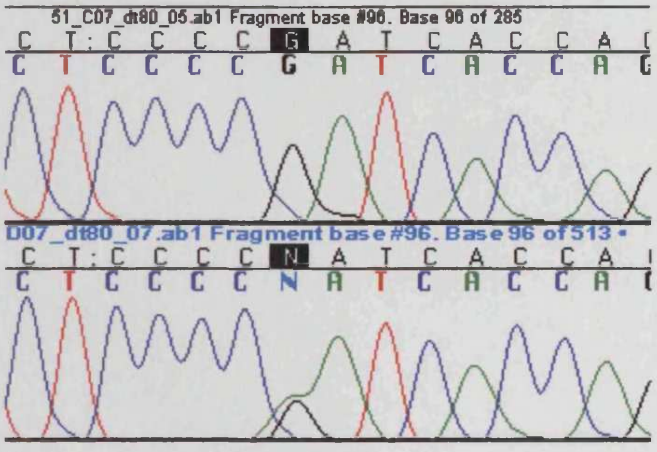
CaMKII-3'UTR-5- SNP 2 (A/G)
Heterozygous individuals: 009R



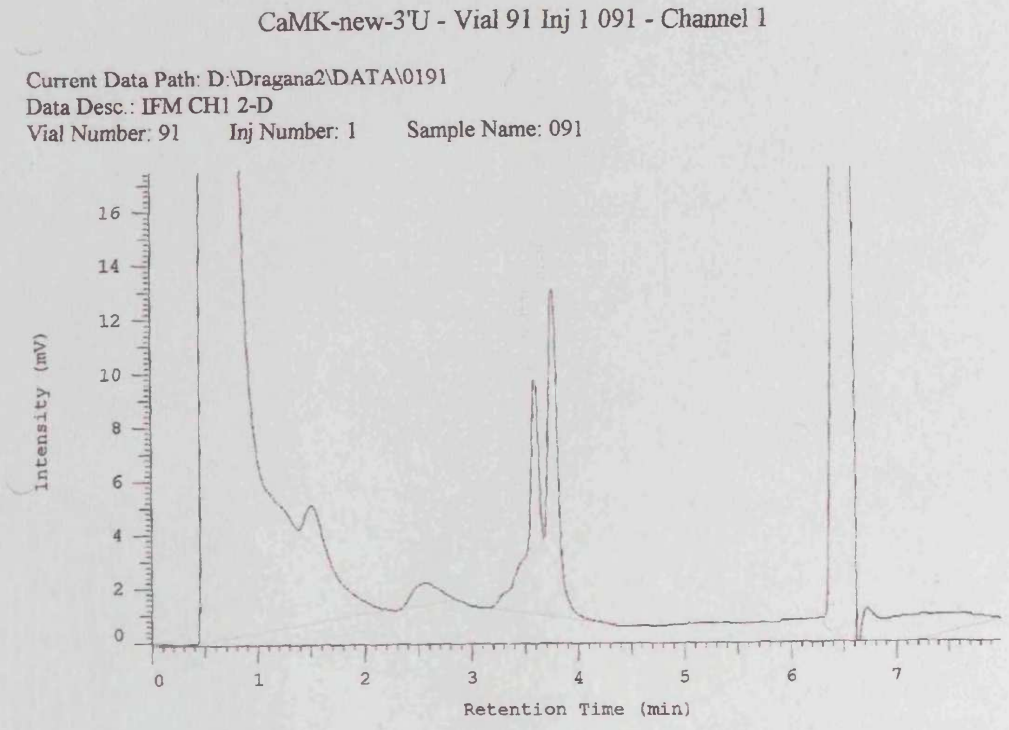



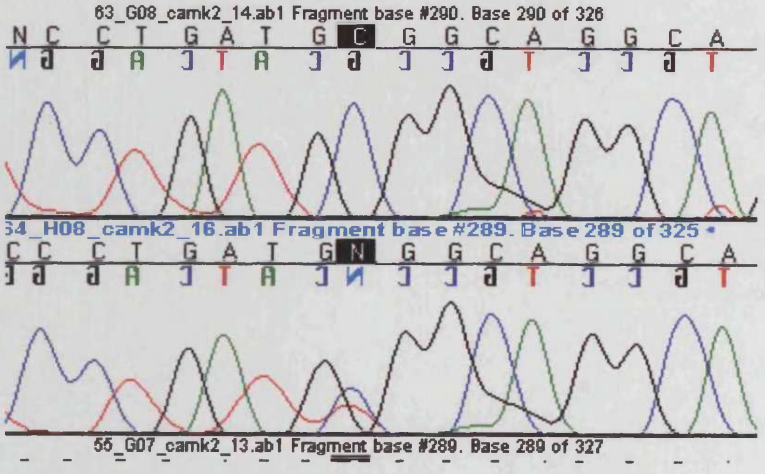
CaMKII-3'UTR-new 6 – SNP 1 (G/A)
Heterozygous individuals: 080R



CaMKII-3'UTR- 7 – SNP 1 (C/T)
Heterozygous individuals: 001R



# 8 BIBLIOGRAPHY


(1) Goldstein, D. B. Islands of linkage disequilibrium. *Nature Genetics* **2001**, *29*, 109-111.

(2) Falconer, D. S. *Introduction to Quantitative Genetics*; 3rd ed.; Longman Group: New York, 1990.

(3) Bouchard, T. J. Genetic and environmental influences on adult intelligence and special mental abilities. *Human Biology* **1998**, *70*, 257-279.

(4) Terwilliger, J. D.; Weiss, K. M. Linkage disequilibrium mapping of complex disease: fantasy or reality? *Current Opinion in Biotechnology* **1998**, *9*, 578-594.

(5) Lisman, J. E.; Zhabotinsky, A. M. A model of synaptic memory: a CaMKII/PP1 switch that potentiates transmission by organizing an AMPA receptor anchoring assembly. *Neuron* **2001**, *31*, 191-201.

(6) Pritchard, J. K.; Rosenberg, N. A. Use of unlinked genetic markers to detect population stratification in association studies. *American Journal of Human Genetics* **1999**, *65*, 220-228.

(7) Pedersen, N. L.; Plomin, R.; Nesselroade, J. R.; McClearn, G. E. A quantitative genetic analysis of cognitive abilities during the second half of the life span. *Psychological Science* **1992**, *3*, 346-353.

(8) Darwin, C. *Descent of Man*, 1871.

(9) Daniels, J.; McGuffin, P.; Owen, M. J. Molecular genetic research on IQ: Can it be done? Should it be done? *Journal of Biosoc. Science* **1996**, *28*, 491-507.

(10) Jensen, A. R. *The g Factor - The Science of Mental Ability*; Praeger: London, 1998.

(11) Strittmatter, W. J.; Hill, C. B. Molecular biology of apolipopritein E. *Current Opinion in Lipidology* **2002**, *13*, 119-123.

(12) Weedman Molavi, D. The Washington University School of Medicine Neuroscience tutorial: An illustrated guide to the essential basics of clinical neuroscience, 1997.

(13) Winder, D. G.; Sweatt, J. D. Roles of serine/threonine phosphatases in hippocampal synaptic plasticity. *Nature Reviews / Neuroscience* **2001**, *2*, 461 - 474.

(14) Wells, D. G.; Fallon, J. R. Dendritic mRNA translation: deciphering the uncoded. *Nat Neurosci* **2000**, *3*, 1062-1064.

(15) McGue, M.; Bouchard, T. J. Genetic and environmental determinants of information processing and special mental abilities: A twin analysis. *Advances in the Psychology of Human Intelligence*; Erlbaum: Hillsdale, NJ, 1989; pp 7-45.

(16) Artiga, M. J.; Bullido, M. J.; Sastre, I.; Recuero, M.; Garcia, M. A. et al. Allelic polymorphisms in the transcriptional regulatory region of apolipoprotein E gene. *Febs Letters* **1998**, *421*, 105-108.

(17) Lambert, J. C.; Pasquier, F.; Cottel, D.; Frigard, B.; Amouyel, P. et al. A new polymorphism in the APOE promoter associated with risk of developing Alzheimer's Disease. *Human Molecular Genetics* **1998**, *7*, 533-540.

(18) Mackintosh, N. J. *IQ and Human Intelligence*; 1st ed.; Oxford University Press, Inc.: New York, 1998; 419.

(19) Ando, J.; Ono, Y.; Wright, M. J. Genetic structure of spatial and verbal working memory. *Behavior Genetics* **2001**, *31*, 615-624.

(20) Hoogendoorn, B.; Norton, N.; Kirov, G.; Williams, N.; Hamshere, M. L. et al. Cheap, accurate and rapid allele frequency estimation of single nucleotide polymorphisms by primer extension and DHPLC in DNA pools. *Hum Genet* **2000**, *107*, 488-493.

(21) Sternberg, R. J.; Kaufman, J. C. Human abilities. *Annual Review of Psychology* **1998**, *49*, 479-502.

(22) Gardner, H. *Frames of Mind*; Basic Books: New York, 1983.

(23) Gottfredson, L. S. Mainstream science on intelligence: an editorial with 52 signatories, history, and bibliography. *Intelligence* **1997**, *24*, 13-23.

(24) Galton, F. Hereditary talent and character. *Macmillan's Magazine* **1865**, *12*, 318-327.

(25) Galton, F. *Hereditary genius: An inquiry into its laws and consequences*; Macmillan: London, 1869.

(26) Galton, F. *Inquiries into human faculty, and its development*; Macmillan: London, 1883.

(27) Cattell, J. M. Mental tests and measurements. *Mind* **1890**, *15*, 373-381.

(28) Binet, A. New methods for the diagnosis of the intellectual level of subnormals. *L'Annee Psychologique* **1905**, *12*, 191-244.

(29) Herrnstein, R. J.; Murray, C. *The Bell Curve: Intelligence and Class Structure in American Life*; The Free Press: New York, 1994.

(30) Terman, L. M. The uses of intelligence tests. *The measurement of intelligence*; Houghton Mifflin: Boston, 1916.

(31) Laughlin, H. H. Report of the Committee to Study and to Report on the Best Practical Means of Cutting off the Defective Germ Plasm in the American Population. *Eugenics Record Office Bulletin* **1914**, 1-64.

(32) Lombardo, P. A. Eugenic Sterilization in Virginia: Aubrey Strode and the Case of Buck v. Bell; University of Virginia, 1982.

(33) Plomin, R.; DeFries, J. C.; McClearn, G. E.; Rutter, M. *Behavioral genetics*; 3rd ed.; W. H. Freeman and Company: New York, 1997.

(34) Spearman, C. General intelligence objectively determined and measured. *American Journal of Psychology* **1904**, *15*, 202-293.

(35) Spearman, C. *The Abilities of Man: Their Nature and Measurement*; Macmillan: New York, 1927.

(36) Burt, C. The structure of the mind. *British Journal of Educational Psychology* **1949**, *19*, 176-199.

(37) Detterman, D. K.; Daniel, M. H. Correlations of mental tests with each other and with cognitive variables are highest for low IQ groups. *Intelligence* **1989**, *13*, 349-359.

(38) Neisser, U.; Boodoo, G.; Thomas J. Bouchard, J.; Boykin, A. W.; Brody, N. et al. Intelligence: knowns and unknowns. *American Psychologist* **1996**, 77-101.

(39) Brody, N. *Intelligence*; 2nd ed.; Academic Press: New York, 1992; 395.

(40) Horn, W. F.; Packard, T. Early identification of learning problems: a meta-analysis. *Journal of Educational Psychology* **1985**, *77*, 597-607.

(41) McCall, R. B. Childhood IQ's as predictors of adult educational and occupational status. *Science* **1977**, *197*, 482-483.

(42) Cardon, L. R.; DiLalla, L. F.; Plomin, R.; Defries, J. C.; Fulker, D. W. Genetic correlations between reading performance and IQ in the Colorado Adoption Project. *Intelligence* **1990**, *14*, 245-257.

(43) Thompson, L. A.; Detterman, D. K.; Plomin, R. Associations between cognitive abilities and scholastic achievement: genetic overlap but environmental differences. *Psychological Science* **1991**, *2*, 158-165.

(44) Mackintosh, N. J.; Mascie-Taylor, C. G. N. The IQ question. *Personality, cognition and values*; Macmillan: London, 1986.

(45) Jencks, C. *Inequality: A reassessment of the effect of family and schooling in America*; Basic Books: New York, 1972.

(46) Jausovec, N.; Jausovec, K. Differences in event-related and induced brain oscillations in the theta and alpha frequency bands related to human intelligence. *Neuroscience Letters* **2000**, *293*, 191-194.

(47) Rijsdijk, F. V.; Boomsma, D. I.; Vernon, P. A. Genetic analysis of periferal nerve conduction velocity in twins. *Behavior Genetics* **1995**, *25*, 341-348.

(48) Rijsdijk, F. V.; Boomsma, D. I. Genetic Mediation of the correlation between peripheral nerve conduction velocity and IQ. *Behavior Genetics* **1997**, *27*, 87-98.

(49) Vernon, P. A.; Mori, M. Intelligence, reaction times, and peripheral nerve conduction velocity. *Intelligence* **1992**, *16*, 273-288.

(50) Wickett, J. C.; Vernon, P. A. Peripheral nerve conduction velocity, reaction time, and intelligence: an attempt to replicate Vernon and Mori. *Intelligence* **1994**, *15*, 223-228.

(51) Reed, T. E.; Jensen, A. R. Arm nerve conduction velocity (NCV), brain NCV, reaction time, and intelligence. *Intelligence* **1991**, *15*, 33-47.

(52) Haier, R. J.; Nuechterlein, K. H.; Hazlett, E.; Wu, J. C.; Paek, J. Cortical glucose metabolic rate correlates of abstract reasoning and attention studied with positron emission tomography. *Intelligence* **1988**, *12*, 199-217.

(53) Willerman, L.; Schultz, R.; Rutledge, J. N.; Bigler, E. D. *In vivo* brain size and intelligence. *Intelligence* **1991**, *15*, 223-228.

(54) Wickett, J. C.; Vernon, P. A.; Lee, D. H. *In vivo* brain size, head perimeter, and intelligence in a sample of healthy adult females. *Personality and Individual Differences* **1994**, *16*, 831-838.

(55) Tramo, M. J.; Loftus, W. C.; Stukel, T. A.; Green, R. L.; Weaver, J. B. et al. Brain size, head size, and intelligence quotient in monozygotic twins. *Neurology* **1998**, *50*, 1246-1252.

(56) Miller, E. M. Intelligence and brain myelination: a hypothesis. *Personality and Individual Differences* **1994**, *17*, 803-832.

(57) Fidelman, U. The three components of intelligence: evolutionary specialization in mental abilities. *Cybernetica* **1997**, *40*, 103-126.

(58) Thompson, P. M.; Cannon, T. D.; Narr, K. L.; Erp, T. v.; Poutanen, V.-P. et al. Genetic influences on brain structure. *Nature Neuroscience* **2001**, *4*, 1253-1258.

(59) Luo, D.; Petrill, S. A.; Thompson, L. A. An exploration of genetic *g*: hierarchical factor analysis of cognitive data from the Western Reserve Twin Project. *Intelligence* **1994**, *18*, 335-348.

(60) Pedersen, N. L.; Plomin, R.; McClearn, G. E. Is there G beyond *g*? (Is there genetic influence on specific cognitive abilities independant of genetic influence on general cognitive ability?). *Intelligence* **1994**, *18*, 133-143.

(61) Petrill, S. A.; Luo, D.; Thompson, L. A.; Detterman, D. K. The Relationship between Elementary Cognitive Tasks and Wisc-R Subtests - Genetic and Environmental-Influences. *Behavior Genetics* **1995**, *25*, 283-283.

(62) Lander, E. S.; Schork, N. J. Genetic dissection of complex traits. *Science* **1994**, *265*, 2037-2047.

(63) Pritchard, J. K. Are rare variants responsible for susceptibility to complex diseases? *American Journal of Human Genetics* **2001**, *69*, 124-137.

(64) Tanksley, S. D. Mapping polygenes. *Annual Review of Genetics* **1993**, *27*, 205-233.

(65) Owen, M. J.; Craddock, N. Modern molecular genetic approaches to complex traits: implications for psychiatric disorders. *Molecular Psychiatry* **1996**, *1*, 21-26.

(66) Plomin, R.; Owen, M. J.; McGuffin, P. The genetic basis of complex human behaviors. *Science* **1994**, *264*, 1733-1739.

(67) Risch, N.; Spiker, D.; Lotspeich, L.; Nouri, N.; Hinds, D. et al. A genomic screen of autism: Evidence for a multilocus etiology. *American Journal of Human Genetics* **1999**, *65*, 493-507.

(68) Sham, P. Genetic Epidemiology. *Biological Psychiatry*; Royal Society of Medicine Press Ltd.: London, 1996; pp 408-433.

(69) Johnson, G. C. L.; Todd, J. A. Strategies in complex disease mapping. *Current Opinion in Genetics & Development* **2000**, *10*, 330-334.

(70) Strachan, T.; Read, A. P. *Human Molecular Genetics*; 1st ed.; BIOS Scientific Publishers Limited: Oxford, 1997.

(71) McGuffin, P.; Owen, M. J.; O'Donovan, M. C.; Thapar, A.; Gottesman, I. I. *Seminars in Psychiatric Genetics*; GASKELL: London, 1994.

(72) Ott, J. Invited editorial: cutting a Gordian knot in the linkage analysis of complex traits. *American Journal of Human Genetics* **1990**, *46*, 219-221.

(73) Risch, N. Linkage strategies for genetically complex traits. I. Multilocus models. *American Journal of Human Genetics* **1990**, *46*, 222-228.

(74) Kamin, L. J.; Goldberger, A. S. Twin studies in behavioral research: a sceptical view. *Theoretical Population Biology* **2002**, *61*, 83-95.

(75) Joseph, J. The equal environment assumption of the classical twin method: a critical analysis. *Journal of Mind and Behavior* **1998**, *19*, 325-358.

(76) Guo, S.-W. Does higher concordance in monozygotic twins than in dizygotic twins suggest a genetic component? *Human Heredety* **2001**, *51*, 121-132.

(77) Plomin, R.; Willerman, L.; Loehlin, J. C. Resemblance in appearance and the equal environments assumption in twin studies of personality traits. *Behavior Genetics* **1976**, *6*, 43-52.

(78) Kendler, K. S.; Neale, M. C.; Kessler, R. C.; Heath, A. C.; Eaves, L. J. Parental treatment and the equal environment assumption in twin studies of psychiatric illness. *Psychological Medicine* **1994**, *24*, 579-590.

(79) Kendler, K. S.; Neale, M. C.; Kessler, R. C.; Heath, A. C.; Eaves, L. J. A test of the equal-environment assumption in twin studies of psychiatric illness. *Behavior Genetics* **1993**, *23*, 21-27.

(80) Hettema, J. M.; Neale, M. C.; Kendler, K. S. Physical similarity and the equal-environment assumption in twin studies of psychiatric disorders. *Behavior Genetics* **1995**, *25*, 327-335.

(81) Morris-Yates, A.; Andrews, G.; Howie, P.; Henderson, S. Twins: a test of the equal environments assumption. *Acta Psychiatrica Scandinavica* **1990**, *81*, 322-326.

(82) Klump, K. L.; Holly, A.; Iacono, W. G.; McGue, M.; Willson, L. E. Physical similarity and twin resemblance for eating attitudes and behaviors: a test of the equal environments assumption. *Behavior Genetics* **2000**, *30*, 51-58.

(83) Freeman, F. N.; Holzinger, K. J.; Mitchell, B. C. The influence of environment on the intelligence, school achievement, an coduct of foster children. *27th Yearbook of the National Society for the Study of Education* **1928**, *27, Part 1.*

(84) Loehlin, J. C.; Horn, J. M.; Willerman, L. Modeling IQ change: evidence from the Texas Adoption Project. *Child Development* **1989**, *60*, 993-1004.

(85) Plomin, R.; Fulker, D. W.; Corley, R. P.; DeFries, J. C. Nature, nurture, and cognitive development from 1 to 16 years: A parent-offspring adoption study. *Psychological Science* **1997**, *8*, 442-447.

(86) Sorensen, T. I. A.; Holst, C.; Stunkard, A. J. Adoption study of environmental modifications of the genetic influences on obesity. **1998**, *22*, 73-81.

(87) Bouchard, T. J.; McGue, M. Familial studies of intelligence: a review. *Science* **1981**, *212*, 1055-1059.

(88) Chipuer, H. M.; Rovine, M. J.; Plomin, R. LISREL modeling: genetic and environmental influences on IQ revisited. *Intelligence* **1990**, *14*, 11-29.

(89) Devlin, B.; Daniels, M.; Roeder, K. The heritability of IQ. *Nature* **1997**, *388*, 468-471.

(90) Lynn, R. *Dysgenics: Genetic deterioration in modern populations*; Praeger: Westport, CT, 1996.

(91) Bouchard, T. J.; Lykken, D. T.; McGue, M.; Segal, N. L.; Tellegen, A. Sources of human psychological differences: The Minnesota study of twins rared apart. *Science* **1990**, *250*, 223-228.

(92) McClearn, G. E.; Johansson, B.; Berg, S.; Pedersen, N. L.; Ahern, F. et al. Substantial genetic influence on cognitive abilities in twins 80 or more years old. *Science* **1997**, *276*, 1560-1563.

(93) Wright, M.; De Geus, E.; Ando, J.; Luciano, M.; Posthuma, D. et al. Genetics of cognition: outline of a collaborative twin study. *Twin Research* **2001**, *4*, 48-56.

(94) Teasdale, T. W.; Owen, D. R. Heritability and familial environment in intelligence and educational level: a sibling study. *Nature* **1984**, *309*, 620-622.

(95) Fulker, D. W.; DeFries, J. C.; Plomin, R. Genetic influence on general mental ability increases between infancy and middle childhood. *Nature* **1988**, *336*, 767-769.

(96) Finkel, D.; Pedersen, N. L.; McGue, M. Heritability of cognitive abilities in adulte twins: comparison of Minnesota and Swedish data. *Behavior Genetics* **1995**, *25*, 421-431.

(97) Plomin, R.; DeFries, J. C.; Loehlin, J. C. Genotype-environment interaction and correlation in the analysis of human behavior. *Psychol Bull* **1977**, *84*, 309-322.

(98) Thomson, G.; Esposito, M. S. The genetics of complex diseases (Reprinted from Trends in Biochemical Science, vol 12, Dec., 1999). *Trends in Genetics* **1999**, *15*, M17-M20.

(99) Strathdee, C. A.; Gavish, H.; Shannon, W. R.; Buchwald, M. Cloning of cDNAs for Fanconi's anaemia by functional complementation. *Nature* **1992**, *356*, 763-767.

(100) Jagadeeswaran, P.; Lavelle, D. E.; Kaul, R.; Mohandas, T.; Warren, S. T. Isolation and characterization of human factor IX cDNA: identification of Taq I polymorphism and regional assignment. *Somat Cell Mol Genet* **1984**, *10*, 465-473.

(101) Abecasis, G. R.; Cookson, W. O. C.; Cardon, L. R. The power to detect linkage disequilibrium with quantitative traits in selected samples. *American Journal of Human Genetics* **2001**, *68*, 1463-1474.

(102) Slatkin, M. Disequilibrium mapping of a quantitative-trait locus in an expanding population. *American Journal of Human Genetics* **1999**, *64*, 1765-1773.

(103) Nyholt, D. R. All LODs are not created equal. *American Journal of Human Genetics* **2000**, *67*, 282-288.

(104) Burmeister, M. Basic concepts in the study of diseases with complex genetics. *Biological Psychiatry* **1999**, *45*, 522-532.

(105) Curtis, D. Fundamental concepts in genetics, 1995-2000.

(106) Almasy, L.; Blangero, J. Linkage strategies for mapping genes for complex traits in man. *Handbook of Molecular-Genetic Techniques for Brain and Behavior research*; Elsevier Science, 1999; pp 100-112.

(107) Lake, S. L.; Blacker, D.; Laird, N. M. Family-based tests of association in the presence of linkage. *American Journal of Human Genetics* **2000**, *67*, 1515-1525.

(108) Gorwood, P. Genetic association studies in behavioral neuroscience. *Handbook of Molecular-Genetic Techniques for Brain and Behavior Research*; Elsevier Science, 1999.

(109) Risch, N.; Merikangas, K. The future of genetic studies of complex human diseases. *Science* **1996**, *273*, 1516-1517.

(110) Falk, C. T.; Rubinstein, P. Haplotype relative risk: an easy reliable way to construct a proper control sample for risk calculation. *American Journal of Human Genetics* **1987**, *51*, 227-233.

(111) Spielman, R. S.; McInnis, R. E.; Ewens, W. J. Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetus mellitus (IDDM). *American Journal of Human Genetics* **1993**, *52*, 506-516.

(112) Morton, N. E.; Collins, A. Tests and estimates of allelec association in complex inheritance. *Proceedings of the National Academy of Sciences of the USA* **1998**, *95*, 11389-11393.

(113) Mountain, J. L.; Cavalli-Sforza, L. L. Multilocus genotypes, a tree of individuals, and human evolutionary history. *American Journal of Human Genetics* **1997**, *61*, 705-718.

(114) Bland, J. M.; Altman, D. G. Multiple significance tests: the Bonferroni method. *British Medical Journal* **1995**, *310*, 170.

(115) Kidd, K. K. Association of disease with genetic markers: deja vu all over again. *American Journal of Medical Genetics* **1993**, *48*, 71-72.

(116) Owen, M. J.; Cardno, A. G.; O'Donovan, M. C. Psychiatric genetics: back to the future. *Molecular Psychiatry* **2000**, *5*, 22-31.

(117) Kruglyak, L. Prospects for whole-genome linkage disequilibrium mapping of common disease genes. *Nature Genetics* **1999**, *22*, 139-144.

(118) Ardlie, K. G.; Kruglyak, L.; Seielstad, M. Patterns of linkage disequilibrium in the human genome. *Nature Reviews / Genetics* **2002**, *3*, 299-309.

(119) Kruglyak, L. What is significant in whole-genome linkage disequilibrium studies? *American Journal of Human Genetics* **1997**, *61*, 810-812.

(120) Stumpf, M. P. H. Haplotype diversity and the block structure of linkage disequilibrium. *Trends in Genetics* **2002**, *18*, 226-228.

(121) Lewontin, R. C. The interaction of selection and linkage. I. General considerations; heterotic models. *Genetics* **1964**, *49*, 49-67.

(122) Weiss, K. M.; Clark, A. G. Linkage disequilibrium and the mapping of complex human traits. *Trends in Genetics* **2002**, *18*, 19-24.

(123) Pritchard, J. K.; Przeworski, M. Linkage disequilibrium in humans: models and data. *American Journal of Human Genetics* **2001**, *69*, 1-14.

(124) Long, A. D.; Langley, C. H. The power of association studies to detect the contribution of candidate genetic loci to variation in complex traits. *Genome Research* **1999**, *8*, 720-731.

(125) Terwilliger, J. D.; Zollner, S.; Laan, M.; Paabo, S. Mapping genes through the use of linkage disequilibrium generated by genetic drift: 'Drift mapping' in small populations with no demographic expansion. *Human Heredety* **1998**, *48*, 138-154.

(126) Pfaff, C. L.; Parra, E. J.; Bonilla, C.; Hiester, K.; McKeigue, P. M. et al. Population Structure in Admixed Populations: Effect of Admixture Dynamics on the Pattern of Linkage Disequilibrium. *American Journal of Human Genetics* **2001**, *68*, 198-207.

(127) Reich, D. E.; Cargill, M.; Bolk, S.; Ireland, J.; Sabeti, P. C. et al. Linkage disequilibrium in the human genome. *Nature* **2001**, *411*, 199-204.

(128) Stephens, J. C.; Schneider, J. A.; Tanguay, D. A.; Choi, J.; Acharya, T. et al. Haplotype Variation and Linkage Disequilibrium in 313 Human Genes. *Science* **2001**, *293*, 489-493.

(129) Frisse, L.; Hudson, R. R.; Bartoszewicz, A.; Wall, J. D.; Donfack, J. et al. Gene conversion and different population histories may explain the contrast between polymorphism and linkage disequilibrium levels. *American Journal of Human Genetics* **2001**, *69*, 831-843.

(130) Ardlie, K.; Liu-Cordero, S. N.; Eberle, M. A.; Daly, M.; Barrett, J. et al. Lower-Than-Expected Linkage Disequilibrium between Tightly Linked Markers in Humans Suggests a Role for Gene Conversion. *American Journal of Human Genetics* **2001**, *69*, 582-589.

(131) Peltonen, L.; Palotie, A.; Lange, K. Use of population isolates for mapping complex traits. *Nature Reviews / Genetics* **2000**, *1*, 182-190.

(132) Collins, A.; Lonjou, C.; Morton, N. E. Genetic epidemiology of single-nucleotide polymorhisms. *Proceedings of the National Academy of Sciences of the USA* **1999**, *96*, 15173-15177.

(133) Abecasis, G. R.; Noguchi, E.; Heinzmann, A.; Traherne, J. A.; Bhattacharyya, S. et al. Extent and distribution of linkage disequilibrium in three genomic regions. *American Journal of Human Genetics* **2001**, *68*, 191-197.

(134) Gordon, D.; Simonic, I.; Ott, J. Significant evidence for linkage disequilibrium over a 5-cM region among Afrikaners. *Genomics* **2000**, *66*, 87-92.

(135) Eaves, I. A.; Merriman, T. R.; Barber, R. A.; Nutland, S.; Tuomilehto-Wolf, E. et al. The genetically isolated populations of Finland and Sardinia may not be a panacea for linkage disequilibrium mapping of common disease genes. *Nature Genetics* **2000**, *25*, 320-323.

(136) Kaessmann, H.; Zöllner, S.; Gustafsson, A. C.; Wiebe, V.; Laan, M. et al. Extensive linkage disequilibrium in small human populations in Eurasia. *American Journal of Human Genetics* **2002**, *70*, 673-685.

(137) Jorde, L. B. Linkage disequilibrium and the search for complex disease genes. *Genome Research* **2000**, *10*, 1435-1444.

(138) Shifman, S.; Darvasi, A. The value of isolated populations. *Nature* **2001**, *28*, 309-310.

(139) Wilson, J. F.; Goldstein, D. B. Consistent long-range linkage disequilibrium generated by admixture in a Bantu-Semitic hybrid population. *American Journal of Human Genetics* **2000**, *67*, 926-935.

(140) Taillon-Miller, P.; Bauer-Sardiña, I.; Saccone, N. L.; Putzel, J.; Laitinen, T. et al. Juxtaposed regions of extensive and minimal linkage disequilibrium in human Xq25 and Xq28. *Nature Genetics* **2000**, *25*, 324-328.

(141) Daly, M. J.; Rioux, J. D.; Schaffner, S. F.; Hudson, T. J.; Lander, E. S. High-resolution haplotype structure in the human genome. *Nature Genetics* **2001**, *29*, 229-232.

(142) Jeffreys, A. J.; Kauppi, L.; Neumann, R. Intensely punctate meiotic recombination in the class II region of the major histocompatibility complex. *Nature Genetics* **2001**, *29*, 217-222.

(143) Rioux, J. D.; Daly, M. J.; Silverberg, M. S.; Lindblad, K.; Steinhart, H. et al. Genetic variation in the 5q31 cytokine gene cluster confers susceptibility to Crohn disease. *Nature Genetics* **2001**, *29*, 223 - 228.

(144) Patil, N.; Berno, A. J.; Hinds, D. A.; Barrett, W. A.; Doshi, J. M. et al. Blocks of limited haplotype diversity revealed by high-resolution scanning of human chromosome 21. *Science* **2001**, *294*, 1719-1723.

(145) Maniatis, N.; Collins, A.; Xu, C. F.; McCarthy, L. C.; Hewett, D. R. et al. The first linkage disequilibrium (LD) maps: delineation of hot and cold blocks by diplotype analysis. *Proc Natl Acad Sci U S A* **2002**, *99*, 2228-2233.

(146) Tapper, W. J.; Maniatis, N.; Morton, N. E.; Collins, A. A metric linkage disequilibrium map of a human chromosome. *Ann Hum Genet* **2003**, *67*, 487-494.

(147) Xiong, M.; Guo, S.-W. Fine-scale genetic mapping based on linkage disequilibrium: theory and applications. *American Journal of Human Genetics* **1997**, *60*, 1513-1531.

(148) Hugot, J.-P.; Chamaillard, M.; Zouali, H.; Lesage, S.; Cézard, J.-P. et al. Association of NOD2 leucine-rich repeat variants with susceptibility to Crohn's disease. *Nature* **2001**, *411*, 599-603.

(149) Van Eerdewegh, P.; Little, R. D.; Dupuis, J.; Del Mastro, R. G.; Falls, K. et al. Association of the ADAM33 gene with asthma and bronchial hyperresponsiveness. *Nature* **2002**, *418*, 426-430.

(150) Allen, M.; Heinzmann, A.; Noguchi, E.; Abecasis, G.; Broxholme, J. et al. Positional cloning of a novel gene influencing asthma from chromosome 2q14. *Nat Genet* **2003**, *35*, 258-263.

(151) Karban, A. S.; Okazaki, T.; Panhuysen, C. I.; Gallegos, T.; Potter, J. J. et al. Functional annotation of a novel NFKB1 promoter polymorphism that increases risk for ulcerative colitis. *Hum Mol Genet* **2004**, *13*, 35-45.

(152) Horikawa, Y.; Oda, N.; Cox, N. J.; Li, X.; Orho-Melander, M. et al. Genetic variation in the gene encoding calpain-10 is associated with type 2 diabetes mellitus. *Nat Genet* **2000**, *26*, 163-175.

(153) Hutcheson, H. B.; Olson, L. M.; Bradford, Y.; Folstein, S. E.; Santangelo, S. L. et al. Examination of NRCAM, LRRN3, KIAA0716, and LAMB1 as autism candidate genes. *BMC Med Genet* **2004**, *5*, 12.

(154) Brzustowicz, L. M.; Simone, J.; Mohseni, P.; Hayter, J. E.; Hodgkinson, K. A. et al. Linkage disequilibrium mapping of schizophrenia susceptibility to the CAPON region of chromosome 1q22. *Am J Hum Genet* **2004**, *74*, 1057-1063.

(155) Sunnucks, P. Efficient genetic markers for population biology. *Trends in Ecology and Evolution* **2000**, *15*, 199-203.

(156) Xiong, M.; Jin, L. Comparison of the power and accuracy of biallelic and microsatellite markers in population-based gene-mapping methods. *American Journal of Human Genetics* **1999**, *64*, 629-640.

(157) Sham, P.; Zhao, J. H.; Curtis, D. The effect of marker characteristics on the power to detect linkage disequilibrium due to single or multiple ancestral mutations. *Annals of Human Genetics* **2000**, *64*, 161-169.

(158) Chapman, N. H.; Wijsman, E. M. Genome screens using linkage disequilibrium tests: optimal marker characteristics and feasibility. *American Journal of Human Genetics* **1998**, *63*, 1872-1885.

(159) Bonnen, P. E.; Story, M. D.; Ashorn, C. L.; Buchholz, T. A.; Weil, M. M. et al. Haplotypes at ATM identify coding-sequence variation and indicate a region of extensive linkage disequilibrium. *American Journal of Human Genetics* **2000**, *67*, 1437-1451.

(160) John, S.; Shephard, N.; Liu, G.; Zeggini, E.; Cao, M. et al. Whole-genome scan, in a complex disease, using 11,245 single-nucleotide polymorphisms: comparison with microsatellites. *Am J Hum Genet* **2004**, *75*, 54-64.

(161) Middleton, F. A.; Pato, M. T.; Gentile, K. L.; Morley, C. P.; Zhao, X. et al. Genomewide linkage analysis of bipolar disorder by use of a high-density single-nucleotide-polymorphism (SNP) genotyping assay: a comparison with microsatellite marker assays and finding of significant linkage to chromosome 6q22. *Am J Hum Genet* **2004**, *74*, 886-897.

(162) Plomin, R.; McClearn, G. E.; Smith, D. L.; Vignetti, S.; Chorney, M. J. et al. DNA markers associated with high versus low IQ: the IQ Quantitative Trait Loci (QTL) Project. *Behavior Genetics* **1994**, *24*, 107-118.

(163) Plomin, R.; McClearn, G. E.; Smith, D. L.; Skuder, P.; Vignetti, S. et al. Allelic associations between 100 DNA markers and high versus low IQ. *Intelligence* **1995**, *21*, 31-48.

(164) Chorney, M. J.; Chorney, K.; N., S.; Owen, M. J.; Daniels, J. et al. A quantitative trait locus (QTL) associated with cognitive ability in children. *Psychological Science* **1998**, *9*, 1-8.

(165) Hill, L.; Chorney, M. J.; Chorney, K.; Craig, I.; Fisher, P. J. et al. IGF2R and cognitive ability. *Molecular Psychiatry* **1999**, *4*.

(166) Tolman, E. C. *Purposive Behavior in Animals and Men*; Appleton-Century: New York, 1932.

(167) Cooper, R. M.; Zubek, J. P. Effects of enriched and restricted early environments on the learning ability of bright and dull rats. *Canadian Journal of Psychology* **1958**, *12*, 159-164.

(168) Henderson, N. D. Relative effects of early rearing environment on discrimination learning in housemice. *Journal of Comparative and Physiological Psychology* **1972**, *72*, 505-511.

(169) Kandel, E. R. *Behavioral biology of Aplysia, a contribution to the comparative study of opisthobranch mollusks*; Freeman: San Francisco, 1979.

(170) Pittenger, C.; Kandel, E. R. In search of general mechanisms for long-lasting plasticity: Aplysia and the hippocampus. *Philos Trans R Soc Lond B Biol Sci* **2003**, *358*, 757-763.

(171) Tsien, J. Z.; Chen, D. F.; Gerber, D.; Tom, C.; Mercer, E. H. et al. Subregion- and cell type-restricted gene knockout in mouse brain. *Cell* **1996**, *87*, 1317-1326.

(172) Chan, C. S.; Weeber, E. J.; Kurup, S.; Sweatt, J. D.; Davis, R. L. Integrin requirement for hippocampal synaptic plasticity and spatial memory. *J Neurosci* **2003**, *23*, 7107-7116.

(173) Bukalo, O.; Fentrop, N.; Lee, A. Y.; Salmen, B.; Law, J. W. et al. Conditional ablation of the neural cell adhesion molecule reduces precision of spatial learning, long-term potentiation, and depression in the CA1 subfield of mouse hippocampus. *J Neurosci* **2004**, *24*, 1565-1577.

(174) Meng, Y.; Zhang, Y.; Tregoubov, V.; Janus, C.; Cruz, L. et al. Abnormal spine morphology and enhanced LTP in LIMK-1 knockout mice. *Neuron* **2002**, *35*, 121-133.

(175) Geiger, J. R.; Melcher, T.; Koh, D. S.; Sakmann, B.; Seeburg, P. H. et al. Relative abundance of subunit mRNAs determines gating and Ca2+ permeability of AMPA receptors in principal neurons and interneurons in rat CNS. *Neuron* **1995**, *15*, 193-204.

(176) Jia, Z.; Agopyan, N.; Miu, P.; Xiong, Z.; Henderson, J. et al. Enhanced LTP in mice deficient in the AMPA receptor GluR2. *Neuron* **1996**, *17*, 945-956.

(177) Tang, Y. P.; Shimizu, E.; Dube, G. R.; Rampon, C.; Kerchner, G. A. et al. Genetic enhancement of learning and memory in mice. *Nature* **1999**, *401*, 63-69.

(178) Jamot, L.; Matthes, H. W.; Simonin, F.; Kieffer, B. L.; Roder, J. C. Differential involvement of the mu and kappa opioid receptors in spatial learning. *Genes Brain Behav* **2003**, *2*, 80-92.

(179) Avital, A.; Goshen, I.; Kamsler, A.; Segal, M.; Iverfeldt, K. et al. Impaired interleukin-1 signaling is associated with deficits in hippocampal memory processes and neural plasticity. *Hippocampus* **2003**, *13*, 826-834.

(180) Rissman, E. F.; Heck, A. L.; Leonard, J. E.; Shupnik, M. A.; Gustafsson, J. A. Disruption of estrogen receptor beta gene impairs spatial learning in female mice. *Proc Natl Acad Sci U S A* **2002**, *99*, 3996-4001.

(181) Wechsler, D. *Manual for the Wechsler Intelligence Scale for Children - Revised*; Psychological Corporation: New York, 1974.

(182) Lubinski, D.; Benbow, C. The first three decades of a planned 50-year study of intellectual talent. *Beyond Terman: Longitudinal studies in contemporary gifted education*; Ablex: Norwood, NJ, 1994; pp 255-281.

(183) Wechsler, D. *Wechsler Adult Intelligence Scale-III*; Psychological Corporation: New York, 1997.

(184) Don, R. H.; Cox, P. T.; Wainwright, B. J.; Baker, K.; Mattick, J. S. 'Touchdown' PCR to circumvent spurious priming during gene amplification. *Nucleic Acids Res* **1991**, *19*, 4008.

(185) Oefner, P. J.; Underhill, P. A. DNA mutation detection using denaturing high-performance liquid chromatography (DHPLC). *Current Protocols in Human Genetics*; John Wiley and Sons, Inc.: New York, 1998; pp 10.11-10.12.

(186) Sanger, F.; Nicklen, S.; Coulson, A. R. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A* **1977**, *74*, 5463-5467.

(187) Beckman, J. S.; Weber, J. L. Survey of human and rat microsatellites. *Genomics* **1992**, *12*, 627-631.

(188) Liu, Q.; Thorland, E. C.; Sommer, S. S. Inhibition of PCR amplification by a point mutation downstream of a primer. *Biotechniques* **1997**, *22*, 292-294, 296, 298, passim.

(189) Haff, L. A.; Smirnov, I. P. Single-nucleotide polymorphism identification assays using a thermostable DNA polymerase and delayed extraction MALDI-TOF mass spectrometry. *Genome Res* **1997**, *7*, 378-388.

(190) Jorde, L. B.; Watkins, W. S.; Carlson, M.; Groden, J.; Albertsen, H. et al. Linkage disequilibrium predicts physical distance in the adenomatous polyposis coli region. *American Journal of Human Genetics* **1994**, *54*, 884-898.

(191) Barcellos, L. F.; Klitz, W.; Field, L. L.; Tobias, R.; Bowcock, A. M. et al. Association mapping of disease loci, by use of a pooled DNA genomic screen. *American Journal of Human Genetics* **1997**, *61*, 734-747.

(192) Daniels, J.; Holmans, P.; Williams, N. M.; Turic, D.; McGuffin, P. et al. A simple method for analysing microsatellite allele image patterns from DNA pools and its application to allelic association studies. *American Journal of Human Genetics* **1998**, *62*, 1189-1197.

(193) Fisher, P. J.; Turic, D.; Williams, N. M.; McGuffin, P.; Asherson, P. et al. DNA pooling identifies QTLs on chromosome 4 for general cognitive ability in children. *Human Molecular Genetics* **1999**, *8*, 915-922.

(194) Lipkin, E.; Mosig, M. O.; Darvasi, A.; Ezra, E.; A., S. et al. Quantitative trait locus mapping in dairy cattle by means of selective milk DNA pooling using dinucleotide microsatellite markers: analysis of milk protein percentage. *Genetics* **1998**, *149*, 1557-1567.

(195) Darvasi, A.; Soller, M. Selective DNA Pooling for Determination of Linkage between a Molecular Marker and a Quantitative Trait Locus. *Genetics* **1994**, *138*, 1365-1373.

(196) Tautz, D. Hypervariability of simple sequences as a general source for polymorphic DNA markers. *Nucleic Acids Research* **1989**, *25*, 6463-6471.

(197) Hauge, X. Y.; Litt, M. A study of the origin of 'shadow bands' seen when typing dinucleotide repeat polymorphisms by the PCR. *Human Molecular Genetics* **1993**, *2*, 411-415.

(198) Demers, D. B.; Curry, E. T.; Egholm, M.; Sozer, A. C. Enhanced PCR amplification of VNTR locus D1S80 using peptide nucleic acid (PNA). *Nucleic Acids Research* **1995**, *23*, 3050-3055.

(199) Clark, J. M. Novel non-templated nucleotide addition-reactions catalysed by procaryotic and eucaryotic DNA-polymerases. *Nucleic Acids Research* **1988**, *16*, 9677-9686.

(200) Hu, G. X. DNA polymerase-catalysed addition of nontemplated extra nucleotides to the 3' end of a DNA fragment. *DNA and Cell Biology* **1993**, *12*, 763-770.

(201) Smith, J. R.; Carpten, J. D.; Brownstein, M. J.; Ghosh, S.; Magnuson, V. L. et al. Approach to genotyping errors by nontemplated nucleotide addition by Taq DNA polymerase. *Genome Research* **1995**, *5*, 312-317.

(202) Pacek, P.; Sajantila, A.; Syvanne, A.-C. Determination of allele frequencies at loci with length polymorphism by quantitative analysis of DNA amplified from pooled samples. *PCR Methods* **1993**, *2*, 313-317.

(203) Khatib, H.; Darvasi, A.; Plotski, Y.; Soller, M. Determining relative microsatellite allele frequencies in pooled DNA samples. *PCR Methods* **1994**, *4*, 13-18.

(204) Walsh, P. S.; Erlich, H. A.; Higuchi, R. Preferential PCR amplification of alleles: mechanisms and solutions. *PCR Methods and Applications* **1992**, *1*, 241-250.

(205) Mutter, G. L.; Boynton, K. A. PCR bias in amplification of androgen receptor alleles, a trinucleotide repeat marker used in clonality studies. *Nucleic Acids Research* **1995**, *23*, 1411-1418.

(206) Kaiser, R.; Tremblay, P.-B.; Roots, I.; Brockmoller, J. Validity of PCR with emphasis on variable number of tandem repeat analysis. *Clinical Biochemistry* **2002**, *35*, 49-56.

(207) Brownstein, M. J.; Carpten, J. D.; Smith, J. R. Modulation of non-templated nucleotide addition by Taq DNA polymerase: primer modifications that facilitate genotyping. *Biotechniques* **1996**, *20*, 1004-1006, 1008-1010.

(208) Ginot, F.; Bordelais, I.; Nguyen, S.; Gyapay, G. Correction of some genotyping errors in automated fluorescent microsatellite analysis by enzymatic removal of one base overhangs. *Nucleic Acids Research* **1996**, *24*, 540-541.

(209) LeDuc, C.; Miller, P.; Lichter, J.; Parry, P. Batched analysis of genotypes. *PCR Methods and Applications* **1995**, *4*, 331-336.

(210) Perlin, M. W.; Lancia, G.; Ng, S. K. Toward fully automated genotyping: genotyping microsatellite markers by deconvolution. *American Journal of Human Genetics* **1995**, *57*, 1199-1210.

(211) Risch, N.; Teng, J. The relative power of family-based and case-control designs for linkage disequilibrium studies of complex human diseases: I DNA pooling. *Genome Research* **1998**, *8*, 1273-1288.

(212) Sham, P. C.; Chen, S. S.; Purcell, S.; Hewitt, J. K. Power of linkage versus association analysis of quantitative traits, by use of variance-components models, for sibship data. *American Journal of Human Genetics* **2000**, *66*, 1616-1630.

(213) Plomin, R.; Hill, L.; Craig, I. W.; McGuffin, P.; Purcell, S. et al. A genome-wide scan of 1847 DNA markers for allelic associations with general cognitive ability: A five-stage design using DNA pooling and extreme selected groups. *Behavior Genetics* **2001**, *31*, 497-509.

(214) Thompson, E. A.; Neel, J. V. Allelic disequilibrium and allele frequency distribution as a function of social and demographic history. *Am J Hum Genet* **1997**, *60*, 197-204.

(215) Kendler, K. S.; MacLean, C. J.; Ma, Y.; O'Neill, F. A.; Walsh, D. et al. Marker-to-marker linkage disequilibrium on chromosomes 5q, 6p, and 8p in Irish high-density schizophrenia pedigrees. *Am J Med Genet* **1999**, *88*, 29-33.

(216) Wang, G. L.; Paterson, A. H. Assesment of DNA pooling strategies for mapping of QTLs. *Theoretical Applied Genetics* **1994**, *88*, 355-361.

(217) Allsopp, T. E. Ageing brains and ebbing synapses. *Trends in Neurosciences* **1998**, *21*, 503-504.

(218) Owen, M. J.; Liddell, M.; McGuffin, P. Alzheimer's disease: an association with apolipoprotein E4 may help unlock the puzzle. *British Medical Journal* **1994**, *308*, 672-673.

(219) Corder, E. H.; Saunders, A. M.; Strittmatter, W. J.; Schmechel, D. E.; Gaskell, P. C. et al. Gene dose of apolipoprotein E type 4 allele and the risk of Alzheimer's disease in late onset families. *Science* **1993**, *261*, 921-923.

(220) Feskens, E. J. M.; Havekes, L. M.; Kalmijn, S.; Knijff, P. d.; Launer, L. J. et al. Apolipoprotein e4 and cognitive decline in elderly men. *British Medical Journal* **1994**, *309*, 1202-1206.

(221) Salo, A.; Ylikoski, R.; Verkkoniemi, A.; Polvikoski, T.; Juva, K. et al. Does apolipoprotein E influence learning and memory in the nondemented oldest old? *International Psychogeriatrics* **2001**, *13*, 451-459.

(222) Slooter, A. J. C.; van Duijn, C. M.; Bots, M. L.; Ott, A.; Breteler, M. B. et al. Apolipoprotein E genotype, atherosclerosis, and cognitive decline: the Rotterdam study. *Journal of Neural Transmission-Supplement* **1998**, 17-29.

(223) Flory, J. D.; Manuck, S. B.; Ferrell, R. E.; Ryan, C. M.; Muldoon, M. F. Memory performance and the apolipoprotein E polymorphism in a community sample of middle-aged adults. *American Journal of Medical Genetics* **2000**, *96*, 707-711.

(224) Fillenbaum, G. G.; Landerman, L. R.; Blazer, D. G.; Saunders, A. M.; Harris, T. B. et al. The relationship of APOE genotype to cognitive functioning in older African-American and Caucasian community residents. *Journal of the American Geriatrics Society* **2001**, *49*, 1148-1155.

(225) Mahley, R. W.; Nathan, B. P.; Pitas, R. E. Apolipoprotein E - Structure, function, and possible roles in Alzheimer's disease. *Neurobiology of Alzheimer's Disease* **1996**, *777*, 139-145.

(226) Mercier, G.; Lucotte, G. Apolipoprotein E epsilon 4 allele, the major risk factor for Alzheimer's disease, is also involved in temporal lobe epilepsy. *Alzheimers Reports* **2000**, *3*, 265-268.

(227) Van Everbroeck, B.; Croes, E. A.; Pals, P.; Dermaut, B.; Jansen, G. et al. Influence of the prion protein and the apolipoprotein E genotype on the Creutzfeldt-Jakob Disease phenotype. *Neuroscience Letters* **2001**, *313*, 69-72.

(228) Kuusi, T.; Nieminen, M. S.; Ehnholm, C.; Ykijarvinen, H.; Valle, M. et al. Apoprotein-E Polymorphism and Coronary-Artery Disease - Increased Prevalence of Apolipoprotein E-4 in Angiographically Verified Coronary Patients. *Arteriosclerosis* **1989**, *9*, 237-241.

(229) Nieminen, M. S.; Mattila, K. J.; Aaltosetala, K.; Kuusi, T.; Kontula, K. et al. Lipoproteins and Their Genetic-Variation in Subjects with and without Angiographically Verified Coronary-Artery Disease. *Arteriosclerosis and Thrombosis* **1992**, *12*, 58-69.

(230) Gotthardt, M.; Trommsdorff, M.; Nevitt, M. F.; Shelton, J.; Richardson, J. A. et al. Interactions of the low density lipoprotein receptor gene family with cytosolic adaptor and scaffold proteins suggest diverse biological functions in cellular communication and signal transduction. *Journal of Biological Chemistry* **2000**, *275*, 25616-25624.

(231) Stockinger, W.; Brandes, C.; Fasching, D.; Hermann, M.; Gotthardt, M. et al. The reelin receptor ApoER2 recruits JNK-interacting proteins-1 and-2. *Journal of Biological Chemistry* **2000**, *275*, 25625-25632.

(232) Teter, B.; Xu, P.-T.; Gilbert, J. R.; Roses, A. D.; Galasko, D. et al. Defective neuronal sprouting by human apolipoprotein E4 is a gain-of-negative function. *Journal of Neuroscience Research* **2002**, *68*, 331-336.

(233) Ignatius, M. J.; Gebickeharter, P. J.; Skene, J. H. P.; Schilling, J. W.; Weisgraber, K. H. et al. Expression of Apolipoprotein-E During Nerve Degeneration and Regeneration. *Proceedings of the National Academy of Sciences of the United States of America* **1986**, *83*, 1125-1129.

(234) Chen, Y.; Lomnitski, L.; Michaelson, D. M.; Shohami, E. Motor and cognitive deficits in apolipoprotein E-deficient mice after closed head injury. *Neuroscience* **1997**, *80*, 1225-1262.

(235) Teasdale, G. M.; Nicoll, J. A.; Murray, G.; Fiddes, M. Association of apolipoprotein E polymorphism with outcome after head injury. *Lancet* **1997**, *350*, 1069-1071.

(236) Smith, C.; Graham, D. I.; Murray, L.; Nicoll, J. A. Association of APOE polymorphisms and pathological features in traumatic brain injury. *Neuropathology and Applied Neurobiology* **2002**, *28*, 151-152.

(237) Sabo, T.; Lomnitski, L.; Nyska, A.; Beni, S.; Maronpot, R. R. et al. Susceptibility of transgenic mice expressing human apolipoprotein E to closed head injury: the allele E3 is neuroprotective whereas E4 increases fatalities. *Neuroscience* **2000**, *101*, 879-884.

(238) Masliah, E.; Mallory, M.; Ge, N.; Alford, M.; Veinbergs, I. et al. Neurodegeneration in the central nervous system of apoE-deficient mice. *Experimental Neurology* **1995**, *136*, 107-122.

(239) Cambon, K.; Davies, H. A.; Stewart, M. G. Synaptic loss is accompanied by an increase in synaptic area in the dentate gyrus of aged human apolipoprotein E4 transgenic mice. *Neuroscience* **2000**, *97*, 685-692.

(240) Nathan, B. P.; Bellosta, S.; Sanan, D.; Weisgraber, K. H.; Mahley, R. W. et al. Differential effects of apolipoproteins E3 and E4 on neuronal growth *in vitro*. *Science* **1994**, *264*, 850-852.

(241) Mahley, R. W. Neurobiology of Alzheimer's disease. *Annals of the New York Academy of Science* **1996**, *777*, 139-145.

(242) Buttini, M.; Orth, M.; Bellosta, S.; Akeefe, H.; Pitas, R. E. et al. Expression of human apolipoprotein E3 or E4 in the brains of Apoe(-/-) mice: Isoform-specific effects on neurodegeneration. *Journal of Neuroscience* **1999**, *19*, 4867-4880.

(243) White, F.; Nicoll, J. A. R.; Roses, A. D.; Horsburgh, K. Impaired neuronal plasticity in transgenic mice expressing human apolipoprotein E4 compared to E3 in a model of entorhinal cortex lesion. *Neurobiology of Disease* **2001**, *8*, 611-625.

(244) Nathan, B. P.; Chang, K.; Bellosta, S.; Brisch, E.; Ge, N. et al. The inhibitory effect of apolipoprotein E4 on neurite outgrowth is associated with microtubule depolymerisation. *Journal of Biological Chemistry* **1995**, *270*, 19791-19799.

(245) Buttini, M.; Akeefe, H.; Lin, C.; Mahley, R. W.; Pitas, R. E. et al. Dominant negative effects of apolipoprotein E4 revealed in transgenic models of neurodegenerative disease. *Neuroscience* **2000**, *97*, 207-210.

(246) Terry, R. D.; Masliah, E.; Salmon, D. P.; Butters, N.; DeTeresa, R. et al. Physical basis of cognitive alterations in Alzheimer's disease: synapse loss is the major correlate of cognitive impairment. *Annals of Neurology* **1991**, *30*, 572-580.

(247) Bullido, M. J.; Artiga, M. J.; Recuero, M.; Sastre, I.; Garcia, M. A. et al. A polymorphism in the regulatory region of APOE associated with risk for Alzheimer's dementia. *Nature Genetics* **1998**, *18*, 69-71.

(248) Lambert, J. C.; Berr, C.; Pasquier, F.; Delacourte, A.; Frigard, B. et al. Pronounced impact of Th1/E47cs mutation compared with -491 AT mutation on neural APOE gene expression and risk of developing Alzheimer's disease. *Human Molecular Genetics* **1998**, *7*, 1511-1516.

(249) Lambert, J. C.; Perez-Tur, J.; Dupire, M. J.; Galasko, D.; Mann, D. et al. Distortion of allelic expression of apolipoprotein E in Alzheimer's disease. *Hum Mol Genet* **1997**, *6*, 2151-2154.

(250) Wenham, P. R.; Price, W. H.; Blundell, G. Apolipoprotein E genotyping by one-stage PCR. *The Lancet* **1991**, *337*, 1158-1159.

(251) Sham, P. C.; Curtis, D. Monte-Carlo Tests for Associations between Disease and Alleles at Highly Polymorphic Loci. *Annals of Human Genetics* **1995**, *59*, 97-105.

(252) Zhao, J. H.; Curtis, D.; Sham, P. C. Model-free analysis and permutation tests for allelic associations. *Human Heredity* **2000**, *50*, 133-139.

(253) Artiga, M. J.; Bullido, M. J.; Frank, A.; Sastre, I.; Recuero, M. et al. Risk for Alzheimer's disease correlates with transcriptional activity of the APOE gene. *Human Molecular Genetics* **1998**, *7*, 1887-1892.

(254) Barnes, C. A. Memory changes during normal aging: neorobiological correlates. *Neurobiology of Learning and Memory*; Academic Press: London, 1998; pp 247-287.

(255) O'Keefe, J.; Dostrovsky, J. The hippocampus as a spatial map. Preliminary evidence from unit activity in the freely-moving rat. *Brain Res* **1971**, *34*, 171-175.

(256) O'Keefe, J. Two spatial systems in the rat brain--implications for the neural basis of learning and memory. *Prog Brain Res* **1983**, *58*, 453-464.

(257) Scoville, W. B.; Milner, B. Loss of recent memory after bilateral hippocampal lesions. *Journal of Neurology, Neurosurgery, and Psychiatry* **1957**, *20*, 11-21.

(258) Corkin, S. Lasting consequences of bilateral medial temporal lobectomy: Clinical course and experimental finds in H. M. *Seminars in Neurology* **1984**, *4*, 249-259.

(259) Smith, M. L.; Milner, B. The role of the right hippocampus in the recall of spatial location. *Neuropsychologia* **1981**, *19*, 781-793.

(260) Zola-Morgan, S.; Squire, L. R.; Amaral, D. G. Human amnesia and the medial temporal region: Enduring memory impairment following a bilateral lesion limited to field CA1 of the hippocampus. *Journal of Neuroscience* **1986**, *6*, 2950-2967.

(261) Kim, H.-J.; Chae, S.-C.; Lee, D.-K.; Chomy, B.; Lee, S. C. et al. Selective neuronal degeneration induced by soluble oligomeric amyloid beta-protein. *The FASEB Journal Express Article 10.1096/fj.01-0987fje* **2002**, *Published online.*

(262) West, M. J.; Kawas, C. H.; Martin, L. J.; Troncoso, J. C. The CA1 region of the human hippocampus is a hot spot in Alzheimer's disease. *Annals of the New York Academy of Science* **2000**, *908*, 255-259.

(263) Bobinski, M.; de Leon, M. J.; Tarnawski, M.; Wegiel, J.; Reisberg, B. et al. Neuronal and volume loss in CA1 of the hippocampal formation uniquely predicts duration and severity of Alzheimer disease. *Brain Research* **1998**, *805*, 267-269.

(264) Longstaff, A. *Instant Notes in Neuroscience*; Bios Scientific Publishers: Oxford, UK, 2000.

(265) Malenka, R. C.; Nicoll, R. A. Long-term potentiation - A decade of progress? *Science* **1999**, *17*, 1870-1874.

(266) Bliss, T. V.; Lomo, T. Long-lasting potentiation of synaptic transmission in the dentate area of the anaesthetized rabbit following stimulation of the perforant path. *Journal of Physiology* **1973**, *232*, 331-356.

(267) Bear, M. F.; Abraham, W. C. Long-term depression in hippocampus. *Annual Review of Neuroscience* **1996**, *19*, 437-462.

(268) Manahan-Vaughan, D.; Braunewell, K. H. Novelty aquisition is associated with induction of hippocampal long-term depression. *Proceedings of the National Academy of Sciences of the United States of America* **1999**, *96*, 8739-8744.

(269) Nakao, K.; Ikegaya, Y.; Yamada, M. K.; Nishiyama, N.; Matsuki, N. Hippocampal long-term depression as an index of spatial working memory. *European Journal of Neuroscience* **2002**, *16*, 970-974.

(270) Malenka, R. C.; Nicoll, R. A. NMDA-receptor-dependent synaptic plasticity: multiple forms and mechanisms. *Trends in Neurosciences* **1993**, *16*, 521-527.

(271) Zalutsky, R. A.; Nicoll, J. A. Comparison of two forms of long-term potentiation in single hippocampal neurons. *Science* **1990**, *248*, 1619-1624.

(272) Langdon, R. B.; Johnson, J. W.; Barrionuevo, G. Posttetanic potentiation and presynaptically induced long-term potentiation at the mossy fiber synapse in rat hippocampus. *Journal of Neurobiology* **1995**, *26*, 370-385.

(273) Hebb, D. O. *The organisation of behavior*; John Wiley: New York, 1949.

(274) Lisman, J.; Schulman, H.; Cline, H. The molecular basis of CaMKII function in synaptic and behavioural memory. *Nature Reviews / Neuroscience* **2002**, *3*, 175-190.

(275) Bennett, M. K.; Erondu, N. E.; Kennedy, M. B. Purification and characterisation of a calmodulin-dependent protein kinase that is highly concentrated in brain. *Journal of Biological Chemistry* **1983**, *258*, 12735-12744.

(276) Brocke, L.; Chiang, L. W.; Wagner, P. D.; Schulman, H. Functional implications of the subunit composition of neuronal CaM kinase II. *Journal of Biological Chemistry* **1999**, *274*, 22713-22722.

(277) Hudmon, A.; Schulman, H. Structure-function of the multifunctional $Ca^{2+}$/calmodulin-dependent protein kinase II. *Biochemical Journal* **2002**, *364*, 593-611.

(278) Meyer, T.; Hanson, P. I.; Stryer, L.; Schulman, H. Calmodulin trapping by calcium-calmodulin-dependent protein kinase. *Science* **1992**, *256*, 1199-1201.

(279) Hanson, P. I.; Schulman, H. Neuronal $Ca^{2+}$/calmodulin-dependent protein kinases. *Annual Review of Biochemistry* **1992**, *61*, 559-601.

(280) Strack, S.; Barban, M. A.; Wadzinski, B. E.; Colbran, R. J. Differential inactivation of postsynaptic density-associated and soluble Ca2+/calmodulin-dependent protein kinase II by protein phosphatases 1 and 2A. *Journal of Neurochemistry* **1997**, *68*, 2119-2128.

(281) Cohen, P. The structure and regulation of protein phosphatases. *Annual Review of Biochemistry* **1989**, *58*, 453-508.

(282) Blitzer, R. D.; Connor, J. H.; Brown, G. P.; Wong, T.; Shenolikar, S. et al. Gating of CaMKII by cAMP-regulated protein phosphatase activity during LTP. *Science* **1998**, *280*, 1940-1943.

(283) De Koninck, P.; Schulman, H. Sensitivity of CaM kinase II to the frequency of Ca2+ oscillations. *Science* **1998**, *279*, 227-230.

(284) Mayford, M.; Wang, J.; Kandel, E. R.; O'Dell, T. J. CaMKII regulates the frequency-response function of hippocampal synapses for the production of both LTD and LTP. *Cell* **1995**, *81*, 891-904.

(285) Lisman, J. The CaM kinase II hypothesis for the storage of synaptic memory. *Trends in Neurosciences* **1994**, *17*, 406-412.

(286) Malenka, R. C.; Kauer, J. A.; Perkel, D. J.; Mauk, M. D.; Kelly, P. T. et al. An essential role for postsynaptic calmodulin and protein kinase activity in long-term potentiation. *Nature* **1989**, *340*, 554-557.

(287) Malinow, R.; Schulman, H.; Tsien, R. W. Inhibition of postsynaptic PKC or CaMKII blocks induction but not expression of LTP. *Science* **1989**, *245*, 862-866.

(288) Hvalby, O.; Hemmings Jr, H. C.; Paulsen, O.; Czernik, A. J.; Nairn, A. C. et al. Specificity of protein kinase inhibitor peptides and induction of long-term potentiation. *Proceedings of the National Academy of Sciences of the United States of America* **1994**, *91*, 4761-4765.

(289) Silva, A. J.; Stevens, C. F.; Tonegawa, S.; Wang, Y. Deficient hippocampal long-term potentiation in alpha-calcium-calmodulin kinase II mutant mice. *Science* **1992**, *257*, 201-206.

(290) Mayford, M.; Bach, M. E.; Huang, Y.-Y.; Wang, L.; Hawkins, R. D. et al. Control of memory formation through regulated expression of a CaMKII transgene. *Science* **1996**, *274*, 1678-1683.

(291) Giese, K. P.; Fedorov, N. B.; Filipkowski, R. K.; Silva, A. J. Autophosphorylation at $Thr^{286}$ of the αCalcium-Calmodulin Kinase II in LTP and learning. *Science* **1998**, *279*, 870-873.

(292) Kennedy, M. B. The postsynaptic density at glutamatergic synapses. *Trends in Neurosciences* **1997**, *20*, 264-268.

(293) Kennedy, M. B.; Bennett, M. K.; Erondu, N. E. Biochemical and immunochemical evidence that the "major postsynaptic density protein" is a subunit of a calmodulin-dependent protein kinase. *Proceedings of the National Academy of Sciences of the United States of America* **1983**, *80*, 7357-7361.

(294) Leonard, A. S.; Lim, I. A.; Hemsworth, D. E.; Horne, M. C.; Hell, J. W. Calcium/calmodulin-dependent protein kinase II is associated with the N-methyl-D-aspartate receptor. *Proceedings of the National Academy of Sciences of the United States of America* **1999**, *96*, 3239-3244.

(295) Gardoni, F.; Caputi, A.; Cimino, M.; Pastorino, L.; Cattabeni, F. et al. Calcium/calmodulin-dependent protein kinase II is associated with NR2A/B subunits of NMDA receptor in postsynaptic densities. *Journal of Neurochenistry* **1998**, *71*, 1733-1741.

(296) Barria, A.; Muller, D.; Derkach, V.; Griffith, L. C.; Soderling, T. R. Regulatory phosphorylation of AMPA-type glutamate receptors by CaM-KII during long-term potentiation. *Science* **1997**, *276*, 2042-2045.

(297) Zamanillo, D.; Sprengel, R.; Hvalby, Ø.; Jensen, V.; Burnashev, N. et al. Importance of AMPA Receptors for Hippocampal Synaptic Plasticity But Not for Spatial Learning. *Science* **1999**, *284*, 1805-1811.

(298) Hayashi, Y.; Shi, S.-H.; Esteban, J. A.; Piccini, A.; Poncer, J.-C. et al. Driving AMPA receptors into synapses by LTP and CaMKII: requirement for GluR1 and PDZ domain interaction. *Science* **2000**, *287*, 2262-2267.

(299) Luscher, C.; Nicoll, R. A.; Malenka, R. C.; Muller, D. Synaptic plasticity and dynamic modulation of the postsynaptic membrane. *Nature Neuroscience* **2000**, *3*, 545-550.

(300) Rebhan, M.; Chalifa-Caspi, V.; Prilusky, J.; Lancet, D. GeneCards: encyclopedia for genes, proteins and diseases. GeneCard for [CAMK2G] [22 Oct 2002]; Weizmann Institute of Science, Bioinformatics Unit and Genome Center: Rehovot, Israel, 1997.

(301) Appel, R. D.; Bairoch, A.; Hochstrasser, D. F. A new generation of information retrieval tools for biologists: the example of the ExPASy WWW server. *Trends in Biochemical Sciences* **1994**, *19*, 258-260.

(302) Miyashiro, K.; Dichter, M.; Eberwine, J. On the nature and differential distribution of mRNAs in hippocampal neurites: implications for neuronal functioning. *Proc Natl Acad Sci U S A* **1994**, *91*, 10800-10804.

(303) Crino, P. B.; Eberwine, J. Molecular characterization of the dendritic growth cone: regulated mRNA transport and local protein synthesis. *Neuron* **1996**, *17*, 1173-1187.

(304) Gardiol, A.; Racca, C.; Triller, A. Dendritic and postsynaptic protein synthetic machinery. *J Neurosci* **1999**, *19*, 168-179.

(305) Mayford, M.; Baranes, D.; Podsypanina, K.; Kandel, E. R. The 3'-untranslated region of CaMKII alpha is a cis-acting signal for the localization and translation of mRNA in dendrites. *Proc Natl Acad Sci U S A* **1996**, *93*, 13250-13255.

(306) Rook, M. S.; Lu, M.; Kosik, K. S. CaMKIIα 3' untranslated region-directed mRNA translocation in living neurons: visualization by GFP linkage. *J Neurosci* **2000**, *20*, 6385-6393.

(307) Mori, Y.; Imaizumi, K.; Katayama, T.; Yoneda, T.; Tohyama, M. Two cis-acting elements in the 3' untranslated region of alpha-CaMKII regulate its dendritic targeting. *Nat Neurosci* **2000**, *3*, 1079-1084.

(308) Wells, D. G.; Richter, J. D.; Fallon, J. R. Molecular mechanisms for activity-regulated protein synthesis in the synapto-dendritic compartment. *Curr Opin Neurobiol* **2000**, *10*, 132-137.

(309) Miller, S.; Yasuda, M.; Coats, J. K.; Jones, Y.; Martone, M. E. et al. Disruption of dendritic translation of CaMKIIalpha impairs stabilization of synaptic plasticity and memory consolidation. *Neuron* **2002**, *36*, 507-519.

(310) Rozen, S.; Skaletsky, H. J. Primer3 on the WWW for general users and for biologist programmers. *Bioinformatics Methods and Protocols: Methods in Molecular Biology*; Humana Press: Totowa, NJ, 2000; pp 365-386.

(311) Zhang, M. Q. Identification of human gene core promoters in silico. *Genome Res* **1998**, *8*, 319-326.

(312) Frith, M. C.; Hansen, U.; Weng, Z. Detection of cis-element clusters in higher eukaryotic DNA. *Bioinformatics* **2001**, *17*, 878-889.

(313) Norton, N.; Williams, N. M.; Williams, H. J.; Spurlock, G.; Kirov, G. et al. Universal, robust, highly quantitative SNP allele frequency measurement in DNA pools. *Hum Genet* **2002**, *110*, 471-478.

(314) Olson, N. J.; Masse, T.; Suzuki, T.; Chen, J.; Alam, D. et al. Functional identification of the promoter for the gene encoding the alpha subunit of calcium/calmodulin-dependent protein kinase II. *Proc Natl Acad Sci U S A* **1995**, *92*, 1659-1663.

(315) Petrill, S. A.; Plomin, R.; McClearn, G. E.; Smith, D. L.; Vignetti, S. et al. No association between general cognitive ability and the A1 allele of the D2 dopamine receptor gene. *Behav Genet* **1997**, *27*, 29-31.

(316) Turic, D.; Fisher, P. J.; Plomin, R.; Owen, M. J. No association between apolipoprotein E polymorphisms and general cognitive ability in children. *Neuroscience Letters* **2001**, *299*, 97-100.

(317) Plomin, R.; Turic, D.; Hill, L.; Turic, D.; Stephens, M. et al. A functional polymorphism in the succinate-semialdehyde dehydrogenase (aldehyde dehydrogenase 5 family, member A1) gene is associated with cognitive ability. *Molecular Psychiatry* **2003**.

(318) Visscher, P. M.; Tynan, M.; Whiteman, M. C.; Pattie, A.; White, I. et al. Lack of association between polymorphisms in angiotensin-converting-enzyme and methylenetetrahydrofolate reductase genes and normal cognitive ageing in humans. *Neurosci Lett* **2003**, *347*, 175-178.

(319) SLI-Consortium Highly significant linkage to the SLI1 locus in an expanded sample of individuals affected by specific language impairment. *Am J Hum Genet* **2004**, *74*, 1225-1238.