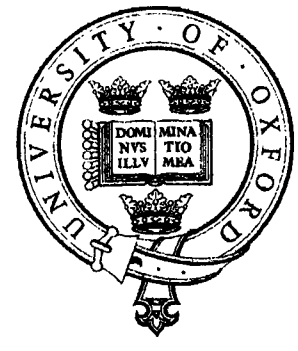


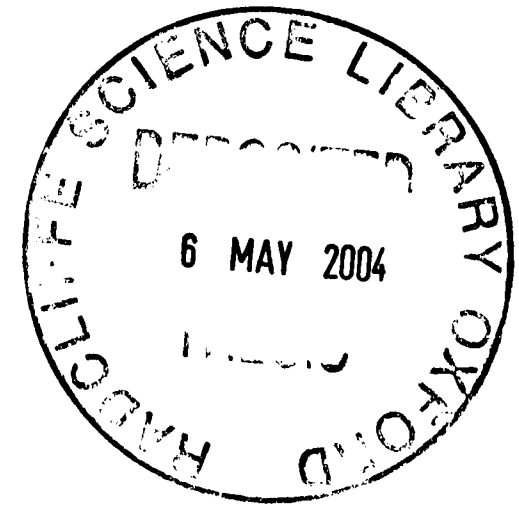
Common Genetic Variants of the *IFN-γ* and *IFNGR1* Regions - Disease Associations and Functional Properties

A thesis submitted for the degree of *Doctor of Philosophy*

Oliver Koch

Keble College, University of Oxford

Trinity 2003



Common Genetic Variants of the *IFN-γ* and *IFNGR1* Regions - Disease Associations and Functional Properties

Oliver Koch

Keble College, University of Oxford

Thesis submitted for the degree of *Doctor of Philosophy*, Trinity 2003

Abstract

There is growing evidence that susceptibility to many inflammatory and infectious diseases may be influenced by our genetic make up. Genetic variants in important immune genes may partially explain variation in susceptibility to common diseases. Interferon-γ (IFNγ) is one of the central mediators of the innate and adaptive immunity and has been implicated in a wide range of infectious and inflammatory disease processes. Severe disruptive mutations in coding regions of the IFN-γ receptor 1 gene (*IFNGR1*) have been found to be associated with fatal but very rare mycobacterial infections. This study looked at common polymorphisms in potentially regulatory non-coding regions of the *IFNγ* gene and the *IFNGR1* gene and investigated their association with susceptibility to severe malaria, a disease for which there have been indications of a genetic component to susceptibility. Malaria is one of the major causes of childhood deaths in Africa. IFNγ and its receptor have been shown to be critically involved in the host response to the malaria parasites. The promoter regions of *IFNGR1* and its neighbouring genes, located on chromosome 6q23, and *IFNγ* and its neighbours, on chromosome 12q14, were screened for polymorphisms. Haplotypes and linkage disequilibrium maps were constructed, signatures of natural selection were investigated, haplotype tagging SNPs were dentified, and association with disease was analysed. One of these preliminary results was a putative association between the *IFNGR1*-470ddel allele and susceptibility to severe malaria in

the Mandinka ethnic group. This allele was in strong linkage disequilibrium (LD) with markers which are a considerable distance away which might represent a signature of natural selection. To assess the potential functional significance of the *IFNGR1*-470 polymorphism, its effects on DNA-protein interactions and gene expression was investigated further in various cell lines. Evidence of tissue-specific nuclear protein binding to this site which seems to be involved in transcriptional regulation was observed.

Kosh 2003

Acknowledgements

The four-year process of taking this project from an idea to a finished thesis has been exhilarating. It has not, however, been painless and I have relied heavily on the support, understanding and expertise of those around me.

Foremost, I am deeply grateful to my supervisor, Dominic Kwiatkowski. Without his support, encouragement and guidance this thesis would have never materialised. I am genuinely indebted to his patience, intellectual rigor, tolerance and enthusiasm establishing an environment which is exciting and a privilege to work in.

I would like to thank all patients and their relatives who so kindly took part in this study and provided their DNA. Dominic Kwiatkowski, University of Oxford, Muminatou Jallow, Stanley Usen and colleagues at the Royal Victoria Hospital, The Gambia, and Margaret Pinder and Melanie Newport at the MRC Laboratories, The Gambia, were the key people for establishing the epidemiological study and collecting the DNA samples.

The logistics and the database for the resequencing project were established by Kirk Rockett. Kirk has been of immense help with so many major and minor things and spent an enormous amount of his own time helping to make things work for which I am tremendously grateful. Irina Udalova guided and encouraged me through the functional analysis. Her in-depth knowledge has been invaluable for me. She also performed the immunohistochemistry experiments. Others have helped in so many ways. Kate Rowling, Anna Richardson, Steve Laurie and Jonathan Wilson have helped archiving the DNA samples. Man-Suen Chan has established the software tools to allow rapid sequence analysis and facilitated access to haplotype analysis via a web-based tool. Neil Hanchard has provided the DNA samples of the UK families for the construction of haplotypes. Brendan Keating provided the CEPH DNA that was used for the resequencing project. My thanks also to Dr Jeremy Hull, for providing the chimpanzee DNA and more importantly an always lively and interesting debate.

There were endless discussions I benefited from and I would like to thank all members of the research group who were not only of great help but also

made me enjoy the last four years enormously.

This work was supported by scholarships of the Gottlieb Daimler- and Karl Benz-Foundation and the Medical Research Council for which I am profoundly grateful.

I am indebted to my parents for their long-standing support and tolerance. Finally, I am deeply grateful for the infinite patience of my wife Pax giving me the freedom to pursue this project.

Kosh 2003

To my parents

Contents

Косн 2003

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Genes and Immunity

Infectious pathogens are humans' greatest natural enemies. Of over 52 million deaths in 1996, over 17 million were due to infectious or parasitic diseases. Infectious and parasitic diseases accounted for 43% of the 40 million deaths in developing countries (WHO, World Health Report 1997). While cheap and safe vaccines are the the only practical way of reducing the global burden of infectious diseases, this has proven to be difficult for the biggest killers, HIV, tuberculosis and malaria. The pathogens for these diseases manage to evade the human immune system.

Although exposed to the same environment and the same pathogens there is great individual variability in how people respond to infection (Hill, 1999; Kwiatkowski, 2000). While some people resist infections successfully others seem far more vulnerable. Understanding the molecular mechanisms of these infections might help us in developing more efficient and better targeted drugs and vaccines.

A number of single gene defects with devastating consequences has been discovered although these diseases are generally extremely rare. Susceptibility to common multifactorial diseases is likely to be influenced by more subtle effects. Genetic factors can explain this to some extent. Although our genetic

make up is roughly 99.9% similar, it are those differences which contribute to each of us being unique (Kruglyak and Nickerson, 2001). The most common form of genetic variation is the substitution of one nucleotide for another - a so-called *single nucleotide polymorphism* (SNP). Polymorphisms of a frequency $\geqslant$1-2% are found on average every 1000-2000 nucleotides, suggesting that there might be several million variants throughout the whole genome (Cargill et al., 1999; Sachidanandam et al., 2001).

While it is widely recognised that genetic factors are important in common, multifactorial diseases such as cancer, ischaemic heart disease and diabetes mellitus, it is, however, a common misconception that the risk of developing a serious infection is entirely due to environmental and social factors. To assess genetic and environmental influences on adult mortality, Sorensen et al. (1988) followed 960 families that included children born during the period 1924 through 1926 who were placed early in life with adoptive parents unrelated to them. They found that the genetic component leading to premature death from infection was greater than that for cancer or for cardiovascular and cerebrovascular causes. Garrod (1931) was one of the first to suggest that infectious diseases may have been a major selective force in human evolution. Haldane (1949) recognised the importance infective pathogens have on the evolution of the human genome. He proposed that red-blood-cell disorders, such as thalassaemia might protect an individual from life-threatening infection with malaria (see also Haldane (1948)). Murphy (1993) compared genes common to both rodents and humans and found that genes involved in the inflammatory host response are much more diverse than other genes, suggesting that selection in many species has resulted from exposure to different infectious agents. There have been several twin studies suggesting heritability of susceptibility to various infectious disease, such tuberculosis (Diehl and Von Verschuer, 1936; Kallmann and Reisner, 1942; Comstock, 1978), leprosy (Chakravarti and Vogel, 1973), poliomyelitis (Herndon and Jennings, 1951) and hepatitis (Lin et al., 1989).
Furthermore, there been several studies showing differences between populations in susceptibility to infectious diseases. The notion that genetic dif-

ferences between West African ethnic groups may make a substantial contribution to malaria susceptibility is bolstered by evidence of strong protection against malaria among the Fulani of Burkina Faso, who seem markedly more resistant to severe malaria than do neighbouring ethnic groups (Modiano et al., 1996). The genetics of susceptibility to infectious disease have also been reviewed in Hill (1996); Weatherall et al. (1997); Abel and Dessein (1997); Cooke and Hill (2001).

There are over 100 examples of rare Mendelian conditions that predispose an individual to a particular infectious disease (see the Online Mendelian Inheritance in Man database at: `http://www.ncbi.nlm.nih.gov/omim/` and Forfar and Arneil's Textbook of Pediatrics (McIntosh et al., 2003) for further reading). Many primary immunodefiencies have been described (Buckley, 1994; Korthauer et al., 1993; DiSanto et al., 1993; Fischer and Arnaiz-Villena, 1995; Puck, 1994).

Several severe disruptive mutations in the coding region of the *IFNGR1* receptor gene leading to severe infections with environmental mycobacteria have been found (Newport et al., 1996; Jouanguy et al., 1996). These conditions with Mendelian inheritance are extremely rare. But there are also examples of more common polymorphisms which might have more subtle effects on disease susceptibility. For example common polymorphisms found in the promoter of the *CCR5* gene lead to acceleration of AIDS progression (Martin et al., 1998).

Recent advances in unravelling the sequence of the human genome (Lander et al., 2001; Venter et al., 2001) have lead to an increasing number of markers that could help mapping common infectious diseases. As resources are limited the challenge will be to determine which of these markers will be worth typing. Less than 1% of all single nucleotide polymorphisms result in variation in proteins (Venter et al., 2001), but many more in regulatory regions of the genome might influence the expression of proteins. One approach

to finding disease influencing mutations is therefore the investigation of candidate genes that are known to be involved in the disease process. When challenged by infection, the host rapidly mounts an innate immunological response involving cytokines and other inflammatory mediators. These immune mediators attack the microbe but are also largely responsible for the clinical symptoms of infection. The evolutionary problem is how to optimise this innate host response such that it is forceful enough to repel infectious pathogens without causing excessive inflammation that may kill the host. In the course of evolution, humans and their ancestors have been exposed to a huge range of different infectious agents, and there is turning out to be great diversity in gene promoter regions that regulate the transcription of cytokines and other inflamnatory mediators. This genetic variation may largely explain why individuals differ in their susceptibility to infection and to inflammatory diseases such as rheumatoid arthritis, inflammatory bowel disease and asthma. Understanding how these genetic variants might influence susceptibility to disease might ultimately lead to improvements in diagnosis, treatment and vaccine development.

1.2 Malaria

1.2.1 Malaria - A Global Problem

Malaria is the most important parasitic disease of man. Until the nineteenth century malaria was found in North America, Russia, and in various parts of Europe. While malaria has been eradicated from these areas and the incidence in areas such as China, the Middle East and India has fallen, a resurgence of the disease in the tropics has been observed (Bruce-Chwatt, 1988). Malaria infects about one tenth of the world's population annually, causing over 2.7 million deaths every year and is responsible for over one million childhood deaths each year in Africa alone (WHO, 1997).
Over 40% of the world population live in areas with significant malaria risk. And every year millions of travellers from countries where malaria is not

prevalent visit countries where malaria is endemic. In 1994, malaria risk of varying degrees existed in 100 countries and territories. In 92 of these, transmission included the malignant (*Plasmodium falciparum*) form of the disease. The parasite has made its way into 52 of 58 countries in Africa. Malaria is Africa's leading cause of under-five mortality (20%) and constitutes 10% of the continent's overall disease burden. It accounts for 40% of public health expenditure, 30-50% of inpatient admissions, and up to 50% of outpatient visits in areas with high malaria transmission (Krishna, 1997).

Malaria is a major constraint to social and economic development. One estimate puts the direct economic cost of malaria in Africa at 1.7 billion USD (data as published on the website of *The United States Agency for International Development*) but there is growing recognition that this is a huge underestimate of the real scale of the economic burden (Gallup and Sachs, 2001; Sachs and Malaney, 2002). Outpatient health care costs for malaria in many African countries account for up to 40% of all public health expenditures. Approximately 20% of disposable income within low-income African households is spent on malaria treatment. The direct costs of malaria include a combination of personal and public expenditures on both prevention and treatment of the disease. Personal expenditures include individual or family spending on insecticide treated mosquito nets, doctors' fees, anti-malarial drugs, transport to health facilities, support for the patient and sometimes an accompanying family member during hospital stays. Public expenditures include spending by government on maintaining health facilities and health care infrastructure, publicly managed vector control, education and research.

By far the greatest impact of malaria is in African children and pregnant women. Pregnant women infected with malarial parasites have an increased risk of maternal anaemia, abortion, stillbirth, prematurity, intrauterine growth retardation, and infants of low birthweight. A child is estimated to die from malaria every 12 or so seconds. This burden of mortality is not equally shared, falling most heavily on sub-Saharan Africa, where >90% of these deaths occur and 5% of children die from the disease before reaching

5 years; among the newborns of Africa an estimated 3 million suffer complications from low birth weight, including death, arising from malaria infection during pregnancy.

However, the spread of malaria is exacerbated by population movements, due to migration or wars, into malaria-endemic regions and countries which are causing major disease outbreaks. Malaria is a major killer of refugees and displaced persons in Africa. Epidemics can occur when malaria attacks vulnerable populations with little or no immunity. In such situations, people of all age groups are at risk of death or severe disease. Because malaria epidemics occur in populations not normally exposed to the disease, or who are exposed for only a short part of the year, local health services are usually unprepared to predict, detect and control such epidemics in time, resulting in severe cases and high death rates.

There are four types of human malaria. *Plasmodium vivax*, *P.malariae*, *P. ovale* and *P. falciparum*. *P. vivax* and *P. falciparum* are the most common and *P. falciparum* the most deadly type of malaria infection. *Plasmodium falciparum* malaria is most common in Africa, south of the Sahara, accounting in large part for the extremely high mortality in this region. It also predominates in Papua New Guinea and Haiti. There are also worrying indications of the spread of *P. falciparum* malaria into new regions of the world and its reappearance in areas where it had been eliminated. *Plasmodium vivax* is more common in Central and parts of South America, North Africa, the Middle East and the Indian subcontinent. *P. vivax* is rare in sub-Saharan Africa, whereas *P. ovale* is rare outside West Africa. *P. malariae* is found in most areas, but is relatively uncommon outside Africa. Almost all deaths and severe disease are caused by *Plasmodium falciparum*.

1.2.2 Clinical Manifestations of Malaria

The clinical manifestations are dependent on the immune status of the host. In areas of high *Plasmodium falciparum* transmission asymptomatic para-

sitaemia is found in all age groups, but the clinical manifestations are mostly seen in children. By far the commonest symptom is fever, which if left untreated usually resolves over a period of weeks or months although it may recur at a later date: this is termed *uncomplicated malaria* because, although the illness can be extremely debilitating, it is not associated with life threatening complications. Such life threatening complications occur in only a minority (in the order of 1%) of infections, and collectively known as *severe malaria.*

The commonest form of severe malaria in African children are profound anaemia and cerebral malaria, which are defined below. Most cases occur in children under 10 years of age: for reasons that are poorly understood, but may relate to the maternal immunity, severe malaria rarely occurs in infants below 6 months of age. There is a tendency for severe anaemia to occur at a younger age than cerebral malaria, but this depends on the population concerned and there is considerable overlap in their age distribution. In highly endemic areas, apart from children the other vulnerable group is women in their first pregnancy, who are more liable to develop malaria fever than other adults, and the main public health consequence is low birth weight (McGregor, 1984; Brabin, 1983).
At lower levels of malaria transmission (such as in South East Asia) the age distribution of severe malaria shifts and is seen in adults as well as children.

1.2.2.1 Uncomplicated Malaria

The clinical features of uncomplicated malaria are common to all four species. The first symptoms of malaria are non-specific and resemble a severe episode of influenza. They are similar for all four species of *Plasmodium.* Headache, muscular ache, vague abdominal discomfort, lethargy and dysphoria are often the first symptoms and proceed fever by up to two days. Then, the body temperature rises with shivering, mild chills, worsening headache and malaise, and loss of appetite. Children are irritable, lethargic and anorexic. If the infection is left untreated the fever in *P. vivax* and *P. ovale* reoccurs

very 2 days (tertian), and *P. malariae* every 3 days (quartan pattern). The fever of *P. falciparum* malaria (*malignant tertian*) has a natural periodicity of 2 days but this is much more erratic than with other malaria species.

In a true, i.e. untreated, paroxysm the temperature usually rises steeply which is associated with headache and muscular discomfort, shivering and chills. The rigor usually last for about 10 to 30 minutes but can last up to 90 minutes. At the end of the rigor peripheral vasoconstriction occurs with profuse sweating and exhaustion. However, true rigors are unusual in naturally acquired *P. falciparum* malaria and are more a feature of *P. vivax* and *P. ovale*. As the infection continues the patient develops hepatosplenomegaly and anaemia. If the infection is left untreated, it gradually resolves over weeks and months. However, the duration of the illness is proportional to the level of immunity and differs between the parasite species. Both *P. vivax* and *P. ovale* have the tendency to relapse after resolution of the primary infection. *P. falciparum*, however, is the usual cause of recrudescent infections.

1.2.2.2 Severe Malaria

One of the following features in the presence of asexual parasitaemia defines severe falciparum malaria as proposed by the World Health Organization (WHO) (WHO, 1990):

1. Cerebral Malaria - unrousable coma not attributable to any other cause in a patient with falciparum malaria. The coma should persist for at least 30 minutes after a generalized convulsion to make the distinction from postictal coma.

2. Severe anaemia - normocytic anaemia with haematocrit <15% or haemoglobin <5/dl in the presence of of parasitaemia more than 10000/μl.

3. Renal failure - defined as a urine output of less than 400ml in 24 hours in adults, or 12ml/kg in 24 hours in children, failing to improve after rehydration, and a serum creatinine of more than 265μmol/l.

4. Pulmonary oedema or adult respiratory distress syndrome.

5. Hypoglycaemia - defined as a whole blood glucose concentration of less than 2.2mmol/l.

6. Circulatory collapse or shock - hypotension (systolic blood pressure <50mmHg in children aged 1-5 years or <70mmHg in adults), with cold clammy skin or core-skin temperature difference >10°C.

7. Spontaneous bleeding from gums, nose, gastrointestinal tract, etc., and/or substantial laboratory evidence of DIC.

8. Repeated generalized convulsions - more than two observed within 24 hours despite cooling.

9. Acidaemia - defined as an arterial pH <7.25, or acidosis defined as a plasma bicarbonate concentration <15mmol/l.

10. Macroscopic haemoglobinuria - if definitely associated with acute malaria infection and not result of treatment with antimalarial drugs.

11. Postmortem confirmation of diagnosis.

All of the above mentioned complications may occur in an adult with severe malaria in South East Asia (Warrell et al., 1990). In contrast, severe malaria in Africa is largely confined to children, who are liable to suffer from cerebral malaria or severe anaemia but rarely develop jaundice, renal failure or pulmonary oedema (Molyneux et al., 1989; Brewster et al., 1990; Marsh et al., 1995). However, regional differences and age alone can not explain the clinical differences to the infection with the malaria parasite. Various parasite and host factors influence the clinical manifestations of the infection. One of these factors is the inflammatory host response to the infection. Crucially pro-inflammatory cytokines represent the first line innate immunological host defence against the malaria parasite. Not only do they play a role in the classical symptom of malaria, the fever paroxysm, but they also play a causal role in the pathogenesis of severe malaria. This study will focus mainly on cerebral malaria and severe anaemia, since these are the major problems

afflicting African children who suffer by far the highest death toll. There is growing evidence that severe disease occurs when parasite sequestration and other pathological processes are aggravated by the excessive or inappropriate production of inflammatory mediators. The coincidence of various factors, including parasite strain, immune status and host genotype, is likely to be responsible for this.

1.2.3 Role of pro-inflammatory cytokines

Cytokines are key mediators of the innate host response that fights the infections. Cytokines may up-regulate the endothelial expression of vascular ligands for *Plasmodium falciparum*-infected erythrocytes and thus promote cytoadherence. They may also be important mediators of parasite killing by activating leukocytes, and possibly other cells, to release toxic oxygen species (Malhotra et al., 1988), nitric oxide, by generating parasiticidal lipid peroxides, and by causing fever. Mice were shown to suppress *Plasmodium berghei* infection more effectively if treated with bacterial endotoxin, which is known to be a potent stimulus for a wide range of cytokines and other inflammatory mediators (Martin et al., 1967; MacGregor et al., 1969).

Tumour necrosis factor (TNF) in particular has been shown to have many protective, i.e. anti-parasitc, properties. For example, it was shown that recombinant TNF could suppress parasite growth in murine malaria models (Clark et al., 1987; Taverne et al., 1987; Stevenson and Ghadirian, 1989) while anti-TNF antibodies had the reverse effect (Neifer et al., 1989). Interferon-γ is a cytokine that is equally important in the host to response to to the infection as will be discussed in detail further on in this chapter.

The disease outcome will furthermore be affected by the acquired immune response - a process in which cytokines play a major role in directing the cellular response. For example, mice who are rendered deficient in the Th1 subset by *IL-2* and *IFN-*γ gene knockout show diminished ability to control both *P. chabaudi* and *P. yoelii* infection (van der Heyde et al., 1997).

But cytokines are also largely responsible for the clinical signs and symptoms

of the malaria infection. However, not all of them are pathological as for example fever is a natural host response to malaria and other infections and in normal circumstances the production of pyrogenic and pro-inflammatory cytokines plays an important part in the anti-malarial host defence. But it has equally become clear that cytokines are involved in the generation of severe complications of malaria, such as cerebral malaria (Grau et al., 1989b) and severe anaemia.

1.2.3.1 Malaria Fever

Unlike most other pathogenic microbes, the malaria parasite does not invade host tissue (except in the liver stage of the parasite at the start of the infection, when parasite numbers are low) but circulate within the intravascular compartment, hence providing less of a focus for a localised inflammatory reaction. The principal manifestation of the pro-inflammatory cytokines produced in malaria is a systemic response, namely fever. Cytokines that are thought to be involved in the development of fever include tumour necrosis factor (TNF), interleukin-1β (IL-1β), interleukin-1α (IL-1α), interleukin-6 and interferon-α, lymphotoxin-α and macrophage inflammatory protein-1 (Endres et al., 1987; Kluger, 1991). These cytokines promote a fever response by acting on the thermoregulatory centre in the hypothalamus.
Many studies have shown high levels of TNF in individuals infected with the malaria parasite (Scuderi et al., 1986; Grau et al., 1989c; Molyneux et al., 1991; Shaffer et al., 1991). These seem to coincide with schizont rupture (Kwiatkowski et al., 1989). Anti-TNF therapy was shown to significantly reduce fever (Kwiatkowski et al., 1993; van Hensbroek et al., 1996).
In addition to those cytokines with direct pyrogenic properties, malaria patients produce a variety of other mediators which may contribute to fever by stimulating the production of pyrogenic cytokines. Interferon-γ levels were significantly higher in patients with hyperpyrexia compared with those without hyperpyrexia (Seoh et al., 2003). The fever of *vivax* malaria was associated with IFN-γ induced macrophage activation (Brown et al., 1991). In

another study, however, production of malaria-specific interferon-γ was associated with reduced risk of fever and clinical malaria (Dodoo et al., 2002). While fever is part of the 'natural' response to the malaria infection and has undoubtedly at least in part a beneficial impact on the resolution of the infection, cytokines are also involved in complications of malaria.

1.2.3.2 Cerebral Malaria

The process whereby erythrocytes containing mature forms of *P. falciparum* adhere to microvascular endothelium (*cytoadherence*) and thus disappear from the circulation is known as *sequestration.* This constitutes one of the major pathological events resulting in organ damage, and the brain in particular, via microcirculatory obstruction with reduced oxygen and substrate supply. Over the years it has become clear that pro-inflammatory mediators play a major role in the development of this complication of the malaria infection. There is substantial clinical evidence that high levels of TNF are associated with cerebral malaria (Grau et al., 1989b; Kwiatkowski et al., 1990). One of the proposed mechanisms is that TNF and other pro-inflammatory mediators promote sequestration by upregulating the expression of endothelial adhesion molecules which mediate parasite cytoadherence, including ICAM-1 and E-selectin.

In a murine model neurological complications could be prevented by the use of anti-TNF antibodies (Grau et al., 1987).

Both the neurological syndrome and the preceding rise in serum TNF levels can be suppressed by depletion of CD4+ lymphocytes (Grau et al., 1986), by administration of monoclonal antibodies to IFNγ (Grau et al., 1989a) or by targeted disruption of the gene for IFNγ receptor (Rudin et al., 1997).

The metabolic milieu created adjacent to the sequestered and highly metabolically active parasites interferes with neurotransmission but how this occurs is unknown. Cytokines increase production of nitric oxide (NO), a potent inhibitor of neurotransmission, by microglia, the vascular endothelium,

smooth muscle cells. Local synthesis of NO might lead to the impairment of consciousness (Clark et al., 1992).

The pathogenesis of cerebral malaria remains a complex issue which is not at least dependent on the parasite strain and the geographical area. However, it has become apparent that cytokines play a significant role in the development of the molecular events leading to this severe complication of the infection.

1.2.3.3 Severe Malarial Anaemia

The pathogenesis of severe malarial anaemia remains poorly understood. Two processes seem to be responsible for the anaemia: destruction of erythrocytes and the failure of the bone marrow to produce an adequate supply of new erythrocytes (Phillips et al., 1986; Abdalla et al., 1980).

While it is obvious that infected erythrocytes get destroyed through schizont rupture or by destruction through the immune system, it was noted that uninfected erythrocytes were equally destroyed (Looareesuwan et al., 1987). The key organ for this process is the spleen. Pro-inflammatory cytokines such as TNF act to promote the clearance of parasitised erythrocytes, a process that may also reduce the survival of uninfected red cells. Furthermore, it has been noted that despite significant haemolysis the reticulocyte response in the malaria infection is often weak or absent. TNF and other pro-inflammatory mediators seem to be at least in part responsible by inhibition of the proliferation of erythroid progenitor cells in the bone marrow (Roodman et al., 1987). In one study, *P. berghei* infection was found to reduce the number of erythroid progenitors and the level of iron incorporation into erythrocytes - a process that that could partially be prevented by treatment with anti-TNF antibodies (Miller et al., 1989).

1.2.4 Interferon-γ and its receptor

Interferon-γ is a central mediator of innate and adaptive immunity and has been implicated in a wide range of infectious and inflammatory disease pro-

cesses (Spellberg and Edwards Jr, 2001). It is best known for its role in the polarisation of the adaptive immune response towards the T helper (Th) type 1 spectrum (Kobayashi et al., 1989; Chehimi and Trinchieri, 1994), and as a primary effector molecule of cytolytic T cells (CTL). However, IFN-γ is also an important mediator of innate immunity. Natural killer (NK) cells secrete large amounts of the cytokine in response to interleukin-12 (IL-12) and the recently described NKT cells, as well as $\gamma\delta$-T cells can produce significant amounts of IFN-γ in response to non-peptidic antigens (Schofield et al., 1999; Tomura et al., 1999; Goodier et al., 1992; Constant et al., 1994). IFN-γ production by NK cells is one of the earliest components of the inflammatory response to infection, while its production by both CD4+ and CD8+ T lymphocytes is critical for effective antibody formation and cytotoxic mechanisms against intracellular pathogens. It activates macrophages and neutrophils to kill microbes by free radical production and phagocytosis, and it influences the overall pattern of acquired immunity by polarising T-helper cells towards a Th1 and away from a Th2 type of response (Kobayashi et al., 1989; Chehimi and Trinchieri, 1994).

The receptor for interferon-γ is made up of α (IFNGRAC or IFNγR1) and β (IFNGRBC or IFNγR2) subunits, both integral membrane proteins. Classically the α subunit (encoded by the gene *IFNGR1* on chromosome 6) is known to play a critical role in ligand binding. Together with the β subunit it is furthermore involved in receptor trafficking and signal transduction by activating various co-factors (Stat1, JAK1, JAK2) (Bach et al., 1997). A recent paper suggested an even more active involvement in signal transduction by IFNγR1 being internalized and translocated to the nucleus upon binding of IFN-γ to the dimeric receptor complex (Larkin et al., 2000). The β chain is encoded on chromosome 21. It plays a minor role in ligand binding but is crucial for signal transduction events. In the absence of the ligand the two chains of the receptor are not associated. Upon binding of IFN-γ homodimer two α-chains dimerize and two β chains are recruited into the complex. After several activation steps phosphorylated Stat1 proteins dissociate from the receptor and translocate to the nucleus where they bind to specific promoter elements.

Disruptive mutations in the *IFNGR1* coding region cause extreme vulnerability to infection with environmental nontuberculous mycobacteria (Newport et al., 1996), such as *Mycobacterium avium*, *Mycobacterium fortuitum*, *Mycobacterium chelonei*, and *Mycobacterium smegmatis*. Furthermore infections were found with *Mycobacterium tuberculosis* (Jouanguy et al., 1997), bacillus Calmette-Guérin (BCG) (Holland et al., 1998; Jouanguy et al., 1997) and *Salmonella* (Casanova et al., 1996; Altare et al., 1998). Rare cases of viral infections (Dorman et al., 1999) and *Mycobacterium avium* osteomyelitis (Villella et al., 2001) have been described. These coding mutations of the *IFNGR1* gene either completely preclude cell surface expression of the receptor or lead to a non-functional abrogated protein leading to complete or partial IFNγR1 deficiencies. These mutations tend to be fatal but are extremely rare. Mutations in *IFNGR2* have also been found to be associated with mycobacterial infections (Doffinger et al., 2000).

1.2.4.1 Protective Role of Interferon-γ in Malaria

Interferon-γ is currently used as one of the main surrogate markers of immunogenicity in malaria vaccine development (Stoute et al., 1998; Perlaza et al., 1998; Hollingdale et al., 1998). Aside from the fact that its production indicates the activation of CD8+ cytolytic T cells and CD4+ Th1 cells, it has also been shown to be directly involved in the development of sterile immunity. In rats, challenge with irradiated sporozoites confers protection to subsequent *Plasmodium berghei* challenge. This protection could be shown to be largely dependent on the production of IFN-γ by CD8+ T-cells (Schofield et al., 1987). In mice the protective effect conferred by IL-12 during *Plasmodium yoelii* sporozoites challenge could be reversed by neutralising IFN-γ antibody (Sedegah et al., 1994). Similar effects were seen in monkeys (Maheshwari et al., 1986; Puri et al., 1988), and in most experiments both T- and NK cells were required for the protective effect. The postulated effector mechanisms involve the killing or developmental inhibition of the liver stages of the parasite (Hoffman et al., 1997). There is also some evidence for a protective effect of IFN-γ against the erythrocytic stages of the parasite

(Favre et al., 1997), particularly in combination with chloroquine treatment (Kremsner et al., 1991). Genetic background seems to play a more significant role here, as suggested by experiments comparing a lethal to a non-lethal *Plasmodium yoelii* strain in BALB/c mice. Exogenous IFNγ provided a dose-dependent protection from the lethal strain but has little effect on the non-lethal infection (Shear et al., 1989). Endogenous IFNγ production from spleen cells was detected only in response to the non-lethal strain, whereas in CBA/J mice, which are resistant to the lethal strain, it was detected in response to both.

1.2.4.2 Role of Interferon-γ in Malaria Pathology

While IFN-γ is undoubtedly a crucial mediator in protection from infection, the anti-parasitic mechanisms it activates may also result in damage to the host. This hypothesis has been explored in a mouse model of cerebral malaria (Grau et al., 1989b). In this model, activated CD4+ cells produce IFN-γ, which in turn is the major stimulus for local TNF production by murine macrophages, microglia, or astrocytes (Grau et al., 1989a; de Kossodo and Grau, 1993). Neurological damage is then due to the release of large amounts of inflammatory mediators and consequent nitric oxide production around parasites sequestered into brain blood vessels (Clark et al., 1992). Antibodies to IFNγ are, indeed, protective in this model (Grau et al., 1989a; Johnson et al., 1993). It was also shown that *IFNGR1* knockout mice did not develop cerebral malaria upon challenge with *Plasmodium berghei* (Amani et al., 2000). Malarial anaemia may also be exacerbated by large amounts of IFNγ, since it has been shown to reduce erythropoietin secretion (Vannucchi et al., 1994), but there is thus far no direct evidence for such an effect. A small number of epidemiological studies confirm the involvement of IFNγ in pathology. Children whose cells produced more IFNγ after in vitro challenge with *Plasmodium falciparum* antigen were more likely to experience the clinical manifestations of malaria infection (Riley et al., 1991). Chizzolini et al. (Chizzolini et al., 1990) have also shown that CD4+ T cell

dependent IFNγ responses were down-regulated in malaria immune individuals. The balance between protective and pathological effects of IFNγ is clearly due to the 'appropriate' magnitude and timing of the effector mechanisms of this cytokine. This will be influenced by two factors: a) the amount of IFNγ produced by the various cell types at any particular time, and b) the amount of IFN-γ receptor present on the respective target cells. The latter is a potentially vital component in the equation, as it will crucially determine the type and the magnitude of effector mechanisms activated in the course of the response. The importance of surface receptor expression for the biological function of any cytokine is becoming increasingly apparent. The 'switch' between Th1 and Th2 responses to *Leishmania major* infection in mice, for example, appears to lie entirely in the up- or downregulation of IL-12 receptor expression (Himmelrich et al., 1998). To understand the modulation of the biological function of IFN-γ it is therefore of paramount importance to not only study the regulation of production of the cytokine itself, but also that of its receptor. Both are likely to be influenced by the genetic background of the host, as suggested by the variability of the observed effects in different mouse and parasite strains.

1.2.5 Malaria and the genetic influence on its outcome

Haldane (1949) recognised the importance infective pathogens have on the evolution of the human genome. He proposed that people who were heterozgous for red-blood-cell disorders, such as thalassaemia and sickle cell disease, might be protected against life-threatening infection with malaria (see also Haldane (1948)). This would explain the high gene frequencies for the haemoglobinopathies in tropical areas. A state of *balanced polymorphism* would exist, whereby the loss of the disadvantaged homozygotes would be offset by the survival advantage in heterzygotes.

Tishkoff et al. (2001) suggested that severe malaria has had a major impact on humans since the introduction of agriculture within the past 10,000 years. Mild forms of the disease may have existed in humans throughout

much of the human evolutionary history. The infection had significant effects on morbidity and mortality before reproductive age and malaria has therefore exerted a strong selective pressure on the human genome.

Although African children suffer repeated episodes of malaria, severe life-threatening disease - cerebral malaria and severe malaria anaemia - develops in only about 1-2%. This suggests that effective immunity can develop and, in combination with the body's innate protective responses, provides a potentially powerful system to resolve the infection. But given its ubiquity, the malaria mortality is still extremely high. Why is it, that one child dies of malaria while another one survives? There is evidence for a genetic basis of the wide variation in outcome, which has fuelled the search for both host and parasite factors, which might modulate the disease process. In a large twin study Jepson et al. (1995) suggested that infection per se is largely determined by environmental factors. Once infected however, both members of a monozygous twin pair were more likely to develop fever than were twins of a dizygous pair, suggesting that genetic factors influence the presentation of clinical disease. Abel et al. (1992) found evidence for the presence of a recessive major gene controlling the degree of infection in human malaria. Garcia et al. (1998a) found an effect of age on infection levels and its interaction with a putative major gene suggesting that genetic related differences are much more important in children than in adults. Both authors tried to fit the proposed genetic susceptibility to malaria into a Mendelian model. While some of their data seems consistent with that, this hardly captures the complexity of the genetic susceptibility to this disease.

Several haemoglobinopathies confer protection against malaria. Normal adult haemoglobin is composed of two pairs of distinct haemoglobin chains ($\alpha_2\beta_2$; called haemoglobin A, HbA). In the variant haemoglobin called haemoglobin S (HbS), position 6 of the β-chain is mutated ($\beta^{6Glu-Val}$). Although *HbS* homozygosity is associated with sickle cell anaemia, individuals heterozygous for the *HbS* variant are strongly protected against severe malaria. The patients' red cells resist invasion, and once invaded the cells sickle readily , thereby facilitating clearance by the reticuloendothelial system. Thalassemias, defined as defective synthesis of one type of globin chain (α or

β), also show strong evidence of having reached their current frequencies by heterzygote selection against malaria (Weatherall, 2001). However, the mechanism by which this protection is conferred is less well understood. Red cells from patients with thalassaemia are invaded normally by *P. falciparum*, but they bind greater numbers of antibody molecules to the erythrocyte surface than infected normal cells and may therefore be cleared from the circulation more readily (Luzzi et al., 1991a,b). Mutations in Glucose-6-Phosphate Dehydrogenase (G6PD), an enzyme crucial for the glucose metabolism in erythrocytes, have also been associated with protection against malaria (Ruwende et al., 1995; Ruwende and Hill, 1998; Tishkoff et al., 2001; Sabeti et al., 2002).

In addition to the well-known erythrocyte polymorphisms, such as sickle haemoglobin and thalassaemia, there is growing interest in genes that control the innate response to infection, particularly cytokines, adhesion molecules and other inflammatory mediators (Hill, 1998). Several genes have already been implicated in this disease. Rihet et al. (1998) and Garcia et al. (1998b) both suggested that *Plasmodium falciparum* blood infection levels are linked to chromsome 5q31-q33, a genetic area encoding for important immune genes such as *IL4*, *IL9*, and *IL13*. Three different single nucleotide polymorphisms (SNP) in the promoter of the tumor necrosis factor (*TNF*) gene have been reported to be associated with forms of severe malaria in Gambian children (McGuire et al., 1994; Knight et al., 1999). In patients from Gabon an SNP in the inducible nitric oxide synthase promoter region (*NOS2*) has been associated with protection against severe malarial anaemia (Kun et al., 1998), and in The Gambia a *NOS2* microsatellite was found to be associated with susceptibility to fatal malaria (Burgner et al., 1998). Intracellular adhesion molecule-1 (*ICAM-1*) has also been implicated in malaria, although results of three different association studies gave conflicting results (Bellamy et al., 1998; Kun et al., 1999; Fernandez-Reyes et al., 1997). Immunological data have shown that the cytokine IFN-γ and its receptor are also critically involved in the disease process of malaria which make them interesting candidate genes to investigate.

1.3 Aims

There is evidence for significant effects of alterations in the IFN-γ receptor on the course of various disseminated mycobacterial infections and fatal bacillus Calmette-Guerin (BCG) infection (Holland et al., 1998; Jouanguy et al., 1997). More subtle effects are seen in point mutations that correlate with systemic lupus erythematosus (Nakashima et al., 1999). The following investigation addresses the question, whether part of the natural variation in susceptibility to common infections might be due to genetic polymorphisms of higher frequency with more modest effects on *IFNGR1* regulation. Since IFN-γ has both beneficial and potentially deleterious effects (e.g. killing microbes vs. causing severe inflammation) natural selection might favour different levels of *IFNGR1* expression depending on the type of infectious pathogens to which a population is exposed. Over time, this might be expected to cause genetic variation in *IFNGR1*. Consistent with this notion, the genomic region around *IFNGR1* has been linked to severe hepatic fibrosis in community-based studies of *Schistosoma mansoni* infection (Dessein et al., 1999) but thus far there is no clear evidence of an association between any specific *IFNGR1* polymorphism and susceptibility to common infections within the general population.

While it becomes clear from immunological studies that both IFN-γ and its receptor play a crucial role in the host defence to pathogens, the knowledge as to which role common variants of these genes play in the susceptibility to disease is limited. While some reports have shown associations of one marker in *IFN-γ* with various conditions, surprisingly little is known about the importance of the wider genetic area, i.e. the relationship to other polymorphisms in this gene, i.e. the haplotypic background. It becomes apparent that studying individual markers can be of limited value as it remains unclear whether any association found represent the *true* functional variant or merely a marker in linkage diseqilibrium with the same.

This study therefore aims to investigate the genetic regions of both the *IFN-γ* and *IFNGR1* gene. Common genetic variants of these genes and their genetic neighbours will be identified, put into their haplotypic context and tested for

association with malaria - a disease in which both IFN-γ and its receptor play a crucial role. Severe malaria affects a vast number of people globally, children in particular, and leads to more than 2 million deaths annually. While all individuals in endemic areas are frequently exposed to the malaria causing parasite, only a small percentage of people develop the severe symptoms of the disease that can subsequently lead to death. Examining the differences in our genetic make-up might contribute to a better understanding of the individual differences in susceptibility to this disease and ultimately help in developing therapeutic strategies benefitting the patients.

Косн 2003

Chapter 2

Methods

2.1 Subjects

This thesis forms part of an ongoing programme of clinical epidemiological research that is being undertaken by the Wellcome Trust Centre for Human Genetics at Oxford University in collaboration with the Medical Research Council Laboratories, Fajara, The Gambia. The epidemiological part of the study was supervised by Prof. Dominic Kwiatkowski (Oxford University) and Dr. Margaret Pinder (MRC Laboratories, The Gambia). Cases of severe malaria and their parents were recruited by Dr. Stanley Usen and Dr. Muminatou Jallow assisted by clinicians and nurses working on the Children's Wards of the Royal Victoria Hospital, Banjul, The Gambia. Cord blood samples were collected by a team led by Dr. Melanie Newport at the MRC Laboratories, Fajara. Various members of the Malaria Laboratory of the MRC Laboratories in The Gambia and the Kwiatkowski laboratory at the Wellcome Trust Centre for Human Genetics have worked on DNA extraction, sample archiving, computer entry of the clinical data and other practical aspects of establishing this epidemiological resource. The study was approved by the MRC / Gambia Government Joint Ethical Committee and the Chief Executive of the Royal Victoria Hospital. This thesis describes a series of genetic investigations carried out on samples that were collected by

this research programme during the period 1996 to 2001.

Children with severe malaria were recruited to the study shortly after admission to the Royal Victoria Hospital, after obtaining informed parental consent. The criteria for recruitment were essentially those defined by WHO (1990) but by far the major categories were cerebral malaria and severe anaemia. Cerebral malaria was defined as a Blantyre coma score of 2 or less (Molyneux et al., 1989) persisting for more than 30 minutes after any convulsions had ceased in a child with *Plasmodium falciparum* parasites on thick blood film and with no evidence of meningitis or any other cause of coma. Severe anaemia was defined as a PCV less than 15% in a child with *Plasmodium falciparum* parasites on thick blood film. Other categories of severe malaria were occasionally identified but the current sample size is too small to form reliable conclusions and so these minor categories were not included in this current research project. Thus for the purposes of this thesis the term *severe malaria* refers to a child with cerebral malaria or severe anaemia or both.

Two approaches were used for genetic association analysis. The first approach was conventional case-control analysis, using cord bloods collected from babies born in clinics in different parts of The Gambia. Several different ethnic groups have long co-existed in this part of West Africa, so pains were taken to record accurate information about the ethnic group of the parents of both cases and controls, and the case-control analysis was stratified by ethnic group. The second approach was intrafamilial association analysis, as described below, where ethnic artefacts are excluded by using parents as genetic controls. Parents were recruited either when the child was admitted to the ward, or on subsequent follow-up by a nurse or fieldworker attached to the Royal Victoria Hospital.

2.2 DNA Quantification and Storage

The DNA concentrations of the samples needed to be established for further use. The initial volume of the DNA solution was measured by filling a pipette tip and taking a reading. The DNA concentration was quantified using PicoGreen double stranded DNA Quantification Kit (Molecular Probes).The PicoGreen reagent allows quantification of double-stranded DNA solution as dilute as 25pg/ml (50pg dsDNA in a 2ml assay volume) with standard spectrofluorometer and fluorescein excitation and emission wavelengths. Using a fluorescence microplate reader, concentrations as little as 250pg/ml dsDNA can be detected (50pg in a 200μl assay volume), which is therefore far more sensitive than traditional methods involving detection of absorbance at 260nm in a spectrofluorometer (typical maximum sensitivity at OD=0.1, equivalent to 5μg/ml DNA). The major disadvantages of the absorbance method are furthermore the large relative contribution of nucleotides and single-stranded nucleic acids to the signal, the interference caused by contaminants and the inability to distinguish between DNA and RNA.

The assay was performed as follows: 1μl of DNA was diluted into 250μl of TE Buffer with a low EDTA content (10mM Tris; 0.1mM EDTA). 50μl of this dilution were than mixed with 50μl of PicoGreen reagent at working concentration (1:200 dilution of the PicoGreen stock in the TE Buffer (low EDTA, see above)) in a flat-bottomed 96-well PS microplate (Greiner). Samples were transferred onto the microplate in duplicate and the resulting fluorescence was detected on a Cytofluor 2350 Fluorometer (Millipore). The fluorescence was calibrated against eight samples in duplicate containing a known concentration of lambda DNA standard ranging from 0 to 1000ng/ml. Using this format (Figure 2.1), 40 samples could accurately be quantified per 96-well plate. The resulting standard curve was calculated using a Microsoft Excel© template (Microsoft). Following quantification 2mg of genomic DNA stock were removed and diluted up to 100μl with TE (low EDTA) to give a 20ng/μl 'working concentration'. Where the original stock did contain less than 4mg DNA, the samples was split in half and the second half diluted to 20ng/μl. Where the original stock already contained less than 2mg of DNA,

	1	2	3	4	5	6	7	8	9	10	11	12
A	STD0		Sample1		Sample 9		Sample 17		Sample 25		Sample 33	
B	STD 25		Sample 2		Sample 10		Sample 18		Sample 26		Sample 34	
C	STD50		Sample 3		Sample 11		Sample 19		Sample 27		Sample 35	
D	STD100		Sample 4		Sample 12		Sample 20		Sample 28		Sample 36	
E	STD300		Sample 5		Sample 13		Sample 21		Sample 29		Sample 37	
F	STD500		Sample 6		Sample 14		Sample 22		Sample 30		Sample 38	
G	STD 700		Sample 7		Sample 15		Sample 23		Sample 31		Sample 39	
H	STD 1000		Sample 8		Sample 16		Sample 24		Sample 32		Sample 40	

Figure 2.1 Illustration of the layout of the 96-well format used for DNA quantification by PicoGreen fluoroscopy. STD 0-1000 represent lambda standard DNA at concentrations ranging from 0-1000ng/ml

a neat aliquot of up to 100μl (but not more than half of the original volume) was removed. The stock DNA tubes and the sub-aliquots were then stored at -70°C. Both sets of samples were stored in 1.5ml screw-top tubes (Greiner) labelled using laser-printed star tag labels on the top and the side of the tubes (Starlab). The labels provided a unique identification code indicating the origin of the sample and status of the individual, i.e. mother father, child with malaria, control individual. At this stage 5μl of the 20ng/μl DNA dilution were placed into a 96-well deep-well plates (ABGene) and 95μl of water was added to each in order to provide a final concentration of 1ng/μl which could consecutively be used for whole genome amplification. These plates were already in the same format as the plates that were used for the genotyping. These plates were stored at -20°C.

2.3 Whole Genome Amplification

In order to increase the amount of DNA to work with the samples were subjected to whole genome amplification using primer extension amplification (PEP)(Zhang et al., 1992). For this purpose 15-base random oligonucleotides from GENPAK Limited were used. 5ng of DNA were amplified using the following conditions: 2.5mMMg+, 0.2mM dNTPs, 8nM N15 primers, 2.5U Biotaq (Bioline); cycling conditions: 94°C for 3 minutes followed by 50 cycles of 94°C for 1minute, 37°C for 2minutes, 0.1°C/sec to 55°C, 55°C for 4 minutes. Final extension for 5 minutes at 72°C.

2.4 Identification of Mutations

2.4.1 *IFNGR1* promoter sequencing (MEGABACE)

36 samples from Gambian individuals with tuberculosis had been sequenced in the host laboratory by Anna Richardson, Jeremy Hull and Agnes Awomoyi using ABI dye primer sequencing. Sequencing was carried out using an

Applied Biosystems ABI377 automated DNA sequencer with the dye primer kit (Perkin-Elmer-Applied Biosystems) and analysed with the Sequence Navigator software package. Fragments for sequencing were generated by nested PCR, i.e. using two rounds of PCR. PCR primers 5' GAG TGC CAA GTA AAG ATG TCA GC and 5' AGA TCC GCG GTG CCC ATC TCA were used to amplify a 1022-bp fragment from position -951 to +71 of the *IFNGR1* gene (in relation to transcriptional start site described in (Merlin et al., 1997). In a second amplification step a 607-bp fragment with incorporated M13 tails was generated from this fragment by using M13 tailed PCR primers 5' TGT AAA ACG ACG GCC AGT AGT AGT TCT TGG TCA AGC CG and 5' CAG GAA ACA GCT ATG ACC GGG TAG GAG AAA GAG GAG AG. Prior to the then following sequencing reaction, samples were purified using spin columns (Qiagen) according to standard protocol. Sequencing was performed using the ABI Dye Primer Sequencing Kit according to the protocol supplied with the kit. It is as yet unknown how long the functional promoter of the *IFNGR1* is. The above sequencing project only looked at the first 500bp 5' of the translational start site and this sequencing was performed in samples of cases with tuberculosis. As part of this thesis the sequenced area was expanded from -1400bp to +100bp (in relation to the translational start site) and Gambian cord blood DNA samples were used for this purpose in order to minimise selection bias. The idea was that this would not only potentially discover new single nucleotide polymorphism further upstream, but would also provide positive controls for the genotyping of the SNPs that had already been found. For this purpose 34 samples were sequenced using MegaBACE Technology (Amersham). The MegaBACE 1000 DNA sequencer is a high-throughput, fluorescence-based DNA sequencer that uses capillary electrophoresis. It has the capability of sequencing 96 samples in 2 hours. Read lengths for a 2 hour run average 600 bp and with longer 5 hour runs 800-900 bases. Samples for the MegaBACE are prepared in a similar way to ABI sequencing. In a first round of PCR a fragment of DNA is generated. In a second round of PCR a dye-terminator kit is used to sequence off this fragment. Samples are then analysed on the sequencing machine.

Overview of the different steps:

1. First Round PCR
2. Spin Column Purification of PCR product
3. DNA Quantification using PicoGreen
4. Sequencing PCR with fluorescently labelled ddNTPs
5. Running samples on the MegaBACE 1000

2.4.1.1 First Round PCR

A 2030bp fragment comprising the *IFNGR1* promoter region was amplified in 34 samples from Gambian cord blood samples. As experiments using Taq DNA polymerase and standard PCR conditions had proven unsuccessful for amplifying this fragment, an Expand Long Template PCR System (Roche) was used. This kit is composed of an enzyme mix containing thermostable Taq and Pwo DNA polymerases. It is designed to amplify fragments up to 27kb from human genomic DNA. It is supplied with three different buffers of which buffer system no.3 (PCR buffer containing $[NH_4]_2SO_4$ + cosolvents + 22.5 mM $MgCl_2$ and detergents) is the buffer of choice when difficulties in amplification are expected. The primer sequences used were as follows: AGC CCC TCA TAC CAA CTC AAG ACA CCT T (*IFNGR1*-1679/-1652) and GGA TTA CAA GAG TCG GAA AAA CAA GCG G (*IFNGR1*+351/+324). Primers were supplied by MWG. After all reagents had been thawed and were briefly centrifuged, the reaction was set up using two different master mixes.

Master Mix 1

10mM dATP (Sigma)	2.5μl
10mM dCTP (Sigma)	2.5μl
10mM dGTP (Sigma)	2.5μl

10mM dTTP (Sigma)	2.5μl
5μM upstream primer (*IFNGR1*-1679/-1652)	3.0μl
5μM downstream primer (*IFNGR1*+351/+324)	3.0μl
DNA (20ng/μl)	4μl
H_2O (Sigma)	5μl

Master Mix 2

Buffer no.3 (Roche)	5μl
enzyme mix (Roche)	0.75μl
H_2O (Sigma)	19.25μl

Both master mixes were kept on ice. Shortly before cycling 25μl of each mix were pipetted together in 96-well PCR plates (ABGene, cat.no.:AB0600), mixed well and overlayed with 30μl of mineral oil (Sigma). Cycling was started immediately afterwards at the following conditions:

93°C 2:00

Followed by 10 cycles of:

93°C 0:10

64°C 0:30

68°C 1:30

Followed by 20 cycles of:

93°C 0:10

64°C 0:30

68°C 1:30 + cycle elongation of 20s for each cycle

and finally:

68°C 7:00

After cycling 2μl of the PCR product were run out on an 1% agarose gel. It seems worth noting that this particular PCR was only found to work, when commercially available H_2O (Sigma) was used, while the reaction failed to amplify the fragment when either Millipore H_2O or autoclaved water were used. After the first round of PCR the samples were purified using spin columns (Qiagen).

2.4.1.2 Spin Column Purification

The PCR products were purified using the standard QIAquick PCR Purification Kit Protocol provided with the spin columns (Qiagen; cat.no.: 28104). After following the various steps the samples were finally eluted in 50μl of H_2O (Sigma).

2.4.1.3 DNA Quantification using PicoGreen

As the amount of DNA used for MegaBACE Sequencing is crucial, the DNA concentration of the above PCR products were quantified using the PicoGreen dsDNA Quantification Kit (Molecular Probes). The procedure was performed as described under DNA Storage and Quantification. The readings for the DNA concentration were ranging from as low as 0.3ng/μl to 62ng/μl.

2.4.1.4 Sequencing using fluorescently labelled ddNTPs

The recommendations for the amount of DNA to use for MegaBACE sequencing are given as 100ng for PCR products bigger than 400bp. It was therefore decided that in cases where the amount of DNA measured by PicoGreen exceeded 20ng, the equivalent of 100ng were used per sequencing reaction. Where concentrations were below 20ng/μl, 6μl of the first round PCR product were used regardless of the concentration measured. The following four primers were used for this dye-terminator sequencing. Forward primers: CTA CTC CAT AGA CGC CTA AAC G (*IFNGR1*-1411/-1390) and TGA CAC TGA ATT GCT GAG AAG G (*IFNGR1*-599/-578). Reverse primers were:

TAC CTG AGG ACG GCC CCA GA (*IFNGR1*+88/+69) and CAT TTG AGG AAC CGT AGA CTT AGC (*IFNGR1*-461/-484). The sequencing reaction was set up as follows:

DYEnamic ET reagent premix (Amersham)	8μl
Primer (5μM)	1μl
DNA	amount as described in text
H_2O (Sigma)	to 20μl

Sequencing was performed in 96-well plates (ABGene; cat.no.:AB0600). Mineral Oil must not be used. Cycling conditions for the sequencing PCR:

95°C 0:30

50°C 0:15

60°C 1:00

for 25 cycles

After cycling samples were purified by ethanol precipitation. 52μl of 100% ethanol and 2μl 7.5 sodium acetate were added to each well. After mixing up and down using a pipette, the plates were centrifuged at 3,000rpm for 30 minutes. The supernatent was carefully taken off using a multichannel pipette. After that 100μl of 70% ethanol was added to each well to wash the pellets and plates were then centrifuged at 3,000rpm for 15 minutes. Following that the ethanol was taken off and the plates were spun inverted for one minute at 1,500rpm to remove the remaining ethanol . The samples were then left to air dry at room temperature for 30 minutes and then stored at -20°C until they were analysed on the MegaBACE. The analysis on the MegaBACE 1000 DNA Sequencing System was then performed using an overnight (i.e. 5 hours) run. The MegaBACE software allows the sequencing data to be stored in various file formats. Traces were exported in an ABI

compatible format and were analysed using the ABI Sequence Navigator software. The average read length was between 600-850bp.

2.4.2 Resequencing promoters of *IFNGR1*, *IL22BP* & *IL20RA*

For the extended resequencing project ABI Dye Primer Sequencing was used. Samples were analysed on an ABI3700 Sequence Analyzer. This approach was chosen as the procedure was optimised as part of a larger project within the research group which increased throughput and partially automated the process of analysis. For this project samples from 24 unrelated Caucasian individuals (from the CEPH repository) (see table 2.1) and 24 unrelated African (Gambian) individuals were used. All DNA was extracted by Brendan Keating, Wellcome Trust Centre for Human Genetics, from immortalised cell lines.

Table 2.1: CEPH samples

Reposit No	Relation	Family	Ind.	Assigned No.
GM11879	paternal grandfather	1347	12	CEPH 5
GM11880	paternal grandmother	1347	13	CEPH 6
GM11881	maternal grandfather	1347	14	CEPH 7
GM11882	maternal grandmother	1347	15	CEPH 8
GM07031	maternal grandmother	13292	13	CEPH 1
GM07051	maternal grandfather	13292	12	CEPH 4
GM13047	paternal grandfather	13292	10	CEPH 2
GM13050	paternal grandmother	13292	11	CEPH 3
GM13042	maternal grandfather	13293	12	CEPH 13

continued on next page

continued from previous page

REPOSIT NO	RELATION	FAMILY	IND.	ASSIGNED NO.
GM13044	maternal grandmother	13293	13	CEPH 14
GM13048	paternal grandfather	13293	10	CEPH 17
GM13051	paternal grandmother	13293	11	CEPH 19
GM11992	paternal grandfather	1362	13	CEPH 9
GM11993	paternal grandmother	1362	14	CEPH 10
GM11994	maternal grandfather	1362	15	CEPH 11
GM11995	maternal grandmother	1362	16	CEPH 12
GM13045	paternal grandfather	13294	9	CEPH 15
GM13046	paternal grandmother	13294	10	CEPH 16
GM13049	maternal grandfather	13294	11	CEPH 18
GM13052	maternal grandmother	13294	12	CEPH 20
GM06986	maternal grandfather	13291	12	CEPH 21
GM07037	paternal grandmother	13291	11	CEPH 22
GM07045	maternal grandmother	13291	13	CEPH 23
GM07435	Paternal Grandfather	13291	10	CEPH 24

2.4.2.1 Preparation of Products

Dye primer sequencing was chosen in preference to dye terminator as initial comparisons with dye terminator revealed dye primer results to have a superior consistency of peak height. This aids differentiation of heterozygotes - an important advantage in sequencing for SNP detection. First round products of the target region were initially generated and from these, nested second round fragments were amplified with M13-tailed primers. For primer sequences used see table 2.2.

To generate first round products, 200ng of genomic DNA was added to

0.5μM of forward primer, 0.5μM of reverse primer, 0.8μM of dNTP, 2.5mM $MgCl_2$, 0.25μl Bio-X-Act DNA polymerase (Bioline), 2μl of 10x Opti-Buffer, 4μl of 5x Enhancer, final volume 20μl. PCR cycling conditions were 94°C for 2 minutes; then 10 cycles of 94°C for 15 seconds, 60°C for 30 seconds, 68°C for 2 minutes; then 20 cycles of 94°C for 15 seconds, 60°C for 30 seconds, 68°C for 2 minutes adding 20 seconds per cycle; and finally 68°C for 5 minutes.

To generate second round products a 1:10 dilution was made of the first round amplification products. 2μl of the first round template was added to 0.5μM of forward primer with an M13 tail, 0.5μM of reverse primer with an M13 tail, x1 Taq Gold Buffer (PE Applied Biosystems), 0.8μM dNTP, 2.5mM $MgCl_2$, 1 unit of AmpliTaq Gold DNA Polymerase (PE Applied Biosystems), final volume 20ul. Cycling conditions were 94°C for 14 minutes; then 5 cycles of 94°C for 30 seconds, 57°C 30 seconds, 72°C 45 seconds; then 30 cycles of 94°C for 1 minute, 72°C for 1 minute, then 72°C for 10 minutes. To check that the PCR had been successful, 2μl of product were run on a 1% agarose gel stained with ethidium bromide and viewed under UV light.

Second round PCR products were cleaned using a Montage PCR Cleanup plate (Millipore). The PCR product was transferred to the clean up plate and a vacuum manifold (Millipore) dried the sample leaving the nucleic acid bound to the membrane. The clean product was then reconstituted in a 1:3 dilution of the original volume.

The sequencing reaction was performed using the ABI BigDye v3.0 Ready Reaction Cycle Sequencing Kit (PE Applied Biosystems). Two different kits were used for M13 forward or M13 reverse primers according to which strand was to be sequenced. All strands were sequenced in forward and reverse direction. Four separate reactions incorporating a different dideoxynucleotide were performed per sample according to the manufacturer's instructions. 1μl of clean second round DNA product was mixed with 2μl of the ready reaction mix, 2mM $MgCl_2$, 80mM TRIS pH 9.0, final volume 5μl. Cycling conditions were 14 cycles of 96°C for 10 seconds, 55°C for 5 seconds, 70°C for 1 minute; then 14 cycles of 96°C for 10 seconds, 70°C for 1 minute.

Each ddNTP specific reaction for a given sample was pooled. To this, 60μl of 100% ethanol and 0.2μl of pellet paint (Novagen) were mixed and

incubated at 4°C for 15 minutes before centrifuging at 2500 rcf for 30 minutes at 4°C. Ethanol was drained from each well by inverting the 96 well plate, pulse centrifuging to 61 rcf and leaving the samples to dry for 1 hour in darkness. Samples were resuspended in 20μl of a 50% dilution of Hi-Di-formamide (ABI) and electrophoresed on an ABI 3700 Sequencer by the core genomics facility at the Wellcome Trust Centre for Human Genetics.

2.4.2.2 Sequence Analysis

Sequence assembly and analysis was assisted by Dr Man-Suen Chan, Wellcome Trust Centre for Human Genetics. The computer programme `PHRED` was used to call nucleotide bases and assign confidence values based on trace quality and peak size. `PHRAP` (*phragment assembly programme*) was then used to assemble sequence data into contigs. `PolyPhred` was used in conjunction with `PHRED` and `PHRAP` to identify SNPs, by comparing traces in the sequence assembly to the expected sequence pattern downloaded from Ensembl (`http://www.ensembl.org`). Default parameters were used for all computer programmes. `PHRAP` assemblies were viewed using the programme `CONSED` using an x-windows interface and each suggested nucleotide variation was visually checked.

Table 2.2: Resequencing Primers. Sequences of 1^{st} and 2^{nd} round PCR primers for resequencing the promoter regions of *IFNGR1*, *IL22BP* and *IL20RA* are shown. "fw" and "rv" indicates the direction of the primers used to amplify the respective fragment.

Gene	PCR round (Fragment)	Direction	Sequence
IFNGR1	1^{st}	fw	TCTTCCCATTGCTCTCCTTG
IFNGR1	1^{st}	rv	CTGTCCCACATTGACCAGC
IFNGR1	2^{nd} (1)	fw	TGTAAAACGACGGCCAGTAGTCCGCCTCCTGCGGCTTC
IFNGR1	2^{nd} (1)	rv	CAGGAAACAGCTATGACCCCTTCTGCAGCGCATAATCG
IFNGR1	2^{nd} (2)	fw	TGTAAAACGACGGCCAGTCACTGAATTGCTGAGAAGG
IFNGR1	2^{nd} (2)	rv	CAGGAAACAGCTATGACCGGTCTGTGCCGGCGTCGG
IFNGR1	2^{nd} (3)	fw	TGTAAAACGACGGCCAGTACAAGTAAAATGATCTGGCTG
IFNGR1	2^{nd} (3)	rv	CAGGAAACAGCTATGACCCAAATCAAATCGGCTTGACC

continued on next page

continued from previous page

GENE	PCR ROUND (FRAGMENT)	DIRECTION	SEQUENCE
IFNGR1	2^{nd} (4)	fw	TGTAAAACGACGGCCAGTCCTTCACAGACGTTTATTCC
IFNGR1	2^{nd} (4)	rv	CAGGAAACAGCTATGACCGAAAATCCTTATAAAGACAC
IFNGR1	2^{nd} (5)	fw	TGTAAAACGACGGCCAGTCATGTCTGTCTTGCTCAGG
IFNGR1	2^{nd} (5)	rv	CAGGAAACAGCTATGACCGATTAAGGTATAAGGCATTG
IL20RA	1^{st}	fw	AAGCAACTACCACCGTCAGG
IL20RA	1^{st}	rv	CCAGGCTTCCCCAGAAAC
IL20RA	2^{nd} (1)	fw	TGTAAAACGACGGCCAGTCCATTGTGTTACAATTGCTTGC
IL20RA	2^{nd} (1)	rv	CAGGAAACAGCTATGACCACTCGCGGAGCCCCCACG
IL20RA	2^{nd} (2)	fw	TGTAAAACGACGGCCAGTGGGCACACACATGTAATCC
IL20RA	2^{nd} (2)	rv	CAGGAAACAGCTATGACCCTAGGCTACAAACCTGTACG
IL20RA	2^{nd} (3)	fw	TGTAAAACGACGGCCAGTCTTTTCAGTCAGTGACAGACTGC
IL20RA	2^{nd} (3)	rv	CAGGAAACAGCTATGACCCCTTCCGGTTTTGATGATTG
IL20RA	2^{nd} (4)	fw	TGTAAAACGACGGCCAGTCATCTTCTATCTCCTTACTCC

continued on next page

continued from previous page

Gene	PCR round (Fragment)	Direction	Sequence
IL20RA	2^{nd} (4)	rv	CAGGAAACAGCTATGACCATCGTATGAGACCACCTTTG
IL22BP	1^{st}	fw	TATTTTCTGCCCCCAAATTC
IL22BP	1^{st}	rv	CAGATGCTGTCTTCCTTTTGG
IL22BP	2^{nd} (1)	fw	TGTAAAACGACGGCCAGTCCAGTTTTTAAACACTTCTCTG
IL22BP	2^{nd} (1)	rv	CAGGAAACAGCTATGACCGCCATTTTAGTGTCAATAATAAG
IL22BP	2^{nd} (2)	fw	TGTAAAACGACGGCCAGTTAACCTCATTAAGCCCTATG
IL22BP	2^{nd} (2)	rv	CAGGAAACAGCTATGACCCAGCATTTTCTTCAACCAGG
IL22BP	2^{nd} (3)	fw	TGTAAAACGACGGCCAGTAGGCAGACCAGTAAGGAG
IL22BP	2^{nd} (3)	rv	CAGGAAACAGCTATGACCGCACAATGACTAATCTGAG
IL22BP	2^{nd} (4)	fw	TGTAAAACGACGGCCAGTTAAAATGAAACATAGCCTGC
IL22BP	2^{nd} (4)	rv	CAGGAAACAGCTATGACCTTTGGTAACCTGCTATCATC

2.5 Genotyping

2.5.1 Genotyping of *IFNGR1*+95, *IFNGR1*-270 and *IFNGR1*-470 by ARMS

PEP DNA samples were genotyped blind to disease outcome or relation (in case of the family study) by ARMS (amplification refractory mutation system) (Newton et al., 1989) using allele specific PCR. For the T to C polymorphism at position +95 the following two allele specific primers were used: 5'GCC GCA GCC CTG CCA CGA A and 5'GCC GCA GCC CTG CCA CGA G. Note that these primers carry an additional mismatch at position 5 from the 3'end to increase specificity. The conserved primer was: 5'AGG CTC CAA GAC AAC CAG G. For the T to C polymorphism at position -270 the following two allele specific primers were used: 5' CAA AGG TTA AGG TCT TTC AA and 5' CAA AGG TTA AGG TCT TTC AG. The conserved primer was 5' TTC ACT AGC TAA GTC TCA GG. For the double mutation at position -470 5'-CCC TGC TTT TTC ATT TGA GGA A and 5'-CCC TGC TTT TTC ATT TGA GGC C were used as allele specific primers and 5'-CAA GCG CTG AAG GAC TTA GC as the conserved primer.

10X PCR Buffer (Bioline)	1.5μl
dNTPs 100mM (25mM each; Sigma)	0.06μl
$MgCl_2$ 50mM (Bioline)	0.57μl
forward primer 3μM (MWG)	2.5μl
reverse primer 3μM (MWG)	2.5μl
BIOTAQ DNA Polymerase (Bioline)	0.4U
Milli-Q H_2O	2.87μl
PEP DNA	5μl

Reactions were performed with 1.9mM $MgCl_2$, 0.6mM dNTPs, 0.4U Bio-Taq (Bioline). Cycling conditions were: 96°C for 1 minute, then five cycles

of 96°C for 35sec, 68°C for 45sec, 72°C for 35sec, followed by 21 cycles of 96°C for 25sec, 64°C for 50sec, 72°C for 40sec, and finally six cycles of 96°C for 35sec, 56°C for 1 minute, 72°C for 1 minute 30 seconds. The reactions for both alleles were run out adjacently on a 1% agarose gel stained with ethidium bromide. In order to assure that each PCR had worked a set of control primers, which amplifies a 796-bp fragment on chromosome 6 was included in each reaction (5'-TGC CAA GTG GAG CAC CCA A and 5'-GCA TCT TGC TCT GTG CAG AT).

2.5.2 Genotyping of *IFNGR1*-56 by LDR

The T to C polymorphism at position -56 was genotyped using Ligation Detection Reaction (LDR) (Barany, 1991). Ligation Detection Reaction (LDR) can be used for detection of single nucleotide polymorphisms (SNP), although it is rather costly compared to ARMS. First of all the region of interest is amplified by PCR. In the next step a common probe that is phosphorylated at the 5' end and a labelled allele specific probe are hybridized to the PCR product adjacent to each other, whereby the common probe binds immediately 5' of the SNP. The allele specific probe will bind 5' of the common probe with its 3' end at the site of the SNP (Figure 2.3). If the 3' end of the probe matches the site of interest, the two probes can then be ligated by the ligase in the next step (Figure 2.4). If there is a mismatch at this site, though, the ligase will not be able to ligate the two probes (Figure 2.2).

The design of the probes is limited by the position of the SNP for obvious reasons. Probes were designed to have very similar melting temperatures (T_m), which was aimed to lie around 65°C. The ligated LDR product was designed so that it was bigger than the two smallest size standards on the genotyping gel (i.e. $\geqslant$50 nucleotides)in order to facilitate detection. The probes were therefore polyA-tailed, the tails, though, were not included in the T_m calculation. When using an Applied Biosystems ABI3700 DNA analyzer one only has a limited choice of labels (HEX, FAM, NED), since ROX will be used for the standards.
Overview of the different steps:

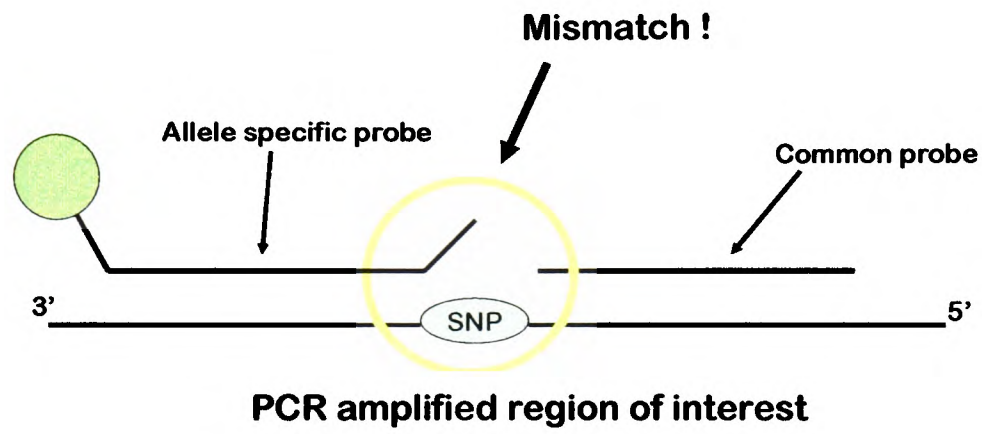


Figure 2.2 Common probe and allele specific probe are annealed to the PCR product. There is a mismatch at the 3' end of the allele specific probe. The two probes will therefore not be ligated.

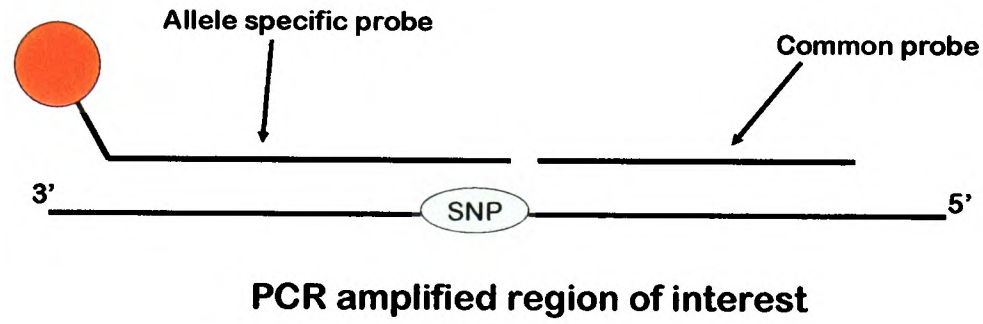


Figure 2.3 Common probe and allele specific probe are annealed to the PCR product. There is a perfect match at the 3' end of the allele specific probe. The two probes can therefore be ligated in the next step.

Koch 2003

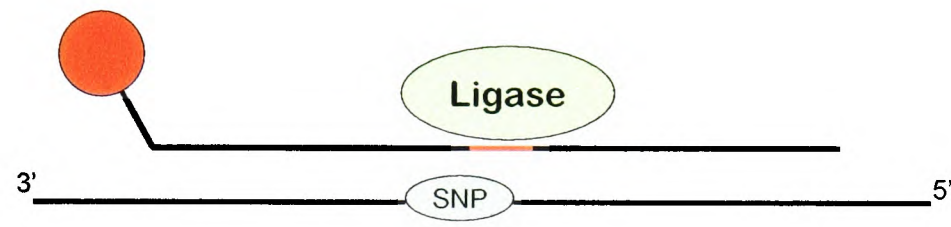


Figure 2.4 If there is a perfect match between allele specific probe and the PCR product at the 3' end of the allelic probe, allelic probe and common probe will be ligated by the ligase and the full length product is yielded.

Koch 2003

1. Amplification of PCR product
2. Removal of remaining BioTaq DNA polymerase
3. Phosphorylation of LDR common probe
4. Ligation Detection Reaction (LDR)
5. Analysis of ligation products by capillary electrophoresis

2.5.2.1 Amplification of PCR product

For this purpose a 670-bp fragment including the region of the *IFNGR1*-56 was amplified (primers were ordered from MWG and had the following sequences: 5'-TCA GGC TCC AAG ACA ACC AGG and 5'-CCA CGG AGC CCC AGT CTC G). The reaction volume was 15μl (excl. enzyme) and was set up in Costar Thermowell thin-wall polycarbonate 96 well PCR plates (Corning Incorporated).

	per 15μl reaction
10X PCR Buffer (Bioline)	1.5μl
dNTPs 100mM (25mM each; Sigma)	0.06μl
$MgCl_2$ 50mM (Bioline)	0.57μl
forward primer 5μM (MWG)	2.5μl
reverse primer 5μM (MWG)	2.5μl
BIOTAQ DNA Polymerase (Bioline)	0.4U (0.08μl)
Milli-Q H_2O	2.87μl
PEP DNA	5μl

10X Bioline PCR Buffer contains: 160mM $(NH_4)SO_4$, 670mM Tris-HCl (pH 8.8 at 25°C), 0.1% Tween-20.
PCR condition were as follows:

96°C 1:00

followed by 40 cycles of:

96°C 0:35

70°C 0:45

74°C 1:00

and finally

74°C 2:00

2.5.2.2 Removal of remaining BioTaq DNA polymerase

1/10 of the reaction volume of 1 mg/ml proteinase K in 50mM EDTA was added, i.e. 1.5μl per 15μl PCR product. Incubation was as follows:

37°C 15:00

55°C 10:00

90°C 10:00

2.5.2.3 Phosphorylation of LDR common probe

	Per 100μl reaction(μl)
PNK buffer@10X (New England Biolabs)	10
ATP 25mM (Sigma)	1
Common probe 0.1mM	2
Milli-Q H_2O	86
Polynucleotide Kinase(New England Biolabs)	1

Incubate as follows:

37°C 45:00

Add 100μl proteinase K (20μg in 100μl TE pH 8.0)

37°C 30:00

90°C 10:00

Store at −20°C. Final concentration is 1μM. Several common probes can be phosphorylated together if desired.

2.5.2.4 Ligation Detection Reaction (LDR)

Probes were ordered from Applied Biosystems. The following probes were used: 5'-FAM-AAA AAA TTA ACC GGG GCT GGA GGG CAG T-3', 5'-HEX- AAA AAA TTA ACC GGG GCT GGA GGG CAG C-3' and as common probe 5'-GCT GGG CTG GTC CCG CAG AAA AAA AAA A-3'. Set up on ice:

	per 15μl reaction(μl)
Taq DNA Ligase buffer @10X (NEB)	1.5
Phosphorylated common probe @1μM	0.15
Allelic probe(s) @1μM	0.15
Taq DNA Ligase @40U/μl (NEB)	0.05
PCR product	1
Milli-Q H_2O (Millpore)	to 15μl(12.02μl)

Overlay with 30μl mineral oil (Sigma).

Hot Block start (heat block up to 95° before placing samples on it).

95°C 1:00

then 15 cycles as follows:

95°C 0:15

66°C 4:00 (ligation)

When the cycling had finished, samples were placed on ice, 3μl of 100mM EDTA were added to prevent further ligase activity at non stringent temperatures.
The ligation temperature is a crucial step in this reaction. On the one hand it has to be high enough to avoid unspecific ligation, on the other hand enough product needs to be ligated in order to gain a readable peak on the genotyping gel. It should ideally range between 64°-72°C and should be optimised with control samples of known genotype.
In order to determine the optimal ligation temperature 9 samples were subjected to the thermocycling protocol as shown above at four different ligation temperatures: 63.0°C, 66.4°C, 70.2°C and 74.1°C on a gradient thermocycler (MJ Research). The products were initially run out on a ABI 377. Results for a representative example are shown in Figure 2.5. A temperature of 66.0° was found to work best for this reaction, since it gave high enough peaks, while keeping misligation at a minimum.

For the analysis on the Applied Biosystems ABI3700 DNA analyser 1μl of LDR product was diluted in 10μl of a formamide / ROX500 dilution (Genescan ROX500HD (ABI) diluted in Hi-Di Formamide (ABI) 1:100). Samples were stored at -20°C protected from light, before there were run on the ABI3700. Genotypes were called manually by visual inspection of the genotyping traces using the ABI sequence navigator software.

2.5.2.5 Confirmation of LDR results by RFLP

Genotyping results by LDR were confirmed on 19 samples using a digest with BtsI (New England Biolabs). Therefore the same PCR fragment that had been amplified for the LDR was used. BtsI would cut at the site of the SNP only in case of the wild type allele (T). In addition to this potential cut site the fragment also contained a conserved BtsI restriction site serving as an internal control in order to assess whether the digest had worked (Figure 2.6).

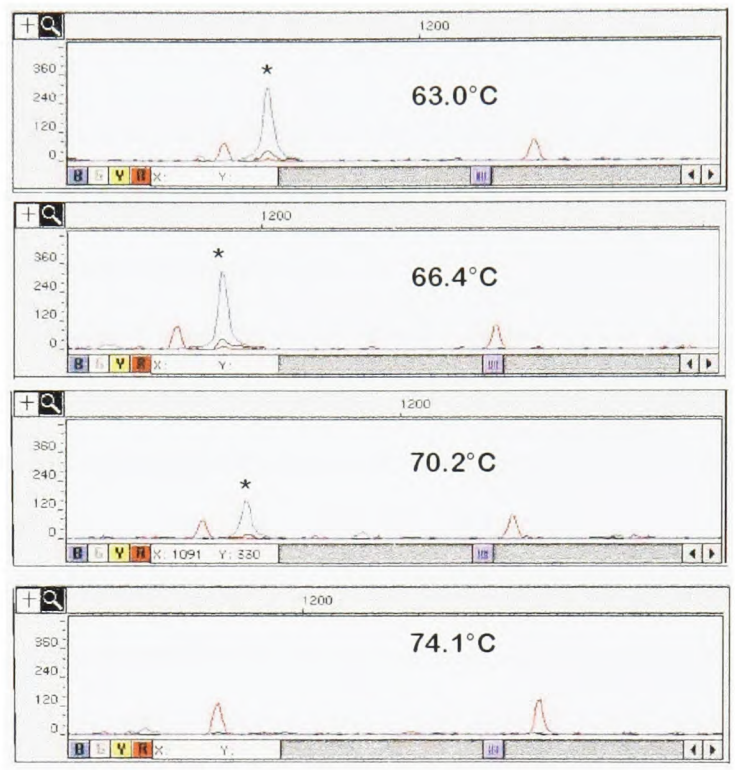


Figure 2.5 Determining the optimal ligation temperature for typing the *IFNGR1*-56 by LDR; figures show one a representative sample of known genotype at four different ligation temperatures. The peak for the allele specific LDR product is is marked by the asterix. The surrounding two peaks are the ROX500 size standards.

tcaggctccaagacaaccaggtgaagtccaagagttagtaaaataaggattgtggctcggct

gtggcctaatgcaaacttgcacaaccccaggaaaccgaaaaaaactggaagaagaattgcag

aatggggtgccaggttgaaagaccttaacctttgcactcaaattcctcccacacccagaagt

ccaggtcccgaccgcacgacgccgtgctcactgctgggtgctgcgcctgagtccgcctcctg

cggcttcccggacttgaccccgcccacgccctggtcccgcctcctgccgacgccggcacaga

ccccggtgacggaagtgacgtaaggccggggctggagggcagTgctgggctggtcccgcagg

cgctcggggttggagccagcgaccgtcggtagcagcatggctctcctctttctcctaccccct

tgtcatgcagggtgtgagcagggctgagatgggcaccgcggatctggggccgtcctcaggta

ccgtcgttcgcggcagggctgcggccgggtcgggacgagagggagggagggatccgccccag

ccgggaagccccgccccgcttctccgaggtcgccctagcccgggacccctgcgtcgggccct

gagcgggcgacggggacgcgaggtggggtcgcgagactggggctccgtgg

SNP

Figure 2.6 Figure shows the sequence of the amplified PCR fragment that was used for LDR and BtsI RFLP. The boxes indicate the BtsI recognition sequence.

The following protocol was used:

	per reaction
DNA	7μl
NEB Buffer No.4 (10X)	4.3μl
BSA (100X)	0.5μl
Milli-Q H_2O	37.2μl
BtsI	0.5μl

Samples were incubated for 5 hours at 37° in an MJ research Thermocycler. All 19 samples gave the same genotyping results as established by LDR.

2.5.3 MALDI-TOF Spectrometry

2.5.3.1 Principle

Two years after the beginning of this project *Matrix-assisted laser desorption-ionization-time-of-flight* (MALDI-TOF) mass spectrometry technology became available. MALDI-TOF mass spectrometry is a high throughput method of SNP analysis which was used for genotyping the majority of SNPs in the latter half of this thesis. A large number of peer reviewed papers describing and assessing this method is available (Tang et al., 1999; Jurinke et al., 2002) and multiple authors have used it for large scale genetic studies (Reich et al., 2001; Gabriel et al., 2002; Sabeti et al., 2002).
Nucleic acid is embedded in a solid matrix of organic molecules and is then irradiated by a short laser pulse at a wavelength close to a resonant absorption band of the matrix. The molecules are ionized and detected by a time-of-flight mass spectrometer. Ions with a larger mass-to-charge-ratio, m/z, travel from the matrix plate to the detector at a lower velocity than those with a higher m/z. With appropriate calibration, molecules differing in mass by 3 Daltons may be detected as two ions with different masses. As mostly

single-charged ions are produced by irradiation, complex mixtures may be analysed and the formation of secondary structure of nucleic acids is avoided. MALDI-TOF mass spectrometry is therefore considered to be more accurate than hybridization methods such as ARMS. SNP analysis is made possible by minisequencing at the SNP under investigation and detecting the resulting species of amplicons. The region of interest that incorporates the polymorphism is amplified and a primer is annealed immediately downstream of the SNP. An extension reaction is performed with DNA polymerase and a mix of deoxynucleotide triphosphates (dNTPs) and dideoxynucleotide triphosphates (ddNTPs). In this way, extension occurs with those nucleotides complementary to the template but terminates when a ddNTP is incorporated. The extension reaction undergoes temperature cycling giving linear amplification of product. Thus two products are produced by sequencing, each differing by as little as one nucleotide according to the single nucleotide variation within the PCR amplicon. This method enables several SNPs to be simultaneously measured by multiplexing several PCR reactions in a single well and relies on careful primer design to enable detection of ionization products of differing m/z ratios.

The process was fully automated with a robot dispenser, MALDI-TOF mass spectrometer and data analysis software all provided by the SEQUENOM company. The automated calling of genotypes was performed by the computer program MassARRAY RT Software thus eliminating manual error.

2.5.3.2 Preparation of PCR products for minisequencing

PEP DNA was used for the amplification reaction (see section 2.3). The PEP reaction was carried out as before in a 25μl reaction in a 384 well plate. Each well was diluted 1:20 with water containing 0.1 mg/ml phenol red solution. 29μl of water/phenol red solution was transferred to each well in a 384 well plate and 1.5μl of PEP PCR product was added to each well. Phenol red is a pH indicator. It is yellow in a 1:20 dilution and turns pink upon addition of DNA. It is chemically inert but visually aids in transferring PEP product to the water-filled wells for dilution.

Primers were designed using the MassARRAY Assay Design and MassARRAY Oligo Check software. PCR products were designed to have at least 25 Dalton difference between each sequenced species. These computer programs allow different assays to be multiplexed and a check of the primer design to be performed, enhancing accuracy and quality control of the final products for analysis.
2ul of diluted PEP was transferred to the corresponding well of a 384 well plate. A 5μl PCR reaction was then performed in each well which comprised of 0.4mM dNTP, 2.0mM $MgCl_2$, PCR buffer (67mM Tris-HCl, 16mM $(NH_4)_2SO_4$, 0.01% Tween-20), 0.125 units of BioTaq polymerase (Bioline) and 0.2μM of each primer. The plate was sealed with a microseal A lid and incubated on a PTC 225 Peltier thermal cycler (MJ Research). Cycling conditions were 96°C for 1 minute, 94°C for 45 seconds, 56°C for 45 seconds (5 cycles), 94°C for 45 seconds, 65°C for 45 seconds, 72°C for 30 secondes (29 cycles), 72°C for 10 minutes. A heated lid was used during cycling.

2.5.3.3 Minisequencing

The extension reaction and handling of the mass spectrometer was carried out by the core genomics facility at the Wellcome Trust Centre for Human Genetics. Arctic shrimp phosphatase (SAP, USB) was added to each sample to dephosphorylate residual dNTPs to prevent their incorporation into the primer extension reaction. To the 5μl of prepared PCR product were added 0.3units SAP, 20mM Tris-HCl (pH8.0) and 10mM $MgCl_2$, to give a final volume of 7μl. This was then incubated at 37°C for 20 minutes and at 85°C for 5 minutes. The extension reaction was performed by adding 0.2μl of termination mix specific for the reaction, 5.4μM of extension primer, and 0.58 units of MassEXTEND enzyme, thermosequenase (Amersham Biosciences), using the automated MassARRAY Liquid Handler. Cycling conditions were 94°C for 2minutes; then 55 cycles of 94°C for 5 seconds, 52°C for 5 seconds, and 72°C for 5 seconds.
0.006g of SpectroCLEAN resin was added to remove ddNTPs and dNTPs. Then 16μl of water were added. The samples were rotated for 10 minutes

so that the resin was well mixed in the samples. The samples were then centrifuged for 5 min at 1600rpm at RT.

2.5.3.4 MALDI-TOF Mass Spectrometry

The MassARRAY Nanodispenser was used to transfer a few nanolitres of extension reaction to the SpectroCHIP bioarray. This chip is provided by SEQUENOM pre-spotted with a matrix designed for DNA mass spectrometry. The sample quickly dries leaving a total of 384 pellets on a single chip. The chip was then read in the Bruker Biflex III Mass Spectrometer system and data collected and analysed using the MassARRAY Analyzer and MassARRAY RT Software (SpectroTYPER). An output is obtained with peaks visible at the respective masses of extension products. The software compiled the genotypes of each plate and the data was transferred to a Microsoft Access database constructed by Dr Kirk Rockett, Wellcome Trust Centre for Human Genetics, for analysis with clinical data.

2.5.3.5 Validation of genotypes

The company SEQUENOM claims MALDI-TOF to be a highly reliable method of genotyping detection with mass spectrometry being sensitive enough to differentiate between masses of 3 Daltons. There are now a large number of peer reviewed publications on this genotyping method indicating its outstanding accuracy and reliability (Tang et al., 1999; Jurinke et al., 2002; Reich et al., 2001; Gabriel et al., 2002; Sabeti et al., 2002). The accuracy of the in-house SEQUENOM technology was tested by Kate Rowlands and Neil Hanchard by direct sequencing of individuals genotyped for SNPs in the MHC region. In all, 12 SNPs were genotyped in 96 chromosomes and a further 20 SNPs were tested in 24 chromosomes by SEQUENOM, in most cases with representation of all 3 genotypes. There was 100% concordance between SEQUENOM genotypes and those found by direct sequencing.

2.5.3.6 Primer Sequences

Косн 2003

Table 2.3: MALDI-TOF Primers. The table shows the name of the SNP typed, the name of the primer (whereby 1^{st} and 2^{nd} refer to the forward and reverse primer and "UEP" to the extension primer), the terminator mix used and the primer sequence. All primers were ordered from METABION, Germany.

SNP	PRIMER	TERMINATOR	SEQUENCE
IFNg-1616	1^{st}		ACGTTGGATGACAGGTAAGGAGACTGAGTC
IFNg-1616	2^{nd}		ACGTTGGATGGGGCAAACTTGATTCCTGAC
IFNg-1616	UEP	ACT	TCTAGCTATATGATTGTGAGTTA
IFNg-308	1^{st}		ACGTTGGATGGTGGGCATAATGGGTCTGTC
IFNg-308	2^{nd}		ACGTTGGATGGTGGCATTTGGGTGTTGTAG
IFNg-308	UEP	CGT	TCCTTTAGACTCCTTGGGTC
IFNg-280	1^{st}		ACGTTGGATGGGTTTTGTGGCATTTGGGTG
IFNg-280	2^{nd}		ACGTTGGATGAATGGGTCTGTCTCATCGTC
IFNg-280	UEP	ACT	AAGGAGTCTAAAGGAAACTCTA
IFNg+874	1^{st}		ACGTTGGATGAGTTCCAAACATGTGCGAGT
IFNg+874	2^{nd}		ACGTTGGATGTATTCAGACATTCACAATTG
IFNg+874	UEP	CGT	ATTCTTACAACACAAAATCAAATC
IFNg+2200	1^{st}		ACGTTGGATGATGAATCACATGGCTGGCTG
IFNg+2200	2^{nd}		ACGTTGGATGTTGGAGCAAAGAAGGTCATC
IFNg+2200	UEP	ACT	GGTCATCAAACTTATACAGTGA
IFNg+3163	1^{st}		ACGTTGGATGTTTAGCTGTAATCCCCCAGC

continued on next page

continued from previous page

SNP	PRIMER	TERMINATOR	SEQUENCE
IFNg+3163	2^{nd}		ACGTTGGATGTTACCTTCCTCTTGGCTCTG
IFNg+3163	UEP	ACG	CTTGGGAACACTATGGCTAC
IFNg+3234	1^{st}		ACGTTGGATGTGTAATGGCAGAGCCAAGAG
IFNg+3234	2^{nd}		ACGTTGGATGTCACTCTAACCAATAGGGCC
IFNg+3234	UEP	ACT	GCCATTTAGATGATGCTTCATAA
IFNg+3686	1^{st}		ACGTTGGATGCCTCTCTATCATCAATCTACC
IFNg+3686	2^{nd}		ACGTTGGATGCAACAGTGAGAATCATTGACC
IFNg+3686	UEP	ACT	TAGTTACTCATTTGATGCATACAG
IFNg+4640	1^{st}		ACGTTGGATGTACTATCCAGTTACTGCCGG
IFNg+4640	2^{nd}		ACGTTGGATGCAAGTTCTGTCTGACATGCC
IFNg+4640	UEP	ACT	GCACTGGCTCAGATTGCAG
IFNg+5173	1^{st}		ACGTTGGATGGTGTTTGCCAGCATTGGATG
IFNg+5173	2^{nd}		ACGTTGGATGATTTCTAGCCCCTTCTCCAC
IFNg+5173	UEP	ACT	CACCTTCCTATTTCCTCCTTC
IFNg+5612	1^{st}		ACGTTGGATGTCCACCATACATAGGACTCC
IFNg+5612	2^{nd}		ACGTTGGATGGCATGGCATTCTCTAAGGTG
IFNg+5612	UEP	ACT	GGTGTCATGGAAAAATAGTTAAAG
IL26-11252	1^{st}		ACGTTGGATGAAGGACCCAGATGAGCAAAG
IL26-11252	2^{nd}		ACGTTGGATGTCTAGCCACACGGACTTGTG
IL26-11252	UEP	ACT	CTTGTGGCCCTCGCCTGA
IL26+28001	1^{st}		ACGTTGGATGGTGCAATGCAAGCGATCATG
IL26+28001	2^{nd}		ACGTTGGATGTTTTCTCAAGGCTCACCTCC

continued on next page

continued from previous page

SNP	PRIMER	TERMINATOR	SEQUENCE
IL26+28001	UEP	ACT	CCACGTCGCTGCACTTTGCAC
IL22-1394	1^{st}		ACGTTGGATGTCTGGCCACCTTCACAAATG
IL22-1394	2^{nd}		ACGTTGGATGATGGCACAGACCTAAGCAAG
IL22-1394	UEP	ACT	ACAGACCTAAGCAAGGTGCCACTGC
IL22-485	1^{st}		ACGTTGGATGTTTGTCTTAGTAGAGTTCAG
IL22-485	2^{nd}		ACGTTGGATGATGAGTCCGTGACCAAAATG
IL22-485	UEP	ACG	CCGTGACCAAAATGCTTACTCAG
IL22+708	1^{st}		ACGTTGGATGGGAGTCAGTGTAAGCTACAG
IL22+708	2^{nd}		ACGTTGGATGAAACCATCATCACCACCACC
IL22+708	UEP	ACG	TGGACGGCACACGGCCC
IL22+1254	1^{st}		ACGTTGGATGCTCAGGTTGCGTAAGATGAG
IL22+1254	2^{nd}		ACGTTGGATGCCCCTCAACAACTTAGACTC
IL22+1254	UEP	ACT	CTCAATGGATCCATTTCCACA
IL22+2611	1^{st}		ACGTTGGATGCCCGATCTCTTTTATACAGG
IL22+2611	2^{nd}		ACGTTGGATGGCTTACCAATTCAGACTTCC
IL22+2611	UEP	CGT	CCAATTCAGACTTCCATTTAACT
IL22+4582	1^{st}		ACGTTGGATGGAAAACATGGCATAAATGCTC
IL22+4582	2^{nd}		ACGTTGGATGCCAGTTCTCCAATTGCTTTG
IL22+4582	UEP	ACG	CAATTGCTTTGATCTCTCCAC
IL22+5697	1^{st}		ACGTTGGATGAAGGAAAGCTGTGTTTTGGG
IL22+5697	2^{nd}		ACGTTGGATGCCATCTTCCACCTATGAGAC
IL22+5697	UEP	CGT	CAGTGCCAAACTTACTAGTAGG

continued on next page

continued from previous page

SNP	PRIMER	TERMINATOR	SEQUENCE
IFNGR1-1882	1^{st}		ACGTTGGATGAACCCAGTCTTCTGTTCCTG
IFNGR1-1882	2^{nd}		ACGTTGGATGCTTCAGAAGGATACGTAGGG
IFNGR1-1882	UEP	ACT	AGGATACGTAGGGTGGTGG
IFNGR1-611	1^{st}		ACGTTGGATGCAAACCCAGAGAGGTAAGAG
IFNGR1-611	2^{nd}		ACGTTGGATGCTTCTCAGCAATTCAGTGTC
IFNGR1-611	UEP	ACG	CAATTCAGTGTCAAATCAGTTTAT
IFNGR1-470	1^{st}		ACGTTGGATGCTGTGGGAATCTGCACAAAC
IFNGR1-470	2^{nd}		ACGTTGGATGAAAGCCATTTCCAGTCCCTG
IFNGR1-470	UEP	ACT	TCCCTGCTTTTTCATTTGAGG
IFNGR1-72	1^{st}		ACGTTGGATGAGAGCCATGCTGCTACCGAC
IFNGR1-72	2^{nd}		ACGTTGGATGACCCCGGTGACGGAAGTGAC
IFNGR1-72	UEP	ACG	GACGGAAGTGACGTAAGGC
IFNGR1-56	1^{st}		ACGTTGGATGTGACGGAAGTGACGTAAGGC
IFNGR1-56	2^{nd}		ACGTTGGATGAGAGCCATGCTGCTACCGAC
IFNGR1-56	UEP	ACT	TGCGGGACCAGCCCAGC
IFNGR1+130	1^{st}		ACGTTGGATGCCGTCCTCAGGTACCGTCG
IFNGR1+130	2^{nd}		ACGTTGGATGTAGGGCGACCTCGGAGAAG
IFNGR1+130	UEP	ACG	GGGCGGATCCCTCCCTC
IFNGR1+20876	1^{st}		ACGTTGGATGTCTTCAGTAGTCACCACTTC
IFNGR1+20876	2^{nd}		ACGTTGGATGGCATACCGAAGACAATCCAG
IFNGR1+20876	UEP	ACT	TGGAACATACAGAAGAACTTTC
IL22BP-1755	1^{st}		ACGTTGGATGTCACCTCCTTCATGTCTCTG

continued on next page

continued from previous page

SNP	PRIMER	TERMINATOR	SEQUENCE
IL22BP-1755	2^{nd}		ACGTTGGATGGGACTGCTATTTTGGGTGTC
IL22BP-1755	UEP	ACT	GGGTGTCATAGAATTCACTGATAAG
IL22BP-1311	1^{st}		ACGTTGGATGCCCATGCACCAGTATGTTTG
IL22BP-1311	2^{nd}		ACGTTGGATGACAGTGGAGGAGGTGAAAAC
IL22BP-1311	UEP	ACG	GAGGTGAAAACCAGTTGGATT
IL22BP-1119	1^{st}		ACGTTGGATGATTACTCGGGTAAGAGATGC
IL22BP-1119	2^{nd}		ACGTTGGATGCTGTGCAATAGTGCAAGTGG
IL22BP-1119	UEP	ACT	TGGTTACCTGGATGGATTGC
IL22BP-1054	1^{st}		ACGTTGGATGAACCACTTGCACTATTGCAC
IL22BP-1054	2^{nd}		ACGTTGGATGACCAAAACATAAAGGAGTGG
IL22BP-1054	UEP	ACT	GGAGTGGTAACATTTAAAAAATAT
IL22BP-1014	1^{st}		ACGTTGGATGCTGGATAAAAGATGCTATTG
IL22BP-1014	2^{nd}		ACGTTGGATGAATGTTACCACTCCTTTATG
IL22BP-1014	UEP	ACG	CCACTCCTTTATGTTTTGGTTT
IL22BP-679	1^{st}		ACGTTGGATGCTGAATTCAAGTTCACAAACC
IL22BP-679	2^{nd}		ACGTTGGATGCATGTCCTTTAAGCAGTTCG
IL22BP-679	UEP	CGT	AGCAGTTCGGCATTCTGGAA
IL22BP-308	1^{st}		ACGTTGGATGTCCATTGGCTTCTCGAATCG
IL22BP-308	2^{nd}		ACGTTGGATGATGTCTAACACAGTGCCTGG
IL22BP-308	UEP	CGT	CTAACACAGTGCCTGGCATATA
6:144153554	1^{st}		ACGTTGGATGTAGGGTGCTGCTTCAGCTCAG
6:144153554	2^{nd}		ACGTTGGATGTGAGCTGGAGTGGTGTACTG

continued on next page

continued from previous page

SNP	PRIMER	TERMINATOR	SEQUENCE
6:144153554	UEP	ACT	GTGGTGTACTGCTTCCCAG
IL20RA-1620	1^{st}		ACGTTGGATGCTCTTAATTGCACCCTTCCC
IL20RA-1620	2^{nd}		ACGTTGGATGTGTGTGGACTTCCACACAAG
IL20RA-1620	UEP	ACG	CCACACAAGGCGGTGGGT
IL20RA-1554	1^{st}		ACGTTGGATGTTGTGTGGAAGTCCACACAC
IL20RA-1554	2^{nd}		ACGTTGGATGGGCAAGGGAAACTCAGAAAG
IL20RA-1554	UEP	ACT	CTCAGAAAGTCTAGGAATCATCATG
IL20RA-786	1^{st}		ACGTTGGATGCTGTATTACACTTAGGCTATG
IL20RA-786	2^{nd}		ACGTTGGATGAAAGTAGAAAGTAAAAAATC
IL20RA-786	UEP	CGT	GTAGAAAGTAAAAAATCATAGTAAG
IL20RA-310	1^{st}		ACGTTGGATGATACAGTTAGGCACTCAGCC
IL20RA-310	2^{nd}		ACGTTGGATGGTTCACACAAGAAACTCGCC
IL20RA-310	UEP	ACT	CGCCTAAGACATTTCTCAGGAG

2.5.4 Microsatellite Typing

The microsatellite in the first intron of the Interferonγ gene was typed by PCR amplifying approximately 110bp surrounding the multiallic marker using fluorescently labelled primer. The product was then run on an ABI3700 Sequence Analyzer and analysed using the Genescan software (ABI). The following two primers were used:

1. 5'FAM-GTAGGGTATTATTATACGAGC
2. ATTTTATTCTTACAACACAAAATC

Primers were ordered from ABI and kept at a 100μM stock concentration. 5μl of each sample was amplified in 384 well plates (ABGene) by adding 5μl

of the following buffer mix and 5μl of primer mix.

Buffer mix (1ml)

PCR gold buffer	300 μl
50mM MgCl	180μl
20mM dNTPs (5mM of each)	120μl
PCR water	360μl
TaqGold	40μl

Primer Mix

Water	940μl
fw primer (100μM)	30μl
rv primer (100μM)	30μl

PCR program

94°C 10 mins

30 of cycles of

94°C 30sec

56°C 30 sec

72°C 30 sec

then

72°C 5 mins

After amplification samples were diluted 1:30 in water. 1μl of the diluted sample was then transferred into 10μl of Hi-Di Formamide (ABI) and run on an ABI3700 Sequence Analyzer. Samples of known repeat length were

included for validation. The analysis was then performed using the ABI Genescan software.

2.6 Cell Culture

The following cell lines were used: A549 (Giard et al., 1973), a lung carcinoma cell line, RPMI 8226 (Matsuoka et al., 1967), a B lymphocyte cell line, Jurkat (Schneider et al., 1977), a T lymphocyte cell line, Raji (Pulvertaft, 1964), a Burkitt's lymphoma B cell line, and MonoMac6 (Ziegler-Heitbrock et al., 1988), a monocytic cell line. All cell lines were obtained from the American Type Culture Collection (ATCC). All cells were grown in 75 cm^2 flasks at 37.0°C in a humidified 5% CO_2 atmosphere.
The MonoMac6 cell line was cultured in RPMI 1640, containing 10% fetal calf serum, 100u/ml penicillin, 100μg/ml streptomycin, 2mM L-glutamine, 1mM pyruvic acid, 1mM oxaloaceitic acid, 1x nonessential amino acids and 9μg/ml insulin.
The A549 cells were cultured in F-12K (Kaighn's modification) Media, containing containing 10% fetal calf serum, 100u/ml penicillin, 100μg/ml streptomycin and 2mM L-glutamine. Cells were passaged as follows. The medium was removed and the cells were washed with warm PBS. The cells were then incubated with 1ml trypsin-EDTA for 10 minutes at 37.0° until they detached from the flask. Then 10ml warm PBS was added and the cells spun down at 1100rpm for 5 minutes. The supernatant was taken off and the cells resuspended in warm medium.
Jurkat, Raji and RPMI8226 cells were cultured in RPMI 1640, containing 10% fetal calf serum, 100u/ml penicillin, 100μg/ml streptomycin, 2mM L-glutamine.

2.7 Cell Stimulation

For Northern Blot analysis activation of cells was attempted with various stimulants. This was done as a screening experiment in order to find an

inducible system rather than investigating the response of each cell line to various concentrations of stimulants in great detail. A549 cells were stimulated with 40ng/ml Tumour Necrosis Factor-Alpha (TNF-α) (Sigma). RPMI 8226, Jurkat and Raji cells were stimulated with either 1μg/ml Bacterial Lipopolysaccharide (Serotype E.coli 055:B5) (LPS) (Sigma), 40ng/ml TNF-α (Sigma), 300U/ml IFNγ (Sigma), 10ng/ml IL1β (Roche), 250nM PMA / 150nM Ionomycin (Sigma), 1μg/ml Prostaglandin E2 (PGE2) (Sigma) or 50μg/ml Concanavaline A (ConA) (Sigma). MonoMac6 were stimulated with all of the above apart from PMA/Ionomycin. Unless where mentioned otherwise cells were grown as described in section 2.6. Twenty four hours after resuspending the cells in fresh medium the stimulant was added and the cell eventually harvested a further 24 hours after that.

2.8 RNA Extraction

RNA extraction was carried out using the RNeasy Mini kit (Qiagen) according to the manufacturer's instructions. Cells were harvested by centrifuging at 2000 rpm for 5 minutes and then suspended in 600μl of buffer RLT (Qiagen). Homogenization of the lysate was performed by loading the lysate directly onto a QIAGEN shredder spin column and centrifuging for 2 minutes at 12500 rpm. 600μl of 70% ethanol was mixed with the lysate before being loaded onto an RNeasy mini column (Qiagen) and centrifuged at 12500 rpm for 15 seconds. To wash the RNA bound to the mini column membrane 500μl of Buffer RPE (Qiagen) was added and the column centrifuged for 15 seconds at 12500 rpm; a further 500μl of buffer RPE was added to the column which was centrifuged for 2 minutes at 12500 rpm. To reduce the risk of Buffer RPE contamination the column was centrifuged for a further 1 min at 12500 rpm. RNA was eluted by adding 50μl of RNase-free water to the membrane and centrifuging at 12500 rpm for 1minute.
DNA contaminant was removed by DNase treatment by addition of 40000 units of DNase I (Roche), 2.5mM Tris-HCl and 5%Glycerol, pH7.6, final volume 60μl. RNA was precipitated by adding 150μl of ice cold 100% ethanol

with 0.5M NH_4Acetate, 0.18μg of glycogen, final volume 160μl, and incubating for 2 hours at −80°C. Samples were centrifuged at 15000 rpm for 15 minutes at 4°C, the supernatant aspirated and the pellet washed with 1ml of ice cold ethanol. The samples were centrifuged for 5 minutes at 15000 rpm, the supernatant aspirated and the resulting pellets were air dried before reconstitution in 10μl of RNase-free water containing 8 units of RNAguard (Amersham Pharmacia Biotech), an RNase inhibitor. Samples were stored at −80°C.

2.9 Northern Blot

2.9.1 Design of *IFNGR1* Probe

2.9.1.1 PCR

Two primers were designed that would amplify most of exon 7 of the *IFNGR1* gene. Primers were designed to incorporate an EcoRI site on the 5' end of one of the primers and a BamHI site on the 5' end of the other primer.

1. 5'-AATGGATCCTGGTCTATAACCAATCAAGG

2. 5'-AATGAATTCCTGTGGTAAGAAGTGCATCT

The PCR was performed using 2μl of each primer (5μM), 2μl of 10x Buffer, 2μl 8mM dNTPs, 2μl 25mM $MgCl_2$, 5.75μl H_2O, 0.25μl TagGold and 2μl of 100ng/μl genomic DNA. The PCR program was as follows: 94°C for 14 minutes, followed by 5 cycles of 94°C for 30 seconds, 55°C for 30 seconds and 72°C for 45 seconds; followed by 5 cycles of 94°C for 30 seconds, 60°C for 30 seconds and 72°C for 45 seconds; and finally 25 cycles of 94°C for 30 seconds, 65°C for 30 seconds and 72°C for 45 seconds. The PCR product was run out on a 1% agarose gel. The band was cut out and the fragment purified from the gel using the QIAEX II kit as per manufacturer's instructions (QIAGEN).

2.9.1.2 Cloning

The fragment was cloned into DH5α competent E.coli cells using the TOPO Cloning Kit (Invitrogen) as per manufacturer's instructions. The cells were plated out on XGal agar plates and incubated at 37°C overnight. Positive colonies were picked and grown overnight in LB Broth (Sigma) containing ampicillin. The cells were harvested and the plasmid DNA extracted using the QIAPREP Spin Kit (Qiagen) as per manufacturer's instructions. A small amount of plasmid was digested with EcoRI and BamHI and run out on a 1% agarose gel to check for the abundance of the expected fragment. The plasmid was was then sequenced using M13 dye primer sequencing to confirm that the correct fragment had been obtained.

2.9.2 Northern Blot

2.9.2.1 Formaldehyde Gel

2.1g Agarose were dissolved in 112 ml of RNAse free water and melted in a microwave oven. 15 ml of 10xMESA and 22.5 ml formaldehyde were added and the gel pored in a RNAse free gel mold. 5μl of RNA were mixed 1:4 with RNA sample loading buffer (Total volume 25 ml). The mix was heated at 65°C for 10 minutes and then immediately placed on ice for 5 minutes. The samples were then loaded onto the gel which was run at 25V overnight.

2.9.2.2 RNA Transfer

200 ml of 10xSSC were placed in both reservoirs of a BioRad gel tank (Bio-Rad). Three large 3M papers (30cm x 15cm) were wetted and placed across the bridge of the tank with both ends in SSC fluid. The gel was placed up side down on the large 3M papers. A Nylon transfer membrane (Hybond-N), 10 cm x 15 cm, wetted in sterile Rnase free water was placed on top of the gel. Several small 3M papers were placed on the membrane and 15 blotting paper sheets (Sigma) added on top. A 0.3-0.5 kg weight was placed on top

and the transfer left overnight. Following the transfer the membrane was autocrosslinked in a Stratagene crosslinking oven.

2.9.2.3 Labelling *IFNGR1* Probe

5μl of 25ng/μl *IFNGR1* DNA fragment was labelled with ^{32}P-dCTP using the MEGAPRIME DNA labelling kit (Amersham) as per manufacturer's instructions. The surplus free probe was removed by using Spin Columns TE-10 (Sigma). It was aimed for at least 70% incorporation.

2.9.2.4 Hybridisation

The membrane was placed in Hybaid bottle with RNA side facing inwards and with the edges not over lapping. 0.125ml/cm^2 (about 15 ml) of Rapid-hyb buffer (Amersham) were pre-warmed at 65°C. The membrane was then immersed in rapid-hyb buffer ensuring that the membrane was completely covered. The labelled DNA probe was denatured at 95°C for 5 minutes and placed on ice immediately for 5 minutes. It was then added to the rapid-hyb buffer in the Hybaid bottle. The bottle was then stirred and incubated in a hybridisation oven at 65°C. The buffer was removed an the membrane washed in subsequent washes of 2x SSC (with 0.1% SDS), 1xSSC (with 0.1% SDS), 0.1xSSC (with 0.2% SDS) all at 65°C for 20 minutes respectively. The membrane was exposed to a phosphor screen. Images were then analysed using a Cyclone Storage Phosphor Screen (Packard) using Optiquant v4.00 software (Packard). The membrane was stripped after use with boiling 0.1% SSC (with 0.2% SDS) and re-hybridised with a β-Actin probe for normalisation.

2.10 EMSA

2.10.1 Preparation of Nuclear Extracts

Cells were grown as described before. One batch was left unstimulated while the second batch was activated with a stimultant. A549 cells were stimulated

with 40ng/ml TNF-α for 24 hours. Raji, RPMI 8226 and Jurkat cells were activated with 250nM PMA / 150nM Ionomycin. MM6 were stimulated with 1μg/ml LPS. $5X10^6$ cells were harvested and washed twice with chilled PBS. They were then resuspended in Buffer A (10mM Hepes, pH 7.9, 10mM KCl, 0.1mM EDTA, 0.1mM EGTA) with 1mM DTT, 0.5mM PMSF, 0.5μg/ml leupeptin and 0.5μg/ml aprotinin, and divided into 1ml aliquots. These were incubated on ice for 15 minutes and 66μl of 10% NP-40 (Sigma) was added to each 1ml of suspension and the tube vortexed for 10 seconds. Following incubation on ice for 2 minutes the samples were centrifuged for 30 seconds at 12000 rpm at 4°C.

Supernatant was aspirated and a further 500μl of Buffer A was added. The pellet was vortexed for 10 seconds and a further 66μl of 10% NP-40 was added, agitated for 10 seconds, incubated on ice for 2 minutes and again centrifuged for 30 seconds at 12000rpm at 4°C.

The supernatant was aspirated and the pellet suspended in three times its volume Buffer B (20mM Hepes, pH7.9, 0.4M NaCl, 1mM EDTA, 1mM EGTA) with 1mM DTT, 0.5mM PMSF, 0.5μg/ml leupeptin and 0.5μg/ml aprotinin. The volume of pellet was estimated by subtracting the weight of the empty tube. The suspension was centrifuged at 12000 rpm at 4°C and the supernatant collected in a pre-cooled tube. Protein concentration of the resulting nuclear extract was measured by spectrophotometer at wavelength 595nM using the Biorad protein assay kit (Biorad) with bovine serum albumin standards at 0.5μg/μl, 1.0μg/μl and 2.0μg/μl. Samples were stored at −80°C.

2.10.2 DNA Probe Labelling

An oligonucleotide duplex was constructed as probe for the EMSA. Two complementary nucleotides were designed to incorporate the nucleotide variant under investigation (see table 2.4) and to have 5' overhang ends which would allow endlabelling with Klenow enzyme (DNA polymerase I) (Roche). 10μl of each 15μM complementary oligonucleotide was mixed with a buffer

containing 2mM Tris-acetate, 1mM magnesium acetate, 5mM potassium acetate, 0.1mM DTT, pH 7.9, to a final volume of 100μl. 50ml of mineral oil was then added to cover the surface. The tube containing the oligoduplex mix was then incubated in a water bath of boiling water and left to cool overnight in an insulating polystyrene container box. The resulting unlabelled annealed duplexes were stored at $-20°$C.

To radioactively label the annealed duplex, 3pmol (2μl) was mixed with 30μCi ^{32}P-dCTP, 0.15M each of dATP, dGTP, dTTP, 2μl of Klenow buffer (130mM potassium phosphate, 6.5 mM $MgCl_2$, 33μM dTTP, poly[d(A-T)], 0.833 A_{260}/ml, 33μM[^{14}C]dATP, 1mM dithioerythritol, 0.032mg/ml bovine serum albumin, pH 7.4) and 5 units of Klenow enzyme. The reaction mixture was incubated at room temperature for 40 minutes. The labelled probes were purified by spin column (Sigma S-1295) and incorporation of radioactivity by the probe was measured by scintillation counting (LS 1800 counter, Beckman). Probes were diluted to a final concentration of 2000 CPM/μl and stored at $-20°$C.

2.10.3 Binding Reaction and Electrophoresis

2μl of nuclear extract was mixed with 2μl of radiolabelled probe, 10mM HEPES, pH7.9, 1mM EDTA, 1mM EGTA, 12.5% glycerol and 0.5*mu*g dI-dC polymer (Amersham Pharmacia Biotech) and left at room temperature for 15 minutes. To perform competition assays, 10x and 100x molar excess of unlabelled probe was incubated with the crude nuclear extract for 10 minutes before addition of radiolabelled probe. Samples were then electrophoresed using a 5% polyacrylamide gel. This was made by adding 6.25 ml of 29:1 acrylamide:bis-acrylamide solution (Bio-RAD), 0.5M TBE, 500 ml of 10% ammonium persulphate and 37μl of TEMED, final volume 50ml. The gel was pre-run with 0.5M TBE buffer at 200V for 10 minutes before the samples were electrophesed at 200V for 2 hours at 4°C. Gels were vacuum-dried at 80°C on 3M chromatography paper (Whatmann) and exposed to a phosphor screen. Images were then analysed using a Cyclone Storage Phosphor Screen (Packard) using Optiquant v4.00 software (Packard).

Table 2.4: EMSA Probes. The polymorphic site is underlined. All probes carry a four base 5' overhang (AGCT)

SITE	ALLELE	DIRECTION	PROBE
-1882	G	fw	AGCTGTAGGGTGGTGGGTGCTGTGATTGT
		rv	AGCTACAATCACAGCACCCACCACCCTAC
	A	fw	AGCTGTAGGGTGGTGGATGCTGTGATTGT
		rv	AGCTACAATCACAGCATCCACCACCCTAC
-1804	G	fw	AGCTTGATGCTAGGAGGGCTGCCAGTAGA
		rv	AGCTTCTACTGGCAGCCCTCCTAGCATCA
	T	fw	AGCTTGATGCTAGGAGTGCTGCCAGTAGA
		rv	AGCTTCTACTGGCAGCACTCCTAGCATCA
-611	G	fw	AGCTGAGGCTGCCTGATAAACTGATTTGA
		rv	AGCTTCAAATCAGTTTATCAGGCAGCCTC
	A	fw	AGCTGAGGCTGCCTAATAAACTGATTTGA
		rv	AGCTTCAAATCAGTTTATTAGGCAGCCTC
-470	ins	fw	AGCTCTAAGTCTCAGGTTTCCTCAAATGAAA
		rv	AGCTTTTCATTTGAGGAACCTGAGACTTAG
	ddel	fw	AGCTCTAAGTCTCAGGCCTCAAATGAAA
		rv	AGCTTTTCATTTGAGGCCTGAGACTTAG
-270	T	fw	AGCTTGGGGTGCCAGGTTGAAAGACCT
		rv	AGCTAGGTCTTTCAACCTGGCACCCCA
	C	fw	AGCTTGGGGTGCCAGGCTGAAAGACCT
		rv	AGCTAGGTCTTTCAGCCTGGCACCCCA

continued on next page

continued from previous page

SITE	ALLELE	DIRECTION	PROBE
-72	C	fw	AGCTGTGACGTAAGGCCGGGGCTGGAGGG
		rv	AGCTCCCTCCAGCCCCGGCCTTACGTCAC
	T	fw	AGCTGTGACGTAAGGCTGGGGCTGGAGGG
		rv	AGCTCCCTCCAGCCCCAGCCTTACGTCAC
-56	T	fw	AGCTCTGGAGGGCAGTGCTGGGCTGGT
		rv	AGCTACCAGCCCAGCACTGCCCTCCAG
	C	fw	AGCTCTGGAGGGCAGCGCTGGGCTGGT
		rv	AGCTACCAGCCCAGCGCTGCCCTCCAG
+95	T	fw	AGCTCAGGTACCGTCGTTCGCGGCAGGGCT
		rv	AGCTAGCCCTGCCGCGAACGACGGTACCTG
	C	fw	AGCTCAGGTACCGTCGCTCGCGGCAGGGCT
		rv	AGCTAGCCCTGCCGCGAGCGACGGTACCTG
+130	G	fw	AGCTCGGGACGAGAGGGAGGGAGGGATCCG
		rv	AGCTCGGATCCCTCCCTCCCTCTCGTCCCG
	A	fw	AGCTCGGGACGAGAGAGAGGGAGGGATCCG
		rv	AGCTCGGATCCCTCCCTCTCTCTCGTCCCG

2.11 Western Blot

2.11.1 Protein gel running using NOVEX NuPAGE system

A 10% NuPAGE Bis-Tris gel (NP0301, Novex/Invitrogen) and 1xNuPAGE MOPS SDS running buffer (NP0001) were used. 10μl of nuclear protein

extract in 1xNuPAGE LDS Sample Buffer (NP0007) with NuPAGE Sample Reducing Agent (NP0004) was loaded and the gel run at 125V for 1.5 hours at room temperature.

2.11.2 Protein transfer to the membrane

1l of the protein transfer buffer containing 1x NOVEX Transfer buffer (LC3675) and 20% methanol was prepared. Blotting pads and pre-cut filter paper were soaked in the protein transfer buffer until saturated. The gel was soaked in the protein transfer buffer for 10-20 minutes. Hybond-P membrane (RPN1416 Amersham) was Pre-wetted in 100% methanol for 10 seconds, washed in distilled water for 5 minutes and equilibrated in the protein transfer buffer for 10 minutes. The transfer was then performed in an electroblotting cassette according to the manufacturer's instructions (Novex) at 4°C for 1-2 hours at 25 Volts in the protein transfer buffer. After that the membrane was rinsed in PBS.

2.11.3 Immunodetection

Non-specific binding sites on the membrane were blocked using 5% Marvel dried skimmed milk in PBST (PBS with 0.1% Tween) for 1 hour at RT. The blot was washed in PBST three times for 5 minutes each time. Incubation with primary antibodies in 1% milk in PBST was then performed. The antibody concentration was found to work optimally at a dilution of 1:1000. The blot was washed in PBST three times for 5 minutes each time. This was followed by incubation with secondary antibodies in 0.2% milk in PBST [anti-rabbit Ig horseradish peroxidase (NA934, Amersham), antibody dilution 1:5000]. The blot was again washed in PBST for three times 5 minutes. The signal was detected using cheiluminescense ECL Plus kit (Amersham) according to the manufacturer's instruction. A X-ray film was exposed for various times (starting 10 seconds) to obtain optimal signal intensity.

2.12 Immunocytochemistry

2.12.1 Cell culture

Adherent A549 cells were grown in sterile slide chambers (Lab-Tek Cat no. 177372). An Aliquot of 1 ml of trypsinized cells were added to each chamber and 1 ml of fresh medium added (F12K, penicillin-streptomycin, glutamine, 10% FCS) and placed at 37°C, 5%CO2. The cells were incubated for 48-72 hours.

2.12.2 Permeability procedure

The procedure was carried out at RT. 2 ml of medium were discarded and cells wash in PBS with gentle shaking for 5 minutes. The PBS was discarded and 4% formaldehyde in PBS (freshly made) added for 10 minutes. The formaldehyde was discarded and two washes in PBS for 5 minutes were performed. The PBS was discarded and 0.5% Triton-X100 in PBS (freshly made) added for 20 min. Then the Triton-X100 was discarded and washed twice in PBS for 5 minutes.

2.12.3 Immunocytochemistry

The PBS was discarded and the chamber and glue carefully removed. Incubation in 100μl of 10% goat serum (G9023 Sigma) in PBST (PBS +0.1% Tween) under cover slip followed. The slides were transfered to a coplin jar and washed in PBST three times for 5 minutes at room temperature. The slides were then incubated in 100 μl of primary antibodies at 37°C for 1 hour. [1:200 dilution in 2% goat serum PBST of STAT-1 antibody raised in rabbit and ATF-1 antibody raised in mouse (Antibodies from Sigma)]. The slides were transferred to a coplin jar and washed in PBST three times for 5 minutes at room temperature. This was followed by incubation in 100 μl of first secondary antibodies at 37°C for 30 minutes [1:200 dilution in 2% goat serum PBST of anti-rabbit-A488 raised in goat (A-11008 Molecular probes)].

The slides were transferred to a coplin jar and washed in PBST three times for 5 minutes at room temperature. Then incubation in 100 μl of second secondary antibodies at 37°C for 30 minutes was performed [1:100 dilution in 2% goat serum PBST of anti-mouse-Cy5 raised in rabbit (315-175-003 Jackson Immunoresearch Laboratories Inc)]. The slides were again transferred to a coplin jar and washed in PBST three for 5 minutes at room temperature. This was followed by counterstaining in Vectorshield mounting medium with propidium iodide (H-1300 Vector Laboratories Ins). The slides were then assessed by confocal microscopy.

2.13 *IFNGR1* Promoter Reporter Assay

2.13.1 Amplification of promoter constructs

Based on results by Merlin et al. (1997) indicating the minimal length promoter for full activity, constructs spanning from position -2 to -724 of the *IFNGR1* promoter were generated. Using the following two primers the *wild type* construct, i.e. with no deletion at -470/-471 and the T allele at -56, was amplified: *IFNGR1*-724/-707 `AGC TAG CGA GCA CAA GCG CTG AAG G` which contains a 5' NheI site and *IFNGR1*-2/-19 `AAT AGA TCT CTG CTA CCG ACG GTC GCT` containing a 5' BglII site. Using 20ng of Caucasian genomic DNA the fragment was amplified using the Expand Long Template Kit (Roche) according to the manufacturer's protocol using the following PCR program: 94°C for 14 minutes, followed by 5 cycles of 94°C for 30 seconds, 55°C for 30 seconds and 72°C for 45 seconds, followed by followed by 5 cycles of 94°C for 30 seconds, 60°C for 30 seconds and 72°C for 45 seconds, followed by followed by 25 cycles of 94°C for 30 seconds, 65°C for 30 seconds and 72°C for 45 seconds; finally 72°C for 2 minutes. The PCR product was then run out on a 1% agarose gel, cut out and purified with the QIAX II kit (Qiagen). The eluted fragment was then digested with NheI and BglII (both: New England Biolabs) for 2 hours at 37°C and subsequently purified using PCR purification spin columns (Qiagen). The fragment was then ligated into a

pGL3-Basic Vector (Promega) using the Rapid DNA Ligation Kit (Roche) according to the manufacturer's protocol (see Figure 2.7).

Косн 2003

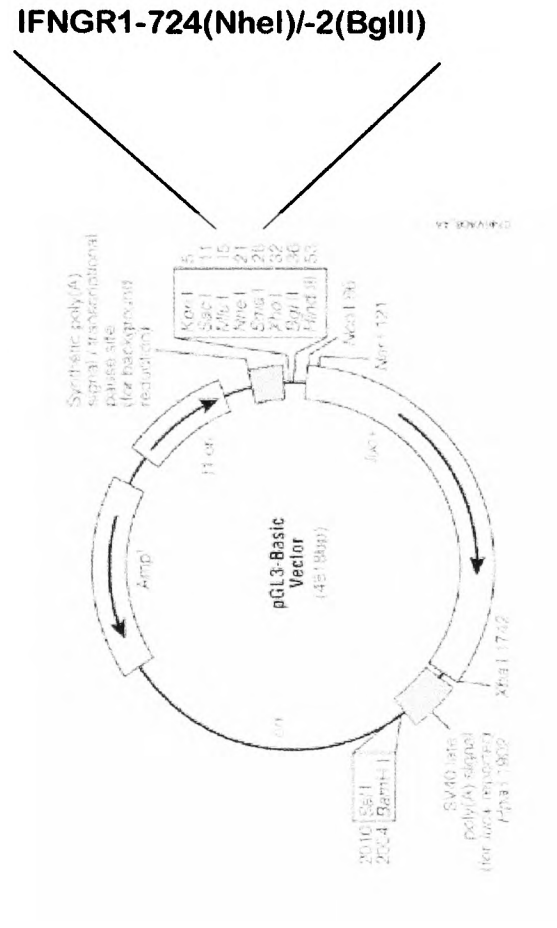


Figure 2.7 Basic pGL3 vector (Promega) in to which the *IFNGR1* promoter construct was ligated at the BglII and NheI site.

After Ligation the plasmids were transformed into competent DH5α E.coli cells. 50μl of competent cells were incubated with 5μl of ligation product on ice for 30 minutes, followed by a heat shock for 2 minutes at 37°C. The mixture was put back on ice for a few minutes. Then, 500μl of SOC medium were added at the tube incubated in a shaker for 1 hour at 37°C. The tube was then spun at 13.000rpm for 30 seconds, the 50μl of supernatant discarded and the pellet resuspended in the remaining 5μl of medium and plated out on an agar plate. The plate was left overnight at 37°C. Colonies were picked and again grown overnight at 37°C in LB broth containing 50μg/ml ampicillin. Minipreps were performed using the Qiaprep Spin Kit (Qiagen) according to the manufacturer's protocol and plasmid diluted to a concentration of 0.5μg/μl.

Using this construct as a template a construct containing the C allele at position -56 was amplified by using the *IFNGR1*-724/-707 as above and the following primer: *IFNGR1*-2/-61 `AAT AGA TCT CTG CTA CCG ACG GTC GCT GGC TCC AAC CCC GAG CGC CTG CGG GAC CAG CCC AGC GCT GCC`. The Expand Long Template Kit (Roche) was used with the following PCR program: 94°C for 10 minutes followed by 6 cycles of 94°C for 30 seconds, 56°C for 30 seconds and 68°C for 45 seconds; followed by 31 cycles of 94°C for 1 minute and 68°C for 1 minute.

A fragment containing the double deletion at -470/471 was constructed by initially amplifying two fragments using the conditions above, one by using the primer *IFNGR1*-724/-707 and the primer *IFNGR1*-454/-487 `GCT TTT TCA TTT GAG GCC TGA GAC TTA GCT AG`, and the other one by using the primer *IFNGR1*-2/-19 and the primer *IFNGR1*-487/-454 `CTA GCT AAG TCT CAG GCC TCA AAT GAA AAA GC`; or using *IFNGR1*-2/-61 and *IFNGR1*-487/-454 in case of the fragment containing both the -56 C allele and the double deletion. These two fragments were then used as templates to amplify the larger fragment containing the double deletion by using the primers *IFNGR1*-724/-707 and *IFNGR1*-2/-19 and the following PCR program: 92°C for 2 minutes, followed by 6 cycles of 92°C for 1 minute, 55°C for 45 seconds and 68°C for 90 seconds; followed by 26 cycles of 92°C for 30 seconds, 55°C for 30 seconds and 68°C for 90 seconds. All products were then gel purified, ligated

and transformed as before. All fragment were sequenced using DyePrimer to confirm the sequences.

2.13.2 Transfection of promoter constructs into human cells lines

All cells were transfected were with three different *IFNGR1* promoter constructs inserted into a pGL3 vector described above. Different transfection methods were optimised for the three different cell lines used.

2.13.2.1 A549 cells

A549 cells were plated out in 6 well plates (ABgene) at a density of 100.000/ml. After 24 hours the cells were washed once with PBS and 1.5ml of F12K Kaighn modified medium (GIBCO) replaced. The cells were then transfected using the Fugene Kit (Roche). Therefore, 0.5μg of DNA, 0.01μg pRL-TK Renilla control vector and 1μl of reagent were incubated in 50μ basic medium for 30 minutes, after which 450μl full medium (incl. 10% FCS) were added and the mixture added to the cells. The cells were incubated overnight. At this point half of the cells were stimulated with 40ng/ml TNF-α (SIGMA) and incubated for an additional 24 hours. The cells were harvested with the cell lysis buffer as described in the Dual-Luciferase Reporter Assay System (Promega) manual. The cells were processed and Renilla and Luciferase activities measured as described in the manufacturer's manual.

2.13.2.2 Jurkat cells

Jurkat cells were plated out in 6 well plates (ABgene) at a density of 2Mio cells/ml. The cells were transfected using the Superfect Kit (Qiagen). Therefore, 1.0μg of DNA, 0.02μg pRL-TK Renilla control vector and 10μl of reagent were incubated in 50μ basic medium for 30 minutes, after which 450μl full medium (incl. 10% FCS) were added and the mixture added to the cells.

The cells were incubated overnight. At this point half of the cells were stimulated with 250nM PMA and 150nM Ionomycin (SIGMA) and incubated for an additional 24 hours. The cells were harvested with the cell lysis buffer as described in the Dual-Luciferase Reporter Assay System (Promega) manual. The cells were processed and Renilla and Luciferase activities measured as described in the manufacturer's manual.

2.13.2.3 RPMI 8226 B-cells

The RPMI8226 B-cells were tranfected using the DEAE-Dextran method. Cells were fed 24 hours prior to transfection ($10x10^6$ cells per transfection). PBS and DEAE-dextran solutions were pre-warmed to 37°C. The cells were harvested in 50ml falcons. Cells were pelleted by centrifuging at 1200rpm for 5 minutes. The cells were washed by resuspending the pellet in 5ml warm (37°C) PBS. A second wash step was performed with PBS and the cells aliquot into 15ml Falcon tubes. The cells were again spun down at 1200RPM for 5 min. 2μg DNA (luciferase construct) plus 0.2μg pRL-TK DNA (renilla construct) in 0.5ml PBS-TR (Sigma) solution for each transfection point were added. 1.0ml of 1mg/ml DEAE-dextran solution (Sigma) was added and mixed gently by tapping the tube. The cells were resuspended at $10x10^6$ per ml by using 1ml of the 0.5mg/ml DEAE-dextran/DNA solution. The tubes were placed at an angle on their sides in an incubator at 37°C for 30 minutes and shaken gently every 15minutes. Then, cells were pelleted by centrifugation and the supernatant removed. Cells were washed with 5ml full medium (incl. 10%FCS). Cells were again pelleted by centrifugation and the supernatant removed. The cells were then resuspended in 5ml full medium and transferred to a medium tissue culture flask containing another 25ml full medium. Cells were incubated at 37°C. Half of the cells were stimulated with 250nM PMA and 150nM Ionomycin (SIGMA) after 24 hours and the cells incubated at 37°C and harvested after an additional 24 hours. The cells were processed and Renilla and Luciferase activities measured as described in the manufacturer's manual for the Dual-Luciferase Reporter Assay System (Promega).

2.14 Haplotype Construction

Haplotypes from unrelated individuals were constructed using PHASE, a program written by Stephens et al. (2001b). PHASE is based on a Bayesian method, which uses ideas from population genetics and coalescent theory that make predictions about the patterns of haplotypes to be expected in natural populations. By continuously sampling from known haplotypes the program makes predictions about unknown haplotypes. Xu et al. (2002) found this approach to outperform other methods such as the subtraction method. PHASE has been extensively used by various authors since (Newton et al., 2003; Lin et al., 2002). Family data were analysed by entering the family genotypes into a program called PHAMILY written by Ackerman (2001) as part of a DPhil thesis. PHAMILY examines the genotypes of the family trios and assigns phase were this is unambiguously possibly. While this process would be possible by simple inspection this can be tedious for a large data set and PHAMILY automates this process. The data for the two parents were then treated as data from unrelated individuals and entered into PHASE. In addition PHASE was provided with the information about haplotypic phase that was gained by examining the family trios. The run parameters used were number of iterations = 10000, thinning interval = 100 and burn-in = 10000. It was found that running the program longer using the data presented in this thesis did not significantly alter the final results. For the case control analysis case and control groups were run both separately and together which yielded similar results. PHASE is available at: `http://www.stats.ox.ac.uk/mathgen/software.html`

For the family based association study the Transmission Disequilibrium Test (TDT) and haplotype construction were performed using the TRANSMIT program by Clayton (Clayton, 1999), which can test for association between genetic markers and disease by examining the transmission of markers from parents to affected offspring even in the case of incomplete genotypic information. Documentation for this program can be found at the following website: `http://www.well.ox.ac.uk/docs/transmit.html`.

2.14.1 Fixation Index F_{ST}

The fixation index F_{ST} of Wright's F Statistic is a useful measure to estimate genetic divergence amongst subpopulations (Hartl and Clark, 1997b; Wright, 1921). F_{ST} is based on the reduction of the average heterozygosity H_S compared to the total heterozygosity H_T:

$$F_{ST} = \frac{H_T - H_S}{H_T}$$

According to Wright the range 0 to 0.05 may be considered as indicating little genetic differentiation for animal and plant species in general (Wright, 1978).

2.14.2 HKA Test

The Hudson, Kreitman and Aguade's test (HKA test) (Hudson et al., 1987) is based on the neutral theory of molecular evolution (Kimura, 1983) prediction that regions of the genome that evolve at high rates will also present high levels of polymorphism within species. The test requires data from an interspecific comparison of at least two regions of the genome, and also data of the intraspecific polymorphism in the same regions of at least one species. If two genes are undergoing neutral evolution, so that the fate of new mutations is determined entirely by genetic drift, and if their levels of interspecific divergence are similar (suggesting that the rates of mutation are the same), then their levels of intraspecific polymorphism likewise are expected to be similar. This is because under neutral model polymorphism is determined by the mutation rate and the effective population size. The HKA test was performed using the program `DnaSP` (Rozas and Rozas, 1999).

2.15 Tests for Deviation from the Neutral Model

Several methods are used to estimate the population parameter $\theta = 2M\mu$ where M is equal to $2N$ for diploid populations of the effective size N_e, and μ is the overall neutral mutation rate per generation.

2.15.1 Nucleotide Polymorphism

The level of nucleotide polymorphism, symbolised θ, is the proportion of nucleotide sites that are expected to be polymorphic in any suitable sample from this region of the genome. The estimate $\hat{\theta}$ equals the proportion of nucleotide polymorphism observed in the sample, symbolised as S divided by

$$a_1 = \sum_{i=1}^{n-1} \frac{1}{i}$$

where n is the size of the sample. The variance of $\hat{\theta}$, per nucleotide site, is given by

$$V\hat{\theta} = \frac{\hat{\theta}}{ka_1} + \frac{a_2\hat{\theta}^2}{a_1^2}$$

where a_1 is defined as above, k is the number of nucleotides in each sequence, and a_2 is a function of the number of alleles n in the sample, namely

$$a_2 = \sum_{i=1}^{n-1} \frac{1}{i^2}$$

2.15.2 Nucleotide Diversity

Nucleotide Diversity, denoted as π, is the average proportion of nucleotide differences between all possible pairs of sequences in the sample. In a sample of n sequences there are $n(n-1)/2$ pairwise comparisons. The average proportion of pairwise differences between the sequences in the sample is the estimate of the nucleotide diversity, $\hat{\pi}$. The variance of $\hat{\pi}$ is estimated as

$$V\hat{\pi} = \frac{b_1}{k}\hat{\pi} + b_2\hat{\pi}^2$$

where k is the number of nucleotides in each sequence and where

$$b_1 = \frac{n+1}{3(n-1)}$$

$$b_2 = \frac{2(n^2+n+3)}{9n(n-1)}$$

2.15.3 Tajima's D

The first test aimed specifically at testing neutrality in the context of infinite-sites models of sequence evolution was that of Tajima (1989). Two possible estimators of the population mutation rate, θ , are the average pairwise differences in a sample, π , and the number of segregating sites divided by the Watterson constant S/a_1 , where $a_1 = \sum_{i=1}^{n-1} 1/i$. Tajima suggested that the difference between these estimators could be used as the basis of a neutrality test

$$D = \frac{\pi - S/a_1}{\sqrt{Var(\pi - S/a_1)}}$$

Where the difference is normalised by the expected standard deviation of the difference. The test has two particularly desirable properties; the expectation of the numerator is zero, and (because the difference is scaled by the variance) the critical region of the test is little affected by the sample size or number of segregating sites. Tajima (1989) derived the variance analytically, under the assumption of no recombination, and showed how it could be estimated from the data. Watterson's estimator of θ, is only influenced by the number of segregating sites. In contrast, π is sensitive to allele frequencies at segregating sites, such that alleles at intermediate frequencies contribute more than alleles at low frequencies. Consequently, if a sample has an excess of rare variants, π will be less than Watterson's estimator and D will be negative. In contrast, if there is an excess of intermediate frequency variants, π will be greater than Watterson's estimator and D will be positive. While Tajima's D statistic was not designed to pick up any particular deviation from the mull model, it will tend to be negative under selective sweeps (and population growth) and positive under balancing selection (or population structure with sampling from many populations).

2.15.4 Fu and Li D and F Statistic

A second test was derived by Fu and Li (1993) who showed that the expected number of derived mutations that are present only once in a sample, η_e , is equal to θ . Consequently, it is possible to construct a test statistic in a similar manner to Tajima's

$$D = \frac{S - a_1\eta_e}{\sqrt{Var(S - a_1\eta_e)}}$$

and a similar statistic (called D*) if the direction of mutations is not known (in which case the statistic is based on the number of segregating sites at which the rare allele is only represented once - often called singletons - the expectation of which is $\theta n/(n-1)$. A negative value for the Fu and Li's D statistic indicates an excess of singletons (which would also give a negative Tajima D), and a positive value indicates a lack of singletons (which would typically, though not necessarily, give a positive Tajima D). However, certain population genetic scenarios, particularly selective sweeps, tend to generate an excess of singletons, to which this test is more sensitive than Tajima's D. While the D statistic compares the number of derived nucleotide variants observed only once in a sample with the total number of derived nucleotide variants, the F statistic compares the number of derived nucleotide variants observed only once in a sample with the mean pairwise difference between sequences.

Tajima's D, Fu and Li's D* and F*, θ and π were calculated using `Arlequin` (Schneider et al., 2000) available at: `http://lgb.unige.ch/arlequin/`;and `DnaSP` (Rozas and Rozas, 1999) available at: `http://www.ub.es/dnasp/`.

2.16 Linkage Disequilibrium

Linkage Disequlibrium, LD, is defined as the non-random association of alleles at linked loci. The first commonly used LD measure, D, was developed by Lewontin (1964). For a pair of diallelic loci, A and B, this

statistic measures the difference between two quantities:(1)the observed frequency of the co-occurence of an allele of A (A_1) and an allele of B (B_1) on the same chromosome and (2) the expected frequency of the co-occurence under linkage equilibrium. The observed frequency, denoted by P_11, is the proportion of chromosomes on which alleles A_1 and B_1 co-occur in a population. The expected value of P_11 under linkage equilibrium is the product of the alleles frequencies of A_1 and B_1 in the population. Thus, $D = P_11 - p_1q_1$, where the allele frequencies are symbolised as follows: $p_1 = f(A_1); p_2 = 1 - p_1 = f(A_2); q_1 = f(B_1); q_2 = 1 - q_1 = f(B_2)$. If D differs significantly from zero, LD is said to exist. The degree of LD between two loci is dependent on both the recombination fraction, θ, and time in generations, t since the first occurence of the mutation.

D is dependent on the allele frequencies in the population: its maximum value is given by $D_max = min(p_1q_2, p_2q_1)$, whereas the minimum value, D_min, is given by $max(-p_1q_1, -p_2q_2)$. D can be scaled as $D' = D/D_max$ (Lewontin, 1964). Another common scaling of D is to devide it by $\sqrt{p_1p_2q_1q_2}$ (Hill and Robertson, 1968). This quantity, commonly labelled as R or Δ, is equal to $\sqrt{\chi^2/N}$, where the χ^2 statistic can be obtained from the 2x2 table of the haplotype frequencies and N is the total number of haplotypes in the sample. In practice r^2 or δ^2 are commonly used:

$$\Delta^2 = \frac{(P_11 - p_1q_1)^2}{p_1p_2q_1q_2}$$

Linkage disequilibrium values were calculated by `HaploXT`, a program written by G.Abecasis, Wellcome Trust Centre for Human Genetics (Oxford, UK), which generates calculates observed linkage disquilibrium by inputting a set of haplotypes for each individual. These haplotypes were generated by `PHASE` (Stephens et al., 2001b). Thompson et al. (1988) and Akey et al. (2001) describe that haplotypic data increase the power of LD calculation. The output file of `HaploXT` can then be used as an input for `GOLD` (Graphical Overview of Linkage Disequilibrium) (Abecasis and Cookson, 2000) or `MARKER` version 1.0 , a graphical output program written by D.Kwiatkowski, Wellcome Trust Centre for Human Genetics (Oxford, UK). `HaploXT` is available at: `http://archimdes.well.ox.ac.uk/pise/haploxt-simple.html`.

GOLD is available at:
`http://www.sph.umich.edu/csg/abecasis/GOLD/index.html`.

2.17 Marker Selection

2.17.1 Entropy

In order to choose those markers which are best at capturing haplotypic diversity, marker selection by maximum entropy was used. The program ENTROPY, written by Dr Richard Mott (Wellcome Trust Centre for Human Genetics, Oxford, UK), examines each possible subset of markers that make up the haplotypes. The programme conveys numerically the entropy of each marker subset as an approximation to the haplotypic structure of the data. Each SNP contributes a varying amount of information to a haplotype according to its allele frequency and uniqueness to the haplotype. The aim is, given a set of SNPs and associated haplotypes with haplotype population frequencies, to choose a subset of SNPs that best approximate the haplotypic diversity in the population. The informativeness of different combinations of SNPs within a haplotype can be formulated as an entropy value and values for various combinations of SNPs can be compared to determine which is optimal. For a haplotype composed of 5 SNPs the programme calculates the entropy value for the various combinations of 1 SNP, then 2 SNPs, etc to 5 SNPs. The idea is to compute the SNP subset with maximum entropy. k is the number of diallelic markers typed across n individuals, producing $2n$ haplotypes, which are assumed known. There are $2k$ possible haplotypes, each of which may be represented as a binary number of length k. Haplotype i occurs with frequency $f(i)$ in the sample. The objective is to find the subsets of markers which capture as much of the haplotypic diversity as possible, weighted by the population frequencies of the haplotypes. If $p_i = f_i/2n$, the entropy of the data is defined as

$$E = -\sum(i)p_i log_2(p_i)$$

The maximum entropy is achieved if all haplotypes are present at equal quantity. The program calculates the entropy if only a subset of the available markers are used and therefore gives a measures how well haplotypic diversity can be captures by using less than the maximum number of available markers. When graphically presented (see results), the minimal number of SNPs necessary to capture all common diversity can be determined. By incrementally adding SNPs to form haplotypes of increasing diversity, the entropy value of the marker subsets will increase according to the combination of SNPs utilised. If after adding successive SNPs entropy fails to increase further, then using more SNPs to construct haplotypes adds no more information then sufficient SNPs have been used in haplotype construction. The point at which most haplotype diversity is captured, is when the entropy graph reaches a plateau, i.e. adding further SNPs fails to increase the value of entropy for the haplotype. The minimal number of SNPs that makes entropy plateau can therefore be found and then the combination of SNPs that has the maximal entropy value. Haplotypes had been constructed using `PHAMILY` and `PHASE`. Only haplotypes with a minimal frequency of $f \geq 0.05$ were used. `ENTROPY` is available at: `http://www.well.ox.ac.uk/~rmott/SNPS/download.shtml`.

2.17.2 HaploBlockFinder

This is a web-based utility for the identification of haplotype blocks based on a greedy algorithm. A greedy algorithm based approach was taken in Patil et al. (2001). The greedy algorithm is a step-by-step recipe for solving a problem. A greedy algorithm repeatedly executes a procedure which tries to maximize the return based on examining local conditions, with the hope that the outcome will lead to a desired outcome for the global problem. In some cases such a strategy is guaranteed to offer optimal solutions, and in some other cases it may provide a compromise that produces acceptable approximations. The program `HaploBlockFinder` is available at: `http://cgi.uc.edu/cgi-bin/kzhang/haploBlockFinder.cgi`.

Table 2.5 General 2x2 frequency table.

	DISEASE		TOTAL
	+	-	
Allele +	a	b	a+b
Allele -	c	d	c+d
Total	a+c	b+d	N

2.18 Microsatellite Analysis

The distribution of the IFNγ microsatellite alleles was analysed using the computer program `CLUMP`, written by Dave Curtis, Genetics Section, Institute of Psychiatry, London. It is available at `diamond.gene.ucl.ac.uk` in `/pub/packages/dcurtis`. `CLUMP` uses a Monte Carlo simulation whereby repeated, random simulations of the n x 2 data are carried out. The χ^2 value from the real data is compared to the number of times it is reached in the simulated data. The resulting χ^2 value termed *T1* compares the real data to that which may be expected by chance. 1000 simulations were performed in each case. A second statistic generated by `CLUMP` is the *T4* value. This groups the columns of data into a 2x2 table by clumping the first unexpectedly high value into one group and all those with the second high value into another. A χ^2 statistic is obtained by comparing differences between groups following Monte Carlo simulation.

2.19 2 x 2 Frequency Tables

All χ^2 values were calculated with the program Epi Info version 6 (available from the Center for Disease Control, USA, at `http://www.cdc.gov`) using 2x2 tables.

The χ^2 is calculated as $\chi^2 = \sum[(O-E)^2/E]$, whereby O and E represent the observed and expected frequencies respectively. This formular is mathematically identical to:

$$\chi^2 = \frac{N(ad-bc)^2}{(a+b)(a+c)(b+d)(c+d)}$$

In this thesis most χ^2 are given as Yates' corrected values. This is obtained with the following formula:

$$\chi^2_Y = \frac{N(|ad-bc| - \frac{N}{2})^2}{(a+b)(a+c)(b+d)(c+d)}$$

2.20 Statistical Analysis Software

Genotype frequency comparisons were carried out using the Statistical Package for the Social Sciences (SPSS© version 10.0) software and the EPI6 program (available on the web at `http://www.cdc.gov`).

Chapter 3

IFNGR1 promoter polymorphisms and susceptibility to severe malaria

3.1 Introduction

3.1.1 Background

1. As outlined in the introduction to this thesis, the cytokine IFNγ and its receptor are two key molecules in the host defence against an infection with the malaria parasite. Their roles are both in protection and development of the sequelae of the infection.

2. Previous studies have described coding polymorphisms in the *IFNGR1* gene that cause an abrogated receptor molecule to be transcribed and thereby lead to severe, often fatal infections with environmental mycobacteria. These mutations, however, are rare.

3. In comparison, severe malaria, however, is common and a Mendelian

pattern for susceptibility inheritance unlikely. The question therefore is whether there are common non-coding polymorphisms which are involved in susceptibility to this disease. These common polymorphisms would be expected to contribute to disease susceptibility rather than being the sole determining factor.

4. There is only limited knowledge about the variability of the *IFNGR1* promoter. Resequencing this area might therefore uncover new variant sites in an area which has been found to be important for the transcriptional regulation of this gene.

3.1.2 Approach

The promoter region of the *IFNGR1* was resequenced in order to search for variant sites in this area. An association study approach was used in order to test the variant sites found for their association with severe malaria. A case control study was carried out using DNA samples of 562 Gambian children with severe malaria. 569 DNA samples collected from cord blood were used as controls. The rational for using these samples as controls was the idea that they are likely to reflect the allele frequency of the general population. The samples were roughly matched for ethnicity although analysis was carried out for each ethnic group individually. Parental DNA for all cases was also available which did allow to carry out a family analysis with the Transmission Disequilibrium Test (TDT). It is worth noting that therefore the two studies were overlapping. As the TDT is a different form of analysis it would nevertheless be possible to compare the two study designs to some extent. All samples were genotyped for the four *IFNGR1* polymorphisms found by resequencing by the Amplification Refractory Mutation System (ARMS) method or by the Ligation Detection Reaction (LDR).
Out of a total of 562 cases, 68 had cerebral malaria (CM) plus severe malarial anaemia (SA), 308 had CM alone, and 186 had SA alone. 238 were Mandinka, 112 Jola, 62 Fula, 67 Wollof and 83 of other ethnic origin. For case-control

analysis, the population controls were 569 umbilical cord blood samples collected at Gambian health clinics: 174 were Mandinka, 89 Jola, 138 Fula, 92 Wollof and 76 of other ethnic origin. For TDT analysis the controls were both parents of the affected child.

3.1.3 Aims

The aim of this preliminary study was to screen the part of the *IFNGR1* promoter that Merlin et al. (1997) suggested as being functionally important for common genetic variants. These were then tested for their association with susceptibility to severe malaria. While this localised approach has the risk that it initially might not identify any *true* functional variants, these variants might nevertheless be in linkage disequlibrium with the markers found in the *IFNGR1* promoter. Further extendend studies would then be necessary to localise any putative effect found.

Kosh 2003

3.2 Results

3.2.1 Mutation Detection

The first 500bp 5' of the translational start site of the *IFNGR1* gene had been screened for common mutations in 72 chromosomes of individuals with tuberculosis from The Gambia prior to my arrival in the host laboratory. Using ABI Dye Primer Sequencing three nucleotide polymorphisms had been found in this region: a T to C single nucleotide polymorphisms (SNP) at positions -56 and -270 and a double deletion of the bases at position -470/-471 (TT). As part of my project the sequencing was extended to the area between about -1400 of the promoter region and up to shortly beyond the first exon (+88). Therefore 68 chromosomes (34 individuals) from Gambian Cord Blood samples were sequenced using MegaBACE 1000 technology. This approach confirmed the three polymorphisms that had been described previously and detected a new single nucleotide polymorphism at position +95 (T to C) which is located in the first intron close to the splice site. Furthermore all samples differed at position -611 from the sequence that is published on the database of the National Center for Biotechnology Information (NCBI). They all had an A instead of a G at this position. No other polymorphisms were found. All polymorphisms were initially confirmed by sequencing of the reverse strand, and further validated by genotyping a larger number of individuals as described below.

3.2.2 Allele Frequencies in Cord Blood controls

The allele frequencies of the polymorphisms observed in our control population (cord blood) are given in Table 3.1. The allele frequencies in the four main ethnic groups in The Gambia, the Mandinka, the Jola, the Fula and the Wollof, vary for the *IFNGR1*+95C between 0.46 and 0.53, for the *IFNGR1*-56C between 0.48 and 0.52, for the *IFNGR1*-270C between 0.01 and 0.03 and for the *IFNGR1*-470delTT between 0.10 and 0.14. The overall control population was in Hardy Weinberg equilibrium (HW) for the *IFNGR1*+95

(0.48), for the *IFNGR1*-56 (p=0.64) and the *IFNGR1*-470 (p=0.06). For the rare *IFNGR1*-270, though, it was out of HW (p= 5 x 10^{-15}) due to excess of homozygotes, though the numbers were too small to infer true deviation from equilibrium.

The population of the Gambia consists of several ethnic groups, the largest of which are the Mandinka, Jola, Fula and Wollof. Although they are clearly distinct groups, sources of information on their population history are limited. The Mandinka are thought to originate from Kangaba, a state of the ancient empire of Mali. About 700 years ago they expanded in the region which today is The Gambia and The Senegal. The Fula are phenotypically distinct in that they have lighter skin and somewhat Caucasoid facial features. The Fula are believed to be of North-East African origin and are a subgroup of the nomadic pastoral Fulani who can be found throughout West-Africa. Although today the different ethnic groups live in close proximity to one another, the extent of intermarriage between them (Allsopp et al., 1992) is limited, particularly in rural areas where they often live in separate villages.

The fixation index F_{ST} of Wright's F Statistic is a useful measure to estimate genetic divergence amongst subpopulations (Hartl and Clark, 1997b; Wright, 1921). F_{ST} is based on the reduction of the average heterozygosity H_S compared to the total heterozygosity H_T:

$$F_{ST} = \frac{H_T - H_S}{H_T} \tag{3.1}$$

If F_{ST} is calculated on this basis the values for the four different SNPs at this locus range from 0.002 to 0.007. According to Wright the range 0 to 0.05 may be considered as indicating little genetic differentiation for animal and plant species in general (Wright, 1978) but one would want to use more subtle criteria in assessing the differentiation of human populations. It is

clear from physical appearance that the ethnic groups that make up the Gambian population have significant genetic differences, and therefore the following analyses consider ethnicity as a significant factor.

Косн 2003

Table 3.1 Allele frequencies for four *IFNGR1* promoter polymorphisms in different Gambian ethnic groups. Allele frequencies (f) are shown for the variant allele. H_S and H_T represent the heterozygosity of each ethnic group for the particular allele and total heterozygosity in all ethnic groups respectively. F_{ST} values for *IFNGR1*-470, -270, -56 and +95 are 0.004, 0.007, 0.002 and 0.002 respectively.

	IFNGR1-470DEL		*IFNGR1*-270C		*IFNGR1*-56		*IFNGR1*+95	
	f	H_S	f	H_S	f	H_S	f	H_S
all ethnic groups	0.111	0.197(H_T)	0.018	0.036(H_T)	0.499	0.500(H_T)	0.494	0.500(H_T)
Mandinka	0.140	0.241	0.013	0.026	0.525	0.499	0.525	0.499
Jola	0.096	0.174	0.034	0.066	0.477	0.499	0.475	0.499
Fula	0.097	0.175	0.008	0.016	0.504	0.500	0.492	0.500
Wollof	0.115	0.204	0.011	0.022	0.483	0.499	0.483	0.499
other groups	0.095	0.172	0.033	0.064	0.480	0.499	0.463	0.497

3.2.3 Case Control Study

The allelic association with susceptibility to severe malaria was investigated (Table 3.2). Considering all ethnic groups together in a stratified analysis, there was no significant disease association for any of the four polymorphisms. In a subgroup analysis, the four main Gambian ethnic groups were considered separately, i.e. Mandinka, Jola, Fula and Wollof. Amongst the Mandinka, who were the largest single group in this study, a loss of heterozygotes in the case group was observed and in fact heterozygosity for *IFNGR1*-56 was associated with protection against cerebral malaria (p=0.02; OR 0.54(0.31-0.92)) (Table 3.3). It is worth noting that the group of cases was out of Hardy-Weinberg Equilibrium (p=0.0049). Among Mandinka patients with cerebral malaria, *IFNGR1*-56 heterozygosity appeared to protect against fatal outcome (heterozygote frequency in survivors 33/76, in fatal cases 4/28: P=0.01; OR 0.22 (95% CI 0.06-0.75))(Table 3.4). When one compares those individuals who develop severe malaria, i.e. cerebral malaria or severe malaria anaemia, heterzygotes were less likely to develop cerebral malaria rather than severe anaemia (P=0.004; OR 0.40 (95%CI 0.21-0.75)). A similar trend of heterzygote reduction in the case group can also be seen in the Jola and Fula group. With the *IFNGR1*+95 a similar trend can be observed and heterozygosity of this polymorphism is associated with protection from death following cerebral malaria (P=0.01; OR 0.32 (95%CI 0.12-0.85)). In the Mandinka the double deletion at position *IFNGR1*-470 was associated with protection against severe malaria in general (p=0.02; OR 0.58 (95%CI 0.36-0.93)). No significant association of the *IFNGR1*-270 C allele with severe malaria in any of the ethnic groups could be found. The allele frequency for the *IFNGR1*-270 C allele is very low, though, and in order to detect any effect a far larger sample size would be necessary.

Table 3.2 Genotype frequencies for the four polymorphisms at the Interferon-γ receptor *IFNGR1* promoter gene locus in all ethnic groups. This study included 562 individuals with severe malaria (308 with cerebral malaria only, 186 with severe malarial anaemia only and 68 with both phenotypes) and 569 Mandinka umbilical cord blood samples (controls). Discrepancies in sample size are due to genotyping failures.

	Cord Blood Controls *Genotypes%*	Cerebral Malaria Only *Genotypes%*	Severe Anaemia Only *Genotypes%*	Cerebral Malaria And Severe Anaemia *Genotypes%*
IFNGR1+95TT	136(26.4)	64(25.6)	25(20.7)	19(45.2)
IFNGR1+95TC	250(48.4)	117(46.8)	68(56.2)	14(33.3)
IFNGR1+95CC	130(25.2)	69(27.6)	28(23.1)	9(21.4)
IFNGR1-56TT	139(25.6)	74(28.8)	36(20.4)	22(35.5)
IFNGR1-56TC	266(49)	112(43.6)	95(54.0)	26(41.9)
IFNGR1-56CC	138(25.4)	71(27.6)	45(25.6)	14(22.6)
IFNGR1-270TT	519(96.6)	257(96.6)	153(93.9)	59(98.3)
IFNGR1-270TC	14(2.6)	9(3.4)	10(6.1)	1(1.7)
IFNGR1-270CC	4(0.7)	0(0.0)	0(0.0)	0(0.0)
IFNGR1-470ii	438(77.8)	233(82.0)	147(82.6)	53(82.8)
IFNGR1-470delTT/insTT	122(21.7)	48(16.9)	30(16.9)	11(17.2)
IFNGR1-470ddel	3(0.5)	3(1.0)	1(0.6)	0(0.0)

Table 3.3 Genotype frequencies for the four polymorphisms at the Interferon-γ receptor *IFNGR1* promoter gene locus in Mandinka. This study included 238 ethnic Mandinka individuals with severe malaria (123 with cerebral malaria only, 81 with severe malarial anaemia only and 34 with both phenotypes) and 174 Mandinka umbilical cord blood samples (controls). Discrepancies in sample size are due to genotyping failures.

	Cord Blood Controls	Cerebral Malaria Only	Severe Anaemia Only	Cerebral Malaria And Severe Anaemia
	Genotypes%	*Genotypes%*	*Genotypes%*	*Genotypes%*
IFNGR1+95TT	37(22.7)	31(33.3)	9(16.1)	9(42.9)
IFNGR1+95TC	81(49.7)	38(40.9)	35(62.5)	6(28.6)
IFNGR1+95CC	45(27.6)	25(26.6)	12(21.4)	6(28.6)
IFNGR1-56TT	36(22.2)	37(35.9)	12(15.2)	9(30.0)
IFNGR1-56TC	82(50.6)	37(35.9)	46(58.2)	12(40.0)
IFNGR1-56CC	44(27.2)	30(28.8)	21(26.6)	9(30.0)
IFNGR1-270TT	151(96.8)	101(95.3)	71(93.4)	28(96.6)
IFNGR1-270TC	3(1.9)	5(4.7)	5(6.6)	1(3.4)
IFNGR1-270CC	2(1.3)	0(0.0)	0(0.0)	0(0.0)
IFNGR1-470ii	125(72.7)	94(81.7)	68(86.1)	26(81.3)
IFNGR1-470delTT/insTT	46(26.7)	20(17.4)	11(13.9)	6(18.7)
IFNGR1-470ddel	1(0.6)	1(0.9)	0(0.0)	0(0.0)

Table 3.4 Relationship of *IFNGR1*-56C with death from severe malaria in Mandinka. Comparison of genotypes from Mandinka patients with cerebral malaria (CM), who died in course of the disease whilst in hospital, versus those with CM who survived. Heterozygotes were protected against fatal outcome following cerebral malaria (OR 0.22, P=0.01).

	CM Deaths	CM Survivors
IFNGR1-56TT	15	22
IFNGR1-56TC	4	33
IFNGR1-56CC	9	21

3.2.4 Haplotype Construction

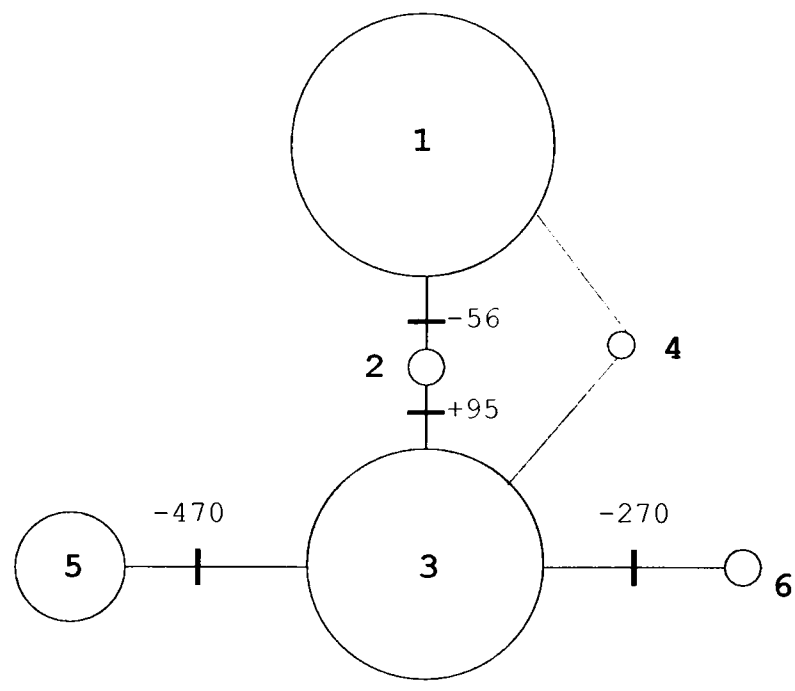
The effect found in the Mandinka raises the possibility of a haplotypic association, for which reason it was necessary to establish the haplotypes for this region. As optimal haplotype construction is facilitated by genotypic information from relatives, a family analysis was performed in order to explore this further. This approach then also allowed to investigate the transmission of both markers and haplotypes from parents to their affected child.
Complete Gambian family trios were typed, in which the offspring was affected by severe malaria, for all four mutations that were found in the *IFNGR1* promoter. Haplotypes were constructed using the TRANSMIT program (Clayton, 1999) which uses the pedigree data to infer haplotypes and test for association between genetic markers and disease by examining the transmission of markers from parents to affected children even in the case of incomplete genotypic information. Families with insufficient genotyping data and cases of non-paternity were excluded. Only six out of sixteen possible haplotypes were found at this locus to be present at a frequency $\geqslant$1%. Two of these account for about 85% of all haplotypes. Haplotype frequencies

Table 3.5 Estimated haplotype probabilities at the *IFNGR1* promoter locus using TRANSMIT.

		Estimated Haplotype Probabilities	
IDENTIFIER	HAPLOTYPE	ALL ETHNIC GROUPS	MANDINKA
1	insTT T T T	0.48	0.47
2	insTT T C T	0.02	0.03
3	insTT T C C	0.37	0.37
4	insTT T T C	0.01	0.02
5	delTT T C C	0.09	0.08
6	insTT C C C	0.02	0.02

based on maximum likelihood calculations are shown in Table 3.5. Haplotype frequencies for Mandinka are similar to the general Gambian population. As chapter 4 will show haplotype 3 is the ancestral haplotype. On its background the -56 and -95 mutations have occurred. The order in which these events occurred remains unclear as both intermediate haplotypes are at very low frequency. Therefore either of them could represent a spurious result due to genotyping error or in fact represent a recombination event between haplotypes 1 and 3 (haplotype 4). The double deletion at *IFNGR1*-470 and the *IFNGR1*-270C allele appear to be in absolute linkage with the *IFNGR1*-56C and the *IFNGR1*+95C alleles and arose on the background of haplotype 3. Figure 3.1 shows a putative haplotype genealogy for this locus.

Figure 3.1 Haplotype genealogy of four polymorphisms at the *IFNGR1* locus. The figure shows a putative tree of the six haplotypes formed by four polymorphisms in the *IFNGR1* promoter region. The haplotype numbering refers to the haplotype identifiers in Table3.5. Only haplotypes with a probability $\geqslant$1% are shown.



Koch 2003

3.2.5 Family analysis: disease association

Parental DNA samples were available for 483 of the severe malaria cases included in the case-control study with either severe malaria anaemia or cerebral malaria, allowing the reanalysis of these cases using the transmission/disequilibrium test (TDT) (Spielman et al., 1993). This provides a way of excluding potential ethnic artefacts that might have confounded the case-control comparison. The TDT was performed for all four polymorphisms individually. Transmission of alleles was not significantly different from what would have been expected by chance alone. Among Mandinka families affected by cerebral malaria, 70 were informative for the *IFNGR1*-56C allele and the ratio of transmitted to non-transmitted alleles was 32:38; while among Mandinka families affected by severe malaria, 54 were informative for the *IFNGR1*-470delTT allele and the ratio was 23:31. i.e. for both alleles there was a weak trend to protection, although this was not statistically significant. Given the strong protective effect that was found for heterozygotes for the *IFNGR1*-56, a variant of the TDT was used to investigate a similar effect in the family study. In cases where there was a theoretical 50:50 chance for the offspring to be either homozygote for either allele or heterozygote given the genotypes of the parents, it was analysed whether the offspring was more or less likely to be heterozygote from what would have been expected by chance alone. Parental genotypes for this scenario would be: TCxTT, CCxTC and TCxTC. When looking at the first combination (TCxTT), a child of these two parents could either carry the genotype TT or CT. For the CCxTC it could be either CT or CC. And for two parents with each being TC the children could carry any of the three genotypes: TT, TC or CC. If we now consider genotypes instead of alleles, i.e. cases were the child is heterozygote, and call these "transmitted" in analogy to the traditional TDT we can calculate the χ^2 using Spielman's formula:

$$\chi^2 = \frac{(T - NT)^2}{(T + NT)} \tag{3.2}$$

were T would be "transmitted" and NT "non-transmitted".

Table 3.6 Comparing the "transmission" of the heterozygote genotype in Mandinka Cerebral Malaria families. The table shows different genotype combinations for parents and the genotype of their child. For all three parental genotype combinations there was a 50% chance of having a heterozygote child. In 19 cases the heterozygote genotype was "transmitted" while it was "non-transmitted" in 33 cases.

		Genotype of Child		
		TT	CT	CC
Genotypes of both Parents	TC X TT	12	5	-
	CC X TC	-	6	11
	TC X TC	6	8	4

Within the Mandinka families with cerebral malaria there were 52 matings with a 50% chance to either produce a heterozygote or homozygote child. Only 19 children were heterozygote, whereas 33 children were homozygote for one of the alleles (OR 0.52, P=0.05) (Table 3.6).

Results of multiple marker haplotypes, i.e. considering all four variants in their genomic context, were analysed for cerebral malaria families using TRANSMIT v2.5. which examines haplotype transmissions from parents to affected offspring, even if phase is unknown and parental genotypes are incomplete. No significant disease association in the overall group and the Mandinka was found for any of the observed haplotypes (Table 3.7).

3.2.6 Adjustment for multiple tests

The question of whether or not one should take into account the number of tests performed on a given sample set when deciding whether a result is

Table 3.7 Chi-squared test of transmission of multi-marker haplotypes using TRANSMIT for all ethnic groups and Mandinka only. The haplotype identifiers refer to the haplotypes given in Table 3.5. The computer program TRANSMIT calculates haplotype transmission and compares the observed number with the expected value based on random transmission.

	Identifier	*Obs*	*Exp*	*O-E*	*Chisq(1df)*
all ethnic groups					
Cerebral Malaria					
	1	230.64	221.59	58.39	1.40
	2	4.99	6.20	2.40	0.60
	3	182.22	183.63	54.57	0.04
	4	4.79	4.18	1.40	0.27
	5	38.03	43.19	16.88	1.58
	6	6.45	7.10	2.46	0.17
Mandinka					
Cerebral Malaria					
	1	118.35	108.88	28.38	3.16
	2	1.27	2.28	0.82	1.26
	3	79.12	82.39	26.62	0.40
	4	2.85	2.66	0.72	0.05
	5	17.90	21.07	7.40	1.36
	6	5.18	5.40	1.83	0.03

statistically significant has been under much debate. It might appear that some studies analyse the available data in so many ways until a significant association can be found. When more than one statistical test is performed in analysing the data from a given study, some statisticians argue that a more stringent criterion be used for *statistical significance* than the conventional $p< 0.05$.

Bonferroni adjustments are based on the following reasoning. If a null hypothesis is true a significant difference ($p< 0.05$) will be observed by chance one in 20 times. This is called type I error, or α. If multiple independent tests are performed and the null hypothesis holds for all of them than the chance of at least one test being significant is higher than 0.05. In order to avoid a lot of spurious positives, the alpha value needs to be lowered to account for the number of comparisons being performed. The formula for the error rate across the study is $1-(1-\alpha)^n$, where n is the number of test performed. However, the Bonferroni adjustment deflates the α applied to each, so the study-wide error rate remains at 0.05. The adjusted significance level is $1-(1-\alpha)^{1/n}$. This is often approximated as α/n.

The current study suggests that Mandinka who carry the *IFNGR1*-470ddel are protected against severe malaria with $p = 0.02$. However, if one corrects for the fact that a total of four markers were typed and the results then analysed in five different ethnic groups (Mandinka, Jola, Fula, Wollof and others), i.e. a total of 9 tests, the alpha level for each test would have to be lowered to approximately 0.006. This correction would suggest that the result found is in fact not statistically significant.

While it seems conceivable that the number of independent tests performed as part of a given study might increase the likelihood of finding a significant result, caution must equally be applied when adjusting for multiple comparisons. The Bonferroni adjustments imply that a given comparison will be interpreted differently according to how many other test were performed. However, this does not always make sense. For example, in a clinical

study the difference in remission rates between two chemotherapeutic treatments could be interpreted as statistically significant or not depending on whether or not survival rates, quality of life scores, and complication rates were also tested. In genetics this is of particular relevance because an increasing number of markers will be tested on the same sample set. Should results be adjusted for this? Should one correct just for the number of markers one investigator has looked at or for all the markers that have ever been typed in this study? Would p-values then need to be updated as more markers get typed? It becomes obvious that a whole genome screen would be likely to fail to yield any significant results if one corrected for several hundred thousand markers. Another problem lies in the fact that genetic markers are not independent of each other as they might be in linkage disequilibrium with each other.

The other problem is that with Bonferroni adjustments type II errors (the probability of accepting the null hypothesis when the alternative is true) are increased. *True* results might therefore be missed. One could imagine a clinical scenario where the doctor has ordered 20 laboratory test and the troponin comes back as being abnormal. Based on the Bonferroni adjustment it was to be expected that one of the 20 tests was abnormal by chance alone and the wrong assumption could be made that the patient did not have a myocardial infarction and should therefore not be further investigated or treated.
The Bonferroni adjustments should therefore be applied with some caution. It seems warranted to use the adjusted values when the same test is repeated in many subsamples, such as when stratified analyses (by age, sex, ethnic group, etc.) are conducted without an *a priori* hypothesis that the primary association should differ between these subgroups.

For the comparisons made for the different ethnic subgroups in the chapter of this thesis this means that adjusting the levels of significance for the number of ethnic groups that were tested might seem warranted. The level of significance would therefore be $p < 0.01$. The association found for protection against severe malaria in Mandinka carrying the *IFNGR1*-470ddel

allele would therefore not be signficant ($p = 0.02$). The protective effect of *IFNGR1*-56 heterozygosity would be borderline significant with $p = 0.01$. An adjustment for the number of markers does not seem appropriate. There was an *a priori* hypothesis that the two major phenotypes of severe malaria (cerebral malaria and severe malarial anaemia) might have different susceptibility factors and an adjustment for looking at these two phenotypes was therefore not made either.

For confident rejection of the null hypothesis, it would be desirable to satisfy the level of significance after applying the Bonferroni correction. However lesser P values are still of interest, as they signal trends that may be worth pursuing. Genetic associations for complex, i.e. multifactorial, diseases can be expected to be much weaker than for diseases of mendelian inheritance. Multiple weak effects might ultimately alter disease outcome. A strict correction for multiple comparisons might make these associations even harder to detect. It will therefore be crucial to establish larger studies of several ten thousand samples in order to eventually overcome these limitations.

Kosh 2003

3.3 Discussion

With the sequencing of the human genome being almost completed, the best strategy for finding disease causing genes has been much debated. In general there are two major stragies: Linkage analysis and association mapping. Although it has been used for complex diseases, linkage analysis has been especially important in detecting disorders that follow a mendelian pattern of inheritance. Disadvantages are inconsistent and ambiguous results and a low resolving power for disease genes with a small relative risk. Association studies using a candidate gene approach have been proposed as a valuable alternative or addition. This approach focuses on genes that are known to play a role in the disease pathology. Here polymorphisms that are located in potentially functionally important areas of the genome are tested for disease association. Furthermore for diseases with a small relative risk the required sample size will be by far lower for an association study to detect the effect compared to a linkage study. Risch and Merikangas calculate that for a disease with a risk ratio of 2.0 and an allele frequency of even 50% one would still need 2498 families using a linkage study compared to only 340 families with an association study in order to detect the effect with 80% power (Risch and Merikangas, 1996). There are several ways of designing an association study. The case control study is the simplest such design. Although it is a potentially very powerful tool in order to detect a disease association, it carries the risk of having confounding factors. Although a well designed study will try to match cases and controls as far as possible, it may be difficult to completely achieve this goal which makes case control studies susceptible to false positive results. A way to circumvent ethnic confounding is to use parents as controls. The Transmission Disequilibrium Test makes use of this. It is less powerful, though, i.e. a larger sample size is needed (Risch, 2000). This study provides the possibility to use both methods in order to investigate the relationship of polymorphisms in the *IFNGR1* promoter with susceptibility to severe malaria. Since IFN-γ has both beneficial effects (on microbial killing) and potentially deleterious consequences (such as severe inflammation) it is conceivable that natural selection might favour

different levels of IFNGRAC expression depending on the type of infectious pathogens to which a population is exposed. Over time, this might be expected to cause genetic variation in *IFNGR1*. Consistent with this notion, the genomic region around *IFNGR1* has been linked to severe hepatic fibrosis in community-based studies of *Schistosoma mansoni* infection (Dessein et al., 1999) but thus far there is no clear evidence of association between any specific *IFNGR1* polymorphism and susceptibility to common infections within the general population. In the present study four polymorphisms at the *IFNGR1* locus in a West African population and their relationship to susceptibility to severe malaria are described. Merlin et al. (1997) have shown that the region extending -700 bp upstream of the *IFNGR1* translation initiation site has full promoter activity and at least two common polymorphisms that were identified in the present study in this region both have features of potential interest. The sequence around -56nt is reminiscent of an AP4 binding site: with the *IFNGR1*-56T allele, the essential `CAGC` motif is observed on the reverse strand only, while with the*IFNGR1*-56C allele it is observed in both directions. It has been noted that the *IFNGR1* gene has multiple transcriptional initiation sites, one of which lies at -55nt and might therefore be perturbed by the *IFNGR1*-56 polymorphism. Other transcription initiation sites lie at -21, -31, -37 -61 and -75nt, so the *IFNGR1*-56 polymorphism would lie within the 5' UTR of certain transcripts. The sequence around -470nt is reminiscent of a STAT1 binding site (`TTCCtcaAA`) and the *IFNGR1*-470delTT allele abolishes the critical first two positions of this binding motif. The *IFNGR1*+95 lies at the beginning of intron 1 and in close proximity to a splice site. These observations suggest possible mechanisms by which these SNPs might exert a direct functional effect and further investigations on a cellular level seem warranted. Both, the *IFNGR1*-470 and the *IFNGR1*-56, show preliminary evidence of association with severe malaria in Mandinka, the major Gambian ethnic group. In a case-control analysis heterozygotes for the *IFNGR1*-56 polymorphism showed a two-fold protection against cerebral malaria and a four-fold protection against death due to cerebral malaria. The *IFNGR1*-470delTT allele was associated with two-fold

protection against severe malaria in general. In case of the *IFNGR1*-470 this association was, however, not significant after Bonferroni correction.

In a subset of these Mandinka cases there also was DNA available from both parents allowing TDT analysis to be performed. For the *IFNGR1*-470 this showed a weak trend in the same direction as the case-control analysis, so that further studies in a larger sample size are warranted. For the *IFNGR1*-56 it was found that the heterozygote genotype was less likely to be "transmitted" from parents to their child with cerebral malaria. Thus both the case-control and family analyses suggest a significant protective effect of the heterozygous genotype at the *IFNGR1*-56 locus. Although this is an unusual finding and associations are generally described with alleles rather than genotypes, a heterozygote advantage is a common phenomenon in nature. Haldane (1949) was the first to propose the concept of a balanced polymorphism, whereby at certain loci heterozygotes may be fitter than either of the homozygous genotypes. This was first proven to be true for the haemoglobin AS genotype in Africa, which provides about tenfold protection against severe malaria than the AA genotype, without incurring the potentially fatal complications of sickle cell disease that are associated with the SS genotype. Heterozygote advantage seems to be a strong selectional force at the major histocompatibility complex (MHC) where it is thought that individuals heterzygote at HLA (human leukocyte antigen) loci are able to present a greater variety of antigenic peptides than are homozygotes (Doherty and Zinkernagel, 1975). More recent examples of heterozygote advantage include HLA class I alleles in HIV-infected individuals (Carrington et al., 1999) , and a non-sense mutation at the Cd36 locus in severe malaria (Pain et al., 2001). It is open to speculation why heterozygote advantage should exist at the *IFNGR1* locus: a possible explanation is that one of homozygous genotypes fails to mount a sufficiently strong immune response to control the parasite, and that the other homozygous genotype has an excessive inflammatory reaction, while the heterozygous genotype has an optimally balanced response. This is clearly an issue that will require detailed functional investigation if the present finding is confirmed in independent studies. It is worth noting that heterozygote advantage can easily be overlooked in standard tests

for allelic association, including the conventional form of the TDT, unless it is specifically examined.

There are other ways in which an apparent loss of heterozygotes in the case group could have occurred. It is conceivable that the genotype frequencies in this group are different from the general population for reasons other than disease association but were introduced by ascertainment of the samples. This scenario could for example arise if within the Mandinka case group there was a population substructure that consisted of two groups. If we consider the following: one group was homozygote for one of the two alleles at a high frequency and the other group mainly homozygote for the other allele and these two groups would not intermarry. If we were not aware of this underlying population substructure and would collect samples from both groups and calculate the general genotype frequency, we would see an apparent loss of heterozygotes. There is a particular risk for this if samples were collected at different locations and if the results were pooled in the analysis. All of the case samples for this study, though, were collected in one hospital in The Gambia which should minimize these effects. With current sample sizes, spurious associations might arise due to sampling error and these data should be interpreted cautiously. However it is important to note that, if there was a true biological difference between allelic or gentoypic associations at a given locus, then it is the largest ethnic group (in this case the Mandinka) which would be most likely to give a significant result. Even when two different ethnic groups share the same disease allele, migration and other population changes may lead to different patterns of linkage disequilibrium with surrounding marker alleles. If this were the case at the *IFNGR1* locus, it would imply that the *IFNGR1*-56 and -470 polymorphisms are not themselves disease alleles, but are in linkage disequilibrium (LD) with neighbouring disease alleles that are either absent, or in different patterns of LD, in other ethnic groups. Alternatively, these polymorphisms may act in conjunction with other genetic or environmental factors possessed by Mandinka but not other ethnic groups. This is not the first study to find that the Mandinka population of the Gambia have specific genetic and phenotypic features: for example, the association of certain

IL10 promoter polymorphisms with trachoma appears to be specific to the Mandinka group (Mozzato-Chamay et al., 2000), while previous studies in rural parts of The Gambia have observed that childhood splenomegaly and malaria are less prevalent in Mandinka than their neighbours (Greenwood et al., 1987). The notion that either genetic or environmental differences between West African ethnic groups may make a substantial contribution to malaria susceptibility is bolstered by evidence of strong protection against malaria among the Fulani of Burkina Faso (Modiano et al., 1996).

Although case control analysis is a powerful method of screening for disease associations (Morton and Collins, 1998; Risch, 2000), the result found for both *IFNGR1*-56 and -470 should be interpreted with caution because, firstly, TDT analysis in a subset of the index cases showed only weak trends in the same direction, and secondly, the association for *IFNGR1*-470 was not significant after applying the Bonferroni correction. These data highlight the need to amass DNA collections that are much larger than currently available, to assess with confidence relatively modest genetic effects that, owing either to gene-environment interactions or population differences in haplotypic structure, may manifest only in a specific ethnic group.
It would be wrong to assume that the null hypothesis could be rejected with confidence. However, despite the weak associations found it seems warranted to pursue these results further as both the *IFNGR1*-470 and -56 are sites of potential regulatory significance.

With the possibility that the markers found are in fact in linkage disequilibrium with functional alleles elsewhere in the region, one possible approach is to analyse the patterns of LD in this genetic region. *IFNGR1* has two neighbouring genes that are potentially involved in the inflammatory host response. As a first step it will therefore be interesting to assess whether LD spans beyond the borders of each gene or whether in fact LD reaches as far as the neighbouring genes. A systematic sequencing approach to each of the three genes will furthermore allow to analyse signs of natural selection based on deviation from the neutral model.

3.4 Summary

1. The region between -1400nt to +100nt of the promoter of the *IFNGR1* gene was resequenced. A total of four variant sites were found at positions -470, -270, -56 and +95.

2. In a case-control analysis of severe malaria in Gambian children, no association was found for the above polymorphisms when all ethnic groups were combined.

3. In Mandinka, the largest ethnic group in The Gambia, heterozygosity at position *IFNGR1*-56 was associated with protection against cerebral malaria and protection against fatal outcome of cerebral malaria.

4. The *IFNGR1*-470ddel allele was associated with protection against severe malaria in Mandinka.

5. Family-based association analysis showed similar trends.

6. If multiple comparisons are taken into account the above associations cannot be considered definitive, but they are of sufficient interest to justify further analysis of the genetic diversity of the genomic region surrounding *IFNGR1*, and of the functional significance of the *IFNGR1*-56 and *IFNGR1*-470ddel polymorphisms.

Chapter 4

Genetic Diversity at the 6q23 locus

4.1 Introduction

The initial association study at of single nucleotide polymorphisms found in the first 1.4kb of the *IFNGR1* promoter region yielded an encouraging association for two of the four polymorphisms found in this region.

The gene for the α chain of the Interferon-γ receptor is located on the long arm of chromosome 6. The gene consists of seven exons and spans a total of 21kb. The nearest neighbouring genes are *IL22BP* (interleukin 22-binding protein) and *IL20RA* (interleukin 20 receptor alpha) spanning a total region of 176kb. Given the association found in the earlier study further investigations in this regions seemed warranted. It was chosen to systematically resequence the promoter regions of all three genes for the following reasons:

1. It was important to establish the extent of linkage disequilibirum (LD) in this region and to further dissect haplotypes on which the associated markers lie on. If sufficient linkage disequilibrium exists between

neighbouring genes associations found in one gene might in fact only pick up an event that is happening in another gene. In order to be able to calculate linkage disequilibrium for a given region one requires information about other genomic markers, i.e. SNPs, in this region. Although the 240kb sequence surrounding the three genes, *IFNGR1*, *IL22BP* and *IL20RA*, contained a total of 296 reported SNPs, only 10 of those were validated. Several papers have reported on the problem of the usefulness of SNPs in the public database that have not been validated (Douabin-Gicquel et al., 2001; Marth et al., 2001). Firstly these SNPs have often been found by investigating overlapping clones, their existence has not been proven by an independent genotyping method. Secondly SNPs that are not validated do not carry any information on allele frequency, which population they were found in and whether they are common to more than one population. In terms of economical feasibility selecting SNPs that are not validated may turn out expensive if subsequent genotyping fails to detect them in the population under investigation. Furthermore although public databases contain a vast number of SNPs the density for a particular region of interest might not be sufficient. Investigation of linkage disequilibrium in the 6q23 region will be the subject of the next chapter.

2. *IL22BP* and *IL20RA* are coding for proteins that are potentially involved in the inflammatory host response and could therefore independently play a role in the infection with the malaria parasite.

3. There is another advantage of having sequencing information. Detailed information obtained from resequencing one or more populations provides a convenient way of measuring genetic diversity for any given locus. Human genetic variation is natural, created by mutation and shaped over history by biological, demographic and historical processes. Natural selection can have a marked effect on levels of nucleotide variability. While a balancing selection might be able to maintain allelic diversity within a population, directional selection may result in reduction in nucleotide variability and increased linkage disequilibrium when

compared to what would be expected under the neutral model (see below). Information for more than one population allows for comparisons between them especially as they might have been exposed to different selective pressures.

4. With the view to facilitate studying the genetic susceptibility to other infectious diseases in Caucasians (RSV bronchiolitis in a UK population) it was important to establish allele frequencies for this population and search for population specific markers.

5. The total region spanning these three genes is 176kb. In the time frame of this project in was neither feasible nor economically possible to resequence the whole region. As an initial approach it therefore seemed reasonable to resequence 2kb of the promoter region of each of the three genes. Firstly there is evidence for all of these three genes to be involved in the inflammatory host response. Secondly one would be able to establish whether significant LD exists between these three gene regions, *IFNGR1*, *IL22BP* and *IL20RA*.

4.1.1 The 6q23 locus

The gene for the α chain of the Interferon γ receptor is located on the long arm of chromosome 6. gene consists of seven exons and spans a total of 21kb. The nearest neighbouring genes are *IL22BP* and *IL20RA*.

4.1.1.1 IL22BP

IL22 encoded on chromosome 12q15 (see chapter on IFNγ and 12q15 locus) has been shown to induce acute phase reactant proteins suggesting that this cytokine contributes to the inflammatory response in vivo (Dumoutier et al., 2000a,c). By DNA database screening for sequences with homology to the extracellular domain of IL10RB and analysis of 6q24 BAC clone, Dumoutier et al. (2001) identified a cDNA encoding IL22BP, which they designated CRF2X. Sequence analysis predicted that the 231-amino acid protein was

approximately 33% identical to the extracellular domains of IL20RA and IL22R. But it seems to lacks a transmembrane domain. RT-PCR analysis detected expression in breast, lungs, and intestinal tract, with lower levels in skin, testis, brain, heart, and thymus. ELISA showed that IL22BP does indeed bind IL22. Functional and luciferase reporter analysis indicated that IL22BP specifically blocked IL22-induced STAT3 activation in an intestinal epithelial cell line. Using similar methods, Kotenko et al. (2001) also cloned and characterized IL22BP, which they initially designated CRF2-10. Western blot analysis showed expression of a 34- to 35-kD glycosylated secreted protein. Crosslinking and autoradiographic analysis confirmed the interaction of IL22 and IL22BP. Functional analysis indicated that IL22BP inhibited IL22-mediated MHC class I antigen expression. EMSA analysis demonstrated inhibition of STAT1 and STAT3 DNA-binding complexes. Northern blot analysis showed blocking of SOCS3 expression.

Xu et al. (2001) obtained a cDNA encoding IL22BP, which they termed IL22RA2. IL22RA2 effectively inhibited proliferation of a mouse pre-B cell line in response to IL22. Northern blot analysis revealed expression of 1.6- and 3.0-kb transcripts in placenta and spleen, but not in other tissues. RT-PCR analysis extended the range of tissues expressing IL22RA2 to skin and lung, with lower levels in heart, pancreas, and prostate, as well as the digestive and female reproductive systems. In situ hybridization analysis demonstrated expression in monocytes, activated B cells, and epithelium. Xu et al. (2001) concluded that IL22RA2 may be important as an IL22 antagonist in the regulation of inflammatory responses.

4.1.1.2 IL20RA

IL20 is homologous to 3 members of the IL10 family, IL19 (40% identity), MDA7 (33% identity), and IL10 (28% identity). Blumberg et al. (2001) found that overexpression of IL20 in transgenic mice caused neonatal lethality with skin abnormalities, including aberrant epidermal differentiation. Blumberg et al. identified the receptor for interleukin-20 as a heterodimer of 2 orphan class II cytokine receptor subunits, IL20RA, also called ZCYTOR7, and IL20RB,

also called DIRS1. Binding assays using radiolabeled ligand demonstrated that IL20 bound to BHK transfectants expressing both IL20RA and IL20RB, but not to untransfected cells nor to transfectants expressing either receptor subunit alone. Both receptor subunits were expressed in skin and were dramatically upregulated in psoriatic skin.

4.1.2 The Neutral Theory

The null model in population genetics describes a world in which natural selection has played no role. A world in which genetic drift, mutation, and recombination are the only forces that influence patterns of genetic diversity (Kimura, 1968). A primary focus of population genetics is to make inferences about each of these evolutionary forces to produce the patterns of molecular sequence variation observed in a given sample set.

If unexpected patterns are observed one can introduce other parameters into the model and see whether the data fit under the new model. For example one can directly compare the likelihood of observing the data under models that ignore and include natural selection. If the model that includes selection is more likely than the one that ignores it, one might claim evidence for the action of natural selection. Several tests have been established to calculate deviation from the neutral theory although it might be difficult to establish which parameter was responsible.

4.1.3 Approach

Diversity was assayed in the 6q23 region using a resequencing approach. A comprehensive picture of total haplotypic sequence diversity was produced by a large-scale sequencing survey of the *IFNGR1*, *IL22BP* and *IL20RA* loci. The first aim was to generate new markers at these three loci to enable LD calculations. The second aim of this investigation was to identify the relative effects that evolutionary processes such as mutation, recombination, natural selection, genetic drift and migration have had on patterns of diversity in

these loci. About 2kb of promoter region of each of the three genes was therefore sequenced in 48 Gambian chromosomes and 48 Caucasian chromosomes. The analysis of a chimp diploid sequence in this region enabled the determination of the ancestral states of each human polymorphism. Comparisons of divergence and diversity data yield valuable insights into the evolutionary process and form the basis of an important test of selective neutrality.

4.2 Results

4.2.1 Sequence Analysis of DNA variations at 6q23

Human *IFNGR1*, located on chromosome 6q23, is neighboured by *IL22BP* and *IL20RA* (Figure 4.1). Approximately 2kb of each of the three promoters were sequenced (total of 6.4kb) in 48 chromosomes from Caucasians from Utah, USA, (CEPH) and in a further 48 chromosomes from Africans (The Gambia). In total 2413bp were sequenced in *IFNGR1*, 1945bp in *IL22BP* and 2005bp in *IL20RA*. Using the `PolyPhred` program a consensus sequence was generated and variants sites tagged. `PolyPhred` identifies heterozygote sites by scanning the assembled sequence traces for: firstly the presence of a drop in normalized fluorescence peak height at the position when compared to the respective peak height for all individuals that are homozygous at the position, and secondly the the presence of another peak that accompanies the drop in fluorescence peak height. Sequencing results were confirmed by sequencing the reverse strand, although full coverage was not always achieved (see Table 4.1). Any discrepancies between the two reads were resolved by manual inspection of the traces. Reads that were of low quality were marked as failed by the software. It is worth noting that the software marked all reads containing heterozygotes for the double deletion at *IFNGR1*-470/-471 as failed most likely because the reads downstream of the deletion would have had a frame shift. Here all genotypes had to be called by manual inspection of the trace files, similarly for the *IFNGR1*-270 which was located on the same sequencing fragment as the *IFNGR1*-470.

A total of 28 variants sites were detected in 6.4kb of sequence. Nine of these markers were in *IFNGR1*, 8 in *IL22BP* and 11 in *IL20RA*. There was one double deletion (*IFNGR1*-470), which had been found earlier, and 27 single nucleotide polymorphisms. For the further analysis the double deletion at *IFNGR1*-470/-471 was regarded as one single event. Seven markers had also been reported in the public database although none of them had been validated. The average coverage of reads, i.e. how many reads for each SNP for each individual were obtained, was 1.4.

In *IFNGR1* 4 markers had been found in the previous sequencing analysis of the promoter. By extending the sequencing region another 5 markers could be found. The *IFNGR1*+130, *IFNGR1*-1804 and *IFNGR1*-1882 lay outside the area previously sequenced. In this sample set the *IFNGR1*-611 was only found in Caucasians. The *IFNGR1*-72 had a low allele frequency in Gambians and had not been found previously. Both the *IFNGR1*+130 and the *IFNGR1*+95 are located in the first intron of the *IFNGR1* gene. In order to facilitate calculating the distance between markers each SNP was named by its chromosomal position according to the September 2002 freeze. Although chromosomal positions change rapidly when the databases are updated the positions of the markers can easily be updated by referencing them to the SNPs that have an entry in the public database. The position of each SNP was also calculated in relation to the start of the first exon (as reported in the Ensembl database) of each gene respectively.

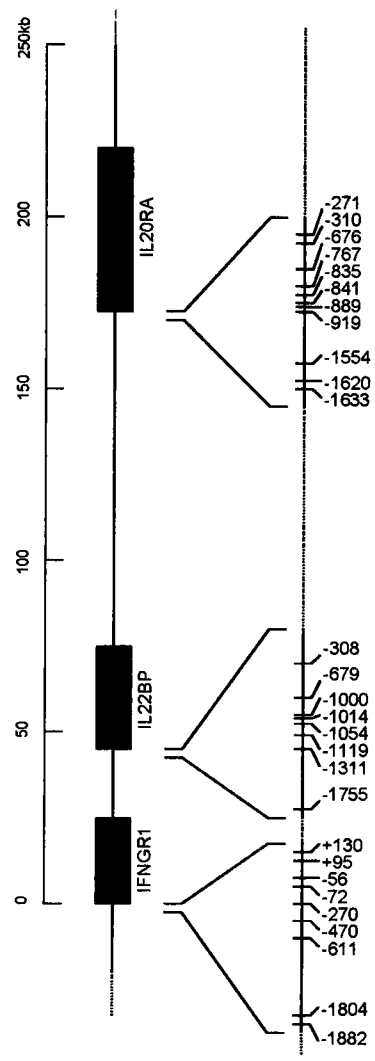


Figure 4.1 The 6q23 region.

Косн 2003

Table 4.1: Sequence Variants identified in 6q23. A total of 28 variants were found, 9 of them in *IFNGR1* (1-9), 8 in *IL22BP* (10-17) and 11 in *IL20RA* (18-28). Each marker was given a chromosomal positions according to the September 2002 freeze. The position in the gene is given in relation to the start of the first exon. Coverage refers to the average number of times each site was covered by a confirming sequence read. Allele frequencies are based on 48 Caucasian and 48 African (Gambian) chromosomes and are marked by an asterisk where deviation from HWE was observed (0.05 level of statistical significance).

							Relative frequency	
Gene	Id.no.	Position	Pos. in Gene	dbSNP ref.	Variant	Coverage	Caucasian	Gambian
IFNGR1	1	6:144241441	-1882		Y	1.6	0.13	0.06*
	2	6:144241363	-1804		K	1.9	1.00	0.91

Gene	Id.No.	Position	Pos. in Gene	dbSNP ref.	Variant	Coverage	Relative frequency	
							Caucasian	Gambian
	3	6:144240170	-611	rs1327474	R	1.8	0.61	1.00
	4	6:144240029	-470		DEL	NA	0.00	0.13
	5	6:144239829	-270		Y	NA	0.00	0.03
	6	6:144239631	-72		Y	1.6	0.85	0.96
	7	6:144239615	-56	rs223411	Y	1.5	0.42	0.61
	8	6:144239464	+95		Y	1.8	0.42	0.61
	9	6:144239429	+130		R	1.8	0.29	0.13
IL22BP	10	6:144195503	-1755		Y	1.0	0.02	0.26
	11	6:144195059	-1311		R	1.0	0.00	0.19
	12	6:144194867	-1119		R	1.0	1.00	0.77
	13	6:144194802	-1054		K	1.0	0.57*	0.25*
	14	6:144194762	-1014		R	1.0	1.00	0.83
	15	6:144194748	-1000		W	1.0	1.00	0.94
	16	6:144194491	-679		M	0.9	0.98	0.71
	17	6:144194056	-308		K	0.8	0.03	0.26

Gene	Id.No.	Position	Pos. in Gene	dbSNP ref.	Variant	Coverage	Relative frequency Caucasian	Relative frequency Gambian
IL20RA	18	6:144066829	-1633		K	0.8	0.98	1.00
	19	6:144066816	-1620	rs276522	R	0.8	0.27	0.71
	20	6:144066750	-1554		Y	0.7	0.27	0.71
	21	6:144066115	-919	rs276519	S	1.6	0.74	0.21
	22	6:144066085	-889		R	1.9	0.98	0.94
	23	6:144066037	-841	rs276518	K	1.9	0.25	0.81
	24	6:144066031	-835	rs276517	R	1.9	0.75	0.19
	25	6:144065963	-767	rs276516	M	1.9	0.24	0.81
	26	6:144065872	-676		Y	1.9	0.02	0.08*
	27	6:144065506	-310		R	1.6	0.19	0.13
	28	6:144065467	-271		Y	1.6	0.00	0.08

4.2.2 Sequence Diversity at the 6q23 locus

The visual presentation of the sequence diversity is shown in Figure 4.3. Homozygotes for the most common allele across the two populations are shown in blue, homozygotes for the alternative allele in yellow and heterozygotes in red. Twenty of the 28 variants were identified in the Caucasian (CEPH) samples while 26 variants were present in the Gambian samples. Two variants were unique to Caucasians while 8 were unique to Gambians (Figure 4.4. In all 20 variant sites were found in Caucasians and 26 in Africans. In Africans 18 out of 26 variants sites had a minor allele frequency above 10%. In Caucasians 14 sites were above 10%. It is interesting to note that the minor allele of the *IFNGR1*-611 had a frequency of 0.39 in Caucasians while it was absent in this set of Gambian samples. The frequency distribution of the minor alleles are shown in Figure 4.2. One way of quantifying the degree of population diversity is with the F_{ST} statistic (Wright, 1921, 1978). The F_{ST} values for *IFNGR1*, *IL22BP* and *IL20RA* were 0.12, 0.17 and 0.41 indicating significant genetic divergence between the two populations, i.e. Africans and Caucasians.

Most of the genotype frequencies fitted the Hardy-Weinberg expectation and the overall distribution of p-values was close to uniform (Figure 4.5B). In only four instances (two with low allele frequencies) was the test for statistical significance of deviation from the Hardy-Weinberg Equilibrium $p < 0.05$.

Two estimates of nucleotide diversity are commonly used. The level of nucleotide polymorphism, θ, is the proportion of nucleotide sites that are expected to be polymorphic in any suitable sample from this region of the genome (Hartl and Clark, 1997a; Watterson, 1975). The second measure of sequence variability is π, the nucleotide diversity. The average proportion of pairwise differences between sequences in a given sample set is the estimate, $\widehat{\pi}$, of the nucleotide diversity. The nucleotide diversity ($\widehat{\pi}$) in the three promoter regions of *IFNGR1*, *IL22BP* and *IL20RA* were 0.00070, 0.00034 and 0.00138 respectively in Caucasians. In Gambians the values were 0.00042,

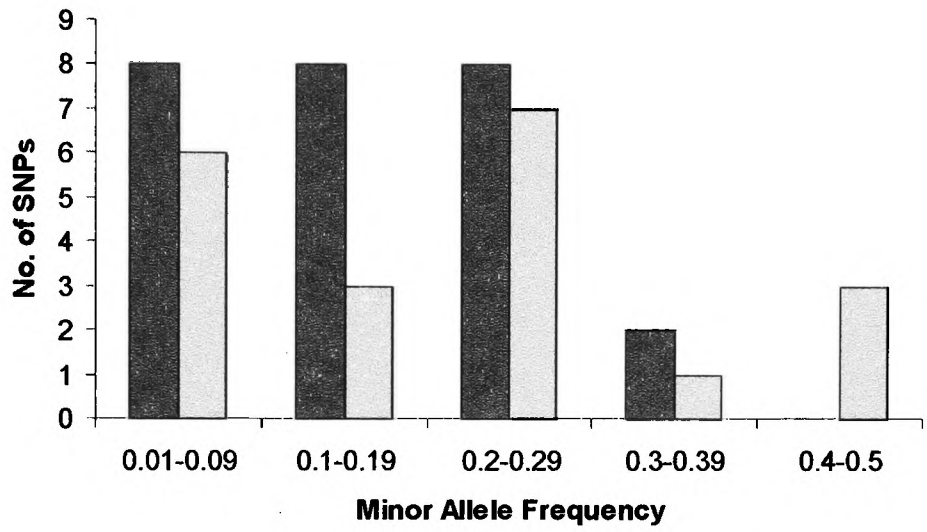


Figure 4.2 Minor allele frequencies of markers found at 6q23. Blacks bars represent the frequencies in the Gambian samples while the grey bars represent the Caucasian samples.

0.00139 and 0.00131 respectively. While the number of segregating sites in the *IFNGR1* promoter was lower in Caucasians $\widehat{\pi}$ was slightly higher and θ slightly lower. This reflects the ancestral population having a higher number of low frequency variants. The most significant difference of nucleotide diversity was found in *IL22BP*. Twice the number of markers were found in Africans and the nucleotide diversity was four times as high as in Caucasians. Three out of the 4 markers found in *IL22BP* in Caucasians occurred only once (singletons). In *IL20RA* both number of segregating sites and nucleotide diversity are similar for both populations.

Косн 2003

Table 4.2 Sequence Diversity at the 6q23 locus in Caucasians (CEPH)

Gene	Region	no.of. bases	no. of segregating sites	θ(S) (per gene)	$\widehat{\theta}$(S) (per site)	Pairwise Differences π	Nucleotide Diversity $\widehat{\pi}$
IFNGR1	promoter	2031	4	0.901 ± 0.249	0.00044 ± 0.00000	1.424 ± 0.783	0.00070 ± 0.00000
	exon1	86	0	0	0	0	0
	intron1	296	2	0.451 ± 0.109	0.00152 ± 0.00000	0.977 ± 0.455	0.00330 ± 0.00000
	total	2413	6	1.352 ± 0.421	0.00056 ± 0.00000	2.401 ± 1.762	0.00099 ± 0.00000
IL22BP	promoter	1872	4	0.901 ± 0.249	0.00048 ± 0.00000	0.628 ± 0.251	0.00034 ± 0.00000
IL20RA	promoter	1961	10	2.253 ± 0.856	0.00115 ± 0.00000	2.705 ± 2.141	0.00138 ± 0.00000

Table 4.3 Sequence Diversity at the 6q23 locus in Africans (Gambian). Note that the deletion found in the *IFNGR1* promoter was excluded from the calculation of θ and π. The total number of segregating sites was therefore only 7 instead of 8.

Gene	Region	no.of. bases	no. of segregating sites	θ(S) (per gene)	$\widehat{\theta}$(S) (per site)	Pairwise Differences π	Nucleotide Diversity $\widehat{\pi}$
IFNGR1	promoter	2031	5(6)	1.127 ± 0.331	0.00560 ± 0.00000	0.848 ± 0.374	0.00042 ± 0.00000
	exon1	86	0	0	0	0	0
	intron1	296	2	0.451 ± 0.109	0.00152 ± 0.00000	0.772 ± 0.330	0.00261 ± 0.00000
	total	2413	7(8)	1.577 ± 0.518	0.00065 ± 0.00000	1.620 ± 0.951	0.00067 ± 0.00000
IL22BP	promoter	1872	8	1.803 ± 0.623	0.00096 ± 0.00000	2.607 ± 2.015	0.00139 ± 0.00000
IL20RA	promoter	1961	10	2.253 ± 0.856	0.00115 ± 0.00000	2.574 ± 1.973	0.00131 ± 0.00000

4.2.3 Genetic Variation and the Neutral Model

The estimators for θ and π assume neutral mutations. Any deviations from what would be expected under this neutral model might indicate natural selection. Tajima (1989) derived a method which utilizes the differences between the average number of pairwise differences (π) and θ estimated from the number of segregating sites. The expectations of π and θ are both $= 4N_e\mu$ under the assumption that mutations are selectively neutral and that the population is at equilibrium, i.e. $\theta = \pi$. Tajima's D tests for deviation from this assumption. Wattersons estimator of θ, is only influenced by the number of segregating sites (Watterson, 1975). In contrast, π is sensitive to allele frequencies at segregating sites, such that alleles at intermediate frequencies contribute more than alleles at low frequencies. Consequently, if a sample has an excess of rare variants, π will be less than Wattersons estimator and D will be negative. In contrast, if there is an excess of intermediate frequency variants, π will be greater than Wattersons estimator and D will be positive. A second test was derived by Fu and Li (1993) who showed that the expected number of derived mutations that are present only once in a sample is equal to θ. For the Gambian chromosomes in this study Tajima's D did not reject the null hypothesis in all of the three investigated genes. In the Caucasian chromosomes, though, Tajima's D bordered on significance ($0.10 > p > 0.05$) when looking at the total sequence length of 2413bp in *IFNGR1*. When investigating the 3' 1500bp (using a sliding window approach) Tajima's D turned significant ($D = 2.28$; $p < 0.05$) indicating potential balancing selection at this this locus.

Other tests were derived by Fu and Li (1993) who showed that the expected number of derived mutations that are present only once in a sample is equal to θ. Their F statistic based on the differences between the number of singletons (mutations appearing only once among the sequences) and the average number of nucleotide differences between pairs of sequences. In Caucasians $F = 1.68$ ($p < 0.05$) was observed for *IFNGR1* indicating a lack of singletons.

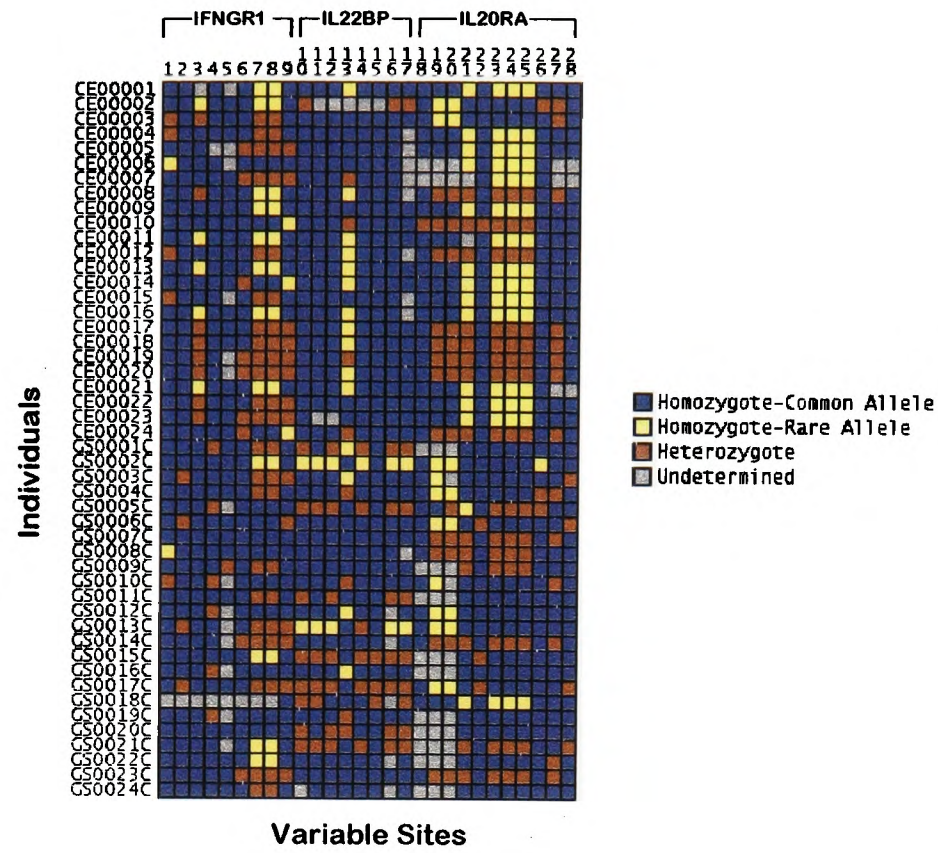


Figure 4.3 Comparison of the 48 individuals to the reference sequence for *IFNGR1*, *IL22BP* and *IL20RA* for the 28 variant sites. Homozygotes for the allele with highest relative frequency across the two populations (most common allele) for the variable site (blue), homozygotes for alternative alleles (yellow) and heterozygotes (red) are indicated. Grey denotes positions were genotypes were not obtained. The first 24 samples are from Caucasians individuals from Utah (CEPH), the other 24 samples are from African individuals from the Gambia. The picture was generated with Visual Genotype (VG), Version 2 (available at: `http://pga.gs.washington.edu/examples/servlet/edu.washington.gs.pga.tools.vg.VG2Web`)

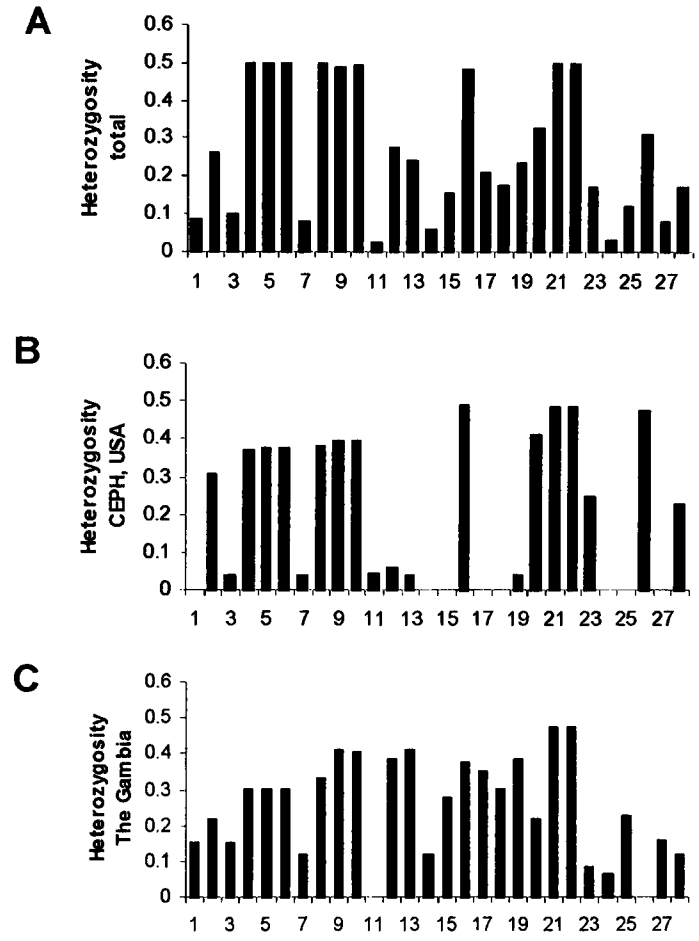


Figure 4.4 A plot of the heterozygosity for the 28 variants sites found in three promoters of *IFNGR1*, *IL22BP* and *IL20RA*. One-half in Caucasians (B) and more than two-thirds of the variant sites in Africans (C) exceed a heterozygosity of 10%. The average heterozygosity across 28 markers is H=0.204 in Caucasians and H=0.255 in Africans.

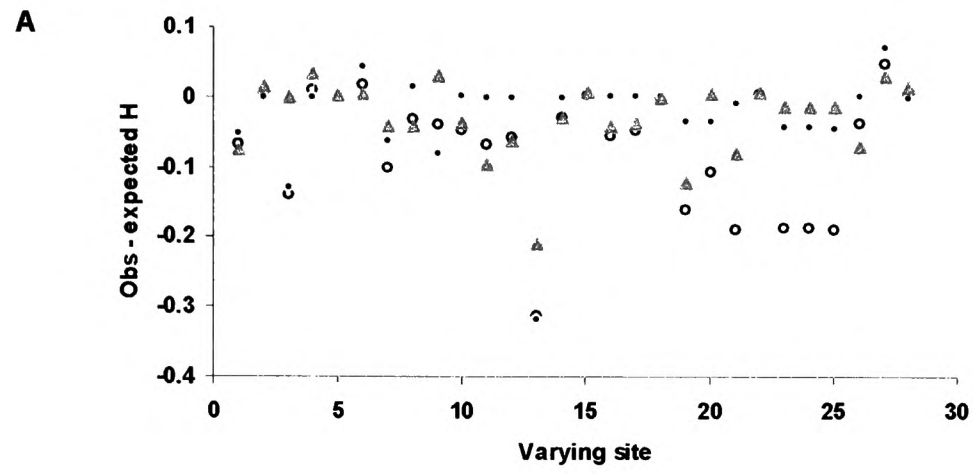


Figure 4.5 **A**. Comparison of the observed and expected heterozygosities. Each dot represents the difference between observed heterozygosity at a nucleotide site and the expected Hardy-Weinberg heterozygosity. Open circles refer to pooled data across populations. Black squares represent the Caucasian (CEPH) samples while the grey triangles refer to African (Gambian) samples. **B**. Distribution of of p-values for the significance tests of departure from Hardy Weinberg Equilibrium.

In Gambians none of the statistical tests rejected the null hypothesis of the neutral model.

Косн 2003

Table 4.4 Test of Selective Neutrality in Caucasians (CEPH). † $0.10 > p > 0.05$; ‡ $p < 0.05$.

Gene	Region	no. of sites	$\widehat{\theta}$	$\widehat{\pi}$	Tajima's D	Fu & Li's D	Fu & Li's F
IFNGR1	promoter	4	0.901 ± 0.249	1.424 ± 0.783	1.327	1.008	1.291
	intron1	2	0.451 ± 0.109	0.977 ± 0.455	$2.104^{\ddagger}$	0.752	1.334
	total non-coding	6	1.352 ± 0.421	2.401 ± 1.762	$1.991^{\dagger}$	1.178	$1.680^{\ddagger}$
IL22BP	promoter	4	0.901 ± 0.249	0.628 ± 0.251	-0.695	$-2.278^{\dagger}$	$-2.094^{\dagger}$
IL20RA	promoter	10	2.253 ± 0.856	2.705 ± 2.141	0.577	-0.425	-0.115

Table 4.5 Test of Selective Neutrality in Africans (Gambian). † $0.10 > p > 0.05$; ‡ $p < 0.05$.

Gene	Region	no. of sites	$\widehat{\theta}$	$\widehat{\pi}$	Tajima's D	Fu & Li's D	Fu & Li's F
IFNGR1	promoter	5	1.127 ± 0.331	0.848 ± 0.374	-0.606	-0.812	-0.875
	intron1	2	0.451 ± 0.109	0.772 ± 0.330	1.286	0.752	1.054
	total non-coding	7	1.577 ± 0.518	1.620 ± 0.951	0.072	-0.306	-0.214
IL22BP	promoter	8	1.803 ± 0.623	2.607 ± 2.015	1.2276	$1.302^{\dagger}$	$1.502^{\dagger}$
IL20RA	promoter	10	2.253 ± 0.856	2.574 ± 1.973	0.409	$1.399^{\dagger}$	1.267

4.2.4 Comparison of Human Polymorphism versus Divergence from Chimpanzee

One Chimpanzee was sequenced for the same genomic regions in order to establish the ancestral haplotype. For *IFNGR1* 2327bp of non-coding sequence was analysed, 1872bp for *IL22BP* and 1961bp for *IL20RA*. In the non-coding regions Chimpanzee and human *IFNGR1* differed at 26 sites, *IL22BP* at 16 sites and *IL20RA* at 17 sites. In *IFNGR1* there were 14 transitions and 11 transversions and one insertion/deletion. In *IL22BP* there were 12 transitions, 3 transversions and one insertion/deletion. And in *IL20RA* there were 14 transitions, 2 transversions and 1 insertion/deletion. Another way of detecting natural selection is by comparing the levels of polymorphism within a species versus the divergence between species (Hudson et al., 1987). The HKA test was therefore applied to *IFNGR1*, *IL22BP* and *IL20RA* sequence data from 24 Caucasians and 24 Africans and one chimpanzee with published data from the *LPL* gene serving as a reference locus (Clark et al., 1998). In Caucasians there were 6 segregating sites for *IFNGR1*, 4 for *IL22BP* and 10 for *IL20RA*. In Africans there were 7 segregating sites for *IFNGR1*, 8 for *IL22BP* and 10 for *IL20RA*. In *LPL* there were 108 fixed differences in 7261bp non-coding sequence. In 94 Caucasian chromosomes there were 61 segregating sites in 8736bp non-coding sequences. In 48 African-American chromosomes there were 74 segregating sites in 8736bp non-coding sequence. In Caucasians the p values for *IFNGR1*, *IL22BP* and *IL20RA* were 0.35, 0.48 and 0.50 respectively. In Africans they were 0.21, 0.82 and 0.95 respectively. The HKA test therefore did not reject the hypothesis that either *IFNGR1*, *IL22BP* or *IL20RA* and LPL have evolved under the same parameters. The genes have similar patterns of polymorphism and diversity: this implies that variation is neutral, or that it is undergoing the same patterns of natural selection in each of the three genes.

Table 4.6 Ancestral alleles at all sites that were polymorphic in humans.

IFNGR1			IL22BP			IL20RA		
Position	Human	Chimp	Position	Human	Chimp	Position	Human	Chimp
-1882	G/A	G	-1755	A/G	G	-1633	G/T	G
-1804	G/T	T	-1311	C/T	C	-1620	G/A	A
-611	G/A	A	-1119	T/C	T	-1554	T/C	T
-470	ins/del	ins	-1054	C/A	A	-919	C/G	G
-270	T/C	T	-1014	C/T	T	-889	A/G	A
-72	C/T	C	-1000	T/A	A	-841	T/G	G
-56	T/C	C	-679	G/T	T	-835	A/G	G
+95	T/C	C	-308	A/C	A	-767	C/A	C
+130	G/A	G				-676	T/C	T
						-310	G/A	G
						-271	T/C	T

GENE	POSITION	CHIMPANZEE	HUMAN
IFNGR1	-1811	T	C
	-1775	A	C
	-1749	G	A
	-1735	T	C
	-1722	T	C
	-1690	T	C
	-1683	G	A
	-1485	C	A
	-1454	G	A
	-1351	A	C
	-1088	G	A
	-861	G	A
	-832	A	G
	-785	-	A
	-778	G	T
	-697	C	T
	-592	C	G
	-393	C	A
	-191	C	T
	-70	A	G
	-58	C	A
	-45	G	T
	-22	C	G
	+177	C	A

GENE	POSITION	CHIMPANZEE	HUMAN
	+201	T	G
	+202	T	C
IL22BP	-1665	G	A
	-1565	G	A
	-1562	G	A
	-1531	T	C
	-1349	C	-
	-1259	C	T
	-1071	T	C
	-872	G	A
	-711	G	A
	-687	G	T
	-304	A	T
	-271	A	G
	-239	G	A
	-134	T	C
	-87	G	C
	-6	G	A
IL20RA	-1891	-	G
	-1866	G	A
	-1821	A	G
	-1807	G	A
	-1803	T	G
	-1773	C	T

GENE	POSITION	CHIMPANZEE	HUMAN
	-1767	G	A
	-1735	T	C
	-1673	T	C
	-1490	T	C
	-1322	G	A
	-1109	A	G
	-1101	C	T
	-925	G	A
	-859	T	G
	-445	T	C
	-246	T	C

4.3 Discussion

In this study approximately 2kb segments of the promoter regions of *IFNGR1*, *IL22BP* and *IL20RA* were sequenced in 48 African and 48 Caucasian chromosomes. In these three regions on chromosome 6q23 a total of 28 variant sites were found. DNA sequencing is the most sensitive method for finding DNA polymorphisms and mutations (Eng and Vijg, 1997). In the present study only 7 of the 28 variant sites had been reported on the public databases. None of these 7 SNPs had been validated. In all 20 variant site were found in Caucasians and 26 in Africans. In Africans 18 out of 26 variants sites had a minor allele frequency above 10%. In Caucasians 14 sites were above 10%. Medium and high frequency markers might potentially be more useful for subsequent association studies as the power to detect an association is not only dependent on the relative risk and the sample size but also on the allele frequency. However, it could equally be the case that rarer alleles

that might have arisen more recently have a larger effect on an individual's risk. Therefore markers that are in LD with the disease causing mutation and have a similar allele frequency would have the greatest power to detect the effect. On the other hand LD is likely to extent a long way from mutations that have arisen recently and it might therefore be more difficult do narrow down the region. Given the cost of current genotyping methods it will not be feasible to type all available makers. It will therefore be crucial to establish the haplotype structure of the region under investigation and to pursue haplotype based association studies. The increased marker density that could be established in the current study will aid further genetic analyses such as calculations of linkage disequilibrium and haplotype construction.

Nucleotide Diversity is hugely variable throughout the genome. The nucleotide diversity in the three promoter regions analysed varied considerably. Nickerson et al. (1998) found a nucleotide diversity of $\widehat{\pi} = 0.0021$ in the non-coding region of the *LPL* gene. Angrist et al. (1998) have described a nucleotide diversity of $\widehat{\pi} = 0.0008$ in the non-coding region of *GFRA1*. In a large analysis Li and Sadler (1991) described a nucleotide diversity of $\widehat{\pi} = 0.0003$. The nucleotide diversities of *IFNGR1*, *IL22BP* and *IL20RA* lie within the range previously found. Several previous studies have shown that Africans are genetically more variable than people from other regions of the world (Cavalli-Sforza et al., 1994). The most significant difference of nucleotide diversity was found in *IL22BP*. Twice the number of markers were found in Africans and the nucleotide diversity was four times as high as in Caucasians. Three out of the 4 markers found in *IL22BP* in Caucasians occurred only once (singletons). However, the nucleotide diversities of the other two genes were similar.

The tests for deviation from the neutral model indicated possible evidence for a balancing selection at *IFNGR1* in Caucasians while the tests were not significant at all other loci and in Africans. It is worth noting, though, that Tajima's test can detect a selective sweep or bottleneck only if it has occurred within a specific interval of time in the recent past. Simulation studies (Fu,

1997; Przeworski, 2002) have shown that tests often have little power to detect the deviations they were designed to pick up. Although Tajima's test is generally considered to be the most powerful test it also relies on a sufficient sample size. Simonsen et al. (1995) suggested a sample size of at least 50-100 in order to have reasonable. Pluzhnikov and Donnelly (1996) suggested on the other hand that an increase in the region sequenced, rather than the number of chromosomes, is more likely to identify new mutations. Therefore the present study which was primarily designed to detect new genetic markers for LD calculations and association studies might not have the power to make a final statement on deviation from neutral genetic drift at this locus.

Early comparative studies of the human and chimpanzee genomes established that the extent of DNA sequence difference is in the order of 1.6% (King and Wilson, 1975; Sibley and Ahlquist, 1984; Goodman et al., 1990; Bailey et al., 1991). More recent two large studies estimated divergence at about 1.24% (Chen and Li, 2001; Ebersberger et al., 2002). The current study investigated only a limited area of 6.4kb non-coding promoter sequence. This showed a sequence divergence for *IFNGR1*, *IL22BP* and *IL20RA* of 1.12%, 0.85% and 0.87% respectively. It has however been shown that different regions of the human genome differ significantly in the extent of divergence from the chimpanzee. For example, the intron of the *ZFY* gene shows only 0.69% divergence between human and chimpanze (Dorit et al., 1995), whereas the olfactory receptor *OR1D3P* pseudogene shows a divergence of 3.04% (Glusman et al., 2000). On comparing variation between specific proteins in humans and chimpanzees, King and Wilson (1975) found a low diversity within each species and a high similarity between them. Given the significant phenotypic differences between human and chimpanzees, they suggested that regulatory mutations may be the prime cause for biological differences. It will be of interest to further investigate the potentially functional implications of divergence in promoter regions and further studies looking at expression using chimp promoter constructs are currently ongoing.

4.4 Summary

1. Approximately 2kb of the promoter regions of *IFNGR1*, *IL22BP* and *IL20RA* were sequenced respectively (a total of 6.4kb).

2. A total of 28 SNPs were found in the three promoter regions sequenced.

3. The level of nucleotide polymorphism (θ) was higher in Africans than in Caucasians.

4. The overall nucleotide diversity across the three gene regions was higher in Africans than in Caucasians.

5. F_{ST} values indicate a significant divergence between Africans and Caucasians.

6. In Africans no deviation from expections under the neutral model, i.e. evidence of natural selection, was found using the standard tests of neutrality.

Chapter 5

Linkage Disequilibrium and Haplotypes in the 6q23 region

5.1 Introduction

1. The initial case-control study looking at the the association of polymorphic markers in the promoter of *IFNGR1* yielded a potentially interesting association. The *IFNGR1*-470ddel allele was associated with protection against severe malaria in the Mandinka ethnic group. In the same ethnic group heterozygosity at position *IFNGR1*-56 was associated with protection against cerebral malaria and fatal outcome of cerebral malaria.

2. However, so far it remains unclear whether this represents an effect due to these markers themselves or due to markers that are in linkage disquilibirum with *IFNGR1*-470 and -56.

3. In order to further localise the associations, and therefore the *true* functional variant, extended haplotypes needed to be created. If multiple haplotypes can be found that carry the associated alleles of *IFNGR1*-470 and -56, only one might be found to be associated with disease,

thereby giving further clues about the location of the functional variant.

4. Calculating linkage disequilibrium (LD) and identifying haplotype blocks can furthermore help in localising the area in which to look for the *true* functional variant.

5. Typing a many SNPs in a large number number of samples is expensive. However, not all SNPs are in fact informative. Identifying haplotype tagging SNPs (htSNP) allows faster and cost efficient typing of a particular region.

5.1.1 Background

Genetic markers that are close to each other are often not randomly associated. When a mutation occurs it is initially associated with all markers on this particular chromosome, i.e. in complete *linkage* with all other markers. Over time this association is broken down by recombination. Depending on the chromosome 0.5 to 2 recombination events occur in each meiosis. A genetic distance at which the chance for recombination to occur equals 1 is called 1 Morgan. Although genetic distance is not equal to physical distance 1 centiMorgan approximately corresponds to 1Mb. Over subsequent generations the association between markers will be broken down until theoretically all markers are randomly associated. In practice one can observe that markers that are close to each other are often not randomly associated, i.e. they are in *linkage disequilibrium.*

Linkage disequilibrium (LD), or the non-random association of alleles, is poorly understood in the human genome. Population genetic theory suggests that LD is determined by the age of the markers, population history, recombination rate, selection and genetic drift. There are an estimated 3 million SNPs throughout the human genome. Given the cost and throughput of current genotyping methods it would be an impossible task to type all of these markers and test their association with a certain disease phenotype. Thanks to the association between markers it will be possible to reduce the number of necessary typings. An understanding of the patterns of linkage

disequilibrium will be crucial in designing association studies and detecting disease causing alleles.

Linkage Disequilibrium can vary significantly throughout the genome (Reich et al., 2001). Simulations using simplified assumptions regarding population demography and the spatial distribution of recombination rates suggest that useful LD is not expected to be found at physical distances over 3 kb in the human genome (Kruglyak, 1999). Empirical studies of LD in human populations often report far higher levels of LD than those predicted by Kruglyak's study, in some cases greater than 100kb (Abecasis et al., 2001; Taillon-Miller et al., 2000; Collins et al., 1999). In addition the distance that LD spans varies in different populations. Reich et al. (2001) found that LD extended 60kb from common alleles in a United States population of European descend, while in Nigerians LD extended markedly less far.

As markers in close physical proximity are correlated due to linkage disequilibrium genetic variability is reduced. While a set of N markers could in theory generate 2^N different haplotypes, in practice this is rarely the case. Often a limited number of haplotypes can be defined in a block of SNPs. These blocks of haplotypes are regions over which there is little evidence for recombination. Several recent studies have described haplotype blocks ranging from about 5kb to 100kb (Patil et al., 2001). The majority of chromosomes in each block could be assigned to a limited number of haplotypes. There is also growing evidence that the haplotype blocks are interspersed with recombination hotspots (Jeffreys et al., 2001). This knowledge has contributed to the identification of genes involved in susceptibility to complex diseases (Rioux et al., 2001; Hugot et al., 2001; Ogura et al., 2001). But given the significant variation of the length of blocks the general properties of haplotype blocks will remain unclear until dense genome wide haplotypic maps are available.

From the above it becomes clear that a mere association for the associations found for the SNPs in the *IFNGR1* promoter does not provide evidence that they are the functional variants themselves. The limited haplotype analysis presented in the initial case control study gave only limited insight into

the haplotype structure at this locus. As the the SNPs themselves might be non-functional and solely act as markers for other functional SNPs in this region it was important to investigate this genetic region more closely.
It was therefore important to establish the extent of LD in this region and to further dissect haplotypes on which the associated markers lie on.
This is of particular interest as the gene under investigation is neighboured by other immune genes. If sufficient linkage disequilibrium exists between neighbouring genes associations found in one gene might in fact only pick up an event that is happening in another gene. As presented in the previous chapter two other genes that seem to be involved in host immune response are located near to *IFNGR1*: *IL22BP* and *IL20RA*.

5.1.2 Approach

By resequencing approximately 2kb of the promoter regions of *IFNGR1*, *Il22BP* and *IL20RA* 28 markers could be identified. For the further study of linkage disequilibrium in this region only markers that had an allele frequency above 10% in at least one of the two populations were selected. This was done for several reasons. First, medium and high frequency markers might potentially be more useful for subsequent association studies as the power to detect an association is not only dependent on the relative risk and the sample size but also on the allele frequency. Second, high frequency markers tend to be of high allele frequency in all populations, facilitating comparison between different populations. Third, low frequency variants tend to be younger and therefore there would have been less time to break down LD. Using higher frequency markers the chance of overestimating LD might be less. Using the SNPs detected by resequencing it was possible to calculate LD between markers, construct haplotypes, define haplotype blocks, and to select haplotype tagging SNPs that were used for a large case control study.

5.1.3 Marker Selection

Of the 28 variant sites that were detected by resequencing the promoter regions of *IFNGR1*, *IL22BP* and *IL20RA* only markers that had an allele frequency above 10% in at least one of the two populations were selected. In the 48 African chromosomes that were sequenced 18 out of 26 variants sites had a minor allele frequency above 10%. In Caucasians 14 sites were above 10%. In total the allele frequencies of 21 markers were above 10% in either one or both of the populations. Two validated markers from the public databases were added. One of these markers (named 6:144218683) is a synonymous change in exon 7 of *IFNGR1*. The other (named 6:144153554) is located in the intergenic region between *IL22BP* and *IL20RA*. Of these 23 markers 4 could not be typed by Sequenom. The assays for *IFNGR1*+95, 6:144066115, 6:144241363 and for 6:144066037 had a failure rate above 25% and were therefore excluded from the analysis. The remaining 19 markers were typed in 32 UK Caucasian family trios and 32 Gambian trios. This approach enabled the integration of family and population data in order to generate efficient haplotypic reconstruction from genotypic data. This was implemented by using the program `PHAMILY` to parse pedigrees for phase information before sending the genotypes to the program `PHASE` to reconstruct haplotypes.

5.1.4 Haplotype Construction

To determine haplotypes in families and unrelated individuals, a novel method for haplotype construction was employed using the programs `PHAMILY` (Hans Ackerman, DPhil thesis, University of Oxford, Oxford, UK, 2001; contact Hans Ackerman (`ackerman@fas.harvard.edu`) for information regarding `PHAMILY`) and `PHASE` (Stephens et al., 2001b)). The `PHASE` program is produced by the Mathematical Genetics Group, University of Oxford, Oxford, UK. The software is available online at:
`http://www.stats.ox.ac.uk/mathgen/software.html`.
`PHASE` uses a Bayesian method which is based on a coalescent model. `PHAMILY`

reconstructs parental haplotypes only at unambiguous sites given the pedigree information. These data can be used as an additional input file for PHASE to increase the accuracy of haplotype reconstruction by using all available information. The remaining ambiguous sites are assigned by PHASE, and the uncertainty associated with each PHASE assignment is calculated. Analyses using PHASE were repeated a minimum of 8 times. No significant interrun variability was observed.

5.1.5 Determination of Haplotype Blocks

The program `HaploBlockFinder version 0.5` by Zhang (available at: `http://cgi.uc.edu/cgi-bin/kzhang/haploBlockFinder.cgi`) was used for this approach. The program uses a greedy algorithm similar to that used previously to discern block structure (Patil et al., 2001; Zhang et al., 2002).

5.1.6 Database Search

The following public databases were searched for validated SNPs:

1. Ensembl
   `http://www.ensembl.org`

2. National Center for Biotechnology Information
   `http://www.ncbi.nlm.nih.gov/SNP/index.html`

3. Human Genome Variation Database
   `http://hgvbase.cgb.ki.se/`

Although the 240kb sequence surrounding the three genes, *IFNGR1*, *IL22BP* and *IL20RA*, contained a total of 296 reported SNPs, only 10 of those were validated.

5.2 Results

5.2.1 Allele Frequencies

Parental alleles were counted to calculate their frequencies in 128 chromosomes. Frequencies of the minor alleles ranged from 0-.42 in Africans and 0.0 - 0.48 in Caucasians (Figure 5.1). In Caucasians 11 SNPs were found at a frequency greater than .10 (Table 5.2). In Africans 12 SNPs had a minor allele frequency greater than .10 (Table 5.1). The *IFNGR1*-611 had not been found in the 24 African samples that were originally sequenced but was found in this larger set of African samples at frequency of .03. The *IFNGR1*-72 had a minor allele frequency less than .10 in both populations, i.e. .05 in Caucasians and .00 in Africans, while it had been found at a frequency of .15 and .04 respectively in the samples that had been sequenced.

5.2.2 Haplotypes

A total of 34 different haplotypes were observed in 128 Gambians chromosomes. Eight of these haplotypes were also found in 128 Caucasian chromosomes in addition to 27 further haplotypes. These eight haplotypes that were present in both populations account for 33% of all haplotypes in Africans and of only 25% of haplotypes present in Caucasians. The allelic composition and frequency of each haplotype is illustrated in tables 5.3 and 5.4.

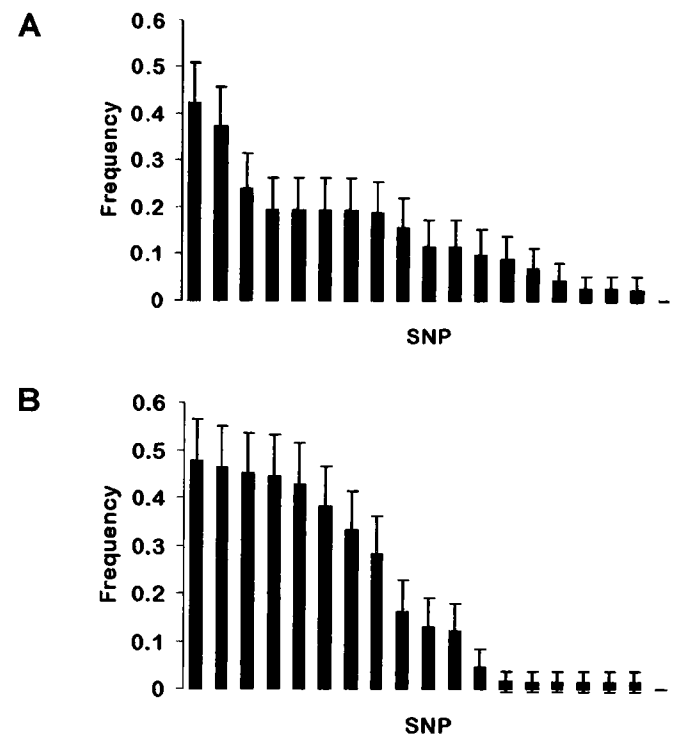


Figure 5.1 Frequency distribution of individual SNPs in Africans (**A**) and Caucasians (**B**). 95% confidence intervals are indicated with error bars.

Table 5.1 Frequencies of 19 SNPs in an African (Gambian) population

IDENTIFIER	POLYMORPHISM	POSITION	FREQUENCY	95% C.I.
1	IFNGR1-1882	6:144241441	0.024	0.027
2	IFNGR1-611	6:144240170	0.026	0.028
3	hIFNGR1-470	6:144240029	0.090	0.050
4	IFNGR1-72	6:144239631	0.000	0.000
5	hIFNGR1-56	6:144239615	0.424	0.086
6	IFNGR1+130	6:144239429	0.045	0.036
7	IFNGR1+20876	6:144218683	0.024	0.027
8	IL22BP-1755	6:144195503	0.194	0.069
9	IL22BP-1311	6:144195059	0.195	0.069
10	IL22BP-1119	6:144194867	0.188	0.068
11	IL22BP-1054	6:144194802	0.242	0.074
12	IL22BP-1014	6:144194762	0.156	0.063
13	IL22BP-679	6:144194491	0.195	0.069
14	IL22BP-308	6:144194056	0.195	0.069
15	*intergenic*	6:144153554	0.373	0.084
16	IL20RA-1620	6:144066816	0.117	0.056
17	IL20RA-1554	6:144066750	0.117	0.056
18	IL20RA-767	6:144065963	0.100	0.052
19	IL20RA-310	6:144065506	0.069	0.044

Table 5.2 Frequencies of 19 SNPs in a UK Caucasian population

IDENTIFIER	POLYMORPHISM	POSITION	FREQUENCY	95% C.I.
1	IFNGR1-1882	6:144241441	0.123	0.057
2	IFNGR1-611	6:144240170	0.479	0.087
3	hIFNGR1-470	6:144240029	0.000	0.000
4	IFNGR1-72	6:144239631	0.046	0.037
5	hIFNGR1-56	6:144239615	0.333	0.082
6	IFNGR1+130	6:144239429	0.163	0.064
7	IFNGR1+20876	6:144218683	0.131	0.059
8	IL22BP-1755	6:144195503	0.016	0.022
9	IL22BP-1311	6:144195059	0.016	0.022
10	IL22BP-1119	6:144194867	0.016	0.022
11	IL22BP-1054	6:144194802	0.381	0.085
12	IL22BP-1014	6:144194762	0.016	0.022
13	IL22BP-679	6:144194491	0.016	0.022
14	IL22BP-308	6:144194056	0.016	0.022
15	*intergenic*	6:144153554	0.463	0.087
16	IL20RA-1620	6:144066816	0.451	0.087
17	IL20RA-1554	6:144066750	0.446	0.087
18	IL20RA-767	6:144065963	0.429	0.086
19	IL20RA-310	6:144065506	0.284	0.079

Table 5.3: Frequencies of 34 haplotypes in a Gambian population sorted by the *IFNGR1*, *IL22BP* and *IL20RA* SNPs. To facilitate viewing allele 2 is symbolised by '-' and allele 1 by 'o'. Haplotypes 1, 4, 10, 16, 24, 25, 33 and 34 were also found in Caucasians.

IDENTIFIER	HAPLOTYPE	NO.	FREQUENCY
1	-----oooo-o--oo----	3	0.02
2	----oo-oo-o--ooooo-	1	0.01
3	o--------oooo--ooo-	1	0.01
4	o------oo-o--o-ooo-	1	0.01
5	o----o---oooo--oooo	3	0.02
6	o----o-oo----o-ooo-	10	0.08
7	o----o-oo----ooooo-	2	0.02
8	o----o-oo-o--o-----	2	0.02
9	o----o-oo-o--o-ooo-	13	0.10
10	o----o-oo-o--ooooo-	3	0.02
11	o----o-oo-o--oooooo	3	0.02
12	o---o----oooo--ooo-	1	0.01
13	o---o--oo-o--o-ooo-	3	0.02
14	o---oo----o-o--ooo-	1	0.01
15	o---oo---oo-o--ooo-	4	0.03
16	o---oo---oooo--ooo-	9	0.07
17	o---oo---oooo--oooo	1	0.01

continued on next page

continued from previous page

IDENTIFIER	HAPLOTYPE	NO.	FREQUENCY
18	o---oo---oooo-oooo-	1	0.01
19	o---oo--o-o--o-ooo-	1	0.01
20	o---oo-o-oooo--ooo-	1	0.01
21	o---oo-oo----o-ooo-	5	0.04
22	o---oo-oo-o--o-----	7	0.05
23	o---oo-oo-o--o-ooo-	11	0.09
24	o---oo-oo-o--oo----	3	0.02
25	o---oo-oo-o--ooooo-	20	0.16
26	o---oo-oo-o--oooooo	1	0.01
27	o-o--o---oooo--ooo-	1	0.01
28	o-o--o-oo----o-ooo-	1	0.01
29	o-o--o-oo----ooooo-	9	0.07
30	o-o--o-oo-o--ooooo-	1	0.01
31	o-o-oo---oooo--ooo-	1	0.01
32	oo---o---oooo--oooo	1	0.01
33	oo--oo-oo----o-ooo-	1	0.01
34	oo--oo-oo----ooooo-	2	0.02

Table 5.4: Frequencies of 34 haplotypes in a UK Caucasian population sorted by the *IFNGR1*, *IL22BP* and *IL20RA* SNPs. To facilitate viewing allele 2 is symbolised by '-' and allele 1 by 'o'. Haplotypes 1, 4, 10, 16, 24, 25, 33 and 34 were also found in Africans.

IDENTIFIER	HAPLOTYPE	NO.	FREQUENCY
35	-----oooo----o-----	1	0.01
36	-----oooo-o--o-----	9	0.07
37	-----oooo-o--o-ooo-	1	0.01
38	-----oooo-o--o-oooo	2	0.02
1	-----oooo-o--oo----	1	0.01
39	----ooooo-o--o-----	1	0.01
40	----ooooo-o--o-oooo	2	0.02
41	-o---oooo-o--o-----	1	0.01
42	o------oo----o-----	3	0.02
43	o------oo----o-ooo-	1	0.01
44	o------oo----oo----	1	0.01
45	o------oo----oooooo	7	0.05
4	o------oo-o--o-ooo-	1	0.01
46	o------oo-o--oo----	6	0.05
47	o----o-oo----oooooo	2	0.02
10	o----o-oo-o--ooooo-	1	0.01
16	o---oo---oooo--ooo-	2	0.02

continued on next page

continued from previous page

IDENTIFIER	HAPLOTYPE	NO.	FREQUENCY
48	o---oo-oo----o-----	2	0.02
49	o---oo-oo----ooooo-	1	0.01
24	o---oo-oo-o--oo----	11	0.09
25	o---oo-oo-o--ooooo-	1	0.01
50	o--o---oo----o-----	1	0.01
51	o--o---oo----o-oooo	2	0.02
52	o--o---oo-o--oo----	2	0.02
53	oo---o-oo----oo----	1	0.01
54	oo--oo-oo----o-----	17	0.13
55	oo--oo-oo----o--o--	1	0.01
33	oo--oo-oo----o-ooo-	4	0.03
56	oo--oo-oo----o-oooo	10	0.08
57	oo--oo-oo----oo----	9	0.07
58	oo--oo-oo----oooo-o	1	0.01
34	oo--oo-oo----ooooo-	9	0.07
59	oo--oo-oo----oooooo	8	0.06
60	oo--oo-oo-o--o-----	4	0.03
61	oo--oo-oo-o--oooooo	2	0.02

5.2.3 Linkage Disequilibrium

Pairwise measures of linkage disequilibrium (D, D', χ^2, Δ^2) were calculated for all pairs of markers using the program `Haploxt`. Figure 5.4 and 5.5 represent linkage disequilibrium between 19 markers. The absolute values of D'

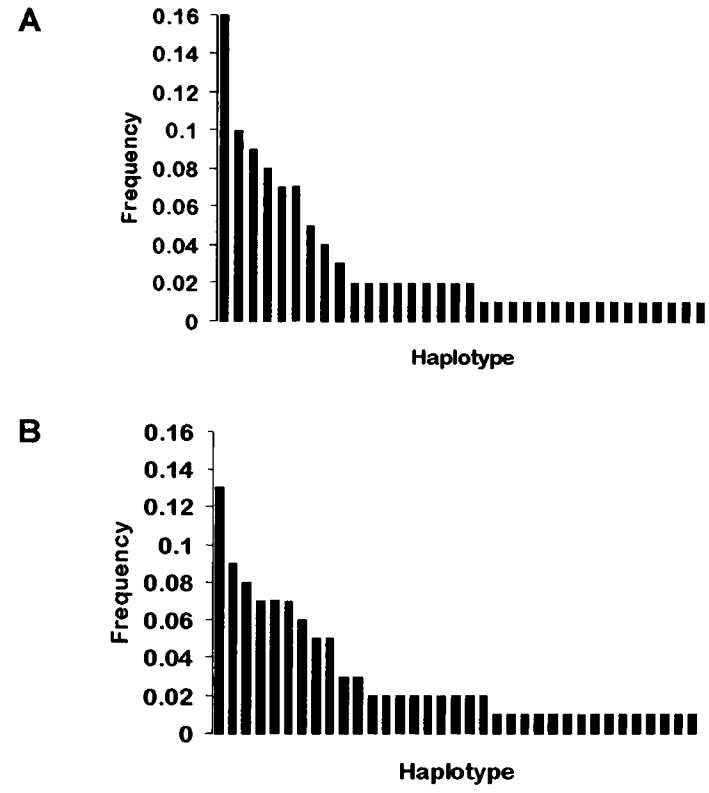


Figure 5.2 Frequency distribution of haplotypes in Africans (**A**) and UK Caucasians (**B**).

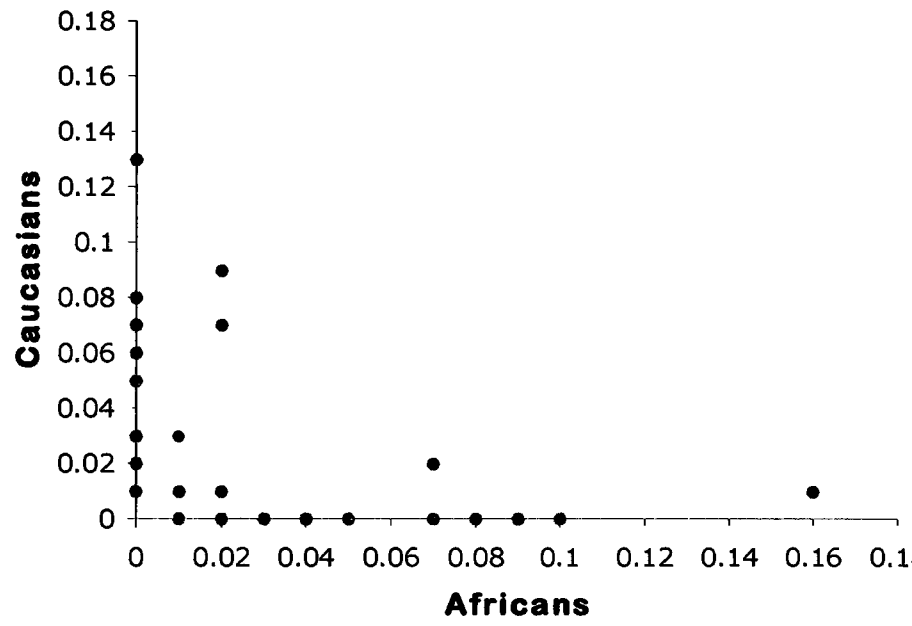


Figure 5.3 Scatterplot of frequency distribution of haplotypes in Africans and UK Caucasians.

and Δ^2 are represented as different colours in the output of the graphical overview of linkage disequilibrium (GOLD) program. The statistical significance of the χ^2 values was calculated by `Haploxt` and significant p-values represented as circles are shown in Figures 5.4A and 5.5A. High values of D' were found across large regions both in Africans and Caucasians while high Δ^2 values which take the sample size into account are mostly confined to each of the three promoter regions. Significant χ^2 values were mostly confined to each promoter region in Caucasians. There are some exceptions, though. Significant LD based on χ^2 was found for some markers that were more than 120kb apart. For example the high frequency marker 6:144194802 was significantly associated with all four markers in *IL20RA* which are located at 127kb distance. Markers 6:144241441 (*IFNGR1*-1882) and 6:144240170 (*IFNGR1*-611) are associated with markers in *IL22BP* and beyond. In Africans the pattern for D' and Δ^2 is similar to that seen in Caucasians. Significant χ^2 values can be found between all markers of *IL22BP* and *IL20RA*, a region spanning 120kb. It is also of note that significant χ^2 values were also seen for *IFNGR1*-470 and *IFNGR1*-56 with markers that are at over 85kb and 170kb distance respectively.

The pairwise measures of linkage disequilibrium D' and Δ^2 were also plotted against physical distance (Figure 5.6). As expected for both D' and Δ^2 a decline can be observed with increasing distance both in Caucasians and Gambians. Nevertheless individual values deviate from this trend and several high D' values could be observed over large distances. It was found that the average D' and Δ^2 facilitate the comparison of LD for Caucasians and Gambians. Therefore all D' and Δ^2 were averaged up to the distance plotted. Here is was seen that the extent and decline of LD was very similar in Caucasians and Gambians.

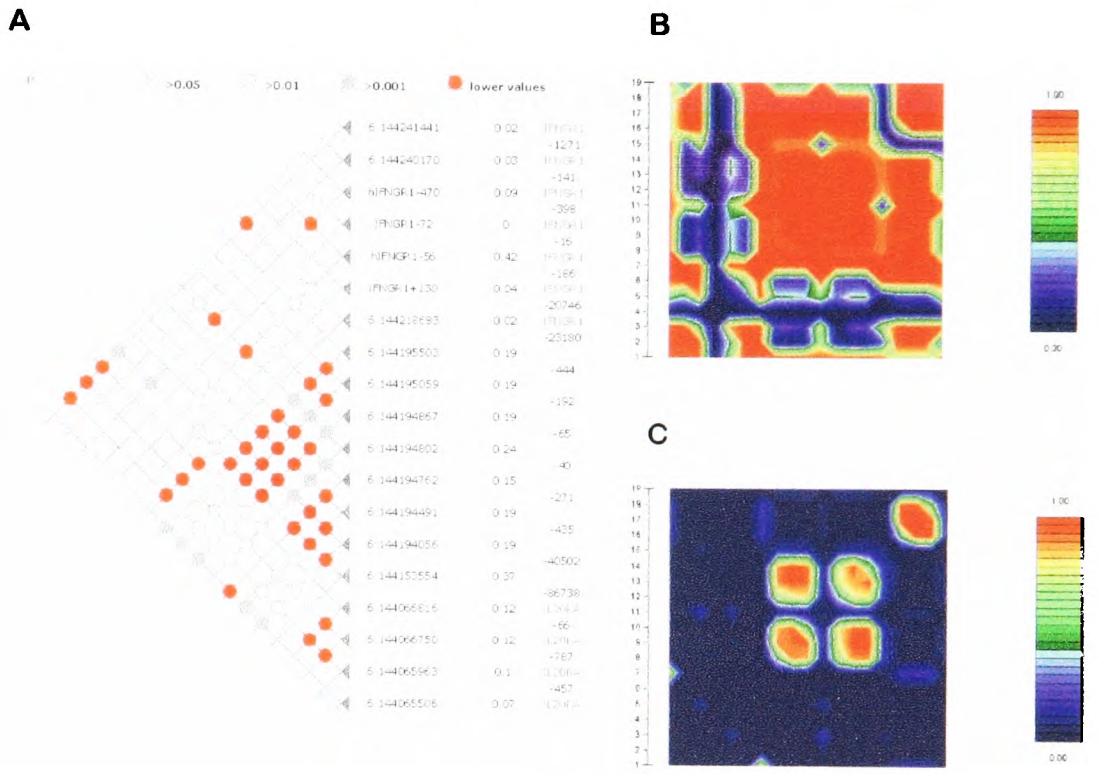


Figure 5.4 Linkage Disequilibrium between 19 polymorphisms in 6q23 in a Gambian population. **A.** Pairwise measure of LD presented by p-value. Various degrees of significance are indicated by differently coloured circles. Open white circles represent p-values $0.05 \geq p > 0.01$, filled grey circles indicate $0.01 \geq p > 0.001$ and filled red circles represent $p \leq 0.001$. **B.** LD measured by D' illustrated with the graphical overview of linkage disequilibrium (GOLD) program. **C.** LD measured as Δ^2.

Koch 2003

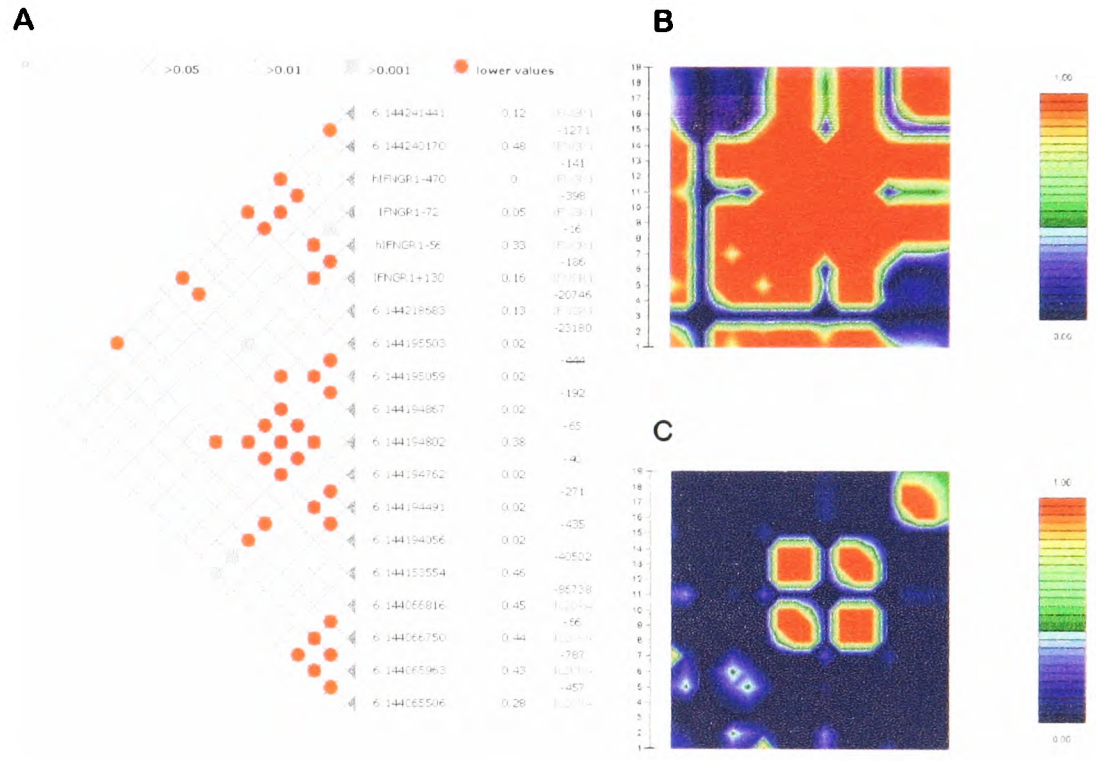


Figure 5.5 Linkage Disequilibrium between 19 polymorphisms in 6q23 in a UK Caucasian population. **A.** Pairwise measure of LD presented by p-value. Various degrees of significance are indicated by differently coloured circles. Open white circles represent p-values $0.05 \geq p > 0.01$, filled grey circles indicate $0.01 \geq p > 0.001$ and filled red circles represent $p \leq 0.001$. **B.** LD measured by D' illustrated with the graphical overview of linkage disequilibrium (GOLD) program. **C.** LD measured as Δ^2.

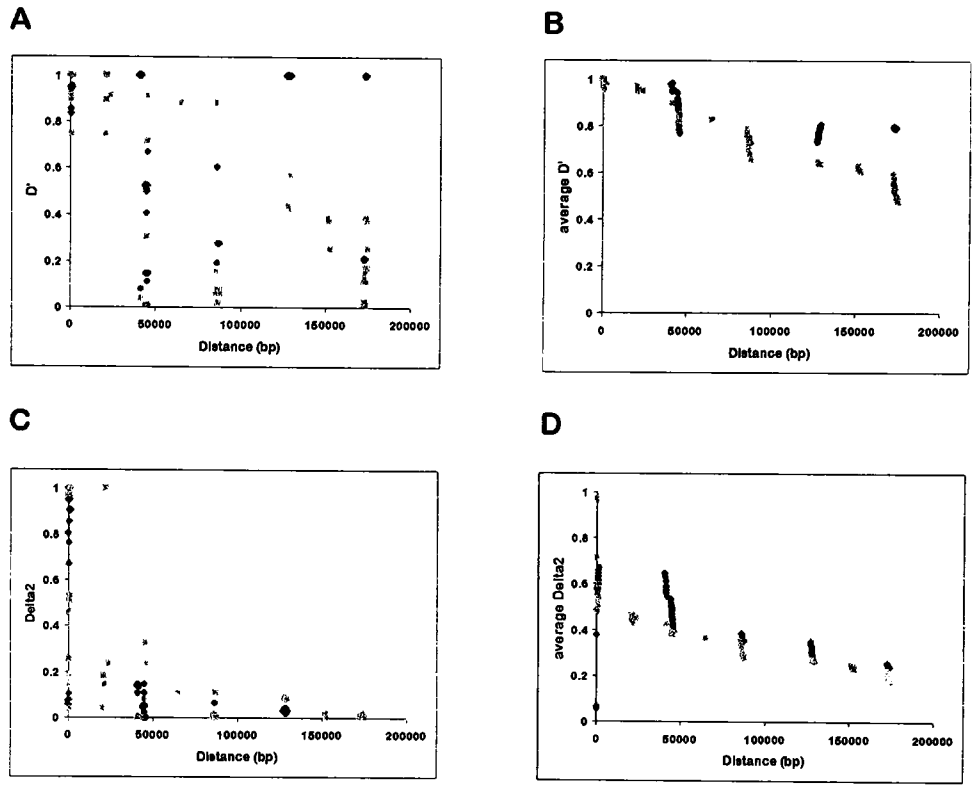


Figure 5.6 Linkage Disequilibrium versus physical distance in 6q23 for all pairs of markers. Only markers with an allele frequency of $f > 0.05$ were used for this analysis. Values for Gambian samples are represented by black squares and Caucasians by grey squares. **A.** LD as measured by D'. **B.** Average D'. **C.** LD measured as Δ^2. **D.** LD measured as average Δ^2.

5.2.4 Haplotype Blocks and Marker Selection

After haplotype reconstruction with PHAMILY / PHASE the haplotype information was analysed using the greedy algorithm as described by Patil et al. (2001) and discrete blocks of limited haplotype diversity were obtained. The program HaploBlockFinder was used for this approach. In order to obtain useful information only variants with an allele frequency of $f \geq 0.05$ were used. The chromosomal coverage that was aimed for in each haplotype block was 0.90. The haplotype blocks obtained were mostly confined to individual promoter regions (Figure 5.7 and 5.8). Choosing chromosomal coverage of ≥ 0.90 or $D' \geq 0.80$ gave similar results (data not shown). Where haplotype blocks with several markers had been created the frequencies of the haplotypes within these block were then analysed by ENTROPY in order to define haplotype tagging SNPs. A cutoff of 0.05 for the minimum haplotype frequency was used. In Gambians the *IL22BP* block consists of 8 SNPs. In order to describe over 90% of haplotypes within this block only 3 of these SNPs (6:144194867, 6:144194802 and the intergenic SNP 6:144153554) need to be typed (Figure 5.9A). Only two out of four SNPs need to be typed in the *IL20RA* block to describe more than 90% of haplotypes (Figure 5.9B). In Caucasians all *IFNGR1* markers need to be typed to achieve maximal haplotype coverage while two out of four *IL20RA* are sufficient for this block (Figure 5.10).

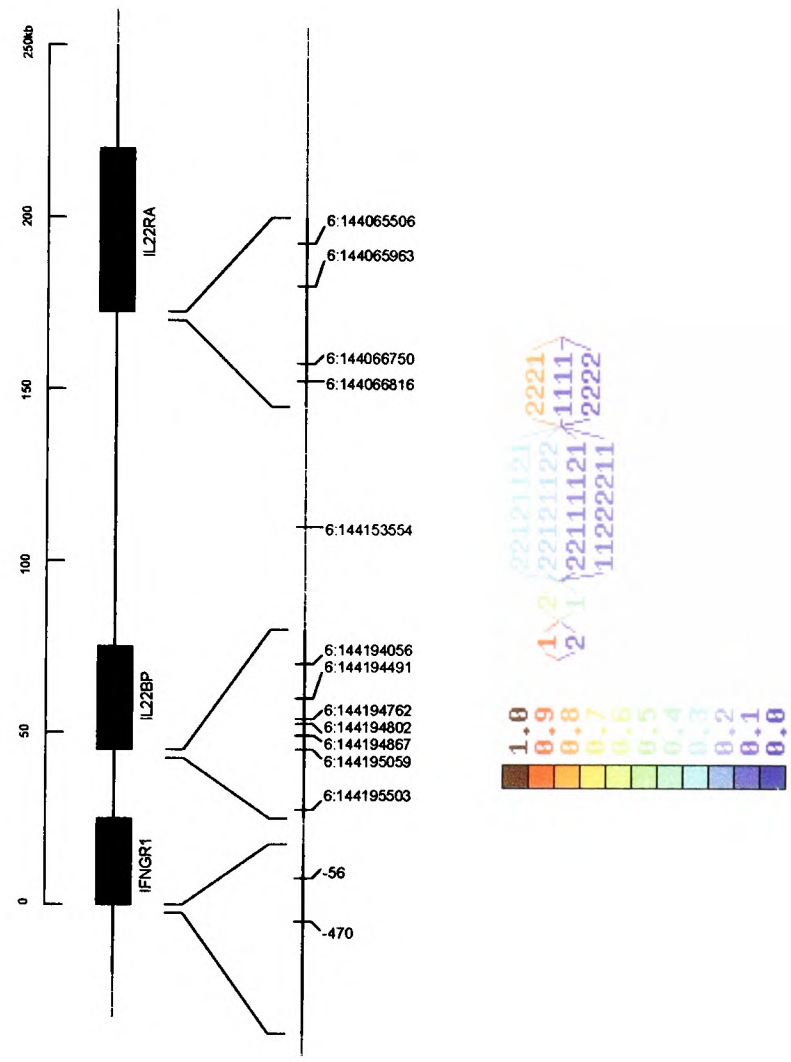


Figure 5.7 Haplotype Blocks of markers in 6q23 in Gambians. Only markers and haplotypes with $f \geq 0.05$ were used. Haplotypes were generated by PHAMILY / PHASE and then analysed by the HaploBlockFinder program with chromosomal coverage set to 0.90.

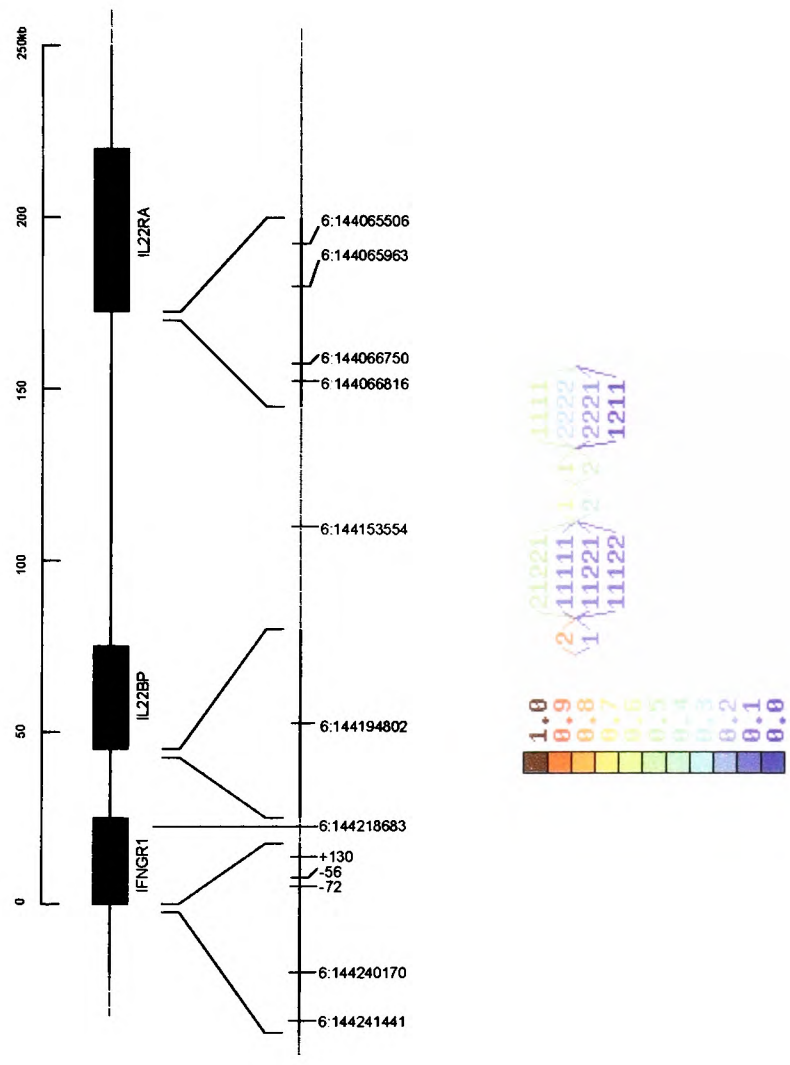


Figure 5.8 Haplotype Blocks of markers in 6q23 in UK Caucasians. Only markers with $f \geq 0.05$ were used. Haplotypes were generated by PHAMILY / PHASE and then analysed by the HaploBlockFinder program with chromosomal coverage set to 0.90.

5.3 Discussion

There has been great interest in describing the patterns of linkage disequilibrium in the human genome. In spite of an increasing number of studies becoming available it turns out that only few predictions can be made from these for other regions. Theoretical predictions and values of LD obtained by *real* data also seem to differ greatly. Choosing the most informative markers to calculate LD is crucial but can also introduce bias to the values of LD observed. Although high frequency markers will be particularly useful for a small sample set in which power is limited, these markers also tend to be the oldest and recombination events will have broken down their relationships with other markers. Linkage disequilibrium might therefore be underestimated. Particularly for markers which have been found to be associated with a given disease this might not be helpful as the actual functional variant might be further away than estimated by calculating LD. Low frequency marker on the other hand tend to be younger and LD might therefore extend a long way from them and it might be harder to localise the area in which the true functional variants should be sought. While Mendelian disease mutations are highly penetrant and under strong selection which keeps allele frequency low, some authors suggest that most of the genetic risk for common, complex diseases is due to disease loci where there is one or few common variants (Reich and Lander, 2001; Lander, 1996; Chakravarti, 1999) as selection and penetrance are less strong. On the other hand, some papers indicate that this hypothesis was dependent on the assumed level of allelic heterogeneity and it might rather be that intermediate frequency variants were responsible for susceptibility to complex diseases (Pritchard, 2001; Pritchard and Cox, 2002).

For the present study a resequencing approach was taken and markers with a range of allele frequencies described. In order to establish a comprehensive assessment of LD in any given genomic region the markers should be evenly spaced throughout the region of investigation and have similar allele frequencies in all populations. However, there were only few validated

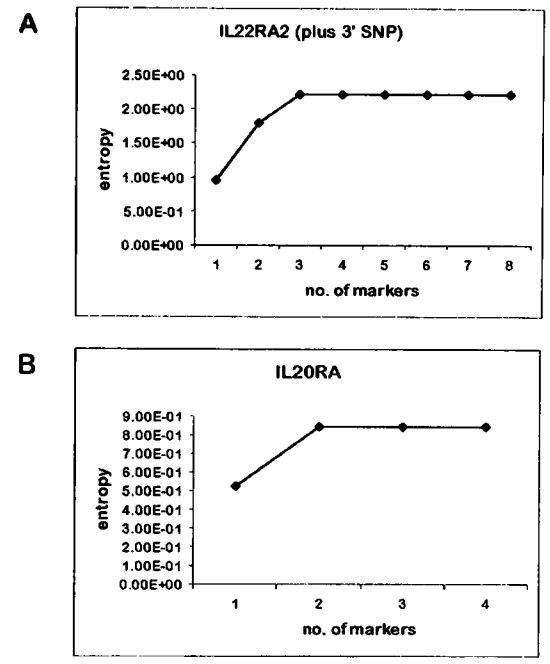


Figure 5.9 Entropy calculations for Gambian haplotypes determined by the greedy algorithm. Only markers and haplotypes with $f \geq 0.05$ were used.

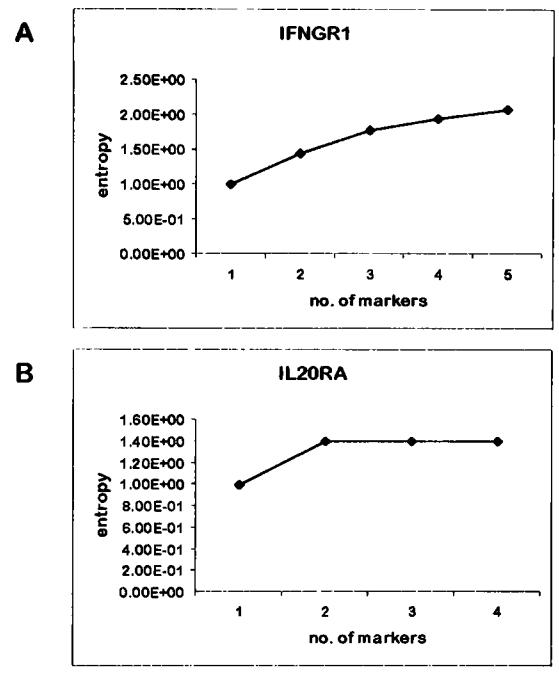


Figure 5.10 Entropy calculations for Caucasian haplotypes determined by the greedy algorithm. Only markers and haplotypes with $f \geq 0.05$ were used.

SNPs in the public databases for the regions surrounding *IFNGR1*. The time frame and financial resources of this project would not have allowed a resequencing approach for the entire 180kb spanning the three genes. The approach to resequence the promoter regions only and calculate LD between genes must therefore be taken as a first approximation to mapping a potential disease locus. But the present study was also designed with two aims in mind in particular. Firstly it should be established whether LD from variants in *IFNGR1* which had been found to be associated with malaria extends as far as the neighbouring genes (*IL22BP* and *IL20RA*). Secondly SNPs in *IL22BP* and *IL20RA*, two genes with potentially important functions in the inflammatory host response, should also be tested for their association with malaria. In order to decrease the number of markers necessary to map an association it was important to establish LD patterns and define blocks of haplotypes. Haplotypes were established by using two steps. First some information about phase was obtained by using family trios. Together with this information the parental genotypes were then run through PHASE which constructs haplotypes from unrelated individuals. Thompson et al. (1988) found that haplotypic data halves the sample size requirements (as compared to genotypic data) for LD calculations.

It was found that different measures for LD give different estimates of the extent of LD. While D' values were generally high throughout the 6q23 region in both populations the measure Δ^2 indicated that LD would not spread across several genes. There has been much debate as to which is the most appropriate representation of LD (Devlin and Risch, 1995; Jorde, 2000; Pritchard and Przeworski, 2001; Weiss and Clark, 2002; Ardlie et al., 2002). Δ^2 is derived by dividing D by $\sqrt{(p_1 p_2 q_1 q_2)}$ where p_1, p_2, q_1 and q_2 represent the allele frequencies of the two markers. In the case of no recombination between two markers *and* equal allele frequency $\Delta^2 = 1$. Δ^2 is also equal to χ^2/N where N is the total number of haplotypes. It therefore takes the sample size into account and provides a means of testing the statistical significance of Δ. Both the p-value (as derived from χ^2) and Δ^2 indicate that

LD is strongest within each of the promoter regions. As indicated by χ^2 several markers, though, are associated with markers in the neighbouring gene. Most strikingly in Gambians where all *IL22BP* markers are strongly associated with markers over 120kb away in the *IL20RA* promoter. Most interestingly *IFNGR1*-56 and *IFNGR1*-470 are strongly associated with markers that are located at a distance of over 80kb and 170kb respectively. Increased linkage disequilibrium can be a sign of recent natural selection whereby associations with all markers on the same haplotype as the one causing selection is maintained. Ardlie et al. (2002) note, however, that p-values can create a misleading impression that LD extends over great distances. Reich et al. (2001) describe that LD tends to extend further in northern Europeans than in Africans. This is coherent with an out-of-Africa bottleneck model whereby a relatively small founding population left Africa approximately 50,000 years ago. In the present study no significant difference in the extent of LD between Africans and Caucasians could be observed. There could be several reasons for this observation. First, the number of available markers in this region is limited. Apart from two markers, one in the last exon 7 of *IFNGR1* and an intergenic SNP between *IL22BP* and *IL20RA*, all markers used were located in the promoter regions of the three genes. Therefore even coverage of the region was not given. Second, the markers used were mostly found by resequencing the promoter regions. Their frequencies in the two populations vary significantly. This might lead to bias of LD calculation. Third, it is conceivable that selective pressure maintains a *beneficial* mutation and all markers on its background at a higher frequency level, therefore leading to higher LD in Africans in this regions than would be expected. Despite the obvious shortcomings of this study in terms of a general statement about LD in this region it is nevertheless interesting that LD could potentially extent far from the variant sites in the *IFNGR1* promoter which were found to be associated with severe malaria. Assuming that these variants were merely markers for truly functional polymorphisms this could make it more difficult to localise the effect. It also adds weight to the need of functional investigations of the *IFNGR1* promoter polymorphisms. By describing the haplotype

structure of this regions in more detail is was also possible to define haplotype blocks and identify *tagging* SNPs within each block that describe the great majority of all haplotypes thereby significantly reducing the number of markers that need to be typed in a subsequent association study.

5.4 Summary

1. A family based method was used to construct haplotypes for an extended region covering *IFNGR1*, *IL22BP* and *IL20RA* and for each gene respectively.

2. Most markers were only in high LD with markers nearby, i.e. LD was highest within each of the three genes. In some instances, however, LD was spanning the entire region.

3. Although haplotypes in Gambians and Caucasians were different (only eight overlapping haplotypes), based on the number of haplotypes observed both populations appear to have similar levels of haplotypic diversity.

4. The range and decline of LD with distance was similar in Caucasians and Africans.

5. Haplotype blocks were defined using the greedy algorithm. None of the blocks reached across two genes.

6. Haplotype tagging SNPs (htSNP) were defined. These capture the majority of haplotypes observed for this region and allow fast and efficient future genotyping.

Chapter 6

Common variants of chromosome 6q23 and susceptibility to severe malaria

6.1 Introduction

1. The previous case control analysis of four polymorphisms in the promoter and first intron of the *IFNGR1* gene had yielded a potential genetic association of the *IFNGR1*-56 variant with susceptibility to cerebral malaria in Mandinka, the largest ethnic group in The Gambia, whereby heterozygote individuals were protected against cerebral malaria. Furthermore a TT double deletion at nucleotide positions -470/-471 was protective against severe malaria, i.e. either cerebral malaria or severe malarial anaemia, in this ethnic group.

2. Resequencing a larger region of the *IFNGR1* promoter and furthermore the promoter regions of the neighbouring genes, *IL22BP* and *IL20RA*, yielded further variant sites which could firstly be tested for their association with malaria and secondly be used to further expand the

haplotypes that were found to be associated with the disease.

3. It was hereby of particular interest on which extended haplotypes the two markers *IFNGR1*-56 and *IFNGR1*-470 lie on as means to further establish whether they might just be markers for other functional variants elsewhere. Linkage disequilibrium analysis had indicated that both markers might be in strong LD with other markers which were located in neighbouring genes, a significant genetic distance away. By analysing extended haplotypes it might therefore become apparent whether the associations found are dependent on these two markers or any particular haplotypes which they can be found on.

6.2 Approach

Meanwhile mass spectrometry had become available for high-throughput genotyping and my own intention to extend the case control analysis to further markers in this region coincided with the group's effort to set up a case-control sample collection that could be used as a resource for other projects as well. Although there was a significant overlap with the sample collection used for the case-control study presented in chapter 3, some of the samples could not be used again due to a lack of DNA while additional samples were added. The new study contained 114 more cases. In order to analyse the clinically distinct phenotypes more clearly, most of the cases used for the new study had either cerebral malaria or severe malarial anaemia alone. While slightly less controls were used all samples now were from subjects belonging to either the Mandinka, Jola, Fula or Wollof ethnic group. This setup would allow rapid screening of a large number of markers.
The analysis of haplotype blocks had indicated that the *IFNGR1*-56 and *IFNGR1*-470 are sufficient to describe 88% of all haplotypes found in the *IFNGR1* promoter using the SNPs available. For further association studies in other diseases it might therefore be sufficient to initially only type these two markers. However, in the case of severe malaria there is already some indication of an association at this locus, therefore all available markers were

typed to further describe and dissect the haplotypes at this locus. It was nevertheless decided to not type *IFNGR1*+95 as it is in complete LD and has the same allele frequency as *IFNGR1*-56 and therefore would not add any additional information. Furthermore *IFNGR1*-270 was not typed due to its low allele frequency. As mentioned in section 5.1.3 the marker *IFNGR1*-1804 could not be successfully typed using mass spectrometry. In the *IL22BP* and *IL20RA* promoter regions two markers were used respectively as defined by the haplotype block analysis. One further marker was used in the intergenic region between *IL20RA* and *IL22BP*.

6.2.1 Subjects

Children with cerebral malaria (CM) and/or severe malaria anaemia were prospectively recruited at the Royal Victoria Hospital, Banjul, The Gambia. Cerebral malaria (CM) was defined as a Blantyre coma score of 2 or less (Molyneux et al., 1989) persisting for more than 30min after any convulsions had ceased in a child with *Plasmodium falciparum* parasites on thick blood film and no evidence of meningitis or any other cause of coma. Severe malaria anaemia (SA) was defined as a PCV less than 15% in a child with *P. falciparum* parasites on thick blood film. Out of a total of 676 cases, 13 had CM plus SA, 400 had CM alone, and 263 had SA alone. 306 were Mandinka, 159 Jola, 104 Fula, 107 Wollof.

The average coma score was 1.5 in the CM patients and 4.5 in the SA patients. The average PCV was 29.6% in the CM group and 12.7% in the SA group. Out of the 400 cases with CM 122 died during hospital admission while out of 263 cases with SA 9 patients died. There were 194 female and 205 male patients in the CM group and 126 female and 137 male patients in the SA group.

For case-control analysis, the population controls were 459 umbilical cord blood samples collected at Gambian health clinics: 174 were Mandinka, 84 Jola, 117 Fula, 84 Wollof.

This sample size would enable detection of an effect of a discrete trait with

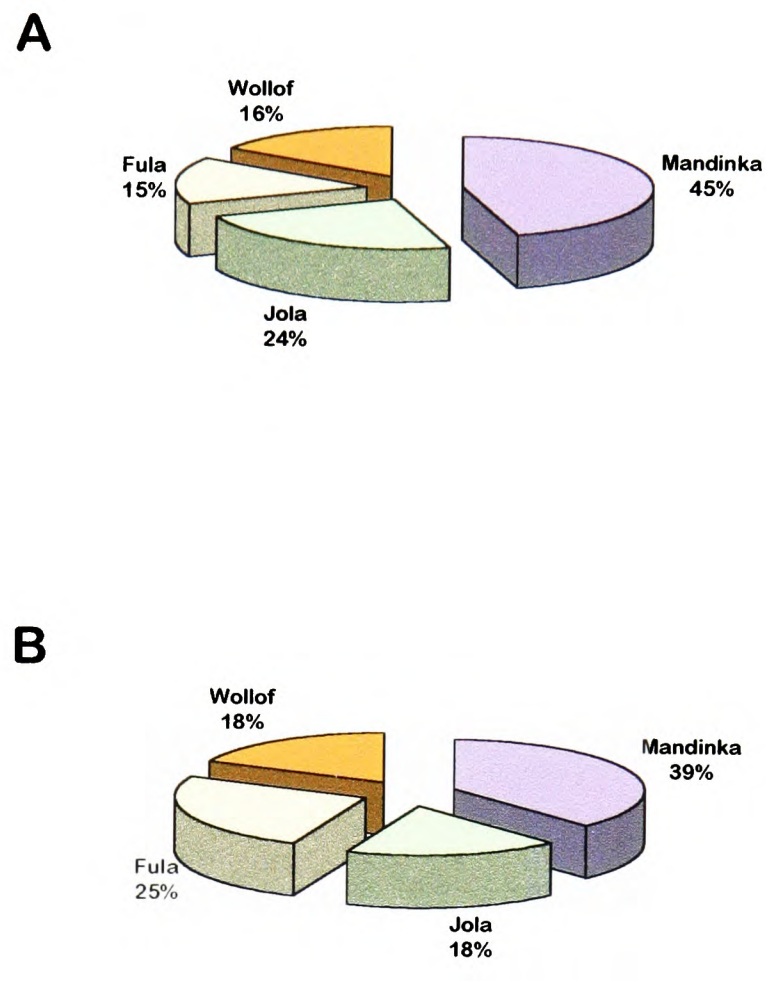


Figure 6.1 Distribution of ethnic groups amongst cases (**A**) and controls (**B**).

a high risk allele frequency (A) of 10% with a relative genotypic risk for AA of 2.0 with 80% power, $p<0.05$.

6.2.2 Genotyping

Case and control samples were plated out together on 384 well plates (AB-gene) and genotyping was performed by mass spectrometry as described before. All typings were performed on the same day with the same batch of reagents.

6.3 Results

6.3.1 Investigation of Association of Individual Markers

This study investigated the association between alleles and susceptibility to severe malaria. Tables 6.1 to 6.11 show the genotypes of individuals with cerebral or severe malarial anaemia or cord blood controls for the 11 markers typed on chromosomes 6q23. Discrepancies in sample size are due to genotyping failures. Within the the total study population there were two associations with severe disease observed. While having a low allele frequency, the *IFNGR1*-611G allele was found to be associated with protection against severe malaria (p=0.03; OR=0.49 (0.25-0.95 95%CI)), and against cerebral malaria in particular (p=0.04; OR=2.32 (1.02-5.40 95%CI)). The frequency of the G allele was 3% in the control group while it was just over 1% in the cerebral malaria group. The *IL20RA*-1620C allele was weakly associated with protection against severe malarial anaemia with an allele frequency of 14% in the controls group and 10% in the case group (p=0.05; OR=0.69 (0.48-1.00)). The *IFNGR1*-1882A allele was only detected once in the whole study population. Out of the 306 Mandinka patients with severe malaria 184 had cerebral malaria and 122 severe anaemia. As found in the previous study the *IFNGR1*-470ddel was associated with protection against severe malaria in this ethnic group (p=0.01; OR=0.56 (0.36-0.88 95%CI), i.e. cerebral malaria (p=0.02; OR 0.54 (0.32-0.92)) and severe anaemia (p=0.07; OR=0.58 (0.32-1.04 95%CI)). Statistical significance is hereby based on *uncorrected* values (for Bonferroni correction see section 3.2.6). This result remained significant after correcting for comparison of four different ethnic groups. The threshold for significance after this correction would be $p \leqslant 0.013$. However, if additionally all markers typed were taken into account the threshold for significance after correction would be $p \leqslant 0.003$ to bring the alpha level overall back to 0.05.

Table 6.1 IFNGR1-1882 (6:144241441); all ethnic groups

	Genotypes				Allele 1 vs. 2	
	aa	ag	gg	HWE	p	OR
cord blood controls	0	1	365	0.98		
cerebral malaria only	0	0	356	NA	NA	NA
severe anaemia only	0	0	222	NA	NA	NA
total no. patients	0	0	586	NA	NA	NA

Unlike previous observations there was no significant association for the *IFNGR1*-56 with severe malaria in Mandinka. There was a trend for the *IFNGR1*-56T allele for susceptibility with cerebral malaria (p=0.12; OR=1.29 (0.94-1.77 95%CI)) which was slightly stronger for TT homozygotes (p=0.07; OR=1.60 (0.96-2.68 95%)). While there were slightly less heterozygotes in the cerebral malaria group than expected this was not statistically significant with p=0.07 testing for deviation from Hardy-Weinberg Equilibrium. In the subgroup of those cerebral malaria Mandinka patients who died of their disease there was a significant reduction (p=0.05; OR 0.47 (0.22-1.00 95%CI)) of heterozygotes (14 out of 45) compared to the control group (79 out of 161) similar to what had been found previously. However, this was not significant after applying the Bonferroni correction. Figures 6.2 and 6.3 give an overview of the p values and odds ratios found for all markers.

Table 6.2 IFNGR1-611 (6:144240170); all ethnic groups

	Genotypes				Allele 1 vs. 2	
	aa	ag	gg	HWE	p	OR
cord blood controls	395	22	1	0.25		
cerebral malaria only	348	9	0	0.81	0.04	2.32
severe anaemia only	225	7	0	0.82	0.18	1.93
total no. patients	580	17	0	0.72	0.03	2.05

Table 6.3 IFNGR1-470 (6:144240029); all ethnic groups

	Genotypes				Allele 1 vs. 2	
	dd	id	ii	HWE	p	OR
cord blood controls	3	93	346	0.22		
cerebral malaria only	4	63	316	0.67	0.23	0.81
severe anaemia only	2	47	203	0.69	0.59	0.89
total no. patients	6	113	528	0.99	0.28	0.85

Table 6.4 IFNGR1-56 (6:144239615); all ethnic groups

	Genotypes				Allele 1 vs. 2	
	cc	tc	tt	HWE	p	OR
cord blood controls	107	194	116	0.16		
cerebral malaria only	102	160	108	0.01	0.96	1.01
severe anaemia only	60	122	65	0.85	0.97	1.00
total no. patients	165	288	177	0.03	0.99	1.01

Table 6.5 IFNGR1+130 (6:144239429); all ethnic groups

	Genotypes				Allele 1 vs. 2	
	aa	ga	gg	HWE	p	OR
cord blood controls	1	43	375	0.84		
cerebral malaria only	0	45	326	0.21	0.63	1.14
severe anaemia only	0	23	219	0.05	0.72	0.88
total no. patients	0	70	555	0.14	0.90	1.05

Table 6.6 IFNGR1+20876 (6:144218683); all ethnic groups

	Genotypes				Allele 1 vs. 2	
	aa	ac	cc	HWE	p	OR
cord blood controls	419	26	0	0.53		
cerebral malaria only	375	15	1	0.05	0.42	1.35
severe anaemia only	236	17	0	0.58	0.77	0.87
total no. patients	621	35	1	0.49	0.99	1.04

Table 6.7 IL22BP-1119 (6:144194867); all ethnic groups

	Genotypes				Allele 1 vs. 2	
	aa	ag	gg	HWE	p	OR
cord blood controls	245	122	18	0.58		
cerebral malaria only	207	101	21	0.08	0.62	0.93
severe anaemia only	140	67	17	0.78	0.45	0.89
total no. patients	355	173	38	0.01	0.48	0.92

Table 6.8 IL22BP-1054 (6:144194802); all ethnic groups

	Genotypes				Allele 1 vs. 2	
	gg	tg	tt	HWE	p	OR
cord blood controls	26	134	275	0.08		
cerebral malaria only	24	119	236	0.09	0.80	1.04
severe anaemia only	11	78	160	0.70	0.62	0.92
total no. patients	35	203	403	0.16	0.99	0.99

Table 6.9 6:144153554; all ethnic groups

	Genotypes				Allele 1 vs. 2	
	aa	ag	gg	HWE	p	OR
cord blood controls	179	184	37	0.29		
cerebral malaria only	172	160	39	0.84	0.99	1.01
severe anaemia only	118	89	33	0.02	0.96	1.00
total no. patients	296	252	74	0.07	0.99	1.00

Table 6.10 IL20RA-1620 (6:144066816); all ethnic groups

	Genotypes				Allele 1 vs. 2	
	cc	ct	tt	HWE	p	OR
cord blood controls	11	97	314	0.29		
cerebral malaria only	6	85	281	0.88	0.59	0.91
severe anaemia only	3	43	195	0.72	0.05	0.69
total no. patients	9	137	479	0.82	0.29	0.86

Table 6.11 IL20RA-310 (6:144065506); all ethnic groups

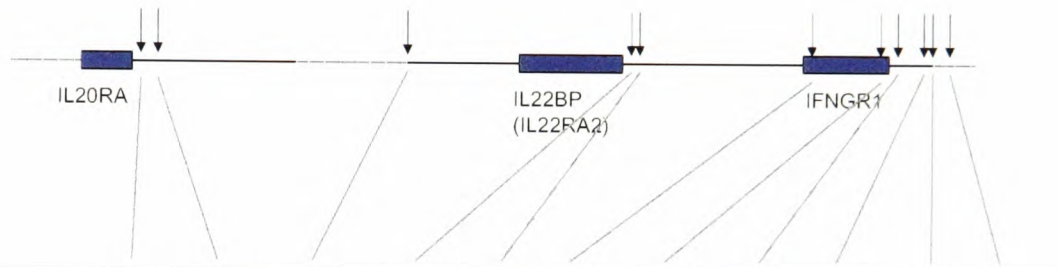
	Genotypes				Allele 1 vs. 2	
	cc	tc	tt	HWE	p	OR
cord blood controls	342	96	10	0.30		
cerebral malaria only	301	84	6	0.96	0.73	1.06
severe anaemia only	204	43	10	0.00	0.77	1.06
total no. patients	515	130	16	0.03	0.68	1.06

Table 6.12 IFNGR1-470; Mandinka

	Genotypes				Allele 1 vs. 2	
	dd	id	ii	HWE	p	OR
cord blood controls	1	43	123	0.18		
cerebral malaria only	1	26	152	0.92	0.02	0.54
severe anaemia only	0	20	101	0.32	0.07	0.58
total no. patients	1	46	253	0.47	0.01	0.56

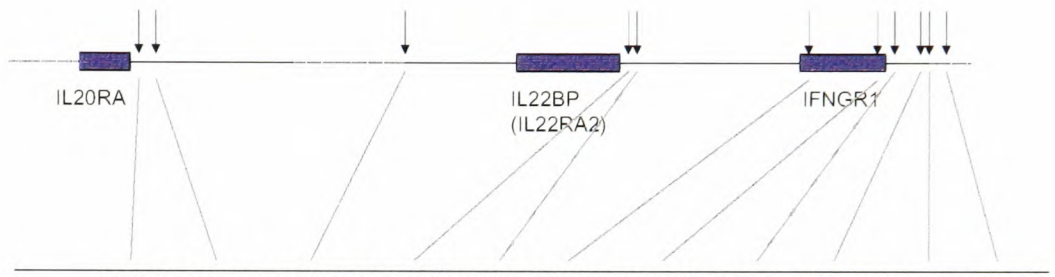
Table 6.13 IFNGR1-56; Mandinka

	Genotypes				Allele 1 vs. 2	
	cc	tc	tt	HWE	p	OR
cord blood controls	44	79	38	0.83		
cerebral malaria only	41	72	56	0.07	0.12	0.78
severe anaemia only	26	66	24	0.14	0.88	0.96
total no. patients	67	138	80	0.62	0.26	0.85



	6:144065506		6:144086816		6:144153554		6:144194802		6:144194867		6:144218683		IFNGR1+130		IFNGR1-56		IFNGR1-470		IFNGR1-611		IFNGR1-1882	
	p (1vs 2)	*OR*	*p (1vs 2)*	*OR*	*p (1vs 2)*	*OR*	*p (1vs 2)*	*OR*	*p (1vs 2)*	*OR*	*p (1vs 2)*	*OR*	*p (1vs 2)*	*OR*	*p (1vs 2)*	*OR*	*p (1vs 2)*	*OR*	*p (1vs 2)*	*OR*	*p (1vs 2)*	*OR*
CM	0.73	1.06	0.59	0.91	0.99	1.01	0.80	1.04	0.62	0.93	0.42	1.35	0.63	1.14	0.96	1.01	0.23	0.81	0.04	2.32	NA	NA
SA	0.77	1.06	0.047	0.69	0.96	1.00	0.62	0.92	0.45	0.89	0.77	0.87	0.72	0.88	0.97	1.00	0.59	0.89	0.18	1.93	NA	NA
all cases	0.68	1.06	0.29	0.86	0.99	1.00	0.99	0.99	0.48	0.92	0.99	1.04	0.90	1.05	0.99	1.01	0.28	0.85	0.03	2.05	NA	NA

Figure 6.2 Case control study 6q23 SNPs for all ethnic groups. P-values are Yates corrected values and are based on comparing the numbers of allele 1 and allele 2 in control and case groups. Odds ratios (OR) are given. A total of 459 cord blood controls, 400 cases with cerebral malaria only and 263 cases with severe malarial anaemia were analysed.



	6:144065506		6:144066816		6:144153554		6:144194802		6:144194867		6:144218683		IFNGR1-130		IFNGR1-56		IFNGR1-470		IFNGR1-611		IFNGR1-1882	
	p (1vs.2)	OR	p (1vs.2)	OR	p (1vs.2)	OR	p (1vs.2)	OR	p (1vs.2)	OR	p (1vs.2)	OR	p (1vs.2)	OR	p (1vs.2)	OR	p (1vs.2)	OR	p (1vs.2)	OR	p (1vs.2)	OR
CM	0.09	1.55	0.84	0.93	0.72	1.08	0.65	0.91	0.61	1.14	0.89	1.21	0.90	1.01	0.12	0.78	0.02	0.54	0.19	2.3	NA	NA
SA	0.72	1.13	0.77	0.89	0.42	1.18	0.38	0.82	0.26	0.76	0.43	0.62	0.79		0.88	0.96	0.07	0.58	0.84	1.26	NA	NA
all cases	0.18	1.35	0.75	0.91	0.50	1.12	0.44	0.87	0.87	0.96	0.92	0.88	0.95	0.94	0.26	0.85	0.01	0.56	0.29	1.73	NA	NA

Figure 6.3 Case control study 6q23 SNPs for Mandinka. P-values are Yates corrected values and are based on comparing the numbers of allele 1 and allele 2 in control and case groups. Odds ratios (OR) are given. A total of 174 cord blood controls, 184 cases with cerebral malaria only and 122 cases with severe malarial anaemia were analysed.

6.3.2 Haplotype Analysis

Haplotypes were constructed for the haplotype blocks as determined in the previous chapter. Table 6.14 shows the haplotypes found using 6 markers in the *IFNGR1* gene, 4 of which in the promoter, one in the first intron and one in the last exon. As described earlier almost 90% of all haplotypes are determined by just the *IFNGR1*-56 and *IFNGR1*-470 polymorphisms. The *IFNGR1*-470ddel lies on the background of the *IFNGR1*-56C as shown earlier. The *IFNGR1*+130A allele lies on the background of the *IFNGR1*-56C allele. The C allele of the the exonic marker 6:144218683 which is located about 21kb downstream has an allele frequency of about 3% and always occurs on the background of *IFNGR1*-56C allele. *The IFNGR1*-611G alleles lies on the background of the *IFNGR1*-56T allele and therefore defines this haplotype. As seen in the analysis of individual markers the haplotype bearing the *IFNGR1*-611G allele is associated with protection against severe malaria (p=0.02; OR=0.44 (0.22-0.88 95%CI)). In Mandinka the only haplotype bearing the *IFNGR1*-470ddel (haplotype 212121 in table 6.15) was associated with protection against severe malaria (p=0.02; OR=0.59 (0.37-0.92 95%CI)). However, this was not significant after correction for multiple comparisons (four ethnic groups; level of significance $p < 0.12$). This haplotype consists of the alleles *IFNGR1*-1882G, -611A, -470ddel, -56C, +130G and +20876A. The haplotype bearing the *IFNGR1*-56T allele (haplotype 211221) was overrepresented in Mandinka patients with cerebral malaria (p=0.05; OR=1.36 (1.00-1.84 95%CI)) which was dose dependent as the effect was stronger in homozygotes for this haplotype. Fifty three out 184 Mandinka cerebral malaria patients were homozygote for this haplotype while there were only 32 out 174 in the control group (p=0.03; OR=1.80 (1.06-3.05 95%CI)).

For the haplotype analysis of *IL22BP* three SNPs were used, *IL22BP*-1119 and -1054 and a SNP located between *IL22BP* and *IL20RA*, 6:144153554. These form five major haplotypes. The *IL22*-1119G allele which has about 21% allele frequency always occurs on the background of the *IL22BP*-1054T

allele (79%) and the A allele of 6:144153554 (68%). As a sign of recombination all combinations of alleles occurred for *IL22BP*-1054 and 6:144153554, a marker 40kb away. Surprisingly in Mandinka the control group has a much higher frequency of the '112' haplotype (18.1%) than in all of the case groups (11.9-13.0%) and is thereby protective against severe malaria (p=0.04; OR=0.67 (0.46-0.98 95%CI)) (table 6.17). This difference was not statistically significant, however after applying the Bonferroni correction. Those Mandinka patients who were carrying the '221' haplotype were more likely to develop severe malarial anaemia (24.2%) than cerebral malaria (16.3%) (p=0.02; OR=0.61 (0.40-0.93 95%CI)).
In *IL20RA* two markers were typed in this case control analysis, *IL20RA*-1620 (6:144066816) and *IL20RA*-310 (6:144065506). The *IL20RA*-1620C allele always occurred on the background of the *IL20RA*-310C allele. A reduction of the *IL20RA*-1620C allele in the severe anaemia group with borderline significance had been observed. This was reflected by a trend for showing fewer '11' haplotypes in the severe anaemia group (10.5%) than in controls (13.7%).

Haplotypes were also constructed using all available 11 markers spanning the entire 176kb. The most common haplotype with $f \geq 0.05$ in at least one of the groups are presented in table 6.19 for all ethnic groups and table 6.20 for Mandinka. The six haplotypes shown account for 55-59% of all haplotypes found. On comparison of cases and controls for the abundance of those six haplotypes, haplotype 2 was found to be significantly underrepesented in the cerebral malaria group. Haplotype 6 on the other hand which differs from haplotype at 2 positions (*IL22BP*-1119 and 6:144153554) was significantly overrepresented in this group. Furthermore, those individuals with severe malaria who carried haplotype 2 were significantly more likely to develop severe malarial anaemia than cerebral malaria (p=0.0002; OR=0.50 (0.34-0.73 95%CI)). Both haplotype 2 and 6 are on the background of the *IFNGR1*-56T allele. In Mandinka haplotype 5 which differs by one allele from haplotype 2 (*IL22BP*-1119) is overrepresented in the cerebral malaria group and underrepresented in the severe malarial anaemia group (p=0.02;

Table 6.14 All ethnic groups. Haplotypes of *IFNGR1* using the following markers: *IFNGR1*-1882 (A=1; G=2), *IFNGR1*-611 (A=1; G=2), *IFNGR1*-470 (ins=1; ddel=2), *IFNGR1*-56 (C=1; T=2), *IFNGR1*+130 (A=1; G=2) and *IFNGR1*+20876 (6:144218683) (A=1; C=2).

HAPLOTYPES	CONTROLS	ALL CASES	CM	SA
211221	441 (48.0%)	657 (48.6%)	393 (49.1%)	256 (48.7%)
211121	283 (30.8%)	444 (32.8%)	262 (32.8%)	172 (32.7%)
212121	97 (10.6%)	125 (9.2%)	72 (9.0%)	51 (9.7%)
211111	44 (4.8%)	70 (5.2%)	46 (5.8%)	23 (4.4%)
211122	26 (2.8%)	30 (2.2%)	14 (1.8%)	14 (2.7%)
221221	23 (2.5%)	15 (1.1%)	8 (1.0%)	7 (1.3%)
212221	2 (0.2%)	0 (0%)	0 (0%)	0 (0%)
111121	1 (0.1%)	0 (0%)	0 (0%)	0 (0%)
221111	1 (0.1%)	0 (0%)	0 (0%)	0 (0%)
211222	0 (0%)	7 (0.5%)	3 (0.4%)	3 (0.6%)
221222	0 (0%)	2 (0.1%)	1 (0.1%)	0 (0%)
211211	0 (0%)	1 (0.1%)	1 (0.1%)	0 (0%)
221121	0 (0%)	1 (0.1%)	0 (0%)	0 (0%)
total	**918**	**1352**	**800**	**526**

Table 6.15 Mandinka. Haplotypes of *IFNGR1* using the following markers: *IFNGR1*-1882 (A=1; G=2), *IFNGR1*-611 (A=1; G=2), *IFNGR1*-470 (ins=1; ddel=2), *IFNGR1*-56 (C=1; T=2), *IFNGR1*+130 (A=1; G=2) and *IFNGR1*+20876 (6:144218683) (A=1; C=2).

HAPLOTYPES	CONTROLS	ALL CASES	CM	SA
211221	158 (45.4%)	303 (49.3%)	195 (53.0%)	112 (45.9%)
211121	102 (29.3%)	196 (32.0%)	109 (29.6%)	83 (34.0%)
212121	45 (12.9%)	49 (8.0%)	29 (7.9%)	20 (8.2%)
211111	21 (6.0%)	36 (5.9%)	22 (6.0%)	13 (5.3%)
211122	9 (2.6%)	15 (2.5%)	7 (1.9%)	8 (3.3%)
221221	11 (3.2%)	11 (1.8%)	5 (1.4%)	6 (2.5%)
111121	1 (0.3%)	0 (0%)	0 (0%)	0 (0%)
211222	0 (0%)	3 (0.5%)	1 (0.3%)	2 (0.8%)
221121	1 (0.3%)	0 (0%)	0 (0%)	0 (0%)
total	**348**	**612**	**368**	**244**

Table 6.16 All ethnic groups. Haplotypes of *IL22BP* using the following markers: *IL22BP*-1119 (A=1; G=2), *IL22BP*-1054 (G=1; T=2) and 6:144153554 (A=1; G=2)).

HAPLOTYPES	CONTROLS	ALL CASES	CM	SA
1 1 2	127 (13.8%)	168 (12.4%)	106 (13.3%)	58 (11.0%)
1 2 1	397 (43.2%)	544 (40.2%)	332 (41.5%)	207 (39.4%)
1 2 2	164 (17.9%)	265 (19.6%)	147 (18.4%)	108 (20.5%)
2 2 1	166 (18.1%)	260 (19.2%)	145 (18.1%)	104 (19.8%)
1 1 1	62 (6.8%)	107 (7.9%)	66 (8.3%)	45 (8.6%)
2 2 2	2 (0.2%)	8 (0.6%)	4 (0.1%)	4 (0.1%)
total	**918**	**1352**	**800**	**526**

Table 6.17 Mandinka. Haplotypes of *IL22BP* using the following markers: *IL22BP*-1119 (A=1; G=2), *IL22BP*-1054 (G=1; T=2) and 6:144153554 (A=1; G=2)).

HAPLOTYPES	CONTROLS	ALL CASES	CM	SA
1 1 2	63 (18.1%)	79 (12.9%)	48 (13.0%)	29 (11.9%)
1 2 1	140 (40.2%)	242 (39.5%)	153 (41.6%)	90 (36.9%)
1 2 2	58 (16.7%)	121 (19.8%)	75 (20.4%)	47 (19.3%)
2 2 1	67 (19.3%)	120 (19.6%)	60 (16.3%)	59 (24.2%)
1 1 1	19 (5.5%)	48 (7.8%)	30 (8.2%)	19 (7.8%)
2 2 2	1 (0.3%)	2 (0.3%)	2 (0.5%)	0 (0.0%)
total	**348**	**612**	**368**	**244**

Table 6.18 All ethnic groups. Haplotypes of *IL20RA* using the following markers: *IL20RA*-1620 (C=1; T=2) and *IL20RA*-310 (C=2; T=2).

HAPLOTYPES	CONTROLS	ALL CASES	CM	SA
2 1	675 (73.5%)	1030 (76.2%)	598 (74.8%)	408 (77.6%)
1 1	126 (13.7%)	160 (11.8%)	105 (13.1%)	55 (10.5%)
2 2	117 (12.7%)	162 (12.0%)	97 (12.1%)	63 (12.0%)
total	**918**	**1352**	**800**	**526**

OR=1.77 (1.08-2.91 95%CI); and p=0.007; OR=0.36 (0.17-0.78 95%CI) respectively). Furthermore, those Mandinka patients who carried haplotype 5 were significantly more likely to develop cerebral malaria than severe malarial anaemia (p=0.000002; OR=4.86 (2.36-10.22 95%CI)). This was statistically significant even after allowing for 11 comparisons (four ethnic groups, six haplotypes and cerebral malaria versus severe anaemia). Haplotype 4 carries the *IFNGR1*-470ddel. It was overrepresented in the control group (p=0.008; OR=0.52 (0.31-0.85 95%CI)). However, if one corrects for comparing four different ethnic groups and six different haplotypes the threshold of significance would be at $p < 0.005$.

As has been noted in the previous chapter long ranging LD was observed from the marker *IFNGR1*-470. Haplotypes were constructed using all 11 markers spanning 176kb in the 6q23 region. It was found that the *IFNGR1*-470ddel allele which has a frequency of about 10-12% was predominantly observed on one haplotype only. This haplotype accounts for more than 86% of all haplotypes bearing the *IFNGR1*-470ddel allele. This is a further indicator that this marker is in high linkage disequilibrium with markers at a considerable physical distance. In order to establish how this relates to other markers in the same region the haplotypic structure of two markers with similar allele frequencies were used for comparison. *IL20RA*-1620 and -310 both have a minor allele frequency of about 12%. The frequency distribution of haplotypes that carry each of these three markers are shown in Figure 6.4. Both the *IL20RA*-1620C allele and the *IL20RA*-310T allele are distributed across several haplotypes with the major haplotypes accounting for just 29% and 30% respectively in cord blood controls. The structure of haplotypes bearing the *IFNGR1*-470ddel is highly conserved for all *IFNGR1* markers in addition to the marker *IL22BP*-1119, a region spanning almost 47kb.

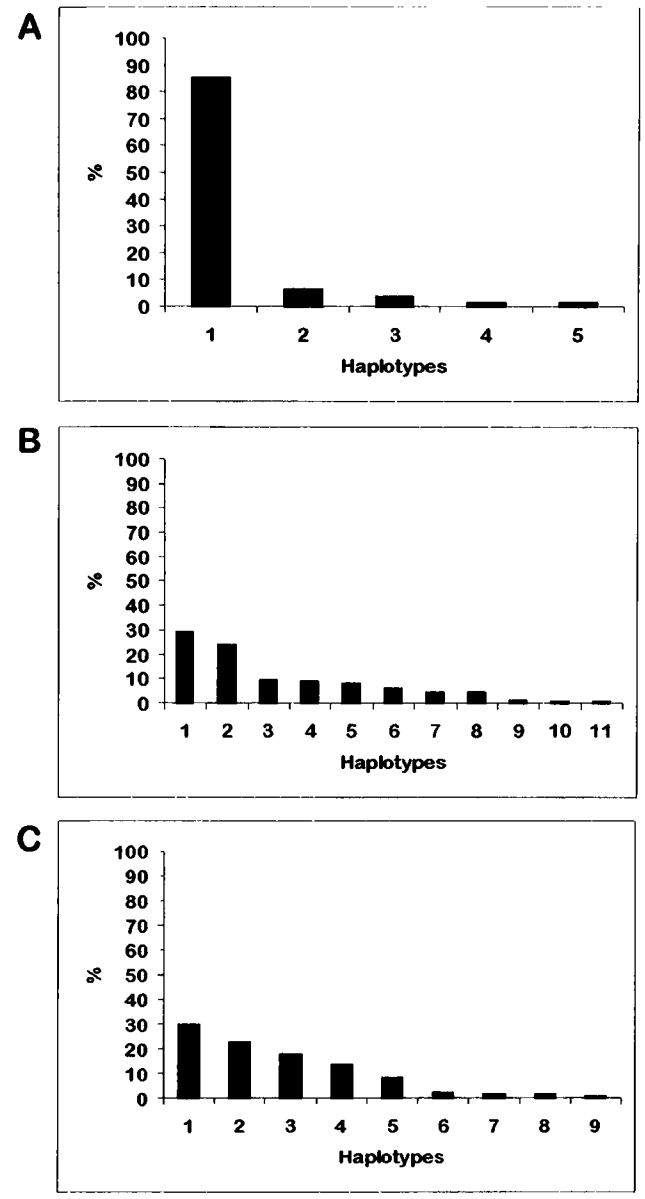


Figure 6.4 Haplotype frequency distribution in the cord blood control sample collection of extended haplotypes bearing the following alleles: *IFNGR1*-470ddel (**A**), *IL20RA*-1620C (**B**) and *IL20RA*-310T (**C**).

6.4 Discussion

The previous case control study had yielded two potentially interesting associations for *IFNGR1*-56 and -470. This additional study extended the area under investigation by using markers in neighbouring genes. This approach allowed to test markers in two further immune genes for association with disease and enabled the construction of longrange haplotypes. It was hoped that this might facilitate localising the association with severe malaria further. The markers had been chosen by constructing haplotype blocks and calculating the minimum number of markers needed in order to achieve the maximum entropy, i.e. information about the haplotype structure of the markers given. It was found that the low frequency allele *IFNGR1*-611G was associated with protection against severe disease in all ethnic groups, i.e. individuals carrying the G allele were half as likely two develop severe malaria. This allele is common in Caucasians but must have arisen before humans left Africa. Its significant increase in frequency in the Caucasian population is of note. While this could be simply due to genetic drift it might equally represent the sign of a selective advantage. Transcription factor prediction analysis using TRANSFAC suggests that the G allele might disrupt a theoretical GATA1 binding site. But given its low allele frequency in Africans the G allele is in high linkage disequilibrium with most other *IFNGR1* markers. Only one *IFNGR1* haplotype carrying the *IFNGR1*-611G allele was observed and not surprisingly this haplotype was found to be protective against severe disease (OR=0.44). The *IFNGR1*-611G allele lies on the background of the *IFNGR1*-56T allele and is therefore on a different background as the *IFNGR1*-470ddel allele.

In the first case control study Mandinka individuals heterozygote for the *IFNGR1*-56 were protected against cerebral malaria. This association could not be replicated in the repeat study. Although there were less heterozygotes in the cerebral malaria group than expected this was not significant. However, in the subgroup who died as a consequence of cerebral malaria there was still a reduction of in the number of heterozygotes. There was a trend for the *IFNGR1*-56T allele to be associated with susceptibility to cerebral

malaria. The two genotyping methods for the *IFNGR1*-56, ligation detection reaction (LDR) and mass-spectrometry, were compared. The first case control study contained 174 Mandinka controls and 123 Mandinka with cerebral malaria. The second case control study contained 174 Mandinka controls and 184 Mandinka with cerebral malaria. There was an overlap of 152 controls and 92 cerebral malaria cases in the two studies, i.e. only half of the cerebral malaria cases in the second study had been used in the first study. It was found that the two genotyping methods showed matching results in 94% and 95% for cases and controls respectively. This was similar to what had been found by other members of the research group when comparing genotyping methods for other assays. It is therefore likely that the altered result is due to the fact that the second study used 92 cerebral malaria cases that had not been used previously.

An association of the *IFNGR1*-470ddel with protection against severe malaria in Mandinka was confirmed in this study. Individuals carrying the deletion were about half as likely to develop the severe form of the disease. This was statistically significant after correcting for comparison of four different ethnic groups, but not if corrected for the 11 different markers used. Therefore, one has to be cautious when rejecting the null hypothesis. These results are, however, of sufficient interest to justify further analysis the functional significance of the *IFNGR1*-56 and *IFNGR1*-470ddel polymorphisms.

The first association study had already indicated that the *IFNGR1* haplotype carrying the *IFNGR1*-470ddel allele was equally associated with protection against severe disease. With the additional markers in the two neighbouring genes haplotypes were now constructed comprising of a region of almost 180kb. Interestingly the *IFNGR1*-470ddel allele was still predominantly found on one haplotype only. This is surprising despite the fact that its allele frequency is only about 10%. While two different markers with a similar allele frequency in the same region fall onto multiple haplotypes there is hardly any sign of the extended *IFNGR1*-470ddel haplotype being broken up by recombination. This could be a signature of natural selection as markers that rise in frequency within a relatively short time span due

to a selective advantage tend to be in higher linkage with their surrounding genetic region. Reich et al. (2001) studied the decay of D' for 19 different genomic regions and found that in a European population the half-length of D' (the distance at which the average D' drops below 0.5) is typically 60kb, while in an African population the half-length of D' is 5kb (Reich et al., 2001). Other studies have also revealed lower levels of linkage disequilibrium in African populations compared with non-African populations (Tishkoff et al., 2001). Sabeti et al. (2002) make use of long-ranging LD by applying their method of extended haplotype homozygosity. This method is unfortunately of only limited help in this case as the *IFNGR1*-470ddel is not a high frequency allele and only few homozygotes would be expected in the sample set used. Furthermore, the haplotypic diversity of the core *IFNGR1* haplotype is limited with the *IFNGR1*-56T and the *IFNGR1*-470ddel alone accounting for almost 90% of all haplotypes at this locus. Nevertheless, a similar phenomenon to the one Sabeti et al. (2002) described for the *G6PD* locus might be present in the *IFNGR1* region with the core haplotype as defined by the *IFNGR1*-470ddel allele extending in fact at least 180kb. It is worth noting, though, that the *IFNGR1*-470ddel might be in greater LD than the two markers used for comparison due to a different genetic history. The other two markers might be much older than the *IFNGR1*-470ddel. Their frequency might now be decreasing due to genetic drift after they might have reached their *peak* frequency some time ago in history. The fact that both markers have high allele frequencies in Caucasians and must therefore have arisen before humans left Africa might add weight to this argument. The *IFNGR1*-470ddel on the other hand could neither be found in Caucasians nor in 400 samples from Vietnam (data not shown). It is therefore possible that it has arisen only recently which would explain its high LD with neighbouring markers. PHASE was used for the constructing of extended haplotypes. It must be noted that it is not clear yet as to how well this method performs over large distances although Stephens et al. (2001b) found their program to perform well in the presence of moderate amounts of recombination.

Nevertheless, this result might be encouraging as there was not only an association with this marker but also evidence of long-ranging LD. If on the other the hand *IFNGR1*-470ddel was not the functional allele itself and just a marker for another allele somewhere on this haplotype it might be hard to localise the association any further. It is therefore crucial to undertake functional analysis of this locus and investigate whether the alleles found thus far might be involved in altering expression of the genes.

Kosh 2003

Table 6.19 All ethnic groups. Long range haplotypes of chromosome 6q23 using the following 11 markers spanning a total of 176kb: *IFNGR1*-1882 (A=1; G=2), *IFNGR1*-611 (A=1; G=2), *IFNGR1*-470 (ins=1; ddel=2), *IFNGR1*-56 (C=1; T=2), *IFNGR1*+130 (A=1; G=2), *IFNGR1*+20876 (6:144218683) (A=1; C=2), *IL22BP*-1119 (A=1; G=2), *IL22BP*-1054 (G=1; T=2), 6:144153554 (A=1; G=2), *IL20RA*-1620 (C=1; T=2) and *IL20RA*-310 (C=2; T=2). Only haplotypes with a frequency of $f \geq 0.05$ in at least one of the groups are shown. An asterisk marks where frequencies in cases are significantly different from controls. CM = cerebral malaria; SA = severe malarial anaemia.

Identifier	Haplotypes	controls	all cases	CM	SA
1	21112112121	119 (13.0%)	195 (14.4%)	120 (15.0%)	81 (15.4%)
2	21122122121	105 (11.4%)	124 (9.2%)	67 (8.4%)*	81 (15.4%)
3	21112122121	29 (3.2%)	72 (5.3%)	32 (4.0%)	31 (5.9%)
4	21212111221	89 (9.7%)	103 (7.6%)	57 (7.1%)	42 (8.0%)
5	21122112121	105 (11.4%)	136 (10.1%)	92 (11.5%)	53 (10.1%)
6	21122112221	68 (7.4%)	144 (10.7%)*	84 (10.5%)*	49 (9.3%)
total		**515 (56.1%)**	**774 (57.3)**	**452 (56.5%)**	**314 (59.7%)**

Table 6.20 Mandinka. Long range haplotypes of chromosome 6q23 using the following 11 markers spanning a total of 176kb: *IFNGR1*-1882 (A=1; G=2), *IFNGR1*-611 (A=1; G=2), *IFNGR1*-470 (ins=1; ddel=2), *IFNGR1*-56 (C=1; T=2), *IFNGR1*+130 (A=1; G=2), *IFNGR1*+20876 (6:144218683) (A=1; C=2), *IL22BP*-1119 (A=1; G=2), *IL22BP*-1054 (G=1; T=2), 6:144153554 (A=1; G=2), *IL20RA*-1620 (C=1; T=2) and *IL20RA*-310 (C=2; T=2). Only haplotypes with a frequency of $f \geq 0.05$ in at least one of the groups are shown. An asterisk marks where frequencies in cases are significantly different from controls. CM = cerebral malaria; SA = severe malarial anaemia.

IDENTIFIER	HAPLOTYPES	CONTROLS	ALL CASES	CM	SA
1	21112112121	47 (13.5%)	96 (15.7%)	54 (14.7%)	32 (13.1%)
2	21122122121	32 (9.2%)	55 (9.0%)	19 (5.2%)*	34 (13.9%)
3	21112122121	7 (2.0%)	28 (4.6%)	11 (3.0%)	22 (9.0%)*
4	21212111221	40 (11.5%)	43 (7.0%)*	26 (7.1%)	16 (6.6%)
5	21122112121	36 (10.3%)	64 (10.5%)	63 (17.1%)*	11 (4.5%)*
6	21122112221	28 (8.0%)	69 (11.3%)	42 (11.4%)	25 (10.2%)
total		**190 (54.5%)**	**355 (58.1)**	**215 (58.5%)**	**140 (57.3%)**

6.5 Summary

1. The association of *IFNGR1*-470 with severe malaria in Mandinka was replicated in this new case-control study using a different genotyping method.

2. The association for *IFNGR1*-56 heterozygosity in Mandinka, however, was not replicated. Although genotyping error cannot be entirely excluded, this might be due to a difference in the sample collections. Haplotype analysis, however, showed a significant association for the haplotype carrying the T allele with cerebral malaria.

3. In the overall group, the low frequency *IFNGR1*-611G allele was associated with protection against severe malaria.

4. Overall no strong association was found for individual markers with severe malaria.

5. Several long range haplotypes showed significant association with disease.

6. The *IFNGR1*-470ddel allele is in strong LD with markers which are a considerable distance away.

7. The *IFNGR1*-470ddel allele was predominantly observed on one haplotype only spanning the entire region of 176kb. This could represent a signature of natural selection.

8. The *IL20RA*-1620C allele was weakly associated with protection against severe malarial anaemia.

9. None of the *IL22BP* SNPs and haplotypes were associated with susceptibility to disease.

10. None of the above results would be statistically significant if full corrections were made for multiple comparisons, and so should be regarded as tentative rather than definitive associations.

Chapter 7

Functional Investigations of *IFNGR1* promoter polymorphisms

7.1 Introduction

Previous chapters have focussed on the genetic analysis of the of common variants, i.e. single nucleotide polymorphisms, in promoter regions of *IFNGR1*, *IL22BP* and *IL20RA*. These genetic regions were screened for common variants. These variants were then genotyped in cases with severe malaria and controls. Frequencies of either individual alleles or haplotypes were compared. While this approach might detect variants that might influence response to and outcome of a given disease, it can however not prove functionality of the respective variant site. This can lead to problems particularly in regions of high linkage disequilibrium (LD). Functional analysis is therefore crucial in advancing the knowledge about the effects a given polymorphism might have on a cellular system. For the associations found for the *IFNGR1* SNPs this seemed a logical way to progress.

1. The association studies presented in chapters 3 and 6 have indicated a possible link between the *IFNGR1*-470 polymorphism and susceptibility to severe malaria.

2. LD studies and haplotype analysis, however, failed to localise the effect as the *IFNGR1*-470ddel occurs on an extended haplotype spanning more than 176kb.

3. Similarly, weak associations were observed for *IFNGR1*-611 in the overall Gambian group and *IFNGR1*-56 in the Mandinka ethnic group.

4. The effects observed may be due to a functional variation of these sites themselves or of other markers located on the same haplotypes.

5. In order to establish whether these sites, and *IFNGR1*-470 in particular, have functional properties themselves, functional analysis seemed warranted.

7.1.1 Background

As outlined in chapter 1 IFNγ plays a central role in the host immune response. There is growing evidence that its action is at least in part determined by the expression of the its receptor, IFNγ receptor1 (IFNγR1) and IFNγ receptor2 (IFNγR2).

While there have been mulitple reports of cases with complete or partial IFNγR1 deficiency due to coding mutations of the *IFNGR1* gene (Newport et al., 1996; Jouanguy et al., 1996, 1997, 1999), little is known about how changes in expression of the non-mutated receptor may alter response to infection and other inflammatory stimuli. IFNγR1 has been shown to be expressed constitutively in many tissues.

Valente et al. (1992) have shown that most B lymphocytes are IFNγR positive, whereas T lymphocytes are negative. However, after *in vitro* activation, peripheral blood T cells become IFNγR+. Although the role of IFNγR1 has been best described in lymphatic cells and macrophages it has

been shown to be expressed in a large variety of cells such as endothelial cells, trophoblastic epithelium of the placenta, glandular cells of stomach, ileum and colon, lung alveolar cells, salivary duct cells, renal tubular cells, and endometrial mucosa cells, hepatocytes (Valente et al., 1992). This suggests that the role of interferon and its receptor is not limited to the lymphatic tissue alone.

Several studies have furthermore shown that IFNγR1 expression underlies significant regulation leading to different cellular responses. A change in the cell membrane expression of both chains greatly transforms the results of IFN-γ's influence on T lymphocyte behavior. Flow cytometry indicates that R1 and R2 can be expressed at low (20% positive cells) or high (from 20 to 100% positive cells) levels. When it interacts with T cells expressing high R1 and low R2 membrane levels, it promotes their proliferation (Novelli et al., 1994, 1996). By contrast, when high levels of both chains are expressed, it induces T cell apoptosis (Novelli et al., 1996, 1997). βgalactoside-binding protein, a negative regulator of the cell cycle, up-regulates the expression of both the alpha- and the beta-chains of the IFNγR on activated T lymphocyte membrane (Allione et al., 1998) which renders T lymphocytes sensitive to IFN-γ induced apoptosis.

The LPS activation of conventional and CD5+ B cells resulted in increased IFNγR1 expression and increased the ability of IFN-gamma to inhibit LPS-induced B cell differentiation correlated with increased IFNγR1 expression. Increased B cell IFNγR1 expression was detected 12 h after activation, and maximal IFNγR1 expression was observed at 24 h (Abed et al., 1994b,a).

Stimulation of human blood monocytes and the monocytic cell line, THP-1, by IL-1 or TNF-alpha leads to the up-regulation of IFN-γ receptors. The data suggest that the immunopotentiating effects of IL-1 and TNF-alpha are mediated in part by enhancing IFN-γ receptor expression on monocytes and macrophages. (Krakauer and Oppenheim, 1993).

In Raji, IM9, and THP-1 cells, IFN-γ had no effect on its receptor gene expression. In Raji and IM9 cells, PMA had no or few effects on the cell surface receptor number and no detectable effect on the receptor protein or on mRNA levels. In contrast, in THP-1 cells, PMA treatment induced a time and dose-dependent five- to sixfold increase of the cell surface receptors (Mao et al., 1990).

Nakagawa et al. (1989) demonstrated IFNγR1 to be present on normal human B-cells as well as four Epstein-Barr virus transformed B-cell lines. In this study activation of normal B-cells in vitro with Staphylococcus aureus Cowan Strain resulted in a decreased receptor number and density.

Tumor necrosis factor-α and IL-6 up-regulate *IFNGR1 gene* expression in human monocytic THP-1 cells by affecting transcriptional regulation and effecting mRNA stability respectively (Sanceau et al., 1992).

Despite intense studies of the general patterns of expression of the *IFNGR1* only little work has been done on the molecular basis of expressional regulation. The exon-intron organisation of the *IFNGR1* resembles the mouse *Ifngr1* (Raval et al., 1995), the human *IFNGR2* (Rhee et al., 1996), and the mouse *Ifngr2* genes (Ebensperger et al., 1996). All these genes have multiple transcription initiation sites preceded by many potential binding sites for different transcription factors including Sp1 and AP-2. The first paper to investigate the 5' flanking region of the gene was done by Raval et al. (1995). They investigated 480bp of DNA upstream from the ATG start codon of the mouse Ifngr1 gene. The 5' sequence was found to lack a TATA box. It was furthermore G+C rich and multiple putative SP1 sites and multiple transcription start points were found. These are characteristics of the promoters of other constitutively expressed genes (Dynan, 1986; Boisclair et al., 1993) including several other receptors with growth or immune regulatory activity, including the p55 tumor necrosis factor receptor (Kemper and Wallach, 1993), the interleukin 1 type-I receptor (Ye et al., 1993), epidermal growth factor (Ishii et al., 1985), nerve growth factor (Sehgal et al., 1988), insulin (Sibley et al., 1989) and thyrotropin (Ikuyama et al., 1992). Raval et al.

(1995) also suggested that the 5' flanking region of *IFNGR1* contained putative responsive elements for transcription factors AP2, NF1 and ATF/CREB that could potentially modulate expression of the *IFNGR1* gene.

Merlin et al. (1997) characterized 4kb of the 5' flanking region of the human *IFNGR1* gene. No TATA or CAAT boxes were found. Potential binding sites for Sp1, AP-1, AP-2, and CREB nuclear factors were identified. Similar to the mouse promoter multiple transcription initiation sites were found. The strongest sites were observed at positions around -21, -31, -37, -55, -61 and -75. In Cos-1 cells the smallest 5' region that still had full promoter activity was 692 bp in length.

Sakamoto and Taniguchi (2001) reported a TPA responsive element in the -128 to -109 base pair region. Induction of this element lead to enhancement of human leukocyte antigen-DRα expression (Sakamoto et al., 1999). They suggested that Sp1 is binding to this site.

7.1.2 Aim

Although studies have looked at theoretical predictions of transcription factor binding sites in the *IFNGR1* promoter, so far only few studies have investigated binding to specific sites in vitro. As presented earlier putative associations with susceptibility to malaria was found for some of the polymorphisms. Although clearly of interest this does not prove that any of those SNPs are functional and play a role in *IFNGR1* gene regulation. The aim of this study was to specifically focus on the polymorphic sites of the *IFNGR1* promoter and investigate their role in the context of *IFNGR1* gene expression in different cell types.

7.2 Results

7.2.1 *IFNGR1* mRNA expression

Expression of *IFNGR1* was initially investigated in five different cell lines. An EBV-immortalized B-cell line derived from Burkitt lymphoma cells (Raji), a B-cell line derived from myeloma cells (RPMI 8226), a lymphoblastic T-cell line (Jurkat), a monocytic line (MonoMac6), and an epithelial cell line derived form lung carcinoma cells (A549) were used to represent a range of different tissue in which IFNGR1 is expressed in vivo. In order to obtain a crude estimate in which cell lines *IFNGR1* gene expression could be induced, all cell lines were subjected to a range of different stimuli and mRNA expression measured by Northern blot hybridization. Four of the cell lines were stimulated with a range of different stimuli for 2, 24 and in some occasions for 48h. A fifth cell line, A549, was stimulated with TNFα over a time course of 48h. mRNA expression was determined by Northern blot hybridisation. All values were normalised by hybridisation with β-actin cDNA. There was significant induction of mRNA expression in Raji 24h after simulation with PMA and ionomycin (Figure 7.1). In the other B-cell line, RPMI8226, a less pronounced induction was seen after 48h of stimulation with PMA/ionomycin while all other stimuli did not induce *IFNGR1* mRNA expression (Figure 7.2). Both in Jurkat T-cells and MonoMac6 no induction of *IFNGR1* mRNA expression was observed at the investigated time points (Figure 7.3, (Figure 7.4)). The epithelial cell line, A549, was stimulated with TNF 10ng/ml over 0, 0.5, 2, 4, 8, 16, 24 and 48 hours. A moderate induction was seen after 8 hours that lasted at least until 48 hours (Figure 7.5).

7.2.2 Protein Binding to Polymorphic Sites of *IFNGR1*

Initially the 5' flanking region of the *IFNGR1* gene was analysed using transcription factor binding site prediction software, `TRANSFAC` (Wingender et al., 2000). This suggested the *IFNGR1*-611 site as a potential GATA binding

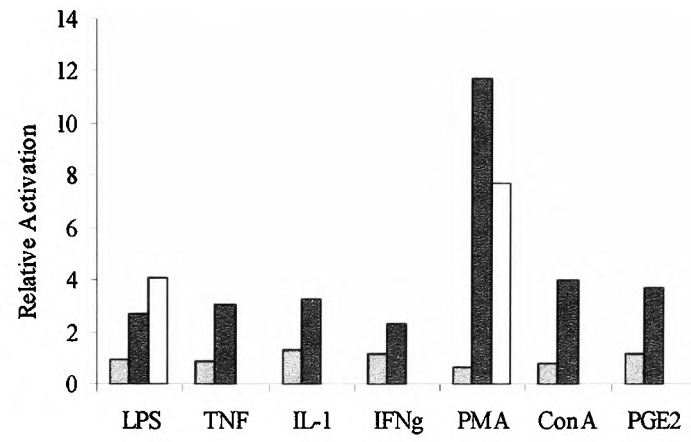


Figure 7.1 *IFNGR1* mRNA expression in Raji B-cells as determined by Northern Blot. Expression is shown as a relative value in comparison with non-activated cells. Light grey bars = 2h after stimulation; dark bars = 24h after stimulation; white bars = 48h after stimulation. Cells were stimulated with either 1μg/ml LPS, 40ng/ml TNF-α, 300U/ml IFNγ, 10ng/ml IL1β, 250nM PMA / 150nM Ionomycin, 1μg/ml PGE2 or 50μg/ml ConA.

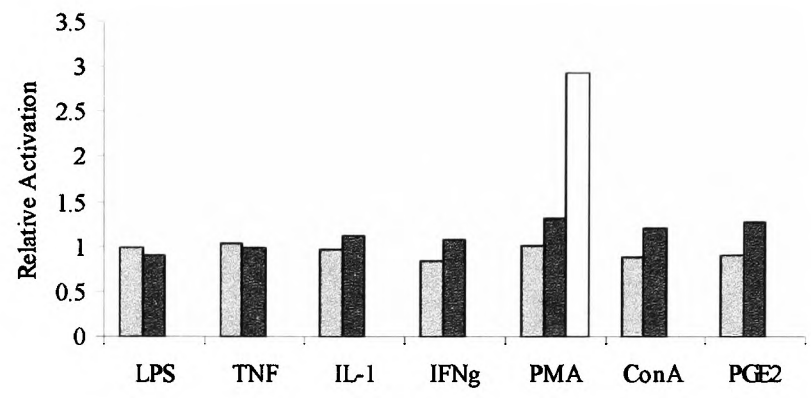


Figure 7.2 *IFNGR1* mRNA expression in RPMI8226 B-cells as determined by Northern Blot. Expression is shown as a relative value in comparison with non-activated cells. Light grey bars = 2h after stimulation; dark bars = 24h after stimulation; white bars = 48h after stimulation. Cells were stimulated with either 1μg/ml LPS, 40ng/ml TNF-α, 300U/ml IFNγ, 10ng/ml IL1β, 250nM PMA / 150nM Ionomycin, 1μg/ml PGE2 or 50μg/ml ConA.

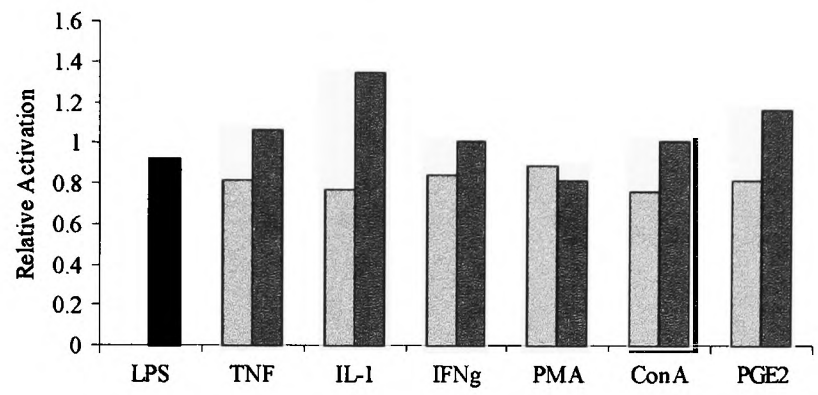


Figure 7.3 *IFNGR1* mRNA expression in Jurkat T-cells as determined by Northern Blot. Expression is shown as a relative value in comparison with non-activated cells. Light grey bars = 2h after stimulation; dark bars = 24h after stimulation; white bars = 48h after stimulation. Cells were stimulated with either 1μg/ml LPS, 40ng/ml TNF-α, 300U/ml IFNγ, 10ng/ml IL1β, 250nM PMA / 150nM Ionomycin, 1μg/ml PGE2 or 50μg/ml ConA.

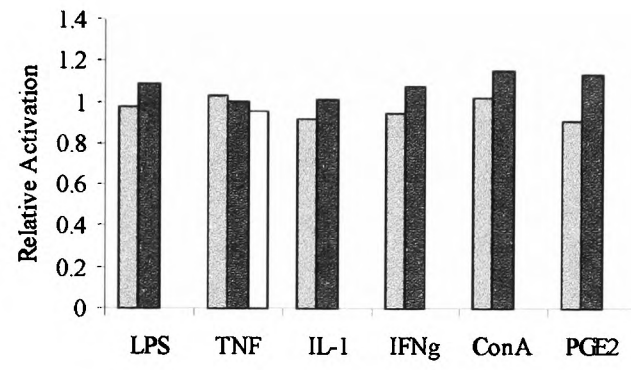


Figure 7.4 *IFNGR1* mRNA expression in MonoMac6 as determined by Northern Blot. Expression is shown as a relative value in comparison with non-activated cells. Light grey bars = 2h after stimulation; dark bars = 24h after stimulation; white bars = 48h after stimulation. Cells were stimulated with either 1μg/ml LPS, 40ng/ml TNF-α, 300U/ml IFNγ, 10ng/ml IL1β, 1μg/ml PGE2 or 50μg/ml ConA.

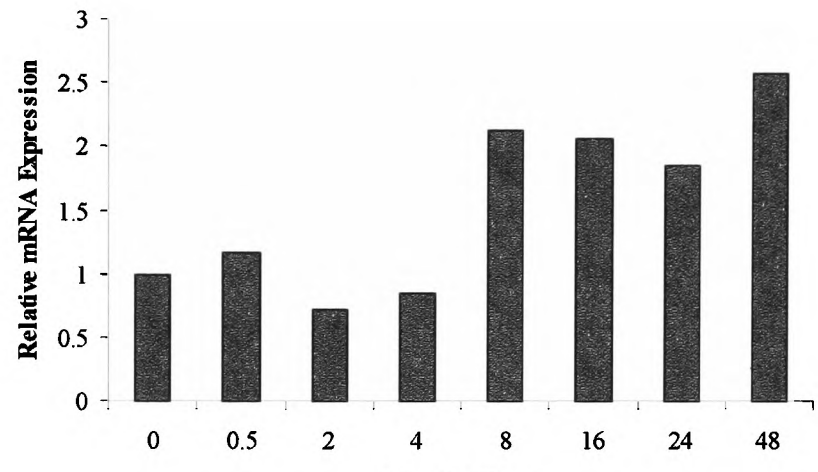


Figure 7.5 *IFNGR1* mRNA expression in A549 stimulated with TNF 40ng/ml as determined by Northern Blot. Expression is shown as a relative value in comparison with non-activated cells.

site. The site surrounding *IFNGR1*-56 resembles an AP4 binding site. It was furthermore predicted that STAT1 might bind to the *IFNGR1*-470ins allele while this binding would be abrogated in the variant containing a double deletion. To investigate the interactions of nuclear proteins with the polymorphic sites found in the *IFNGR1* promoter *in vitro*, electrophoretic mobility shift assays were performed using nuclear extracts from five different cell lines. It was hereby of particular interest whether potential transcription factors might bind to the sites surrounding the polymorphisms and whether binding would be different between the wild type and the mutant allele. For each polymorphism two double stranded oligonucleotide probes that encompassed the SNP of interest were used, one for each allele respectively. Binding to *IFNGR1*-1882, -1804, -611, -270, -72, +95 and +130 is shown in Figures 7.6, 7.7 and 7.8 using nuclear extract from Raji cells. No differential binding was detected for all of these polymorphisms. The seemingly stronger binding to the G-variant of *IFNGR1*-611 was due to slightly unequal amounts of nuclear extract as a consecutive competition assay failed to confirm differential binding (data not shown). Similar binding patterns were obtained for A549, Jurkat, RPMI 8226 and MM6 cells (data not shown). Binding of proteins from all of the five investigated cell lines to *IFNGR1*-56 is shown in Figure 7.9. There was no differential binding to this site. At all sites there were no significant differences in binding between stimulated and unstimulated cells.

Binding of proteins from all of the five investigated cell lines to *IFNGR1*-470 is shown in Figure 7.10. Differential binding was detected in all five cell lines. In A549 a protein/DNA complex was detected running in the upper quarter of the gel (see arrow no. 1 in Figure 7.10) that was strongly present with the *ins* variant probe. This complex was hardly detectable with the *ddel* probe. Binding of this complex was stronger with nuclear extracts from TNF activated A549 cells. In Raji, RPMI8226 and MonoMac6 a protein/DNA complex was seen further down on the gel (see arrow no.2). This complex was present with the *ins* variant while absent with the *ddel* variant. In Jurkat multiple bands were seen amongst them the two bands detected in A549 and Raji/RPMI8226/MM6. Most of these were only present with the *ins* allele. Differential binding to *IFNGR1*-470 was confirmed by competition assays.

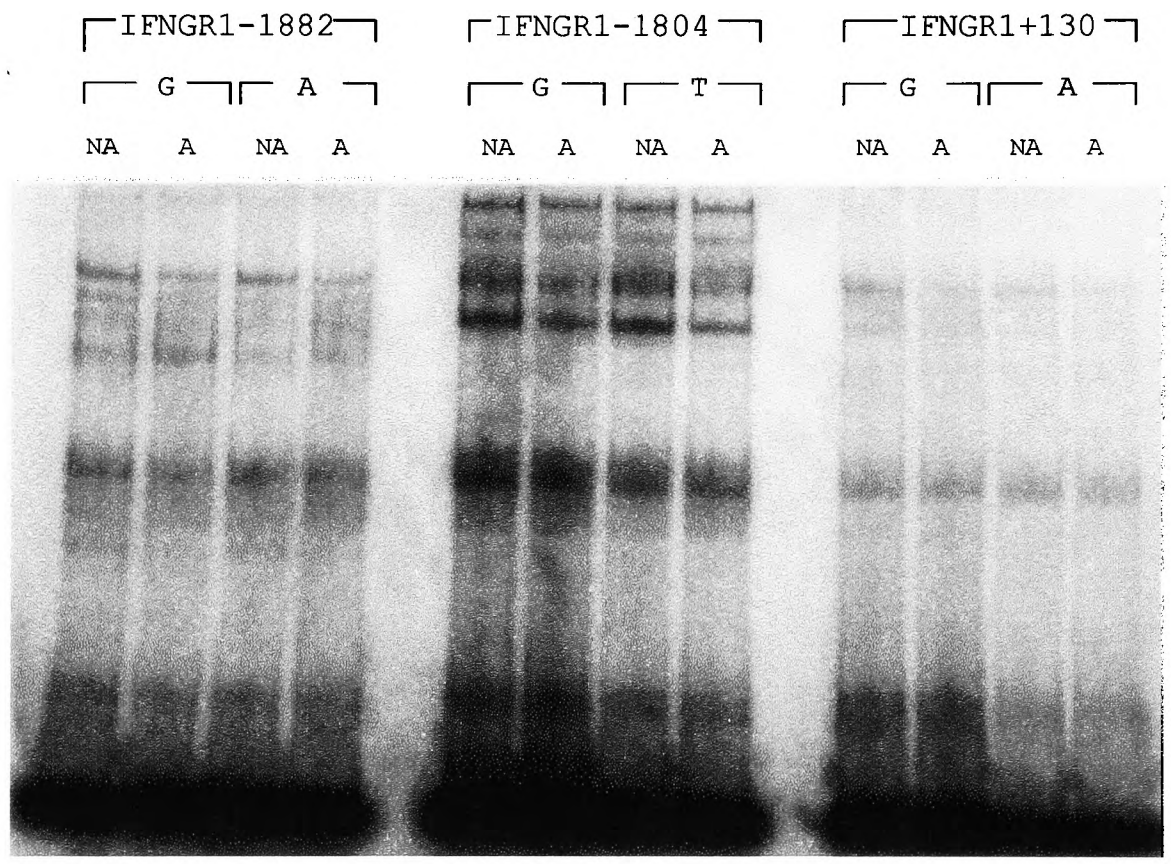


Figure 7.6 EMSA for *IFNGR1*-1882, -1804 and +130 for the two alleles respectively. Nuclear extract from Raji cells. NA = non-activated cells; A = cells activated for 24 hours with 250nM PMA / 150nM Ionomycin.

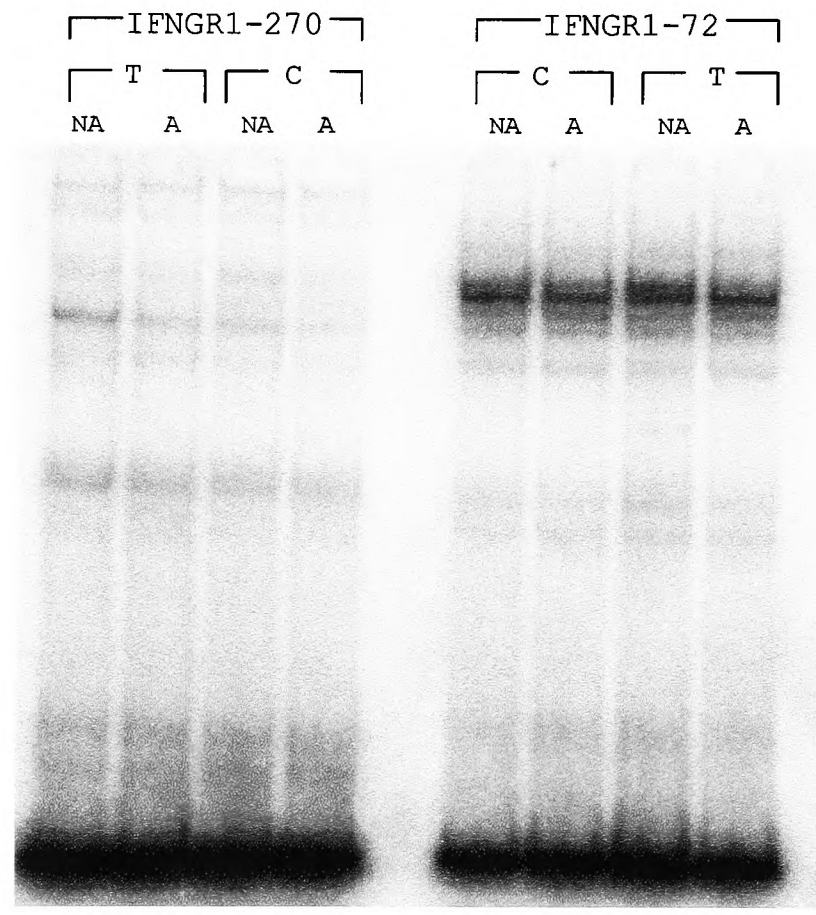


Figure 7.7 EMSA for *IFNGR1*-270 and -72 for the two alleles respectively. Nuclear extract from Raji cells. NA = non-activated cells; A = cells activated for 24 hours with 250nM PMA / 150nM Ionomycin.

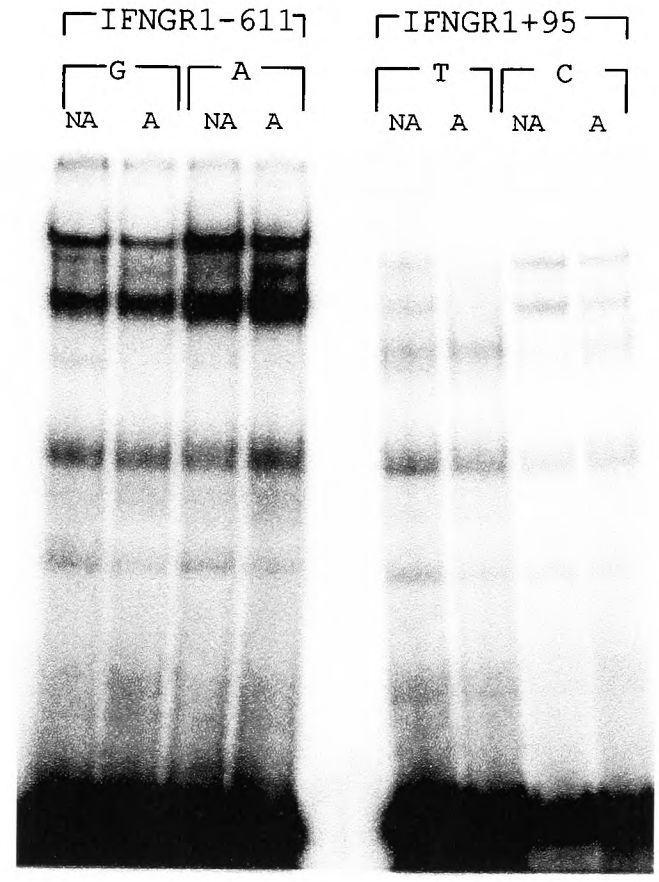


Figure 7.8 EMSA for *IFNGR1*-611 and +95 for the two alleles respectively. Nuclear extract from Raji cells. NA = non-activated cells; A = cells activated for 24 hours with 250nM PMA / 150nM Ionomycin.

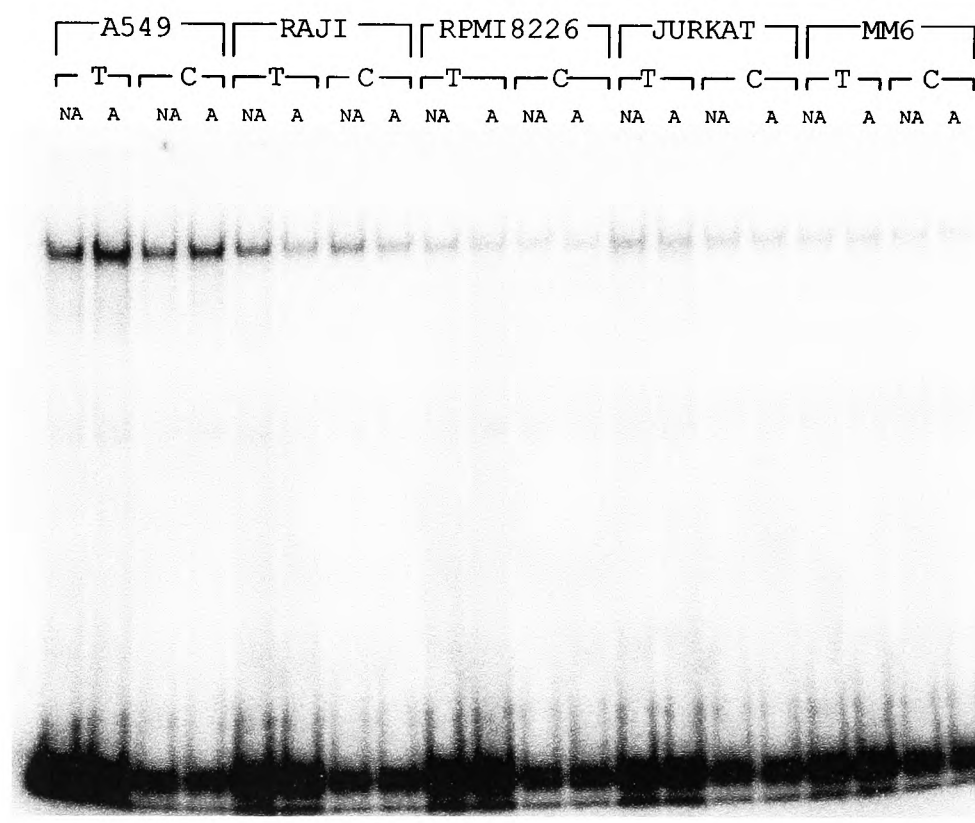


Figure 7.9 EMSA for *IFNGR1*-56 in 5 different cell lines for the two alleles respectively. NA = non-activated cells; A = activated cells (24 hours). A549 cells were stimulated with 40ng/ml TNF-α. Raji, RPMI 8226 and Jurkat cells were stimulated with 250nM PMA / 150nM Ionomycin. MM6 cells were stimulated with 1μg/ml LPS.

Here the ^{32}P-labelled *ins* probe was competed off by unlabelled *ins* allele while the *ddel* allele did compete significantly less. Figure 7.11 shows the competition assay for Raji nuclear extract. A similar picture was seen for RPMI8226 and MM6. Figure 7.12 shows the differential binding in A549.

7.2.3 Functional Analysis of the *IFNGR1* promoter

The influence of the *IFNGR1*-470ddel was further investigated using different reporter gene constructs. Merlin et al. (1997) reported full promoter activity for a fragment consisting of the first 692bp 5' of the transcription start site of the *IFNGR1* gene. In order to investigate the functional impact of the the double deletion at *IFNGR1*-470 three different constructs were generated consisting of -722bp of the *IFNGR1* promoter (from base position -2 to -724). The first had no deletion at position -470 and had the T allele at position -56 (called *470ins/56T* in this section). The second carried the double deletion at position -470/-471 (called *-470ddel* in this section). The third had both the double deletion at -470 plus the C allele at -56 which is the haplotype the -470 exists on naturally (called *-470ddel/-56C* in this section). These variants of the *IFNGR1* promoter were inserted into pGL3 vector carrying a luciferase reporter gene. The resulting plasmids were transiently transfected into A549 cells, RPMI8226 cells and Jurkat cells. Half of the cells were stimulated with either 40ng/ml TNF-α (A549) or 250nM PMA / 150nM Ionomycin (RPMI8226 and Jurkat) for 24h. All experiments were repeated 5 times. In A549 cells luciferase activity in constructs *-470ddel* and *-470ddel/-56* was reduced to about 26% when compared with the *470ins/56T* construct (Figure 7.13). This was the same in both TNF-stimulated and unstimulated cells. Activity was only slightly increased in TNF stimulated cells transfected with the *470ins/56T* construct. Unlike in A549 cells in RPMI8226 B-cells there was a significant increase in luciferase activity of *-470ddel* and *-470ddel/-56* constructs when compared to the *470ins/56T* construct (Figure 7.14). Overall activity was increased in PMA/Ionomycin stimulated cells. In Jurkat no significant change in activity was observed (Figure 7.15). It is worth noting, though, that there was a significant variation between the five transfection

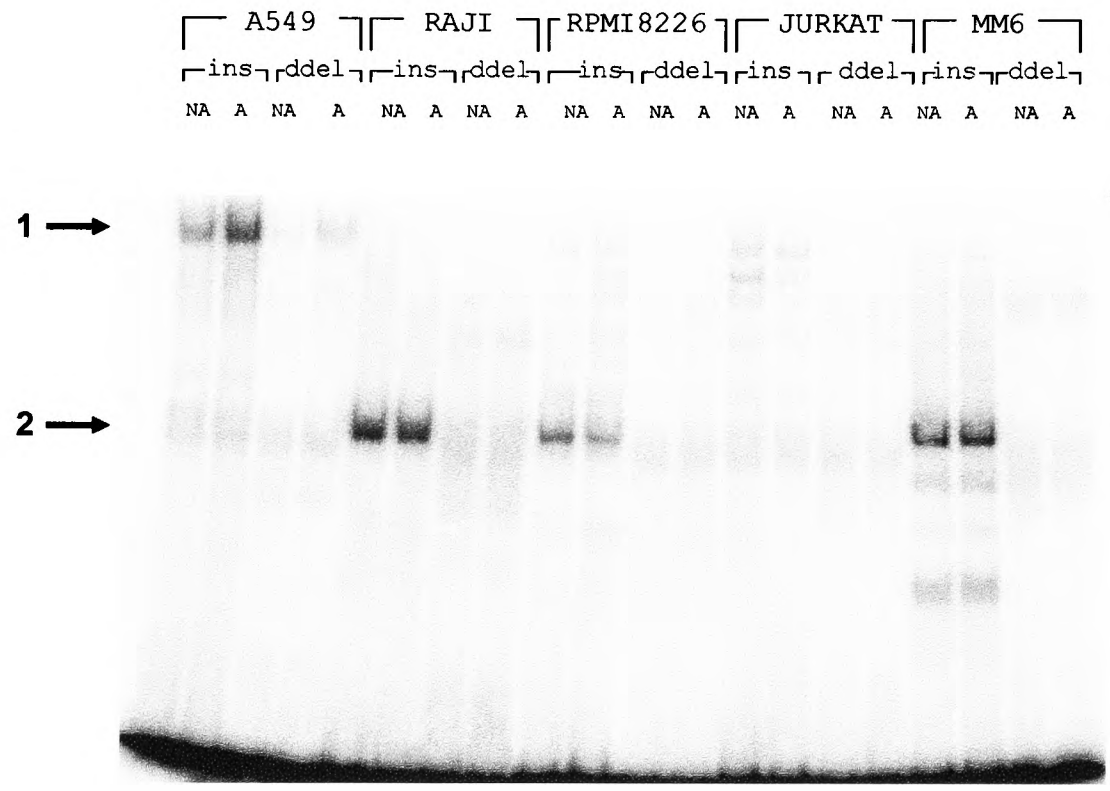


Figure 7.10 EMSA for *IFNGR1*-470 in 5 different cell lines. NA = non-activated cells; A = activated cells (24 hours). A549 cells were stimulated with 40ng/ml TNF-α. Raji, RPMI 8226 and Jurkat cells were stimulated with 250nM PMA / 150nM Ionomycin. MM6 cells were stimulated with 1μg/ml LPS. ins = allele containing the TT allele at positions -470/-471; ddel = allele containing the double deletion at the same site. Arrow 1 indicates the protein/DNA complex in A549 while arrow 2 marks a complex seen in Raji, RPMI8226 and MM6

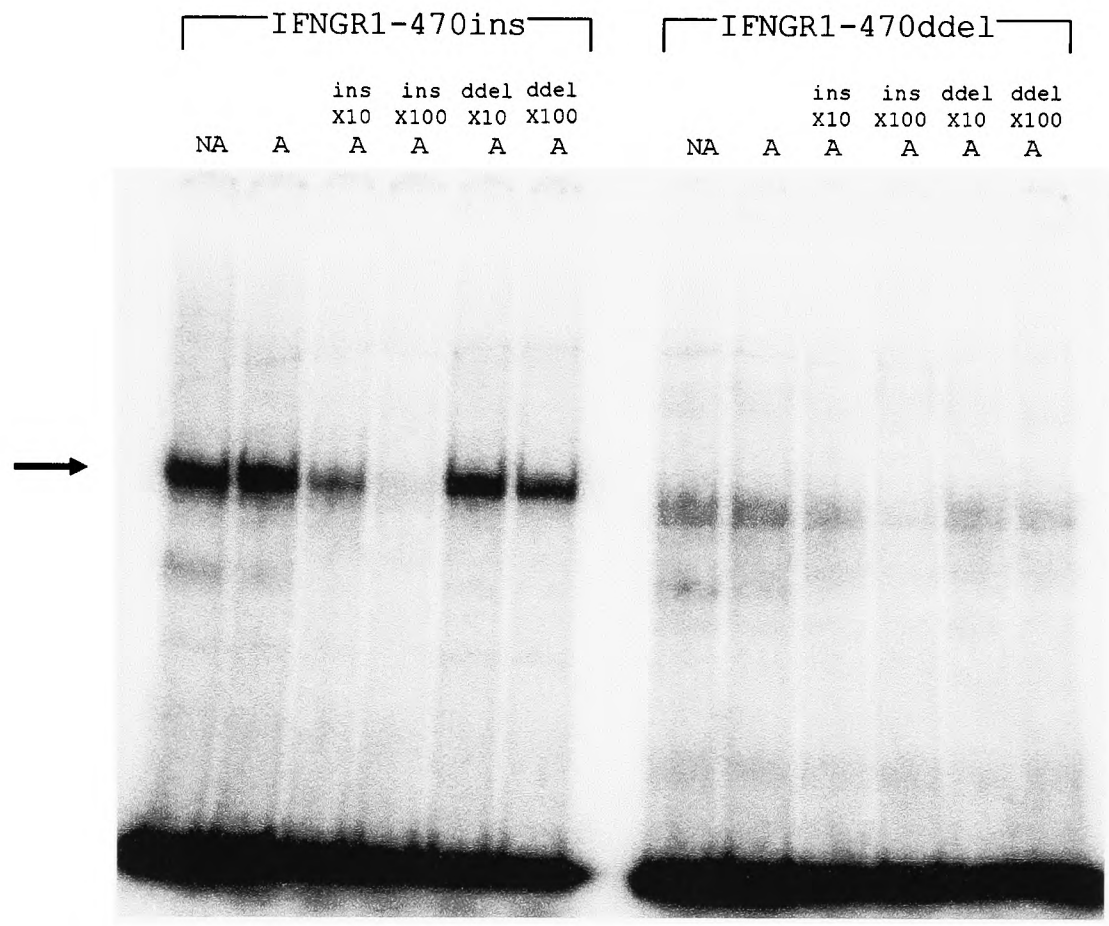


Figure 7.11 Competition EMSA for *IFNGR1*-470 in Raji. ins = allele containing the TT allele at positions -470/-471; ddel = allele containing the double deletion at the same site. NA = non-activated cells; A = cells activated for 24 hours with 250nM PMA / 150nM Ionomycin. Each ^{32}P labelled probe is competed against with unlabelled *IFNGR1*-470ins or *IFNGR1*-470ddel probe at x10 or x100 concentration.

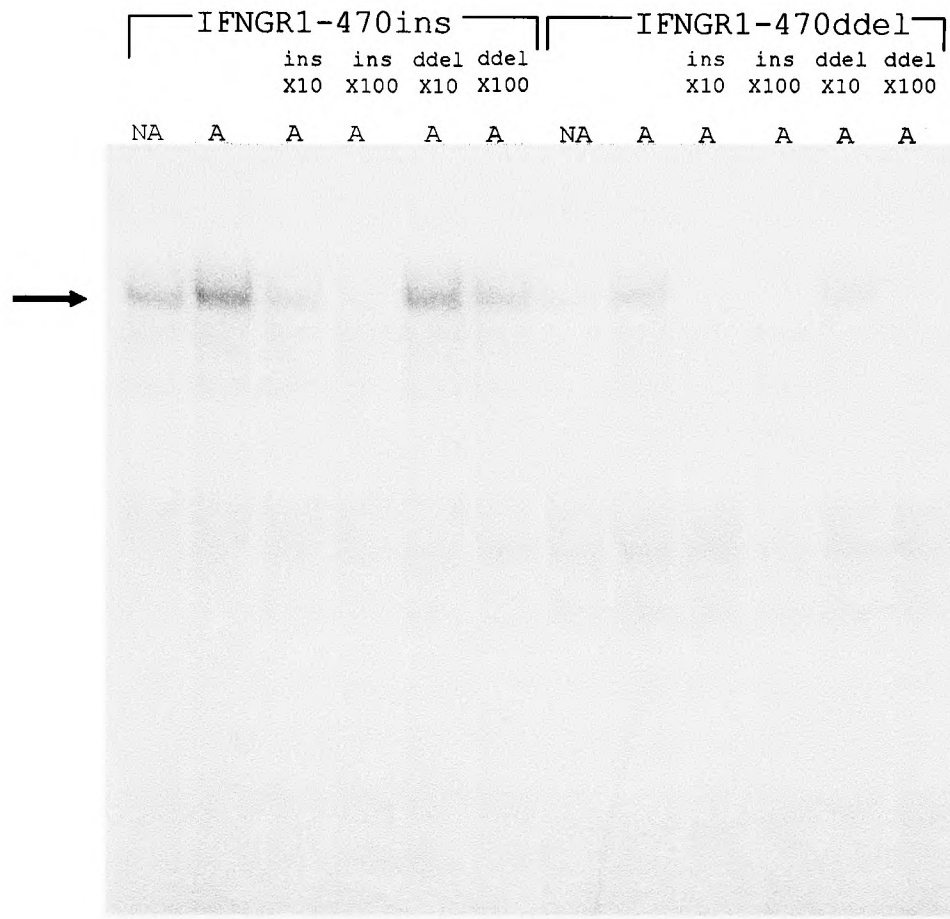


Figure 7.12 Competition EMSA for *IFNGR1*-470 in A549. ins = allele containing the TT allele at positions -470/-471; ddel = allele containing the double deletion at the same site. NA = non-activated cells; A = cells activated for 24 hours with 40ng/ml TNF-α. Each ^{32}P labelled probe is competed against with unlabelled *IFNGR1*-470ins or *IFNGR1*-470ddel probe at x10 or x100 concentration.

experiments in Jurkat cells as indicated by the relatively big standard error of the mean luciferase readings.

7.2.4 Identification of *IFNGR1*-470 Binding Proteins

7.2.4.1 Size Determination

In order to determine the size of the proteins binding to the *IFNGR1*-470 site EMSAs using bromdeoxyuridine labelled *IFNGR1*-470ins probe were performed. UV-crosslinking was done and the band for which differential binding had been observed excised from the gel, eluted and run on a NuPage SDS gel. Size was determined by using a Rainbow size marker. The band representing the protein/DNA complex seen in Raji (see arrow no.2 in Figure 7.10) runs at approximately 50kDA (Figure 7.16). Taking the DNA probe into account the molecular weight of the protein binding to *IFNGR1*-470 in Raji is approximately 35-40kDa in size. The band that was seen in A549 (see arrow no. 1 in Figure 7.10) was approximately 105kDa in size (Figure 7.17. Accounting for the DNA probe this represents a protein of approximately 90-95kDa in size. A second faint complex can be seen at around 50kDa.

7.2.4.2 Competition with STAT Consensus Sites

Analysis of the *IFNGR1*-470 site using the TRANSFAC database had suggested that the signal transducer and activator of transcrition 1 (STAT1) might be binding to this site. STAT1 is similar in size (91kDa) to the protein found to bind in A549 cells. Therefore binding to several known consensus sites for various members of the STAT family were investigated: STAT1, GAS (STAT1/STAT1 homodimer) plus ISRE (STAT1/STAT2/p48), SIE (STAT1/STAT3 heterodimer), STAT3, STAT4, STAT5, and STAT5/6)(Figure 7.18). The same consensus sites were also used for competition against the *IFNGR1*-470ins probe. Binding to the STAT1 consensus site yielded a band that ran slightly higher than the one seen with the *IFNGR1*-470ins probe. The GAS/ISRE probe showed three major bands with the highest of them

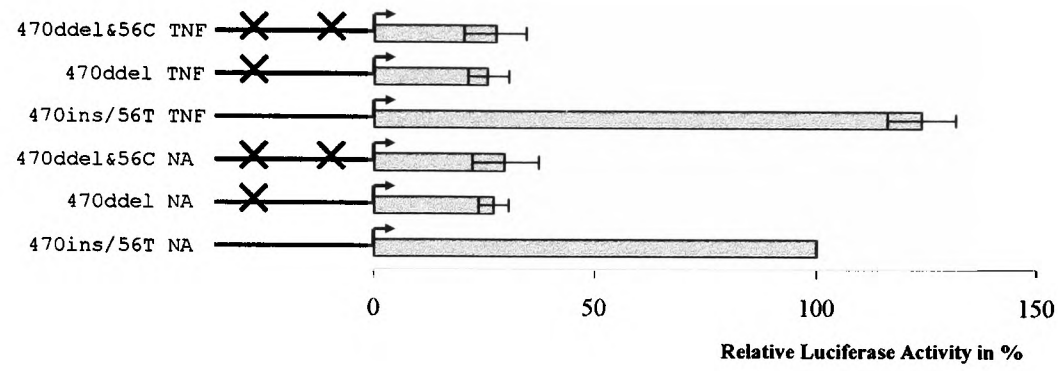


Figure 7.13 Functional analysis of *IFNGR1* promoter pGL3 constructs in A549 cells. Cells were either left unstimulated (NA) or were stimulated with TNF-α 40ng/ml for 24h. Activities were measured by luciferase assays. Experiments were repeated five times.

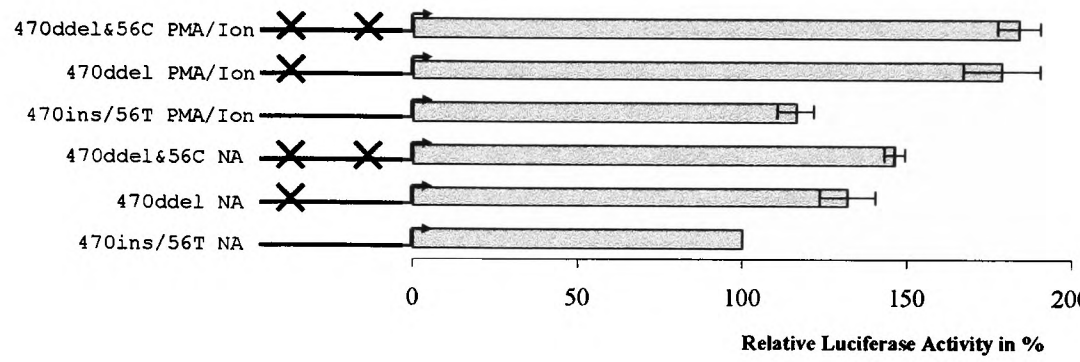


Figure 7.14 Functional analysis of *IFNGR1* promoter pGL3 constructs in RPMI8226 cells. Cells were either left unstimulated (NA) or were stimulated with 250nM PMA / 150nM ionomycin for 24h. Activities were measured by luciferase assays. Experiments were repeated five times.

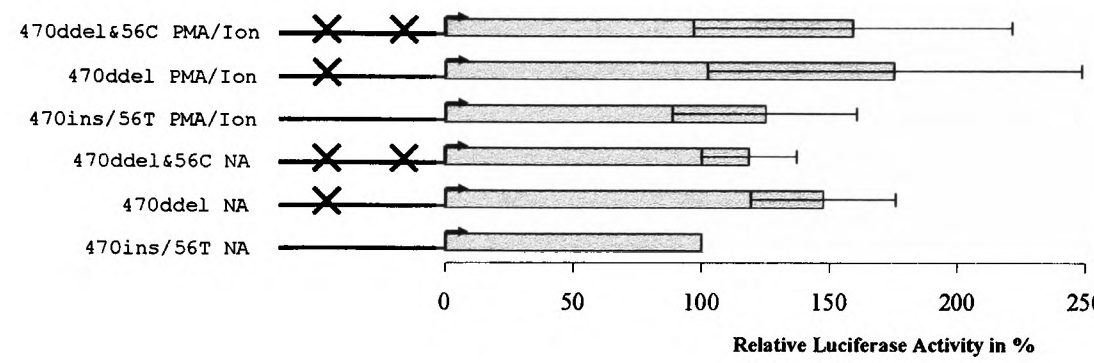


Figure 7.15 Functional analysis of *IFNGR1* promoter pGL3 constructs in Jurkat cells. Cells were either left unstimulated (NA) or were stimulated with 250nM PMA / 150nM ionomycin for 24h. Activities were measured by luciferase assays. Experiments were repeated five times.

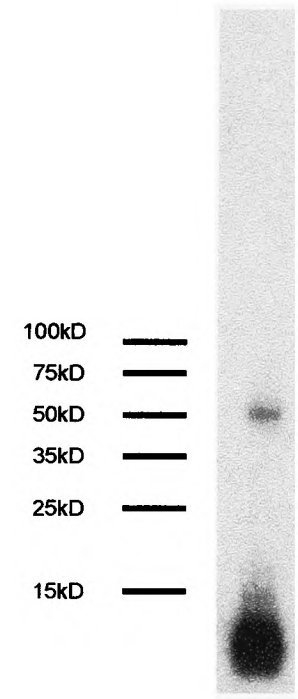


Figure 7.16 SDS protein gel for protein/DNA complex binding to *IFNGR1*-470ins in Raji

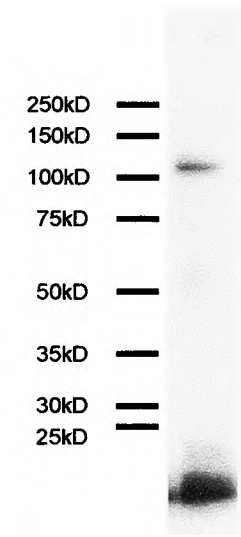


Figure 7.17 SDS protein gel for protein/DNA complex binding to *IFNGR1*-470ins in A549

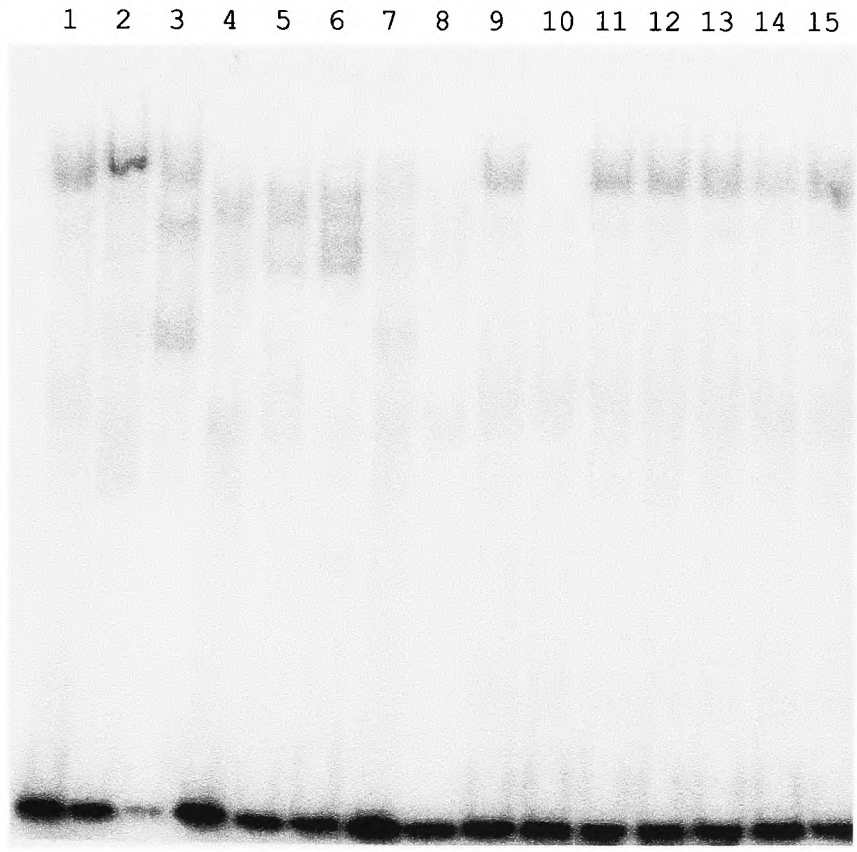
running at the same position as the one seen with *IFNGR1*-470ins. The STAT5 probe also showed a faint band running similar to the one with *IFNGR1*-470ins. All other STAT probes showed a binding pattern that was different compared to the IFNGR1-470ins. On competition the 100x GAS/ISRE probe fully competed off the binding to *IFNGR1*-470ins. There also was minor competition with the STAT5 consensus site. Competition with STAT1 and STAT2 antibodies was attempted but no binding of the antibodies was seen. It is worth noting, though, that competition with antibodies using known consensus sites also failed to work which suggests a failure due to technical reasons.

7.2.4.3 Analysis of STAT1/2 Abundance

In order to assess whether the nuclear extracts of unstimulated Raji, Jurkat, RPMI8226 and A549 contained STAT1 and STAT2, a Western blot was performed using STAT1 and STAT2 antibodies (Figure 7.19). STAT1 and STAT2 were present in all cell lines with STAT1 being most abundant in A549 cells. When the localisation of STAT1 was analysed by immunohistochemistry STAT1 was found in the nucleus of activated (Figure 7.20) and non-activated (data not shown) A549 cells. However, it was most abundant in the cytoplasm of the cell.

7.3 Discussion

The aim of this study was to investigate the impact of polymorphisms on the functional regulation of the promoter of the *IFNGR1* gene. Previous studies have shown that *IFNGR1* is constitutively expressed in many tissues and that its promoter has features similar to other housekeeping genes. There is some evidence, though, that its expression can be induced in several cell lines. In order to assess the inducibility of *IFNGR1* mRNA expression five different cell lines were stimulated with a range of different stimuli and mRNA expression determined by Northern blot. The aim hereby was not



No.	Probe	Competitor
1 =	470ins	-
2 =	STAT1	-
3 =	GAS/ISRE	-
4 =	SIE	-
5 =	STAT3	-
6 =	STAT4	-
7 =	STAT5	-
8 =	STAT5/6	-
9 =	470ins	STAT1 X100
10=	470ins	GAS/ISRE X100
11=	470ins	SIE X100
12=	470ins	STAT3 X100
13=	470ins	STAT4 X100
14=	470ins	STAT5 X100
15=	470ins	STAT5/6 X100

Figure 7.18 Nuclear protein from A549 cells binding to STAT consensus sites. Lane 1-8 show binding to *IFNGR1*-470ins and various STAT consensus binding sites. Lanes 9-15 represent competition of the *IFNGR1*-470ins with unlabelled STAT consensus sites (x100).

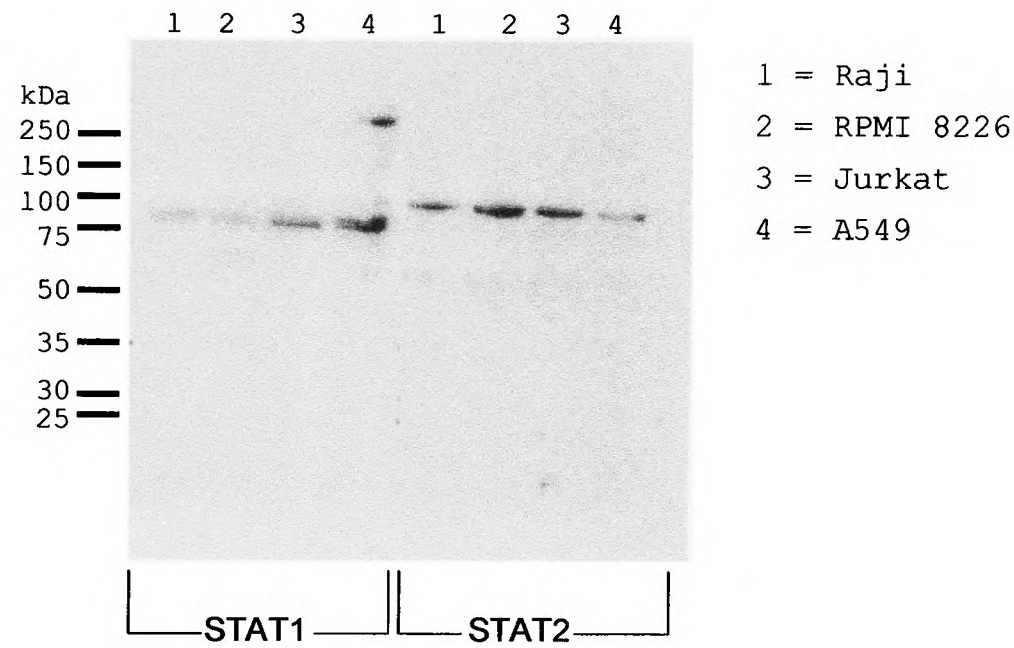


Figure 7.19 STAT1/2 Western Blot. Nuclear Extracts of Raji, RPMI8226, Jurkat and A549 nuclear extract probed with STAT1 or STAT2 antibodies.

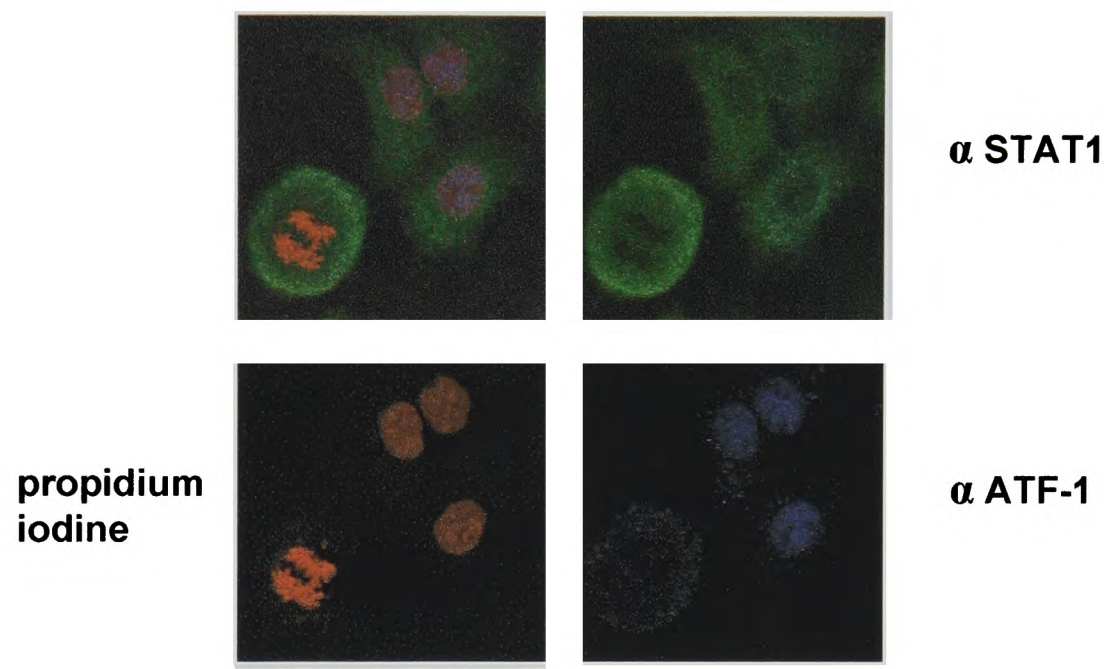


Figure 7.20 STAT1 localisation in A549 cells stimulated with TNF 40ng/ml for 24h. STAT1 (green) was predominately found in the cytoplasm. An antibody against ATF-1 (blue) was included as controls and localisation of this transcription factor was primarily seen in the nucleus. DNA was stained by propidium iodine (red).

to investigate the kinetics of mRNA expression in great detail but rather to establish an inducible cell system that could be used for binding studies and transfection experiments. Previous studies had found that maximal mRNA expression was to be expected 24h after induction (Abed et al., 1994a,b). Although further detailed investigation using varying concentrations of stimuli and detailed time course experiments would be required, it appears from the initial screening that the *IFNGR1* gene is not readily inducible in the cell lines that were investigated. Some induction was achieved by stimulated both B-cell lines with PMA/ionomycin. Mao et al. (1990) had seen increased *IFNGR1* mRNA expression in THP-1 cells induced with PMA but did not observe any change in Raji. Increased mRNA expression was observed in B-cells upon stimulation with LPS (Abed et al., 1994b,a), a finding that was not reproduced in the current study. Sakamoto and Taniguchi (2001) have reported a TPA (=PMA) responsive element in the -128 to -109 base pair region of the *IFNGR1* promoter.

DNA binding to the polymorphic sites of the *IFNGR1* promoter was investigated using EMSAs. Differential binding was found at the *IFNGR1*-470 site. Here a protein/DNA complex was observed that was binding significantly better to the *IFNGR1*-470ins allele (no deletion) than to the *IFNGR1*-470ddel allele (double deletion). In A549 cells this protein/DNA complex was different from the one observed in Raji, RPMI8226 and MM6, while in Jurkat cells both complexes were present. The size of the protein binding to the *IFNGR1*-470ins in A549 cells was determined as approximately 90kDa while in Raji, RPMI8226 and MM6 the protein was only 35kDa in size. Transfection assays suggested that this differences in binding might in fact have a functional correlate. In A549 cells promoter constructs that carried the double deletion at base position -470/-471 luciferase activity was reduced by approximately 75%. In RPMI8226 cells the opposite was observed, i.e. luciferase activity was significantly higher in the constructs with the deletion. This supported the hypothesis that in A549 and B-cells two different factors with different functional properties might be binding to the same site.

In A549 cells the complex binding to *IFNGR1*-470ins was competed off using a GAS/ISRE consensus site. This is a consensus site for STAT1 homodimer (GAS) and a complex of STAT1/STAT2/p48. This supported the hypothesis that a member of the STAT family might be binding to this site as predicted by the transcription factor binding database TRANSFAC. At least seven members of the STAT family have been identified, ranging in size from 90 to 115kDa. STAT1 is activated by IFNγ upon binding to its receptor and forms homodimers that bind to to γ -activated sequence (GAS) elements ($TTN_{5-6}AA$) in the promoter of activated genes. It would be conceivable that STAT1 would bind to the *IFNGR1* promoter in order to form an autofeedback-loop. But it should also be noted that the protein binding in the current study was seen in uninstimulated and TNF stimulated cells. This would suggest constitutive binding to this site. Another possibility could be that there was an autocrine activation by IFNγ. IFNγ in the supernatant was not measured in this study. It could be shown that STAT1 was present in A549, RPMI8226, Jurkat and Raji cells but no significant translocation of STAT1 to the nucleus was seen by immunohistochemistry. STAT1 is best known for its function in IFNγ stimulated gene expression. In addition to function as a transcriptional activator, the transcriptional repression activity of STAT1 has also been documented (Ramana et al., 2000b,a). Furthermore it has also been shown to play a role in untreated cells, where it is required for efficient constitutive expression of many genes, including caspases and the low molecular mass polypeptide 2 (LMP2) (Chatterjee-Kishore et al., 2000b,a,c). It was nevertheless not possible to unambiguously answer the question whether STAT1 (or any other member of the STAT family) was in fact binding in A549. Firstly another consensus site for STAT1 that was used in this study did not compete off the protein. Secondly in the current study the binding of STAT1 unfortunately could not be confirmed using STAT1 antibodies. It is worth noting, though, that competition with antibodies using known consensus sites also failed to work which suggests a failure due to technical reasons.

Although the identity of the factors binding to *IFNGR1*-470 could not be established the above nevertheless suggests a tissue specific binding to the

IFNGR1-470 site with proteins with different functional properties. Liu et al. (1998) showed that STAT1 mediated gene activation was inhibited by a 78kDa protein called protein inhibitor of activated STAT 1 (PIAS1) by blocking DNA binding activity of STAT1. Coccia et al. (2002) showed that this inhibition might be depended on the cell type. It is conceivable that a similar event is happening at the *IFNGR1*-470 site. This is likely to influence constitutive expression of *IFNGR1* as the proteins of interest were present both in stimulated and unstimulated cells. Neither mRNA studies nor transfections assays suggested a large degree of inducibility. Future studies will have to characterise the proteins binding to this site in greater detail.

7.4 Summary

1. The *IFNGR1* gene was not highly inducible in the cell systems used.
2. Altered levels of constitutive expression rather than inducible expression might be important for influencing the inflammatory host response.
3. Differential binding of nuclear protein to the *IFNGR1*-470 site was shown.
4. This differential binding was cell specific with different factors binding in different cell lines.
5. This cell specificity was mirrored in the reporter gene analyses. The deletion of nucleotides -470/-471 of the *IFNGR1* promoter had different effects in different cell lines.
6. An attempt was made to identify the nuclear factor binding to this site. Although STAT seems a likely candidate based on the sequence of the region, the experiments of the current study failed to confirm this hypothesis.

Chapter 8

Linkage disequilibrium analysis of chromosome 12q14

8.1 Introduction

The analysis of the region of the *IFNGR1* gene has yielded some interesting results that require further investigation. However, given the time frame of this thesis this did not seem feasible at this point in time although ongoing work is trying to address some of the issues that have arisen.
At this point efficient high-throughput genotyping methods had become available to the laboratory. This made it possible to investigate not only *IFNGR1* but to expand the investigation to the gene encoding for the ligand binding of this receptor, *IFN*γ.

1. A large amount of SNPs has become available in the public databases for this genetic region.

2. Analysis of this region seemed therefore not only feasible but mandatory given the huge amount of interest in polymorphisms in this gene and their role in various diseases.

3. Furthermore, *IFNγ* has two interesting genetic neighbours, *IL22* and *IL26*.

4. As shown previously for *IFNGR1*, LD studies needed to be undertaken in order to define haplotype blocks and htSNPs that could be used for a further case control analysis.

8.1.1 Background

Despite the central role interferonγ takes in the inflammatory host response surprisingly little work has been done to unravel the genetic variability and haplotype structure of the the interferonγ gene (*IFNγ*) and the surrounding genetic region. While many genetic association studies have looked at *IFNγ* they have focused primarily on one microsatellite found in the first intron.

The first intron of the *IFNγ* gene contains a CA microsatellite repeat that is highly polymorphic, with up to 6 alleles. Allele 2, with 12 CA-repeats, is associated with high levels of interferonγ production in vitro (Pravica et al., 1999). This allele has been associated with higher or lower risk of a variety of diseases, including rheumatoid arthritis (Khani-Hanjani et al., 2000), allograft fibrosis in lung transplant recipients (Awad et al., 1999), and acute graft-versus-host disease (Cavet et al., 2001) in bone marrow transplant recipients. The association with rheumatoid arthritis could not be replicated in two subsequent studies (Constantin et al., 2001; Pokorny et al., 2001). Dabora et al. (2002) found that the frequency of the allele with 12 CA-repeats in patients with tuberous sclerosis with mutations in the TSC2 gene was significantly higher in those without kidney angiomyolipomas than in those with kidney angiomyolipomas. In a cross sectional study Nieters et al. (2001) observed an increased risk of allergies in individuals homozygous for *IFNγ* 13 CA-repeats. Nagarkatti et al. (2002) also found an association for the *IFNγ* microsatellite with asthma in an Indian population.

Pravica et al. (2000) reported an absolute correlation between the 12 CA-repeat allele and the presence of the T allele at a SNP located at position +874 (T to A) from the translational start site, coinciding with a putative NF-κB binding site that might be important in the induction of constitutively high IFNγ production. Lio et al. (2002) found that the +874TT genotype was significantly decreased in Sicilian patients affected by tuberculosis. Similarly Lopez-Maderuelo et al. (2003) found that individuals homozygous for the *IFN*γ+874A allele had a 3.75-fold increased risk of developing tuberculosis.

Bream et al. (2000) screened the *IFN*γ gene from -777 to +5608 for polymorphisms. While the proximal promoter (-777 to transcription start) was invariant in over 100 samples from a predominately Caucasian male population, polymorphisms were found in other regions of the gene including a single base insertion in a poly-T tract in the first intron, three single base pair substitutions in the third intron, and another single base pair substitution in the 3' untranslated region (UTR). Further SNPs were described in the proximal promoter and the 3'UTR in a Sudanese population (Chevillard et al., 2002; Henri et al., 2002).

Deletion analysis studies identified the core *IFN*γ promoter (Chrivia et al., 1990; Penix et al., 1993) in the region -100 to -30 upstream of the translational start. This area is highly conserved through evolution. However, enhancer sites have been located in the first intron (Ciccarone et al., 1990) and several NFκB binding sites and a c-rel protooncogene product (c-Rel) (Sica et al., 1992) site have been described in the intronic regions of the gene. These data indicate that several transcription regulatory regions are distributed along the gene.

The *IFN*γ gene contains 4 exons (Gray and Goeddel, 1982). Bureau et al. (1995) localized this gene close to the D12S335 and D12S313 microsatellites on chromosome band 12q15. By fluorescence in situ hybridization, Zimonjic et al. (1995) mapped the *IFN*γ gene to 12q14. Its nearest genetic neighbours

are *IL22* and *IL26*, two genes with a potentially important role in the immune response.

8.1.1.1 IL22

IL22 or *IL10-related T cell-derived inducible factor* (ILTIF), a cytokine structurally related to IL10, was originally identified in the mouse as a gene induced by IL9 in T cells and mast cells (Dumoutier et al., 2000a). Dumoutier et al. (2000c) cloned human ILTIF cDNA, which encodes a deduced 179-amino acid protein that shares 79% and 25% sequence identity with mouse Iltif and human IL10, respectively. Recombinant human ILTIF was found to activate signal transducer and activator of transcription factors-1 and -3 in several hepatoma cell lines. ILTIF stimulation of HepG2 human hepatoma cells upregulated the production of acute phase reactants such as serum amyloid A, alpha-1-antichymotrypsin, and haptoglobin, suggesting its involvement in the inflammatory response. Xie et al. (2000) cloned *IL22* as well as *IL22R*. IL22 was shown to be upregulated in activated T cells. By genomic sequence analysis, Dumoutier et al. (2000b) determined that *IL22* contains 5 coding exons of similar sizes to the *IL10* gene, as well as an additional noncoding exon. The *IL22* gene was mapped to chromosome 12q15 (Dumoutier et al., 2000b), close to the *IFN*γ gene. They mapped the mouse gene to chromosome 10.

8.1.1.2 IL26

Knappe et al. (2000) have isolated a novel cellular gene, *AK155*, now also called *IL26*, a sequence homolog of the interleukin-10 gene. Specifically herpesvirus saimiri-transformed T cells overexpress *AK155* and secrete the protein into the supernatant. In other T-cell lines and in native peripheral blood cells, but not in B cells, the *AK155* gene is transcribed at low levels. AK155 forms homodimers similarly to interleukin-10. Knappe et al. (2000) determined that the *AK155* gene contains 5 exons mapped it to 12q15, near the *IFN*γ gene.

8.1.2 Aims and Approach

Within a region of 100kb on chromosome 12q14 the *IFNγ* gene is neighboured by two further genes for which there is evidence that they play a significant role in the inflammatory immune response. Despite its crucial role in the inflammatory response to infectious disease only few studies looked at the association of genetic variants in *IFNγ* with susceptibility to infectious diseases. Furthermore despite a large number of SNPs described along the *IFNγ* gene there is no knowledge about the haplotypic structure of these genetic markers. Describing the haplotype structure will facilitate selecting markers to type in subsequent association studies. Unlike on chromosome 6q23, the genomic region of *IFNGR1*, a substantial number of validated SNPs were available in the literature. This study therefore searched public SNP databases and the literature for validated SNPs that had information on allele frequencies. A total of 20 SNPs in *IFNγ*, *IL22* and surrounding *IL26* were selected and typed in 32 Caucasian and 32 Gambian families. For both populations haplotypes were constructed and pairwise linkage disequilibrium was calculated. Blocks of haplotyped were found and haplotype tagging SNPs identified.

8.2 Results

8.2.1 Marker Selection

The public SNP databases and the literature were searched for polymorphic markers within the region of chromosome 12q14 spanning a total of approximately 100kb. Only validated markers or markers for which an allelele frequency had been reported were selected. Marker positions were referenced to Genbank accession number NT_029419. A total of 20 markers were selected, densely covering the *IFNγ* and *IL22* gene. For all markers but four an allele frequency$f > 0.10$ had been reported in the literature. The marker 12:722245636 was selected despite having a reported allele frequency of only $f = 0.02$ as it is located in the coding sequence of the *IL22* gene. It

is a nonsynonymous change of amino acid position 158 (Ser to Gly). Similarly, the markers 12:72156516 (*IFN*γ-280) 12:72156544 (*IFN*γ-308) were selected as a potential functional role had been suggested by Chevillard et al. (2002) (here called *IFN*γ-155 and *IFN*γ-183 respectively). The marker 12:72151597 (*IFN*γ+4640) had been reported by Bream et al. (2000) (here called *IFN*γ+4766). In this study differential binding of an unknown nuclear protein to this site was reported. Only little information was available for markers covering the region between *IFN*γ and *IL22*, i.e. the region surrounding *IL26*. Here only two markers were found in the public databases with a reported allele frequency $f > 0.10$. A total of 11 markers are spanning the *IFN*γ gene. Three of these are located in the 3'UTR, a further three are in the promoter and the remaining SNPs are intronic. A total of seven markers are spanning the *IL22* gene. Two of these markers are promoter SNPs, one is located in the 3'UTR, one is an exonic SNP and the remaining are intronic SNPs. For *IL26* one SNP in the 3'UTR was used. One further SNP is located 11kb upstream of the translational start site. Figure 8.1 is a graphical representation the genomic region of chromosome 12q14 and the selected markers.

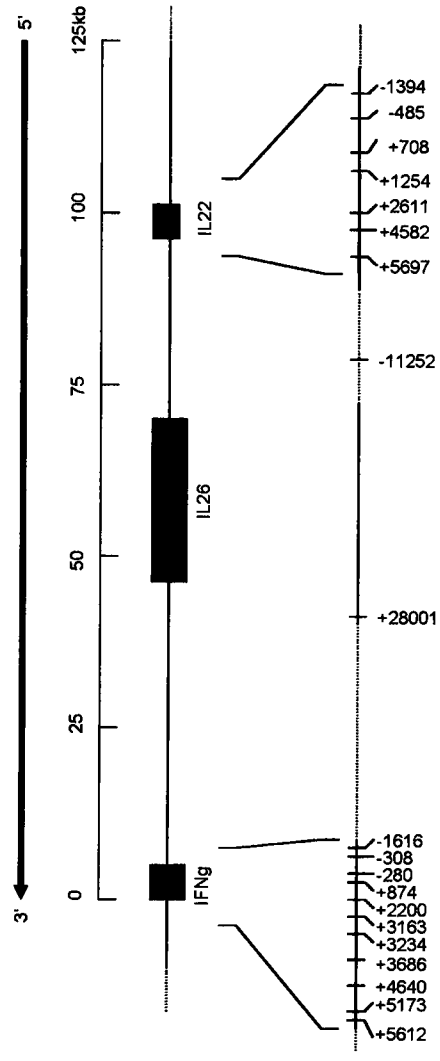


Figure 8.1 SNPs in the 12q14 region. SNP position is referenced to the nearest gene.

8.2.2 Allele Frequencies

As described before parental allele were counted to calculate the allele frequency of each marker. Frequencies of the the minor allele ranged from .01-.48 and from 0-.48 in Caucasians. In Caucasians 14 SNPs had a minor allele frequencies greater than .10 (Table 8.2 and Figure 8.2) but four of the 20 sites were monomorphic. All 20 sites were variant in Africans. 12 SNPs had a minor allele frequency of greater than 0.10. Five of the SNPs were at low frequency in both populations (*IFN*γ+4640, +3686, -280, -308 and *IL22*+4582). All other SNPs had an allele frequency > 0.10 in at least one population. The average minor allele frequency was higher in Caucasians (0.23) than in Africans (0.20).

8.2.3 Haplotypes

Using 20 markers on chromosome 12q14 a total of 67 haplotypes were found in Africans. A total of 47 haplotypes were found in Caucasians using 16 variant sites. Fourteen haplotypes were found both in Caucasians and Africans accounting for 28% of all haplotypes in Africans and 41% of all haplotypes in Caucasians. The allelic composition and frequency of each haplotype is shown in tables 8.3 and 8.4. Most haplotypes were present at very low frequency (8.4) indicating significant recombination in this region as only few haplotypes span the entire region of 100kb in more than one chromosome. None of the haplotypes found in Africans had a frequency $> .10$. In Caucasians one haplotype (no. 54) had a frequency above .10.

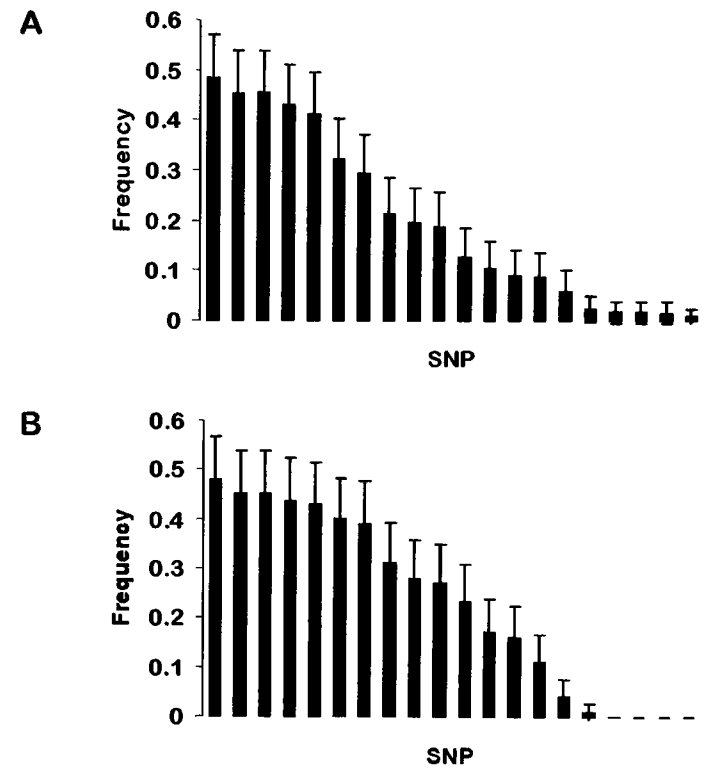


Figure 8.2 Frequency distribution of individual SNPs in Africans (**A**) and Caucasians (**B**). 95% confidence intervals are indicated with error bars.

Table 8.1 Frequencies of 20 SNPs on chromosome 12q14 in an African (Gambian) population. Chromosomal location refers to September 2002 freeze.

NO.	dbSNPref	CHROM.LOC.	REL. TO GENE	FREQ	95% C.I.
1	rs2069728	12:72150625	IFNg+5612	0.105	0.053
2	rs2069727	12:72151064	IFNg+5173	0.188	0.068
3	rs2234687	12:72151597	IFNg+4640	0.008	0.016
4	rs2069720	12:72152551	IFNg+3686	0.017	0.022
5	rs2069718	12:72153003	IFNg+3234	0.322	0.081
6	rs2069733	12:72153074	IFNg+3163	0.016	0.022
7	rs1861493	12:72154037	IFNg+2200	0.087	0.049
8		12:72155363	IFNg+874	0.195	0.069
9	rs2069710	12:72156516	IFNg-280	0.016	0.022
10	rs2069709	12:72156544	IFNg-308	0.024	0.027
11	rs2069705	12:72157852	IFNg-1616	0.484	0.087
12	rs971545	12:72194525	IL26+28001	0.424	0.086
13	rs734871	12:72233777	IL26-11252	0.411	0.086
14	rs1182844	12:72244521	IL22+5697	0.127	0.058
15	rs2227507	12:72245636	IL22+4582	0.059	0.041
16	rs1012356	12:72247607	IL22+2611	0.452	0.087
17	rs2046068	12:72248964	IL22+1254	0.090	0.050
18	rs2227491	12:72249510	IL22+708	0.214	0.071
19	rs2227485	12:72250702	IL22-485	0.453	0.087
20	rs2227478	12:72251611	IL22-1394	0.294	0.079

Table 8.2 Frequencies of 20 SNPs on chromosome 12q14 in a UK Caucasian population. Chromosomal location refers to September 2002 freeze.

NO.	dbSNPref	CHROM.LOC.	REL. TO GENE	FREQ	95% C.I.
1	rs2069728	12:72150625	IFNg+5612	0.271	0.077
2	rs2069727	12:72151064	IFNg+5173	0.430	0.086
3	rs2234687	12:72151597	IFNg+4640	0.000	0.000
4	rs2069720	12:72152551	IFNg+3686	0.000	0.000
5	rs2069718	12:72153003	IFNg+3234	0.392	0.085
6	rs2069733	12:72153074	IFNg+3163	0.172	0.066
7	rs1861493	12:72154037	IFNg+2200	0.234	0.074
8		12:72155363	IFNg+874	0.438	0.086
9	rs2069710	12:72156516	IFNg-280	0.000	0.000
10	rs2069709	12:72156544	IFNg-308	0.000	0.000
11	rs2069705	12:72157852	IFNg-1616	0.313	0.081
12	rs971545	12:72194525	IL26+28001	0.161	0.064
13	rs734871	12:72233777	IL26-11252	0.040	0.034
14	rs1182844	12:72244521	IL22+5697	0.111	0.055
15	rs2227507	12:72245636	IL22+4582	0.010	0.017
16	rs1012356	12:72247607	IL22+2611	0.452	0.087
17	rs2046068	12:72248964	IL22+1254	0.280	0.078
18	rs2227491	12:72249510	IL22+708	0.480	0.087
19	rs2227485	12:72250702	IL22-485	0.452	0.087
20	rs2227478	12:72251611	IL22-1394	0.398	0.085

Table 8.3: Frequencies of 67 haplotypes in a Gambian population sorted by the *IFNγ*, *IL26* and *IL22* SNPs. To facilitate viewing allele 2 is symbolised by '-' and allele 1 by 'o'. Haplotypes 5, 25, 37, 39, 42, 41, 53, 54, 57, 59, 60, 63, 65 and 67 were also found in Caucasians.

IDENTIFIER	HAPLOTYPE	NO.	FREQUENCY
1	oooooo-o---o---o-o-o	2	0.016
2	ooo-oo-o---o-o-o---o	2	0.016
3	ooo-oo-o---o-----o--	2	0.016
4	ooo-oo-o---o---o-o--	2	0.016
5	ooo-oo-o---o-----ooo	1	0.008
6	ooo-oo-o----oo-o-o--	1	0.008
7	ooo-oo-o----o-oo-o--	1	0.008
8	ooo-oo-o----o--oo---	1	0.008
9	ooo-oo------o--oo---	1	0.008
10	ooo--oooo-o-o----ooo	1	0.008
11	ooo--oooo-o-o----o--	1	0.008
12	ooo--ooo--oooo-o---o	1	0.008
13	ooo--ooo--ooo--o-o--	3	0.023
14	ooo--ooo--ooo--o---o	1	0.008
15	ooo--ooo--oo---o---o	1	0.008
16	ooo--o-o-o-o-----ooo	2	0.016
17	ooo--o-o-o--o----ooo	1	0.008

continued on next page

continued from previous page

IDENTIFIER	HAPLOTYPE	NO.	FREQUENCY
18	ooo--o-o--oooo-o---o	1	0.008
19	ooo--o-o--ooo--o-o-o	1	0.008
20	ooo--o-o--ooo----ooo	3	0.023
21	ooo--o-o--ooo--o-o--	1	0.008
22	ooo--o-o--ooo-----oo	1	0.008
23	ooo--o-o--ooo--o-ooo	1	0.008
24	ooo--o-o--oo--oo-ooo	1	0.008
25	ooo--o-o--oo-----ooo	7	0.055
26	ooo--o-o--oo-----o--	4	0.031
27	ooo--o-o--oo---o-o-o	1	0.008
28	ooo--o-o--oo---o-ooo	1	0.008
29	ooo--o-o--o-oo-o---o	2	0.016
30	ooo--o-o--o-o-oo-ooo	3	0.023
31	ooo--o-o--o-o----ooo	5	0.039
32	ooo--o-o--o-o--o-o-o	1	0.008
33	ooo--o-o--o-o--o-ooo	3	0.023
34	ooo--o-o--o-o----o--	2	0.016
35	ooo--o-o--o----o-ooo	3	0.023
36	ooo--o-o--o------o--	1	0.008
37	ooo--o-o--o------ooo	4	0.031
38	ooo--o-o--o----oo---	1	0.008
39	ooo--o-o---oo----ooo	1	0.008

continued on next page

continued from previous page

IDENTIFIER	HAPLOTYPE	NO.	FREQUENCY
40	ooo--o-o---o-----o--	2	0.016
41	ooo--o-o---o-----ooo	3	0.023
42	ooo--o-o---o---oo---	1	0.008
43	ooo--o-o---o---o-ooo	2	0.016
44	ooo--o-o----o-oo-ooo	1	0.008
45	ooo--o-o----o----ooo	2	0.016
46	ooo--o-o----o--oo---	1	0.008
47	ooo--o-o-----o-o----	2	0.016
48	ooo--o-o-------oo---	4	0.031
49	ooo--o-o---------o--	1	0.008
50	ooo---oo--o-o--o-o-o	1	0.008
51	o-o-oo-----oo----ooo	2	0.016
52	o-o-oo-----o-o-o---o	2	0.016
53	o-o-oo-----o-o-oo---	1	0.008
54	o-o-oo-----o-----ooo	5	0.039
55	o-o-oo------o----ooo	5	0.039
56	o-o-oo------o----o--	2	0.016
57	o-o-oo-------o-oo---	2	0.016
58	o-o-oo-------o-o---o	1	0.008
59	o-o-oo----------ooo	3	0.023
60	-oo-oo-o---o-----ooo	2	0.016
61	-oo--o-o--ooo----ooo	2	0.016

continued on next page

continued from previous page

IDENTIFIER	HAPLOTYPE	NO.	FREQUENCY
62	-oo--o-o--oo-o-o---o	1	0.008
63	-oo--o-o--oo-----ooo	3	0.023
64	-oo--o-o--oo---o-o-o	1	0.008
65	-oo---oo--oo-----ooo	2	0.016
66	-o---o-o--oo---o-o-o	1	0.008
67	--o-oo-----o---oo---	1	0.008

Table 8.4: Frequencies of 47 haplotypes in a Gambian population sorted by the *IFNγ*, *IL26* and *IL22* SNPs. To facilitate viewing allele 2 is symbolised by '-' and allele 1 by 'o'. Haplotypes 5, 25, 37, 39, 42, 41, 53, 54, 57, 59, 60, 63, 65 and 67 were also found in Gambians.

IDENTIFIER	HAPLOTYPE	NO.	FREQUENCY
68	ooo-oo-o---o-o-oo---	1	0.008
69	ooo-oo-o---o---oo---	7	0.055
70	ooo-oo-o---o---o----	3	0.023
5	ooo-oo-o---o-----ooo	2	0.016
71	ooo-oo-o----o----ooo	1	0.008
72	ooo-oo-o---------ooo	1	0.008
73	ooo-oo-o-------oo---	3	0.023
74	ooo--ooo--oo----oooo	2	0.016
75	ooo--ooo--oo---oo---	2	0.016

continued on next page

continued from previous page

IDENTIFIER	HAPLOTYPE	NO.	FREQUENCY
25	ooo--o-o--oo-----ooo	5	0.039
76	ooo--o-o--oo---o----	2	0.016
37	ooo--o-o--o------ooo	1	0.008
39	ooo--o-o---oo----ooo	1	0.008
77	ooo--o-o---o-o-o---o	3	0.023
42	ooo--o-o---o---oo---	2	0.016
41	ooo--o-o---o-----ooo	2	0.016
78	ooo--o----oo--o--ooo	1	0.008
79	ooo---oo--oo-----ooo	8	0.063
80	ooo---oo--oo---oo---	1	0.008
81	ooo---oo--o-o----ooo	3	0.023
82	ooo---oo--o----oo---	3	0.023
53	o-o-oo-----o-o-oo---	6	0.047
83	o-o-oo-----o-o-o----	1	0.008
84	o-o-oo-----o---oo---	2	0.016
54	o-o-oo-----o-----ooo	16	0.125
85	o-o-oo-----o---o-o-o	1	0.008
86	o-o-oo-----o---o----	5	0.039
87	o-o-oo-----o----oooo	2	0.016
88	o-o-oo-----o---o---o	2	0.016
57	o-o-oo-------o-oo---	1	0.008
59	o-o-oo----------ooo	5	0.039

continued on next page

continued from previous page

IDENTIFIER	HAPLOTYPE	NO.	FREQUENCY
60	-oo-oo-o---o-----ooo	3	0.023
89	-oo-oo-o---o---oo---	2	0.016
90	-oo--ooo--oo---oo---	2	0.016
91	-oo--ooo--oo----oooo	1	0.008
63	-oo--o-o--oo-----ooo	1	0.008
92	-oo--o-o---o-----ooo	2	0.016
93	-oo---oo--oo-o-oo---	1	0.008
65	-oo---oo--oo-----ooo	5	0.039
94	-oo---oo--oo---oo---	1	0.008
95	-oo---oo--o------ooo	1	0.008
96	--o-oo-----o-o-oo---	1	0.008
97	--o-oo-----o-----ooo	7	0.055
67	--o-oo-----o---oo---	2	0.016
98	--o-oo-----o---o----	2	0.016
99	--o-oo----------ooo	1	0.008
100	--o-o------o---oo---	1	0.008

8.2.4 Linkage Disequilibrium

Pairwise measures of linkage disequilibrium (D, D', χ^2, Δ^2) were calculated for all pairs of markers using the program `Haploxt`. Figure 8.5 and 8.6 represent linkage disequilibrium between 20 markers. The absolute values of D' and Δ^2 are represented as different colours in the output of the graphical overview of linkage disequilibrium (GOLD) program. The statistical significance of the χ^2 values was calculated by `Haploxt` and significant p-values

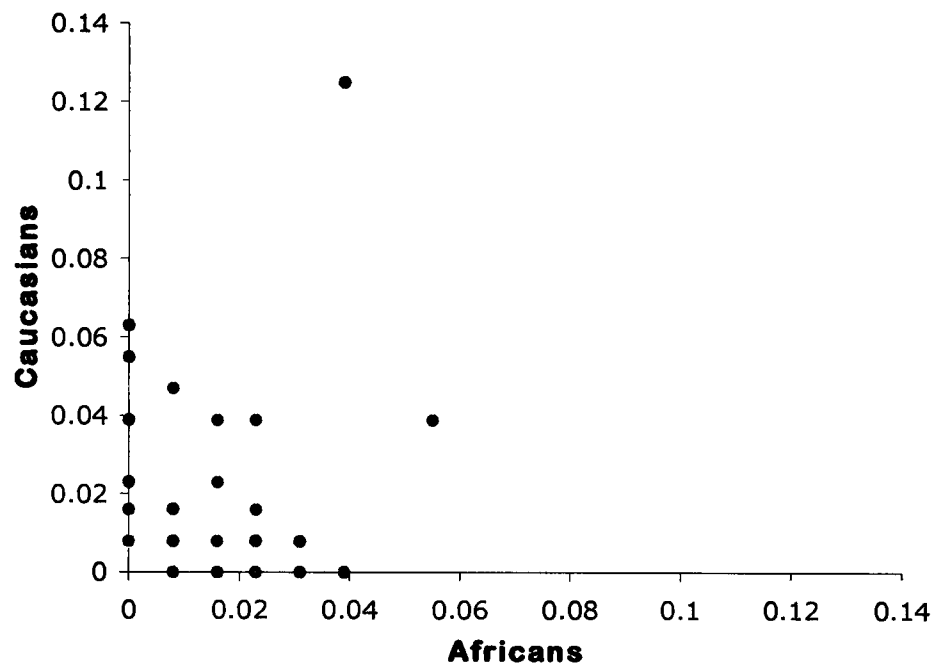


Figure 8.3 Scatterplot of frequency distribution of haplotypes in Africans and UK Caucasians.

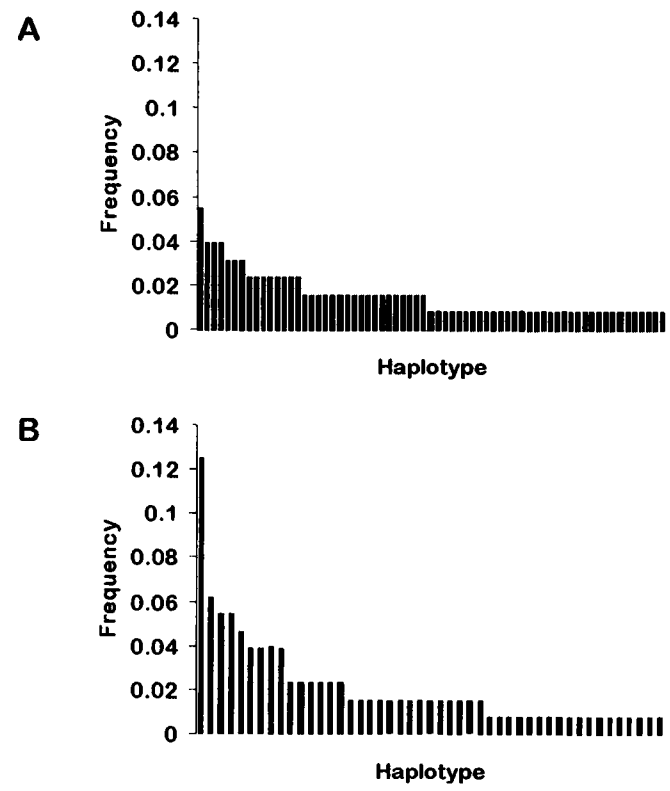


Figure 8.4 Frequency distribution of haplotypes in Africans (**A**) and UK Caucasians (**B**).

represented as circles are shown in Figures 8.5**A** and 8.6**A**. All measures of LD indicate that LD is highest within each of the three gene regions. The Δ^2 values in particular show no significant LD across gene regions. D' values for markers 2 (*IFNγ*+4640) and 9 (12:72156516 or *IFNγ*-280) appear to be high throughout the entire region of 100kb in Gambians although this most definitely reflects the fact that these markers have a very low allele frequency in this population. Both p-value and D' show that LD across the 7kb of the *IFNγ* gene and the 7kb of the *IL22* gene is strong. There also seems to be strong LD between the two markers that are spaced 39kb apart and are surrounding the *IL26* gene. While p-values indicate that some markers might be in linkage with the *IL26* markers there is only little evidence for this from D' and Δ^2. The general patterns of LD are similar in Caucasians and Africans. It is of interest that the variant *IFNγ*+874 which has been implicated in susceptibility to disease previously (Pravica et al., 2000; Lio et al., 2002) was in strong LD with other markers throughout the *IFNγ* gene.

The pairwise measures of linkage disequilibrium, D' and Δ^2, were also plotted against physical distance (Figure 8.7). As expected for both D' and Δ^2 a decline can be observed with increasing distance both in Caucasians and Gambians. Nevertheless individual values deviate from this trend and several high D' values were observed over large distances. As shown before in chapter 5 on page 164 it was found that the average D' and Δ^2 facilitate the comparison of LD for Caucasians and Gambians. Therefore all D' and Δ^2 were averaged up to the distance plotted. While no difference between Caucasians and Africans can be seen for the average D', the average Δ^2 is higher for Caucasians than for Africans in this region of 100kb.

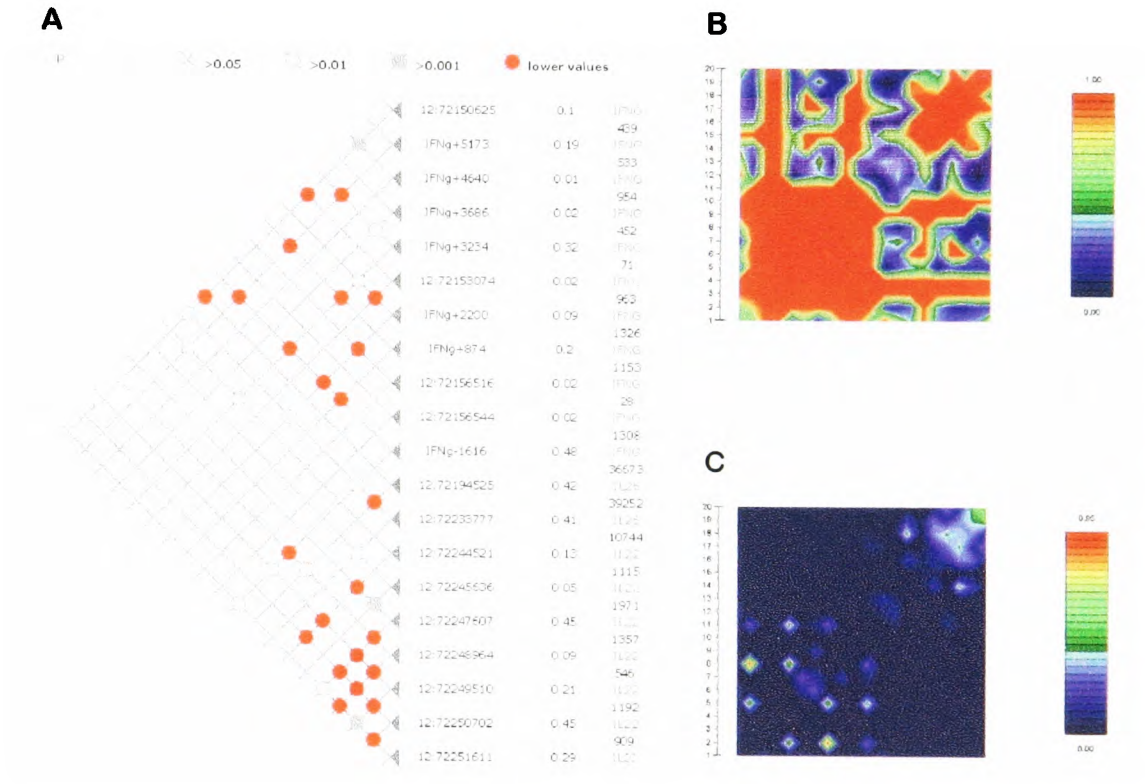


Figure 8.5 Linkage Disequilibrium between 20 polymorphisms in 12q14 in a Gambian population. **A.** Pairwise measure of LD presented by p-value. Various degrees of significance are indicated by differently coloured circles. Open white circles represent p-values $0.05 \geq p > 0.01$, filled grey circles indicate $0.01 \geq p > 0.001$ and filled red circles represent $p \leq 0.001$. **B.** LD measured by D' illustrated with the graphical overview of linkage disequilibrium (GOLD) program. **C.** LD measured as Δ^2

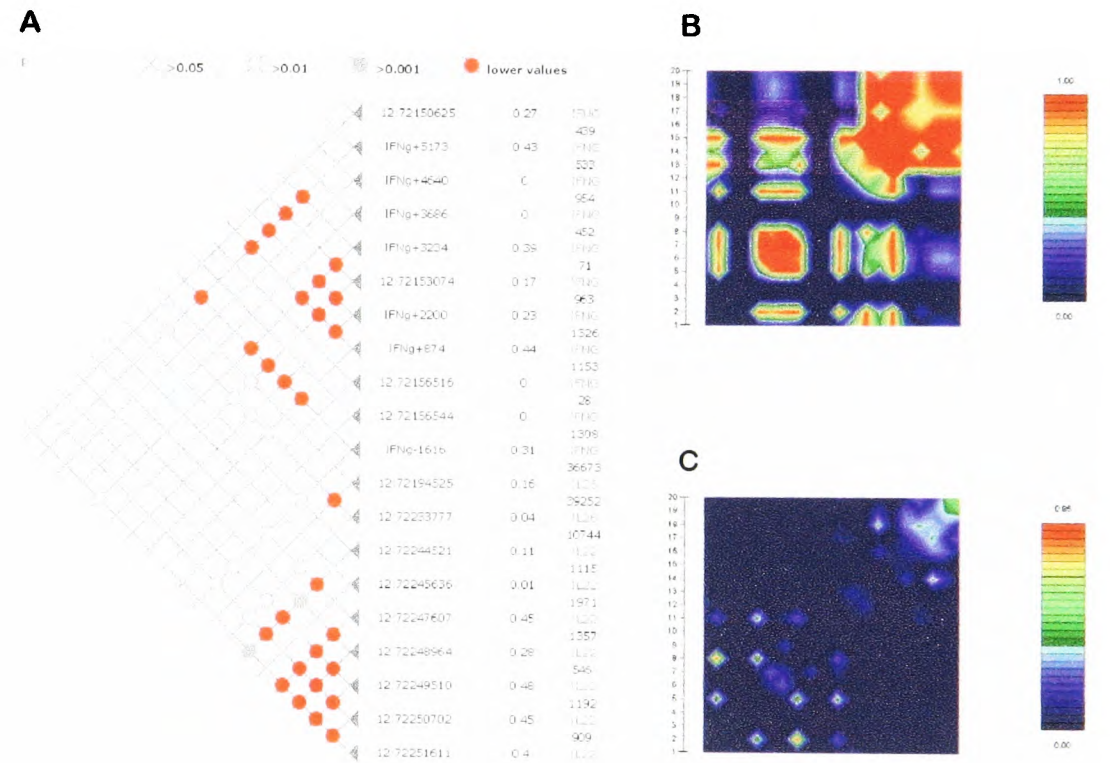


Figure 8.6 Linkage Disequilibrium between 20 polymorphisms in 12q14 in a UK Caucasian population. **A.** Pairwise measure of LD presented by p-value. Various degrees of significance are indicated by differently coloured circles. Open white circles represent p-values $0.05 \geq p > 0.01$, filled grey circles indicate $0.01 \geq p > 0.001$ and filled red circles represent $p \leq 0.001$. **B.** LD measured by D' illustrated with the graphical overview of linkage disequilibrium (GOLD) program. **C.** LD measured as Δ^2.

Koch 2003

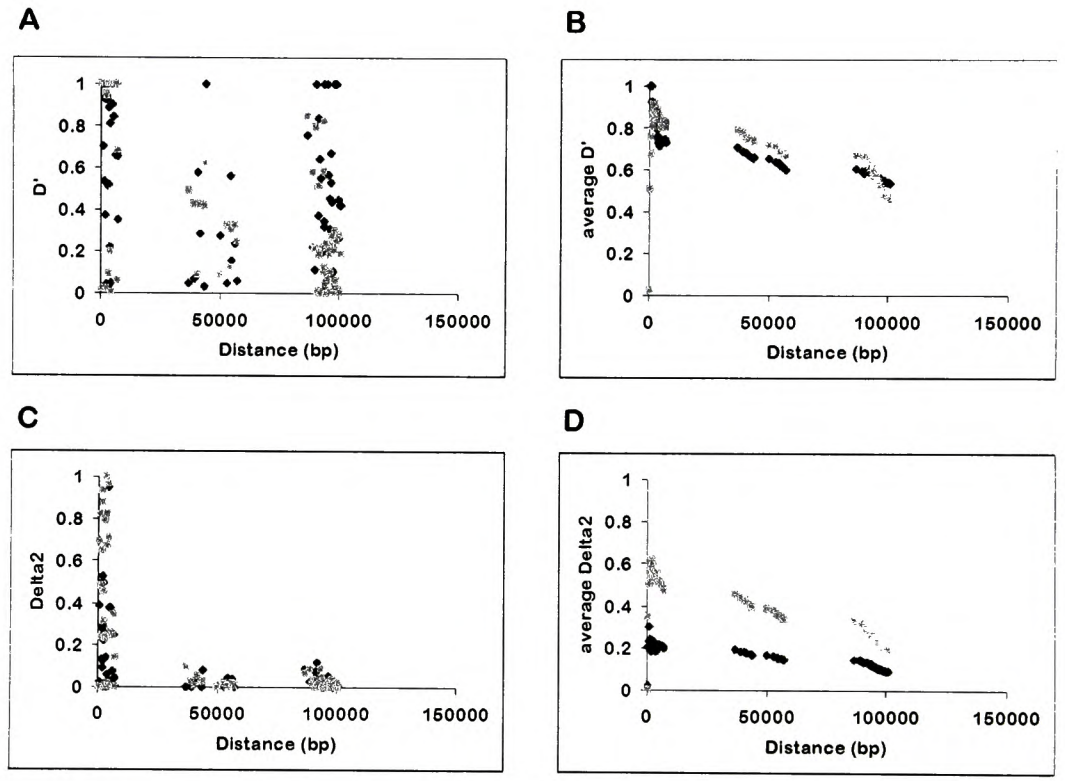


Figure 8.7 Linkage Disequilibrium versus physical distance in 12q14 for all pairs of markers. Only markers with an allele frequency of $f > 0.05$ were used for this analysis. Values for Gambian samples are represented by black squares and Caucasians by grey squares. **A.** LD as measured by D'. **B.** Average D'. **C.** LD measured as Δ^2. **D.** LD measured as average Δ^2.

8.2.5 Haplotype Blocks and Identification of htSNPs

As described previously in section 5.2.4 on page 165, haplotype blocks were constructed by entering the haplotypic information from 32 Gambian and 32 UK Caucasians into the `HaploBlockFinder` program. The chromosomal coverage that was aimed for in each haplotype block was 0.90. Haplotype blocks are confined to each gene region and do not extend beyond each of the three gene regions. Ten out of the 11 SNPs used in the *IFNγ* gene form one haplotype block both in Africans and Caucasians. Only the marker *IFNγ*+5612 is located outside this haplotype block. Similarly, another block is formed solely by *IL22* SNPs. While in Caucsians all 7 SNPs form one block, in Africans the markers *IL22*-485 and *IL22*-1394 lie outside of this block. Between these two genes, *IFNγ* and *IL22*, there is evidence for recombination. Haplotypes were also constructed for the two genes *IFNγ* and *IL22* separately (Tables 8.5 and 8.6). A total of 14 haplotypes were found using all available 11 SNPs for *IFNγ* in Africans. Nine of these haplotypes were also found in Caucasians. Haplotype no. 15 exists at about 8% in Caucasians while it is absent in the Africans samples analysed. Haplotype no. 1 is the dominant haplotype in Africans with a frequency of 38% while it has a frequency of only 6% in Caucasians. In Caucasians haplotypes no. 4 and 12 are the most frequent haplotypes with 33% and 10% respectively. Figure 8.10 shows the minimum mutation network of common *IFNγ* haplotypes, i.e. haplotypes that have a frequency of ≥ 0.05. Both in Africans and Caucasians it is of note that the structure of the network is linear, i.e. there is little sign for recombination. Furthermore almost all of connecting haplotypes exist at intermediate frequencies, maintaining haplotypic diversity in both populations. Six haplotypes make up almost 90% of all haplotypes in Africans. In Caucasians seven haplotypes make up 86% of all haplotypes. In *IL22* (Table 8.6) haplotype 1 is the most common haplotype both in Africans and Caucasians (43% and 52% respectively). While the seven most common haplotypes of *IL22* in Africans make up 88% in Africans, only three haplotypes in Caucasians make up 88% of all chromosomes.

ENTROPY was used to identify the htSNPs (Figures 8.12 and 8.13). In Africans a minimum of 5 SNPs were needed in *IFNγ* to bring the entropy to a plateau. There were two possible combinations of markers to achieve this, both consisting of the four markers *IFNγ*+5612, +3234, +2200 and -1616 and using either *IFNγ*+5173 or *IFNγ*+874 as the fifth marker. In Caucasians the combinations of the following five markers were possible to reach maximum entropy:

1. *IFNγ*+5612, +5173, +3234, +3163, -1616.
2. *IFNγ*+5612, +5173, +3234, +2200, -1616.
3. *IFNγ*+5612, +3234, +3163, +874, -1616.
4. *IFNγ*+5612, +3234, +2200, +874, -1616.

As predicted by the HaploBlockFinder analysis marker 1 needs to be included in all possible combinations. In view of previous studies, it is interesting that the marker *IFNγ*+874 could be substituted for other markers achieving the same coverage of information of this region. For *IL22* a total of 5 markers were needed to achieve maxinum entropy in Africans. The only combination of five markers to achieve this consisted of: *IL22*+4582, +2611, +708, -485 and -1394. In Caucasians only 2 SNPs were sufficient for maximal entropy, consisting of *IL22*+5697 and any of the other *IL22* markers apart from *IL22*+4582. This is a reflection of the limited number of haplotypes in for *IL22* in Caucasians.

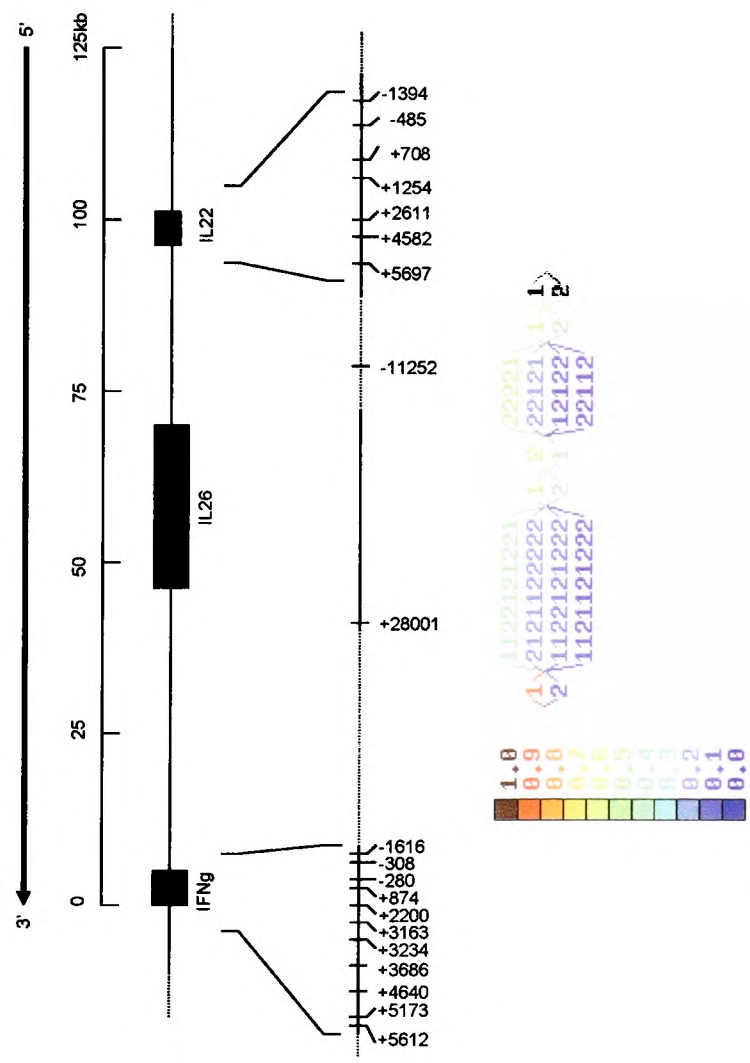


Figure 8.8 Haplotype Blocks of markers in 12q14 in Gambians. Haplotypes were generated by PHAMILY / PHASE and then analysed by the HaploBlockFinder program with chromosomal coverage set to 0.90. Haplotype frequencies are colour coded.

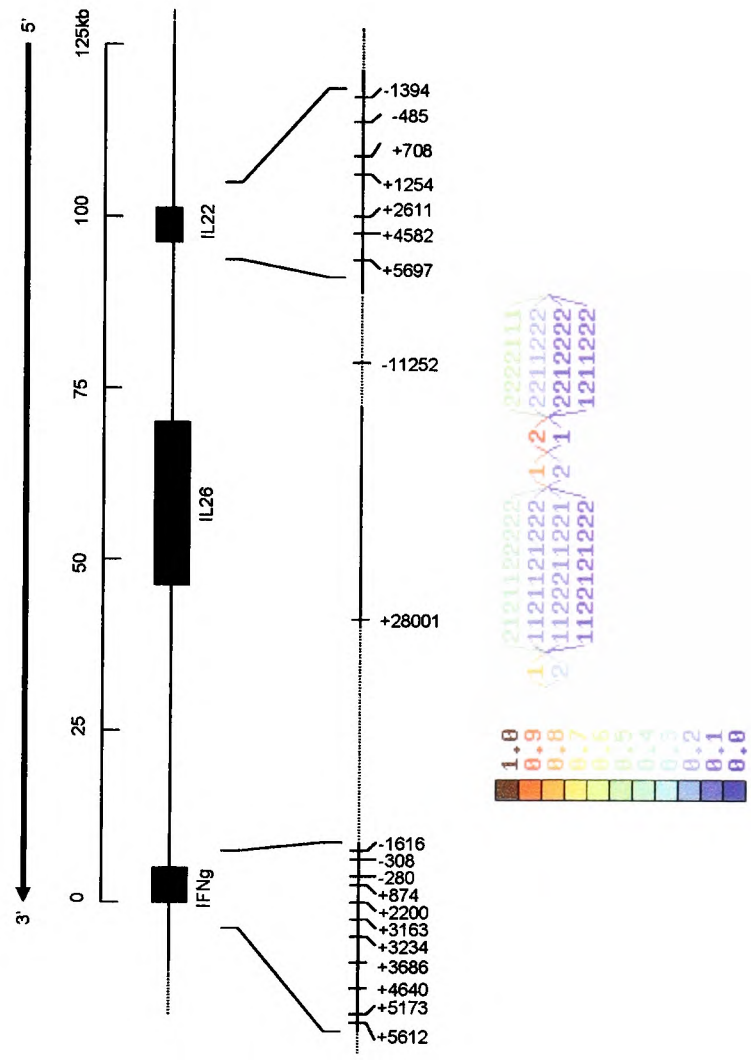


Figure 8.9 Haplotype Blocks of markers in 12q14 in Caucasians. Haplotypes were generated by PHAMILY / PHASE and then analysed by the HaploBlockFinder program with chromosomal coverage set to 0.90. Haplotype frequencies are colour coded.

8.3 Discussion

Interferonγ is a cytokine which has a key role in the immune response to many inflammatory stimuli. It has therefore been of great interest to investigate the role of genetic variants in the gene encoding for IFNγ. In the past these efforts have mainly focussed on one microsatellite in the first intron of the gene, ignoring the overall haplotype structure of other variants in the region. A more detailed approach was taken in order to dissect the haplotypic structure with higher resolution. Thereby the hope was to identify haplotype tagging SNPs (htSNPs) that would allow efficient association studies with a limited number of SNPs which would describe the majority of haplotypes observed in this area. The approach taken hereby relied on validated SNPs that were found in the public databases. Unlike a resequencing approach this can lead to a biased selection of markers as these will generally have higher allele frequencies. These SNPs might also have been identified in a limited set of individuals used in *population panels* to assess the allele frequencies of SNPs. The amount of haplotypes observed might therefore have been smaller than what one might have found using a resequencing approach. For future genetic studies choosing markers from a public databases will nevertheless be the most feasible, fastest and most cost effective way to establish haplotypic information about a particular region. The aim hereby will not be to completely capture the amount of genetic diversity in a region but rather by identifying haplotypes that are associated with disease to narrow down the search by focussing on particular chromosomes on which disease causing functional polymorphisms should be sought.

In this study a large number of validated SNPs were available for the regions immediately surrounding *IFN*γ and *IL22* while only little information was available for the region between these two genes with *IL26* located approximately half way in between. A statement about linkage disequilibrium for the region as a whole therefore has to be cautious. There have been reports about possible functional elements throughout the introns of

Table 8.5 Frequencies of *IFNγ* haplotypes in Africans and Caucasians. To facilitate viewing allele 2 is symbolised by '-' and allele 1 by 'o'. The markers order is in reverse. i.e. *IFNγ*+5612, +5173, +4640, +3686, +3234, +3163, +2200, +874, -280, -308 and *IFNγ*-1616.

IDENTIFIER	HAPLOTYPE	AFRICANS	CAUCASIANS
1	ooo--o-o--o	0.375	0.063
2	-oo--o-o--o	0.063	0.008
3	ooo--o-o---	0.156	0.070
4	o-o-oo-----	0.180	0.328
5	ooo--o-o-o-	0.023	0.000
6	ooo-oo-----	0.008	0.000
7	ooo--oooo-o	0.016	0.000
8	ooo-oo-o---	0.070	0.133
9	--o-oo-----	0.008	0.094
10	-oo-oo-o---	0.016	0.047
11	-o---ooo--o	0.008	0.000
12	ooo---oo--o	0.016	0.102
13	ooo--ooo--o	0.047	0.031
14	oooooo-o---	0.016	0.000
15	-oo---oo--o	0.000	0.078
16	o-o-o------	0.000	0.008
17	-oo--o-o---	0.000	0.008
18	-oo--ooo--o	0.000	0.023
19	ooo--o----o	0.000	0.008

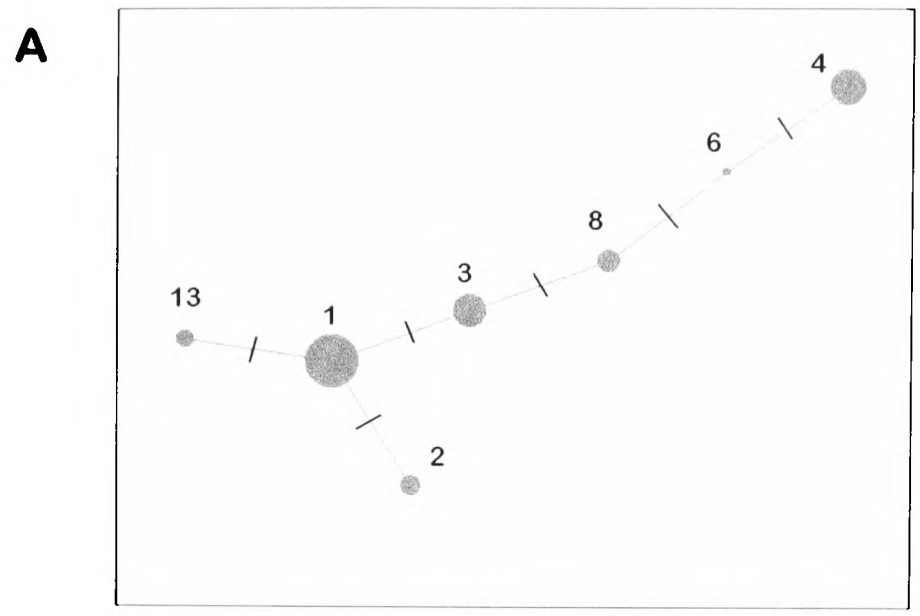


Figure 8.10 Minimum mutation network of common haplotypes in *IFN*γ in Africans (**A**) and Caucasians (**B**). The areas of the circles is proportional to the haplotype frequency. Haplotype identifiers refer to table 8.5. All haplotypes shown had a frequency ≥ 0.05. Haplotypes shown make up 89% and 86% of all haplotypes in Africans and Caucasians respectively.

Table 8.6 Frequencies of *IL22* haplotypes in Africans and Caucasians. To facilitate viewing allele 2 is symbolised by '-' and allele 1 by 'o'. The marker order is in reverse, i.e. *IL22*+5697, +4582, +2611, +1254, +708, -485 and *IL22*-1394.

IDENTIFIER	HAPLOTYPE	AFRICANS	CAUCASIANS
1	----ooo	0.430	0.516
2	-oo-ooo	0.047	0.000
3	--o-ooo	0.063	0.016
4	o-o---o	0.078	0.031
5	----o--	0.109	0.000
6	--oo---	0.078	0.297
7	o-oo---	0.023	0.070
8	--o-o-o	0.078	0.016
9	-o--ooo	0.008	0.008
10	--o-o--	0.047	0.000
11	--o----	0.008	0.008
12	o-o----	0.016	0.023
13	o-o-o--	0.008	0.000
14	--o---o	0.008	0.000
15	-----oo	0.000	0.016

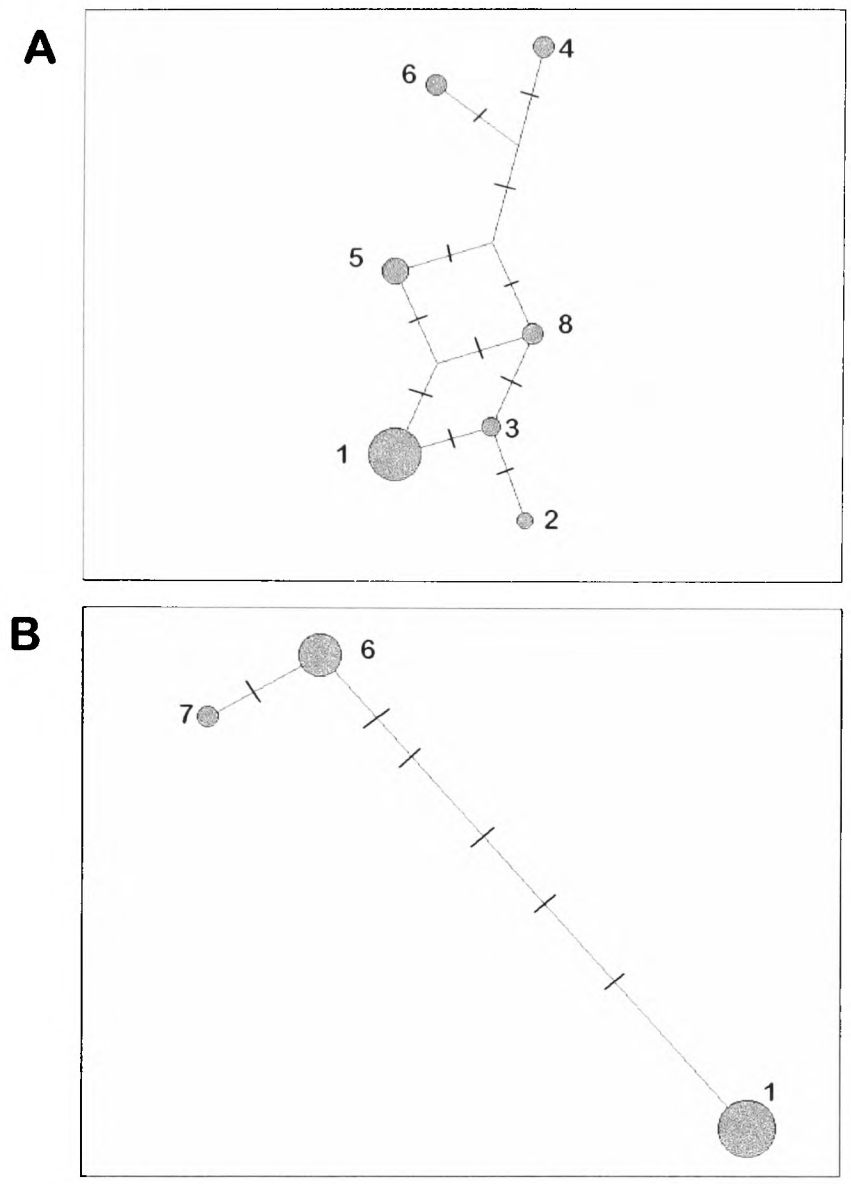


Figure 8.11 Minimum mutation network of common haplotypes in *IL22* in Africans (**A**) and Caucasians (**B**). The areas of the circles is proportional to the haplotype frequency. Haplotype identifiers refer to table 8.6. All haplotypes shown had a frequency ≥ 0.05. Haplotypes shown make up 88% of all haplotypes in Africans and Caucasians respectively.

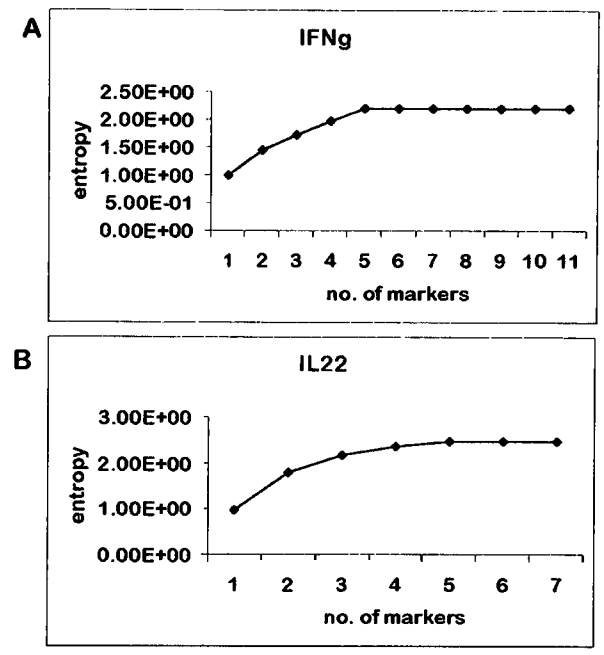


Figure 8.12 Entropy calculations for Gambian haplotypes determined by the greedy algorithm. Only haplotypes with $f \geq 0.05$ were used.

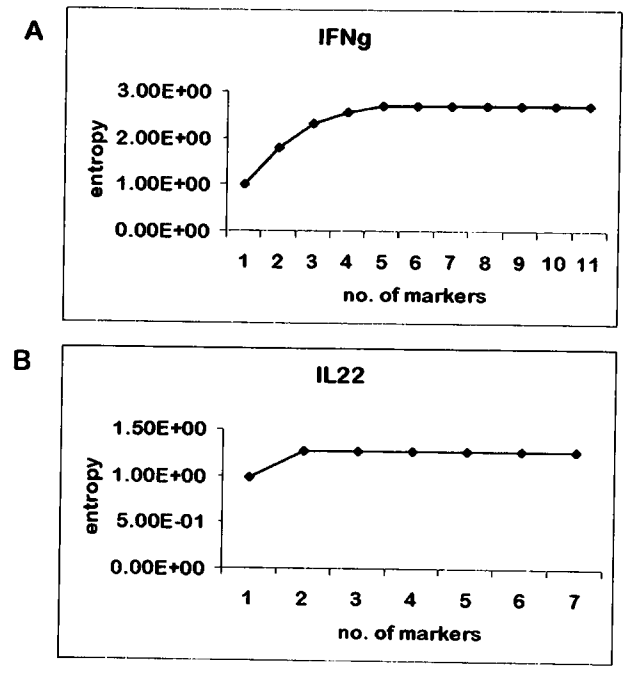


Figure 8.13 Entropy calculations for Caucasian haplotypes determined by the greedy algorithm. Only haplotypes with $f \geq 0.05$ were used.

the *IFNγ* gene. It would be expected that potential functional polymorphisms were located in reasonable proximity of any given gene, although this does not necessarily have to be the case as example of locus control regions show (Grosveld et al., 1987; Loots et al., 2000; Masternak et al., 2003). In the present study dense coverage of both the *IFNγ* and *IL22* gene was achieved. A total of 11 SNPs were spanning 7kb of sequence encompassing the *IFNγ* gene and a further 7 SNPs were selected across the *IL22* gene. The region in between these two genes spans approximately 85kb and two markers where used within this area. There was only little evidence for LD spanning far beyond the boundaries of either *IFNγ* and *IL22* while LD was well conserved within these two genes. Despite the obvious differences in marker ascertainment it is interesting to note that LD on chromosome 6q23 in Africans appears to reach further than in the 12q14 region in the same population. While the average Δ^2 at a distance of 100kb was around 0.4 for Africans in 6q23 it is less than 0.2 in 12q14. Furthermore, while in 6q23 the extend of LD was similar in Caucasians and Africans in 12q14, on the other hand, there is a marked difference in the level of average Δ^2 between these two populations. In the latter region LD seems to extend further in Caucasians which is similar to what other investigators have found (Reich et al., 2001). Furthermore in 12q14 Caucasians shared significantly more of the haplotypes found in Africans (41% compared to 25% in 6q23). This could reflect that haplotypic diversity at 12q14 is being maintained. Interestingly the minimum mutation networks representing the relationship of haplotypes for *IFNγ* are linear with most of the intermediate haplotypes preserved. Although purely speculative this could be a sign of natural selection acting at this locus.

While functionality has not been shown another variant site within the first intron of the *IFNγ* that has been implicated with susceptibility to disease is *IFNγ*+874. The present study, however, shows that this marker is in strong LD with several other markers within the *IFNγ* gene indicating the importance of obtaining a detailed structure of the haplotypes present. This study has identified haplotype tagging markers both in *IL22* and *IFNγ*

thereby making it possible to use a minimum of markers for subsequent association studies while capturing the great majority of genetic diversity at this locus.

8.4 Summary

1. In total 20 markers were selected across the genetic region of 12q14 with 11 markers spanning *IFN*γ alone.

2. Most markers were in high LD with other markers nearby, i.e. LD was highest within each of the three genes.

3. The average Δ^2 is higher for Caucasians than for Africans in this region of 100kb.

4. Haplotype blocks were confined to each gene region and did not extend beyond each of the three gene regions.

5. *IFN*γ+874, a marker that has been much investigated in previous association studies, was in strong LD with other markers in *IFN*γ. This underlines the shortcomings of single SNP analysis and highlights the importance of identifying the haplotypic structure of genes in a given population.

6. Haplotype tagging SNPs were identified. Various combinations of haplotype tagging SNPs in *IFN*γ were possible in order to capture the same amount of haplotypic diversity.

Chapter 9

Case Controls Study for Severe Malaria in 12q15

9.1 Introduction

IFNγ is one of the most crucial cytokines in the host response to the infection with *P.falciparum*. The postulated effector mechanisms involve the killing or developmental inhibition of the liver stages of the parasite (Hoffman et al., 1997). There is already some evidence for the genetic background to be important for IFNγ production and protection against the disease in mice (Shear et al., 1989) (for further details see chapter 3).

The analysis of htSNPs presented in chapter 8 yielded five biallelic markers spanning the *IFNγ* gene, five markers across the *IL22* gene and a further two markers surrounding *IL26*. These markers and additionally one multiallelic marker (microsatellite in the first intron) were used for a case-control study investigating susceptibility to severe malaria in a Gambian population.

The region under investigations spans about 110kb in total. The data presented in chapter 8 indicated that there is only little evidence for LD across the 100kb separating *IFNγ* and *IL22* which allowed a separate analysis of these two genetic areas. The microsatellite in the first intron of *IFNγ* has

been subject of several studies on susceptibility to other diseases previously (Khani-Hanjani et al., 2000; Awad et al., 1999; Cavet et al., 2001; Dabora et al., 2002; Nieters et al., 2001; Nagarkatti et al., 2002). All of these studies have neglected the haplotype structure of this gene in their analysis. Allele 2, with 12 CA-repeats, is associated with high levels of interferonγ production in vitro (Pravica et al., 1999) but is remains unclear whether this multiallelic marker itself has functional properties or is merely a marker for another functional allele. It will therefore be helpful investigate how the alleles of this microsatellite relate to the underlying haplotypic structure of the rest of the gene. The five bi-allelic markers used in this study were identified to represent the vast majority of haplotypes which are made up by the 11 *IFN*γ markers used in the previous chapter. As only half of these markers needed to be typed in the association study this study represents a cost efficient approach for association screening. The sample collection as presented in chapter 6 was used for this purpose.

9.2 Results

9.2.1 Investigation of Association of Individual Markers

9.2.1.1 Microsatellite

The registered DNA sequences of the human *IFN*γ genes deposited in Genbank show the presence of a CA repeat element 875bp downstream from the start of the first exon (Gray and Goeddel, 1982). This microsatellite in the first intron of the *IFN*γ was typed in the present study. The most common repeat number in Gambians is thirteen, with repeat lengths varying from 9 to 18 repeats. The allelic spectrum is far more diverse than in Caucasians where 12 repeats are slightly more common than thirteen (45% versus 44%) while other repeat lengths are far less common with fourteen repeats occurring in about 6% and fifteen in 5% of an Oxfordshire cord blood sample

collection (data not shown). In the present malaria case control study the frequency distribution in cases and controls were not significantly different (frequency distribution shown in Table 9.1 and Figure 9.1). The sample set was analysed using the CLUMP program. The normal χ^2 (T1) was 3.998827 and the χ^2 from the 2x2 table clumped to produce the maximum (T4) was 2.037379 (1000 simulations were run).

9.2.1.2 Bi-allelic Markers

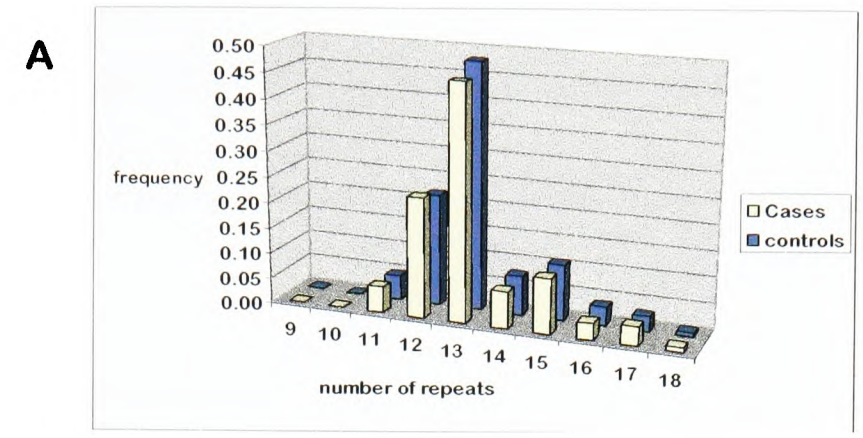
A further 12 SNPs were typed in the 12q14 region. Five of these SNPs were in *IFNγ*, two SNPs were surrounding *IL26* and a further five markers were located throughout the *IL22* gene. None of the *IFNγ* markers was associated with susceptibility to severe malaria in the overall group. In Mandinka the *IFNγ*+2200C allele was associated with susceptibility to severe disease (p=0.01; OR 2.29 (1.16-4.61 95%C.I.)), which was weakly significant after correcting for comparison of four different ethnic groups. The *IFNγ*-1616C allele was weakly associated with susceptibility to cerebral malaria in Mandinka (p=0.05; OR 1.36 (1.00-1.86 95%C.I.)). This was not statistically significant after applying the Bonferroni correction. In the overall group the *IL22*+708T allele was associated with protection against severe anaemia (p=0.01; OR 0.72 (0.56-0.93 95%C.I.)). The *IL22*-1394G allele was weakly associated with susceptibility to cerebral malaria (p=0.05; OR 1.24 (1.00-1.54 95%C.I.)). All markers were in Hardy-Weinberg equilibrium with the only exception of the low frequency marker *IL26*-11252 in the control group due to one more homozygote for the A allele than expected.

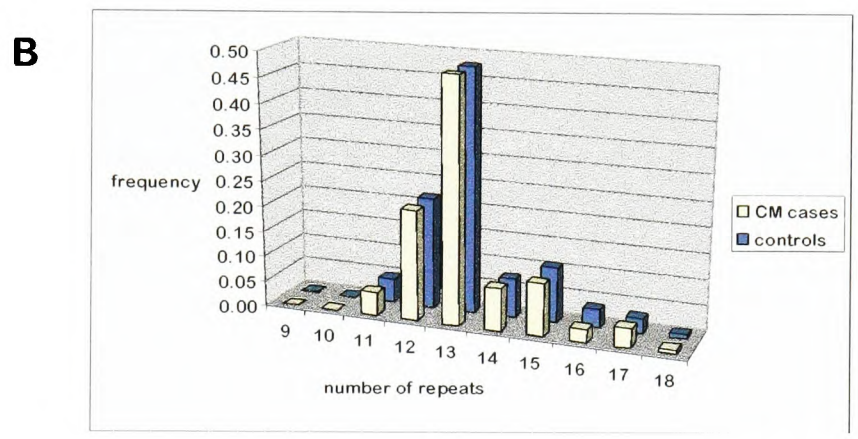
9.2.2 Haplotypes

Haplotypes were constructed with PHASE using the five bi-allelic *IFNγ* markers, *IFNγ*-1616, *IFNγ*+874, *IFNγ*+2200, *IFNγ*+3234 and *IFNγ*+5612, plus the multiallelic marker (*IFNγ*ms) in the first intron of the gene.
The *IFNγ*+3234T allele is in strong linkage disequilibrium with the *IFNγ*ms 12 CA repeat allele. The *IFNγ*ms 12 CA repeat allele falls onto two major

Table 9.1 Number of alleles of *IFNγ* CA-repeats in first intron. A total of 459 cord blood controls and 676 cases with severe malaria of which 400 had cerebral malaria (CM) alone and 263 had severe malarial anaemia (SA) alone were typed. Discrepancies in sample size are due to genotyping failure. Frequency distributions are shown in Figure 9.1.

No. of repeats	controls	All cases	CM cases	SA cases
9R	1(0.1%)	0(0%)	0(0%)	0(0%)
10R	0(0%)	0(0%)	0(0%)	0(0%)
11R	39(4.7%)	62(5.1%)	34(4.7%)	25(5.5%)
12R	181(21.7%)	283(23.3%)	158(21.6%)	116(25.4%)
13R	399(47.9%)	554(45.7%)	349(47.8%)	197(43.2%)
14R	63(7.6%)	88(7.3%)	61(8.4%)	27(5.9%)
15R	90(10.8%)	130(10.7%)	75(10.3%)	54(11.8%)
16R	30(3.6%)	40(3.3%)	19(2.6%)	19(4.2%)
17R	24(2.9%)	45(3.7%)	28(3.8%)	14(3.1%)
18R	6(0.7%)	10(0.8%)	6(0.8%)	4(0.9%)
Total	**833**	**1212**	**730**	**456**





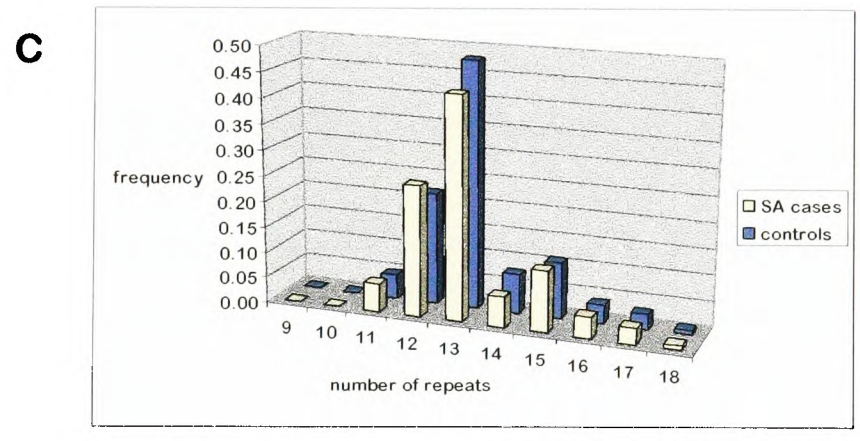


Figure 9.1 Frequency distribution of *IFN*-γ microsatellite alleles in cord blood controls, patients with severe malaria (all cases) (**A**), patients with cerebral malaria alone (CM cases) (**B**) and patients with severe malarial anaemia alone (SA cases) (**C**).

Koch 2003

Table 9.2 *IL22*-1394 (12:72251611); all ethnic groups

	Genotypes				Allele 1 vs. 2	
	aa	ag	gg	HWE	p	OR
cord blood controls	235	176	37	0.61		
cerebral malaria only	180	167	43	0.64	0.05	0.81
severe anaemia only	127	109	23	0.95	0.45	0.91
total no. patients	313	282	67	0.77	0.08	0.84

Table 9.3 *IL22*-485 (12:72250702); all ethnic groups

	Genotypes				Allele 1 vs. 2	
	aa	ag	gg	HWE	p	OR
cord blood controls	120	206	107	0.32		
cerebral malaria only	87	209	91	0.11	0.44	0.92
severe anaemia only	75	123	47	0.79	0.15	1.18
total no. patients	167	339	139	0.18	0.79	1.03

Table 9.4 *IL22*+708 (12:72249510); all ethnic groups

	Genotypes				Allele 1 vs. 2	
	cc	ct	tt	HWE	p	OR
cord blood controls	230	179	43	0.35		
cerebral malaria only	212	157	27	0.78	0.24	1.14
severe anaemia only	151	97	11	0.35	0.01	1.39
total no. patients	369	261	38	0.36	0.04	1.23

Table 9.5 *IL22*+2611 (12:72247607); all ethnic groups

	Genotypes				Allele 1 vs. 2	
	aa	ta	tt	HWE	p	OR
cord blood controls	105	224	124	0.84		
cerebral malaria only	97	182	115	0.14	0.98	0.99
severe anaemia only	50	130	78	0.75	0.25	0.87
total no. patients	147	319	199	0.37	0.42	0.93

Table 9.6 *IL22*+4582 (12:72245636); all ethnic groups

	Genotypes				Allele 1 vs. 2	
	cc	ct	tt	HWE	p	OR
cord blood controls	1	21	418	0.19		
cerebral malaria only	0	20	366	0.60	0.90	0.99
severe anaemia only	0	22	228	0.47	0.10	1.71
total no. patients	0	42	607	0.39	0.48	1.25

Table 9.7 *IL26*-11252 (12:72233777); all ethnic groups

	Genotypes				Allele 1 vs. 2	
	aa	ag	gg	HWE	p	OR
cord blood controls	3	30	400	0.01		
cerebral malaria only	0	21	357	0.58	0.17	0.66
severe anaemia only	1	14	227	0.14	0.53	0.79
total no. patients	1	38	593	0.63	0.27	0.75

Table 9.8 *IL26*+28001 (12:72194525); all ethnic groups

	Genotypes				Allele 1 vs. 2	
	aa	ag	gg	HWE	p	OR
cord blood controls	343	102	7	0.85		
cerebral malaria only	305	81	6	0.82	0.60	1.09
severe anaemia only	193	62	4	0.70	0.78	0.94
total no. patients	505	148	11	0.97	0.97	1.00

Table 9.9 *IFNγ*-1616 (12:72157852); all ethnic groups

	Genotypes				Allele 1 vs. 2	
	cc	tc	tt	HWE	p	OR
cord blood controls	119	206	120	0.12		
cerebral malaria only	105	201	84	0.51	0.27	1.12
severe anaemia only	63	134	59	0.45	0.79	1.04
total no. patients	169	344	146	0.25	0.42	1.08

Table 9.10 *IFNγ*+874 (12:72155363); all ethnic groups

	Genotypes				Allele 1 vs. 2	
	aa	ta	tt	HWE	p	OR
cord blood controls	323	109	11	0.62		
cerebral malaria only	292	88	7	0.90	0.76	1.14
severe anaemia only	169	76	6	0.45	0.20	0.82
total no. patients	467	170	14	0.74	0.83	0.97

Table 9.11 *IFNγ*+2200 (12:72154037); all ethnic groups

	Genotypes				Allele 1 vs. 2	
	cc	tc	tt	HWE	p	OR
cord blood controls	1	35	373	0.85		
cerebral malaria only	3	41	339	0.17	0.19	1.38
severe anaemia only	1	31	214	0.91	0.12	1.52
total no. patients	4	73	563	0.34	0.10	1.43

Table 9.12 *IFNγ*+3234 (12:72153003); all ethnic groups

	Genotypes				Allele 1 vs. 2	
	cc	tc	tt	HWE	p	OR
cord blood controls	37	174	234	0.56		
cerebral malaria only	30	148	211	0.57	0.64	0.94
severe anaemia only	17	109	128	0.33	0.96	1.01
total no. patients	49	265	342	0.81	0.96	0.99

Table 9.13 *IFNγ*+5612 (12:72150625); all ethnic groups

	Genotypes				Allele 1 vs. 2	
	cc	tc	tt	HWE	p	OR
cord blood controls	182	179	49	0.66		
cerebral malaria only	165	152	41	0.51	0.69	1.05
severe anaemia only	118	103	17	0.39	0.07	1.26
total no. patients	292	259	58	0.96	0.17	1.15

haplotypes (no.2 and no.6 in table 9.14). These two haplotypes only differ at position *IFN*γ+874. The *IFN*γms 13 CA repeat allele falls onto four major haplotypes (no. 1, 3, 5 & 7). Haplotypes that have an allele frequency $f \geq 0.05$ are shown in table 9.14. None of the eight major haplotypes found showed any association with either of the two forms of severe malaria investigated. Haplotype analysis was also performed without the microsatellite marker resulting in six major haplotypes (with $f \geq 0.05$). None of these haplotypes were associated with disease.

Haplotypes were also constructed using the five *IL22* markers typed, *IL22*-1394, *IL22*-485, *IL22*+708, *IL22*+2611 and *IL22*+4582 (see table 9.15). These five markers form six major haplotypes with allele frequencies $f \geq 0.05$. Haplotype 3, as defined by the *IL22*-1394A, -485G, +708T, +2611A and +4582T alleles, was associated with protection against severe malaria (p=0.004; OR 0.68 (0.52-0.89 95% C.I.)). This was statistically significant even after correcting for six comparisons (six haplotypes) which changes the threshold for significance to $p < 0.008$. The haplotype analysis presented in chapter 8 showed that this haplotype also carries the major allele (A) of *IL22*+1254 and the minor allele (T) of *IL22*+5697. A second haplotype, haplotype no.4 as defined by the *IL22*-1394G, -485G, +708C, +2611T and +4582T alleles, was associated with susceptibility to severe malaria (p=0.006; OR 1.44 (1.11-1.87 95% C.I.)). This haplotype is also defined by the major allele (A) of *IL22*+1254 and the major allele (A) of *IL22*+5697. The two haplotype therefore differ at four out of seven positions.

9.3 Discussion

The cytokine interferonγ is a crucial mediator of the host response to many inflammatory and infectious stimuli. Many association studies have recognised this and tested the microsatellite marker in intron 1 for association with several diseases. While microsatellite markers might potentially be more informative as they usually consist of more than two allele and therefore poten-

Table 9.14 All ethnic groups. Haplotypes of *IFNγ* using the following markers: *IFNγ*-1616 (C=1; T=2), *IFNγ*+874 (A=1; T=2), *IFNγ*ms, *IFNγ*+2200 (C=1; T=2), *IFNγ*+3234 (C=1; T=2) and *IFNγ*+5612 (C=1; T=2). Only haplotypes with an allele frequency of $f \geq 0.05$ in at least one of the groups are shown.

ID	HAPLOTYPE	CONTROLS	ALL CASES	CM	SA
1	2 1 13 2 2 2	176 (19.2%)	238 (17.6%)	140 (17.5%)	96 (18.3%)
2	2 2 12 2 1 1	125 (13.6%)	197 (14.6%)	100 (12.5%)	89 (16.9%)
3	1 1 13 2 2 2	105 (11.4%)	155 (11.5%)	101 (12.6%)	46 (8.7%)
4	1 1 15 2 2 1	101 (11.0%)	156 (11.5%)	82 (10.3%)	69 (13.1%)
5	1 1 13 2 2 1	96 (10.5%)	124 (9.2%)	78 (9.7%)	53 (9.2%)
6	2 1 12 2 1 1	56 (6.1%)	81 (6.0%)	52 (6.5%)	31 (5.9%)
7	1 1 13 1 2 1	38 (4.1%)	80 (5.9%)	46 (5.8%)	33 (6.3%)
8	1 1 14 2 2 1	37 (4.0%)	61 (4.5%)	42 (5.3%)	18 (3.4%)
other		184 (20.0%)	260 (19.2%)	159 (19.9%)	91 (17.3%)
total		**918**	**1352**	**800**	**526**

Table 9.15 All ethnic groups. Haplotypes of *IL22* using the following markers: *IL22*-1394 (A=1; G=2), *IL22*-485 (A=1; G=2), *IL22*+708 (C=1; T=2), *IL22*+2611 (A=1; T=2) and *IL22*+4582 (C=1; T=2). Significant differences between the case and the control group are marked by an asterisk.

ID	HAPLOTYPE	CONTROLS	ALL CASES	CM	SA
1	1 1 1 2 2	375 (40.8%)	519 (38.4%)	291 (36.4%)	211 (40.1%)
2	2 2 2 1 2	136 (14.8%)	202 (14.9%)	124 (15.5%)	72 (13.7%)
3	1 2 2 1 2	132 (14.4%)	139 (10.3%)*	89 (11.1%)*	49 (9.3%)*
4	2 2 1 2 2	102 (11.1%)	206 (15.2%)*	124 (15.5%)*	80 (15.2%)*
5	1 1 1 1 2	79 (8.6%)	146 (10.8%)	85 (10.6%)	62 (11.8%)
6	1 2 1 1 2	55 (6.0%)	75 (5.5%)	54 (6.8%)	22 (4.2%)
7	1 1 1 1 1	23 (2.5%)	42 (3.1%)	20 (2.5%)	22 (4.2%)
8	2 2 1 1 2	16 (1.7%)	20 (1.5%)	11 (1.4%)	8 (1.5%)
9	1 2 1 2 2	0	2 (0.1%)	1 (0.1%)	0
10	2 1 1 2 2	0	1 (0.1%)	1 (0.1%)	0
total		**918**	**1352**	**800**	**526**

tially represent a greater number of haplotypes, it is precisely this haplotypic structure that has thus far been neglected. At the time of writing these thesis, 35 validated bi-allelic markers, spanning 7kb of the *IFNγ* gene, its promoter and 3'UTR region, were available in the public databases. This allows a rapid and cost efficient approach to screening this gene for association with disease. For this study only markers that had a reported allele frequency of $f \geq 0.05$ in at least one population were chosen. While this approach might miss the *real* disease allele by neglecting low frequency alleles, it possibly presents the most powerful approach given the size of current sample collections. In fact only seven of the 11 *IFNγ* markers originally chosen were found to have the required minimum allele frequency in the two populations used in this study. Haplotype analysis had indicated that five of these markers are sufficient to capture the haplotypic diversity created by the total of 11 markers at this locus. As most of the previous *IFNγ* association studies had reported on the microsatellite only it was crucial to put the haplotypic structure created by the bi-allelic markers in context with the microsatellite that was therefore typed in case control study.

It is interesting that the microsatellite allele with 12 CA-repeats, which is associated with high levels of interferonγ production in vitro (Pravica et al., 1999), is less common in Gambians than in Caucasians. Pravica et al. (1999) have measured IFNγ production in supernatants of the cultures of peripheral blood mononuclear cells (PBMC) from randomly selected (mostly Caucasian) individuals. It would be interesting to repeat this experiment using PBMCs from Africans where the haplotypic background and allelic frequencies of the microsatellite marker is in fact quite different. Ninety percent of Caucasians have either 12 or 13 CA repeats. These two alleles might, however, lie on a variety of genetic backgrounds. Assuming that the microsatellite is not functional itself, the association of the 12 CA repeat allele might therefore only be valid in a particular population.

Based on a non-randomly selected sample set of mostly Caucasians Pravica et al. (2000) suggested an absolute correlation between the presence of

the *IFNγ*+874T allele and the presence of the high-producing 12CA repeat allele, i.e. absolute linkage disequilibrium between these two alleles. They suggested that the *IFNγ*+874 site which they found to be in complete LD with the 12CA repeat allele might in fact be functional as it could be a putative NF-κB site. It is interesting that the present study of African samples could not identify a strong relationship between the *IFNγ*+874 alleles and any of the *IFNγ*ms alleles. This might be due to the greater haplotypic diversity of Africans where the relationship between these two markers might have broken down. While some putative associations were found in one of the ethnic groups, the Mandinka, in the overall group none of the *IFNγ* markers nor any of the haplotypes were found to be associated with severe malaria. This represents an important negative finding. *IFNγ* is one of the major cytokines in the host response to the infection with *P.falciparum*. Many studies have shown that differential expression of this cytokine alters disease outcome. It might therefore be expected that strong selective pressure as represented by malaria might lead to functional variations in the *IFNγ* gene. It is certainly possible that a putative *real* functional polymorphism was missed in this study. However, in this study five bi-allelic markers plus one microsatellite were typed which thereby represents the densest haplotypic analysis of this gene thus far. Both the microsatellite and *IFNγ*+874 have been implicated in functional variation of the gene and no association could be found for either of these two markers.

As the haplotypic analysis of 12q14 showed hardly any LD between the *IFNγ* and *IL22* markers, these two regions were analysed separately. Five markers across the *IL22* gene, encoding for a cytokine structurally related to IL10, were typed. It was found that the *IL22*+708T allele was associated with protection against severe anaemia (OR 0.72) and the *IL22*-1394G allele was weakly associated with susceptibility to cerebral malaria (OR 1.24). The subsequent haplotype analysis showed that this might in fact be due to a haplotypic association. Haplotype 3, which carries the *IL22*-1394A and the *IL22*+708T alleles, was associated with protection against disease while haplotype 4, which carries the *IL22*-1394G and the *IL22*+708C alleles, was

associated with susceptibility to severe disease. There is only limited knowledge about the role of IL22 in the immune response. IL22 was found to activate STAT1 and STAT3 and up-regulate the production of acute phase reactants. The results of this case control analysis might, however, warrant further and more detailed investigations of this region. Further markers on these two haplotypes should be sought in order to identify any potentially functional allele. This could be achieved by sequencing homozygotes for these two haplotypes and identifying further polymorphisms on their background.

9.4 Summary

1. The allelic spectrum of the *IFN*γms marker is more diverse in Africans than in Caucasians.

2. No significant associations of any of the *IFN*γms alleles with severe malaria were found.

3. In the overall group none of the the biallelic *IFN*γ markers was associated with severe malaria.

4. In Mandinka two of the *IFN*γ markers were weakly associated with severe disease. However, these results would not be statistically significant if full corrections were made for multiple comparisons, and so should be regarded as tentative rather than definitive associations.

5. The *IFN*γms 12 CA repeat allele is in strong LD with another biallelic marker in the gene.

6. The *IFN*γms 12 CA repeat allele falls onto two major haplotypes which differ at position *IFN*γ+874, a marker that had previously been shown to be in absolute LD with this ms allele.

7. Two of the *IL22* markers were associated with susceptibility to severe disease. However, these results would not be statistically significant if

the Bonferroni correction was applied taking into account the number of markers typed.

8. Two *IL22* haplotypes were strongly associated with susceptibility to and protection against severe malaria respectively.

Коси 2003

Chapter 10

Discussion

Before considering the statistical and biological significance of the results found in this study, it is worth setting this in the context of current views about the genetic basis of susceptibility to complex diseases. There has been a subtle but important change in our understanding of this problem since work on this thesis was commenced, and it is reflected in the results that were obtained. The prevailing view of four years ago might be summarised as follows.

Firstly, though the full extent of human genomic diversity was unknown, it was widely believed that a relatively small number of polymorphic markers might be adequate to screen for genetic associations with common diseases. One common estimate at the time of foundation of the SNP Consortium was that in the order of 100,000 SNP markers (i.e. an average of 3 SNPs per gene) might suffice to screen the whole human genome for disease association. It is clear from the data presented in this thesis, and from many other recently published studies, that this is a huge underestimate. Not only are there far more common polymorphisms in the human genome than many commentators expected (the total is likely to be in the order of 10 million) but there is also more haplotypic diversity, particularly in African populations, so that a large number of SNPs may be required for comprehensive analysis of a single gene region.

A second prevailing view, at the time when work on this thesis was started, was that functional genetic variants might be expected to have a fairly marked effect on susceptibility to common diseases - for example, at least a doubling or halving of risk. Undoubtedly there exist genetic effects of this magnitude, but the last few years have led to a growing appreciation that smaller effect sizes are both more common and, if considered together, may be of at least as much biological importance. This is a natural consequence of the very large number of different genetic and environmental factors that are now thought to underlie a complex disease such as malaria, which means that even if a gene exerts a functionally important role in the biological process of disease immunity, individual polymorphisms in that gene may have a rather subtle epidemiological effect when analysed against the background of all other genetic and environmental variants that also determine disease outcome. Thus there is growing appreciation of the need to detect genetic effects of relatively modest magnitude. Some leading laboratories are now aiming to detect allelic risk ratios of 1.5, and even smaller effect sizes are likely to be investigated in the future.

A third shift in opinion over the past few years has been an increasing recognition of the phenotypic complexity of most common diseases, and the realisation that functional genetic effects may be inapparent unless the investigators consider precise phenotypic subcategories (e.g. cerebral malaria versus severe malarial anaemia) rather than blanket definitions of a common disease (e.g. severe malaria).

If viewed in the above context, what can be concluded from the results of this thesis about the effect of variation in *IFNGR1* and *IFN*γ on susceptibility to malaria? With a sample size of around 600 cases and 600 controls, we find a number of associations with uncorrected P values in the region of 0.01 and occasionally 0.001. The most interesting associations are with the Mandinka, who are the largest ethnic group in The Gambia, so for these associations it is important to interpret these P values knowing that 4 to 5 ethnic groups were analysed separately. If we additionally consider the total number of SNPs and haplotypes that were analysed, it is clear that we cannot conclude that any particular SNP or haplotype is unequivocally associated with susceptibility

to malaria. However it is equally clear that we cannot exclude a potentially important disease association with this region - these are exactly the sort of results that we might expect if one or more of these markers had a true allelic risk ratio in the region of 1.5.

Thus one of the main conclusions of this thesis is to support the growing view that genetic epidemiological studies of common diseases need to be of sufficient size to be able to detect allelic risk ratios in the order of 1.5 for specific phenotypic subcategories of the disease, after correcting for multiple comparisons with many thousands of genetic markers. This requires very large multicentre studies: a typical sample size estimate would be 4000 cases and 4000 controls. Such epidemiological work on malaria susceptibility is in progress, but for reasons of timescale and other practical factors it was not feasible to take this investigation to that scale of analysis during this project. Instead in this thesis scientific efforts have focused on two of the other fundamental requirements for association mapping of a gene region, namely to get a comprehensive picture of its true diversity in the population in question, and to try to identify variants with functional effects on gene regulation.

Many known factors, such as incomplete penetrance and environmental exposures, greatly hamper association mapping for complex diseases. Current evidence suggests that common complex diseases are affected by many loci, at each of which multiple alleles typically have only small effects. Alleles with with large, i.e. fatal, effects usually have very low alleles frequencies, at least in part due to natural selection (Pritchard, 2001).

If any one SNP leads to a significant increase for developing a certain disease, an association might also be seen between that risk and SNPs in linkage disequilibrium with the causal SNP. SNPs that are located in relatively close proximity to each other might form a limited number of haplotypes, i.e. less than theoretically possible, due to linkage disequilibrium. Thanks to linkage disequilibrium not every single SNP found in the genome will need to be typed. Because co-occurring sites carry redundant information, the use of haplotypes defined by a subset of the sites can reduce the number of SNP markers that have to be genotyped to detect disease associated variants.

Thus far, various studies have given very different estimates for the extent of LD in different chromosomal regions. But the differences also depend on the population studied, on the number of sample analysed, on the parameters used to define LD, on the allele frequencies of the markers used. All these factors currently make it difficult to give any *a priori* estimate of LD in any given region. The availability of an ever increasing number of markers in public databases and systems allowing fast and cheap ways of typing these markers might make it possible in future to create LD maps on larger scales. Understanding the patterns of LD and the haplotypic structure of the human genome will be crucial in designing future studies. Until then an informed approach to certain candidate regions for which there is prior knowledge that the gene product is involved in pathogenesis seems likely to be most fruitful. The candidate gene approach, as chosen in this study, can be defined as the study of the genetic influences on a complex trait. Hypotheses are generated about the aetiology of the disease, and genes identified that might play a role in it. Variants in or near those genes are identified that either cause change in the protein or its expression, or are in linkage disequilibrium with these changes. These variants are then typed in a population and the correlation between genotype and phenotype analysed. This approach assumes, however, that functional changes are located in reasonable genetic proximity to these genes. In future, evolutionary cross-species comparisons (Sidow, 2002) may help in identifying regulatory regions important for gene function that are located at some distance to known genes.

One of the major challenges remains how to identify and choose markers. While it is easier (and cheaper) to find and use common genetic variants, these SNPs also tend to be older and therefore had more time to recombine which can lead to LD breaking down.
While resequencing areas of interest has the advantage of generating information that can be used to study any deviation from the neutral model, e.g. due to natural selection, this is currently still time consuming and expensive. Unlike in Mendelian diseases, causal alleles for complex diseases might be expected to contribute only a small amount of added risk. The sensitivities

of tests for natural selection, such as Tajima's D, might not be high enough to pick up the effects these low risk alleles have. On the other hand a resequencing approach will yield most of the markers. It is also not clear how much information of any given region is needed.
If alleles with large effects on disease risk turn out to be rare and geographically localised, rather than common and global, a haplotype map based on common SNPs might not be particularly powerful for mapping those alleles. It might furthermore be that maps have to be far denser, i.e. markers more numerous and denser and even rarer SNPs must be included, to capture the haplotypic diversity of Africans with the same effectiveness as in Europeans. A recent comprehensive sequencing effort analysed 313 genes and 720kb of genomic sequence (Stephens et al., 2001a). Only 2% of the SNPs identified are in the public dbSNP database, suggesting that there are many more SNPs than can be found in this database.

Two slightly different approaches were chosen for the *IFNGR1* region on chromosome 6q23 and the *IFN*γ region on chromosome 12q14. The availability of a dense collection of validated SNPs for 12q14 in the public database made it possible to initially approach this region without resequencing. It remains unclear whether any information was missed by this approach. Nevertheless, the haplotypic structure obtained in this study is the most detailed for these regions to date and might be helpful for future association studies. In fact, we were able to use this haplotypic information by typing only the haplotype tagging SNPs in a study on the genetic susceptibility to trachoma in The Gambia that was set up in collaboration with the London School of Hygiene and Tropical Medicine. A major haplotype that confers protection to the disease was thereby identified. This haplotype can now be further analysed, e.g. by sequencing carriers and identifying further, possibly functional polymorphisms along this haplotype. Similarly to the approach to chromosome 6q23, there were relatively large gaps between some of the SNPs. While it might seem conceivable that more detailed information about the behaviour of LD in this region could be gained by typing evenly spaced marker, detecting LD over large region as found for the *IFNGR1*-470 in 6q23

might raise some suspicion which can be further analysed in future studies. It might also help in determining if further work can focus on one candidate gene for which an association has been found or whether LD with markers in neighbouring genes needs to be taken into consideration.

While the associations found for 6q23 are encouraging, they have to be taken with caution. Analyses of further populations are desirable. To date the *IFNGR1*-56 and *IFNGR1*-470 have been typed in a large case control study in Malawi consisting of roughly 880 cases and 500 cord blood controls controls. The same phenotype definitions as for the Gambian study were used. The *IFNGR1*-56C allele was associated with protection against malaria (p=0.004; OR=0.79 (95% C.I. 0.66-0.93)). This result is encouraging and is currently being further pursued. The *IFNGR1*-470, on the other hand, was not associated with susceptibility to malaria (p=0.66).

One of the problems with a candidate gene approach has been the failure to replicate results. Different studies might use different study designs and different definitions of phenotypes. The same candidate gene or SNP might be associated with different relative risks in different populations. A non-replicating result might therefore stem from a true biological difference. The failure to replicate might also be due to the small magnitude of the relative risks that are likely to be detected in candidate gene studies of complex diseases. If the linkage disequilibrium does not exist or varies across populations, different studies might have different findings for the same gene.

Genetic associations for complex, i.e. multifactorial, diseases can be expected to be much weaker than for diseases of mendelian inheritance. Multiple weak effects might ultimately alter disease outcome. A strict correction for multiple comparisons might make these associations even harder to detect. It will therefore be crucial to establish larger studies of several ten thousand samples in order to eventually overcome these limitations.

The association study in this thesis was followed up by functional investigations. It appears that the site surrounding the *IFNGR1*-470 is of regulatory importance. The two alleles could have different biological roles depending on their cell environment. Interferon-γ and its receptor have a role both in developing immunity/protection against the malaria parasite *and* in the progression of the pathology of the disease. While many studies have described this differential role of the cytokine and its receptor, only very little is known about what might regulate these events on a molecular level. The results of this study indicate that there might be molecular mechanisms that lead to differential expression of the molecule depending on the tissue type under investigation. This is particularly interesting in the case of the *IFNGR1*-470 site as it is polymorphic and therefore adds an additional degree of variability.

The completion of the first draft of the human genome was a major milestone in unravelling our genetic make-up. But major challenges still lie ahead in order to make sense of this vast amount of information. Despite the apparent complexity of the susceptibility to complex diseases, careful optimism seems warranted. Recent advancements in technology for genotyping and analysis of functional regions of the genome have dramatically increased the speed at which further investigation can be carried out. Our growing understanding of genetic diversity will help us to understand why people can differ so significantly in their response to the same pathogen. Not only might this help in predicting outcome, but ultimately aid drug targeting or development of vaccines against diseases such as malaria that still kills more than a million children every year.

Appendix A

Abbreviations

Abbreviation	Explanation
ARMS	Amplification refractory mutation system
BCG	Bacillus Calmette-Guerin
CEPH	Centre d'Etude du Polymorphisme Humain
CI	Confidence interval
CM	Cerebral Malaria
Con A	Concanavalin
CTL	Cytotoxic T-lymphocytes
ddNTP	dideoxynucleotide
DEAE	Diethylaminoethyl
DIC	Disseminated Intravascular Coagulopathy
DNA	Deoxyribonucleic acid
dNTP	deoxynucleotide
dsDNA	Double stranded DNA

continued on next page

continued from previous page

ABBREVIATION	EXPLANATION
EBV	Epstein Barr Virus
EDTA	Ethylenediaminetetraacetic acid
EMSA	Electro mobility shift assay
FCS	Fetal calf serum
F_{ST}	Fixation index
GAS	Gamma activated site
Hb	Haemoglobin
HKA-Test	Hudson Kreitman and Aguade Test
HLA	Human leukocyte antigen
HSP	Heat shock protein
htSNP	Haplotype tagging SNP
HWE	Hardy-Weinberg Equilibrium
ICAM	InterCellular Adhesion Molecule
IFNγ	Interferonγ (protein)
IFNγ	gene encoding for Interferonγ
IFNGR1	gene encoding for interferonγreceptor 1
IFNγR1	Interferon gamma receptor 1 (protein)
IFNGRAC	Interferon gamma receptor 1 (protein)
Ig	Immunglobulin
IL	Interleukin
ISRE	Interferon -stimulated response element
JAK	Janus kinase

continued on next page

continued from previous page

ABBREVIATION	EXPLANATION
kb	Kilobase
LD	Linkage Disequilibrium
LDR	Ligase detection reaction
LPL	Lipoproteinlipase
LPS	Lipopolysaccharide
MALDI-TOF	Matrix assisted laser desorption ionization time of flight mass spectrometry
MHC	Major histocompatibility complex
MRC	Medical Research Council
mRNA	Messenger ribose nucleic acid
ms	Microsatellite
NK cell	Natural killer cell
OR	Odds ratio
PBS	Phosphate-buffered saline
PCR	Polymerase chain reaction
PCV	Packed cell volume
PEP	Primer extension preamplification
PGE2	Prostaglandin E2
PMA	Phorbol 12-myristate 13-acetate
RNA	Ribose nucleic acid
rRNA	Ribosomal RNA
RT-PCR	Reverse transcriptase PCR

continued on next page

continued from previous page

ABBREVIATION	EXPLANATION
SA	Severe malarial anaemia
SDS	sodium dodecyl sulphate
SNP	Single nucleotide polymorphism
STAT	Signal transducer and activator of transcription
TAE	Tris-acetate-EDTA
TBE	Tris-Borate-EDTA
TDT	Transmission disequilibrium test
TE Buffer	Tris-Cl / EDTA Buffer
TEMED	Tetramethylethylenediamine
TNF	Tumour necrosis factor
USD	US Dollar
UTR	untranscribed region
UV	Ultra-violet
WHO	World Health Organization
WTCHG	Wellcome Trust Centre for Human Genetics

Appendix B

List of Suppliers

Company	Address
Abgene UK	ABgene House, Blenheim Road, Epsom, KT19 9AP, UK
American Type Culture Collection (ATCC)	P.O. Box 1549, Manassas, VA 20108, USA
Amersham plc	Amersham Place, Little Chalfont, Buckinghamshire, HP7 9NA, U.K.
Applied Biosystems (ABI)	Lingley House, 120 Birchwood Boulevard, Warrington WA3 7QH, U.K.
Beckman Coulter (UK) Ltd.	Oakley Court, Kingsmead Business Park, London Road, High Wycombe, HP11 1JU, U.K.

continued on next page

continued from previous page

COMPANY	ADDRESS
Bioline Ltd.	16 The Edge Business Centre, Humber Road, London NW2 6EW, U.K.
Bio-Rad Laboratories	1000 Alfred Nobel Drive, Hercules, CA 94547, USA
Corning Incorporated	45 Nagog Park, Acton, MA 01720, USA
Eppendorf AG	Barkhausenweg 1, 22339 Hamburg, Germany
Genetix Limited (GenPak)	Queensway, New Milton, Hampshire, BH25 5NN, U.K.
Gibco	see *Invitrogen*
Greiner Bio-One Ltd. (Greiner)	Brunel Way, Stroudwater, Business Park, Stonehouse, Glos., GL10 3SX U.K.
Invitrogen Ltd (Invitrogen)	3 Fountain Drive, Inchinnan Business Park, Paisley, U.K.
Labtek	91 Lost Lake Lane, Campbell, California 95008, USA
Merck Biosciences Ltd. (Novagen)	Boulevard Industrial Park, Padge Road, Beeston, Nottingham NG9 2JR, U.K.

continued on next page

continued from previous page

COMPANY	ADDRESS
Metabion	Lena-Christ-Strae 44, 82152 Planegg-Martinsried, Germany
Microsoft Limited	Microsoft Campus, Thames Valley Park, Reading, RG6 1WG, U.K.
Millipore (U.K.) Limited	Units 3&5 The Courtyards, Hatters Lane, Watford, WD18 8YH , U.K.
MJ Research, Inc.	590 Lincoln Street, Waltham, MA 02451, USA
MolecularProbesEuropeBV	Poortgebouw, Rijnsburgerweg 10, 2333 AA Leiden, The Netherlands
MWG Biotech AG	Anzinger Str. 7, 85560 Ebersberg, Germany
New England Biolabs (UK) Ltd. (NEB)	73 Knowl Piece, Wilbury Way, Hitchin, SG4 0TY, U.K.
PerkinElmer UK	Chalfont Road, Seer Green, Beaconsfield, Bucks HP9 2FX, U.K.
Promega UK Ltd	Delta House, Chilworth Science Park, Southampton, SO16 7NS, U.K.

continued on next page

continued from previous page

COMPANY	ADDRESS
Qiagen ltd. (Quiagen)	Boundary Court, Gatwick Road, Crawley, RH10 9AX, U.K.
Roche Diagnostics Ltd.	Bell Lane, Lewes, BN7 1LG, U.K.
Sequenom	EUROPEAN OFFICE, Mendelssohnstrasse 15D, 22761 Hamburg, Germany
Sigma-Aldrich Company Ltd.	Dorset, England
STARLAB GmbH	International HQ, Ewige Weide 7, 22926 Ahrensburg, Germany
The Jackson Laboratory	600 Main Street, Bar Harbor, Maine 04609 USA
Vector Laboratories	30 Ingold Road, Burlingame, CA 94010, USA
Whatman plc	Whatman House, St Leonard's Road, 20/20 Maidstone, Kent, ME16 0LS, U.K.

Bibliography

Abdalla, S., Weatherall, D. J., Wickramasinghe, S. N., and Hughes, M. (1980). The anaemia of P. falciparum malaria. *Br J Haematol*, 46(2):171–83.

Abecasis, G. R. and Cookson, W. O. (2000). GOLD–graphical overview of linkage disequilibrium. *Bioinformatics*, 16(2):182–3.

Abecasis, G. R., Noguchi, E., Heinzmann, A., Traherne, J. A., Bhattacharyya, S., Leaves, N. I., Anderson, G. G., Zhang, Y., Lench, N. J., Carey, A., Cardon, L. R., Moffatt, M. F., and Cookson, W. O. (2001). Extent and distribution of linkage disequilibrium in three genomic regions. *Am J Hum Genet*, 68(1):191–197.

Abed, N. S., Chace, J. H., and Cowdery, J. S. (1994a). T cell-independent and T cell-dependent B cell activation increases IFN-gamma R expression and renders B cells sensitive to IFN-gamma-mediated inhibition. *J Immunol*, 153(8):3369–77.

Abed, N. S., Chace, J. H., Fleming, A. L., and Cowdery, J. S. (1994b). Interferon-gamma regulation of B lymphocyte differentiation: activation of B cells is a prerequisite for IFN-gamma-mediated inhibition of B cell differentiation. *Cell Immunol*, 153(2):356–66.

Abel, L., Cot, M., Mulder, L., Carnevale, P., and Feingold, J. (1992). Segregation analysis detects a major gene controlling blood infection levels in human malaria. *Am J Hum Genet*, 50(6):1308–17.

Abel, L. and Dessein, A. J. (1997). The impact of host genetics on susceptibility to human infectious diseases. *Curr Opin Immunol*, 9(4):509–16.

Ackerman, H. C. (2001). *Disease association mapping: methods and markers.* Dphil, Oxford.

Akey, J., Jin, L., and Xiong, M. (2001). Haplotypes vs single marker linkage disequilibrium tests: what do we gain? *Eur J Hum Genet*, 9(4):291–300.

Allione, A., Wells, V., Forni, G., Mallucci, L., and Novelli, F. (1998). Beta-galactoside-binding protein (beta GBP) alters the cell cycle, up- regulates expression of the alpha- and beta-chains of the IFN-gamma receptor, and triggers IFN-gamma-mediated apoptosis of activated human T lymphocytes. *J Immunol*, 161(5):2114–9.

Allsopp, C. E., Harding, R. M., Taylor, C., Bunce, M., Kwiatkowski, D., Anstey, N., Brewster, D., McMichael, A. J., Greenwood, B. M., and Hill, A. V. (1992). Interethnic genetic differentiation in Africa: HLA class I antigens in The Gambia. *Am J Hum Genet*, 50(2):411–21.

Altare, F., Lammas, D., Revy, P., Jouanguy, E., Doffinger, R., Lamhamedi, S., Drysdale, P., Scheel-Toellner, D., Girdlestone, J., Darbyshire, P., Wadhwa, M., Dockrell, H., Salmon, M., Fischer, A., Durandy, A., Casanova, J. L., and Kumararatne, D. S. (1998). Inherited interleukin 12 deficiency in a child with bacille Calmette- Guerin and Salmonella enteritidis disseminated infection. *J Clin Invest*, 102(12):2035–40.

Amani, V., Vigario, A. M., Belnoue, E., Marussig, M., Fonseca, L., Mazier, D., and Renia, L. (2000). Involvement of IFN-gamma receptor-medicated signaling in pathology and anti-malarial immunity induced by Plasmodium berghei infection. *Eur J Immunol*, 30(6):1646–55.

Angrist, M., Jing, S., Bolk, S., Bentley, K., Nallasamy, S., Halushka, M., Fox, G. M., and Chakravarti, A. (1998). Human GFRA1: cloning, mapping, genomic structure, and evaluation as a candidate gene for Hirschsprung disease susceptibility. *Genomics*, 48(3):354–62.

Ardlie, K. G., Kruglyak, L., and Seielstad, M. (2002). Patterns of linkage disequilibrium in the human genome. *Nat Rev Genet*, 3(4):299–309.

Awad, M., Pravica, V., Perrey, C., El Gamel, A., Yonan, N., Sinnott, P. J., and Hutchinson, I. V. (1999). CA repeat allele polymorphism in the first intron of the human interferon-gamma gene is associated with lung allograft fibrosis. *Hum Immunol*, 60(4):343–6.

Bach, E. A., Aguet, M., and Schreiber, R. D. (1997). The IFN gamma receptor: a paradigm for cytokine receptor signaling. *Annu Rev Immunol*, 15:563–91.

Bailey, W. J., Fitch, D. H., Tagle, D. A., Czelusniak, J., Slightom, J. L., and Goodman, M. (1991). Molecular evolution of the psi eta-globin gene locus: gibbon phylogeny and the hominoid slowdown. *Mol Biol Evol*, 8(2):155–84.

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

Bellamy, R., Kwiatkowski, D., and Hill, A. V. (1998). Absence of an association between intercellular adhesion molecule 1, complement receptor 1 and interleukin 1 receptor antagonist gene polymorphisms and severe malaria in a West African population. *Trans R Soc Trop Med Hyg*, 92(3):312–6.

Blumberg, H., Conklin, D., Xu, W. F., Grossmann, A., Brender, T., Carollo, S., Eagan, M., Foster, D., Haldeman, B. A., Hammond, A., Haugen, H., Jelinek, L., Kelly, J. D., Madden, K., Maurer, M. F., Parrish-Novak, J., Prunkard, D., Sexson, S., Sprecher, C., Waggie, K., West, J., Whitmore, T. E., Yao, L., Kuechle, M. K., Dale, B. A., and Chandrasekher, Y. A. (2001). Interleukin 20: discovery, receptor identification, and role in epidermal function. *Cell*, 104(1):9–19.

Boisclair, Y. R., Brown, A. L., Casola, S., and Rechler, M. M. (1993). Three clustered Sp1 sites are required for efficient transcription of the TATA-less promoter of the gene for insulin-like growth factor-binding protein-2 from the rat. *J Biol Chem*, 268(33):24892–901.

Brabin, B. J. (1983). An analysis of malaria in pregnancy in Africa. *Bull World Health Organ*, 61(6):1005–16.

Bream, J. H., Carrington, M., O'Toole, S., Dean, M., Gerrard, B., Shin, H. D., Kosack, D., Modi, W., Young, H. A., and Smith, M. W. (2000). Polymorphisms of the human IFNG gene noncoding regions. *Immunogenetics*, 51(1):50–8.

Brewster, D. R., Kwiatkowski, D., and White, N. J. (1990). Neurological sequelae of cerebral malaria in children. *Lancet*, 336(8722):1039–43.

Brown, A. E., Teja-Isavadharm, P., and Webster, H. K. (1991). Macrophage activation in vivax malaria: fever is associated with increased levels of neopterin and interferon-gamma. *Parasite Immunol*, 13(6):673–9.

Bruce-Chwatt, L. (1988). History of malaria from prehistory to eradication. In Wernsdorfer, W. and McGregor, I., editors, *Malaria: Principle and Practice of Malariology*, pages 1–59. Churchill Livingstone, Edingburgh.

Buckley, R. H. (1994). Breakthroughs in the understanding and therapy of primary immunodeficiency. *Pediatr Clin North Am*, 41(4):665–90.

Bureau, J. F., Bihl, F., Brahic, M., and Le Paslier, D. (1995). The gene coding for interferon-gamma is linked to the D12S335 and D12S313 microsatellites and to the MDM2 gene. *Genomics*, 28(1):109–12.

Burgner, D., Xu, W., Rockett, K., Gravenor, M., Charles, I. G., Hill, A. V., and Kwiatkowski, D. (1998). Inducible nitric oxide synthase polymorphism and fatal cerebral malaria. *Lancet*, 352(9135):1193–4.

Cargill, M., Altshuler, D., Ireland, J., Sklar, P., Ardlie, K., Patil, N., Shaw, N., Lane, C. R., Lim, E. P., Kalyanaraman, N., Nemesh, J., Ziaugra, L., Friedland, L., Rolfe, A., Warrington, J., Lipshutz, R., Daley, G. Q., and Lander, E. S. (1999). Characterization of single-nucleotide polymorphisms in coding regions of human genes. *Nat Genet*, 22(3):231–8.

Carrington, M., Nelson, G. W., Martin, M. P., Kissner, T., Vlahov, D., Goedert, J. J., Kaslow, R., Buchbinder, S., Hoots, K., and O'Brien, S. J. (1999). HLA and HIV-1: heterozygote advantage and B*35-Cw*04 disadvantage. *Science*, 283(5408):1748–52.

Casanova, J. L., Blanche, S., Emile, J. F., Jouanguy, E., Lamhamedi, S., Altare, F., Stephan, J. L., Bernaudin, F., Bordigoni, P., Turck, D., Lachaux, A., Albertini, M., Bourrillon, A., Dommergues, J. P., Pocidalo, M. A., Le Deist, F., Gaillard, J. L., Griscelli, C., and Fischer, A. (1996). Idiopathic disseminated bacillus Calmette-Guerin infection: a French national retrospective study. *Pediatrics*, 98(4 Pt 1):774–8.

Cavalli-Sforza, L., Menozzi, P., and Piazza, A. (1994). *The History and Geography of Human Genes*. Princeton University Press, Princeton.

Cavet, J., Dickinson, A. M., Norden, J., Taylor, P. R., Jackson, G. H., and Middleton, P. G. (2001). Interferon-gamma and interleukin-6 gene polymorphisms associate with graft-versus-host disease in HLA-matched sibling bone marrow transplantation. *Blood*, 98(5):1594–600.

Chakravarti, A. (1999). Population genetics–making sense out of sequence. *Nat Genet*, 21(1 Suppl):56–60.

Chakravarti, M. and Vogel, F. (1973). *A Twin Study on Leprosy*, volume 1. Thieme, Stuttgart.

Chatterjee-Kishore, M., van Den Akker, F., and Stark, G. R. (2000a). Adenovirus E1A down-regulates LMP2 transcription by interfering with the binding of stat1 to IRF1. *J Biol Chem*, 275(27):20406–11.

Chatterjee-Kishore, M., van den Akker, F., and Stark, G. R. (2000b). Association of STATs with relatives and friends. *Trends Cell Biol*, 10(3):106–11.

Chatterjee-Kishore, M., Wright, K. L., Ting, J. P., and Stark, G. R. (2000c). How Stat1 mediates constitutive gene expression: a complex of unphosphorylated Stat1 and IRF1 supports transcription of the LMP2 gene. *Embo J*, 19(15):4111–22.

Chehimi, J. and Trinchieri, G. (1994). Interleukin-12: a bridge between innate resistance and adaptive immunity with a role in infection and acquired immunodeficiency. *J Clin Immunol*, 14(3):149–61.

Chen, F. C. and Li, W. H. (2001). Genomic divergences between humans and other hominoids and the effective population size of the common ancestor of humans and chimpanzees. *Am J Hum Genet*, 68(2):444–56.

Chevillard, C., Henri, S., Stefani, F., Parzy, D., and Dessein, A. (2002). Two new polymorphisms in the human interferon gamma (IFN-gamma) promoter. *Eur J Immunogenet*, 29(1):53–6.

Chizzolini, C., Grau, G. E., Geinoz, A., and Schrijvers, D. (1990). T lymphocyte interferon-gamma production induced by Plasmodium falciparum antigen is high in recently infected non-immune and low in immune subjects. *Clin Exp Immunol*, 79(1):95–9.

Chrivia, J. C., Wedrychowicz, T., Young, H. A., and Hardy, K. J. (1990). A model of human cytokine regulation based on transfection of gamma interferon gene fragments directly into isolated peripheral blood T lymphocytes. *J Exp Med*, 172(2):661–4.

Ciccarone, V. C., Chrivia, J., Hardy, K. J., and Young, H. A. (1990). Identification of enhancer-like elements in human IFN-gamma genomic DNA. *J Immunol*, 144(2):725–30.

Clark, A. G., Weiss, K. M., Nickerson, D. A., Taylor, S. L., Buchanan, A., Stengard, J., Salomaa, V., Vartiainen, E., Perola, M., Boerwinkle, E., and Sing, C. F. (1998). Haplotype structure and population genetic inferences from nucleotide-sequence variation in human lipoprotein lipase. *Am J Hum Genet*, 63(2):595–612.

Clark, I. A., Hunt, N. H., Butcher, G. A., and Cowden, W. B. (1987). Inhibition of murine malaria (Plasmodium chabaudi) in vivo by recombinant interferon-gamma or tumor necrosis factor, and its enhancement by butylated hydroxyanisole. *J Immunol*, 139(10):3493–6.

Clark, I. A., Rockett, K. A., and Cowden, W. B. (1992). Possible central role of nitric oxide in conditions clinically similar to cerebral malaria. *Lancet*, 340(8824):894–6.

Clayton, D. (1999). A generalization of the transmission/disequilibrium test for uncertain- haplotype transmission. *Am J Hum Genet*, 65(4):1170–7.

Coccia, E. M., Stellacci, E., Orsatti, R., Benedetti, E., Giacomini, E., Marziali, G., Valdez, B. C., and Battistini, A. (2002). Protein inhibitor of activated signal transducer and activator of transcription (STAT)-1 (PIAS-1) regulates the IFN-gamma response in macrophage cell lines. *Cell Signal*, 14(6):537–45.

Collins, A., Lonjou, C., and Morton, N. E. (1999). Genetic epidemiology of single-nucleotide polymorphisms. *Proc Natl Acad Sci U S A*, 96(26):15173–7.

Comstock, G. W. (1978). Tuberculosis in twins: a re-analysis of the Prophit survey. *Am Rev Respir Dis*, 117(4):621–4.

Constant, P., Davodeau, F., Peyrat, M. A., Poquet, Y., Puzo, G., Bonneville, M., and Fournie, J. J. (1994). Stimulation of human gamma delta T cells by nonpeptidic mycobacterial ligands. *Science*, 264(5156):267–70.

Constantin, A., Navaux, F., Lauwers-Cances, V., Abbal, M., van Meerwijk, J. P., Mazieres, B., Cambon-Thomsen, A., and Cantagrel, A. (2001). Interferon gamma gene polymorphism and susceptibility to, and severity of, rheumatoid arthritis. *Lancet*, 358(9298):2051–2.

Cooke, G. S. and Hill, A. V. (2001). Genetics of susceptibility to human infectious disease. *Nat Rev Genet*, 2(12):967–77.

Dabora, S. L., Roberts, P., Nieto, A., Perez, R., Jozwiak, S., Franz, D., Bissler, J., Thiele, E. A., Sims, K., and Kwiatkowski, D. J. (2002). Association between a high-expressing interferon-gamma allele and a lower frequency of kidney angiomyolipomas in TSC2 patients. *Am J Hum Genet*, 71(4):750–8.

de Kossodo, S. and Grau, G. E. (1993). Profiles of cytokine production in relation with susceptibility to cerebral malaria. *J Immunol*, 151(9):4811–20.

Dessein, A. J., Hillaire, D., Elwali, N. E., Marquet, S., Mohamed-Ali, Q., Mirghani, A., Henri, S., Abdelhameed, A. A., Saeed, O. K., Magzoub, M. M., and Abel, L. (1999). Severe hepatic fibrosis in Schistosoma mansoni infection is controlled by a major locus that is closely linked to the interferon-gamma receptor gene. *Am J Hum Genet*, 65(3):709–21.

Devlin, B. and Risch, N. (1995). A comparison of linkage disequilibrium measures for fine-scale mapping. *Genomics*, 29(2):311–22.

Diehl, K. and Von Verschuer, O. (1936). Der Erbeinfluss bei der Tuberkulose. *Beitr Klin Kunsch*, 92:275.

DiSanto, J. P., Bonnefoy, J. Y., Gauchat, J. F., Fischer, A., and de Saint Basile, G. (1993). CD40 ligand mutations in x-linked immunodeficiency with hyper-IgM. *Nature*, 361(6412):541–3.

Dodoo, D., Omer, F. M., Todd, J., Akanmori, B. D., Koram, K. A., and Riley, E. M. (2002). Absolute levels and ratios of proinflammatory and anti-inflammatory cytokine production in vitro predict clinical immunity to Plasmodium falciparum malaria. *J Infect Dis*, 185(7):971–9.

Doffinger, R., Jouanguy, E., Dupuis, S., Fondaneche, M. C., Stephan, J. L., Emile, J. F., Lamhamedi-Cherradi, S., Altare, F., Pallier, A., Barcenas-Morales, G., Meinl, E., Krause, C., Pestka, S., Schreiber, R. D., Novelli, F., and Casanova, J. L. (2000). Partial interferon-gamma receptor signaling chain deficiency in a patient with bacille Calmette-Guerin and Mycobacterium abscessus infection. *J Infect Dis*, 181(1):379–84.

Doherty, P. C. and Zinkernagel, R. M. (1975). Enhanced immunological surveillance in mice heterozygous at the H-2 gene complex. *Nature*, 256(5512):50–2.

Dorit, R. L., Akashi, H., and Gilbert, W. (1995). Absence of polymorphism at the ZFY locus on the human Y chromosome. *Science*, 268(5214):1183–5.

Dorman, S. E., Uzel, G., Roesler, J., Bradley, J. S., Bastian, J., Billman, G., King, S., Filie, A., Schermerhorn, J., and Holland, S. M. (1999). Viral infections in interferon-gamma receptor deficiency. *J Pediatr*, 135(5):640–3.

Douabin-Gicquel, V., Soriano, N., Ferran, H., Wojcik, F., Palierne, E., Tamim, S., Jovelin, T., McKie, A. T., Le Gall, J. Y., David, V., and Mosser, J. (2001). Identification of 96 single nucleotide polymorphisms in eight genes involved in iron metabolism: efficiency of bioinformatic extraction compared with a systematic sequencing approach. *Hum Genet*, 109(4):393–401.

Dumoutier, L., Lejeune, D., Colau, D., and Renauld, J. C. (2001). Cloning and characterization of IL-22 binding protein, a natural antagonist of IL-10-related T cell-derived inducible factor/IL-22. *J Immunol*, 166(12):7090–5.

Dumoutier, L., Louahed, J., and Renauld, J. C. (2000a). Cloning and characterization of IL-10-related T cell-derived inducible factor (IL-TIF), a novel cytokine structurally related to IL-10 and inducible by IL-9. *J Immunol*, 164(4):1814–9.

Dumoutier, L., Van Roost, E., Ameye, G., Michaux, L., and Renauld, J. C. (2000b). IL-TIF/IL-22: genomic organization and mapping of the human and mouse genes. *Genes Immun*, 1(8):488–94.

Dumoutier, L., Van Roost, E., Colau, D., and Renauld, J. C. (2000c). Human interleukin-10-related T cell-derived inducible factor: molecular cloning and functional characterization as an hepatocyte-stimulating factor. *Proc Natl Acad Sci U S A*, 97(18):10144–9.

Dynan, W. (1986). Promoters for housekeeping genes. *Trends Genet.*, 2:196–197.

Ebensperger, C., Rhee, S., Muthukumaran, G., Lembo, D., Donnelly, R., Pestka, S., and Dembic, Z. (1996). Genomic organization and promoter analysis of the gene ifngr2 encoding the second chain of the mouse interferon-gamma receptor. *Scand J Immunol*, 44(6):599–606.

Ebersberger, I., Metzler, D., Schwarz, C., and Paabo, S. (2002). Genomewide comparison of DNA sequences between humans and chimpanzees. *Am J Hum Genet*, 70(6):1490–7.

Endres, S., van der Meer, J. W., and Dinarello, C. A. (1987). Interleukin-1 in the pathogenesis of fever. *Eur J Clin Invest*, 17(6):469–74.

Eng, C. and Vijg, J. (1997). Genetic testing: the problems and the promise. *Nat Biotechnol*, 15(5):422–6.

Favre, N., Ryffel, B., Bordmann, G., and Rudin, W. (1997). The course of Plasmodium chabaudi chabaudi infections in interferon- gamma receptor deficient mice. *Parasite Immunol*, 19(8):375–83.

Fernandez-Reyes, D., Craig, A. G., Kyes, S. A., Peshu, N., Snow, R. W., Berendt, A. R., Marsh, K., and Newbold, C. I. (1997). A high frequency African coding polymorphism in the N-terminal domain of ICAM-1 predisposing to cerebral malaria in Kenya. *Hum Mol Genet*, 6(8):1357–60.

Fischer, A. and Arnaiz-Villena, A. (1995). Immunodeficiencies of genetic origin. *Immunol Today*, 16(11):510–4.

Fu, Y. X. (1997). Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics*, 147(2):915–25.

Fu, Y. X. and Li, W. H. (1993). Statistical tests of neutrality of mutations. *Genetics*, 133(3):693–709.

Gabriel, S. B., Schaffner, S. F., Nguyen, H., Moore, J. M., Roy, J., Blumenstiel, B., Higgins, J., DeFelice, M., Lochner, A., Faggart, M., Liu-Cordero, S. N., Rotimi, C., Adeyemo, A., Cooper, R., Ward, R., Lander, E. S., Daly,

M. J., and Altshuler, D. (2002). The structure of haplotype blocks in the human genome. *Science*, 296(5576):2225–9.

Gallup, J. L. and Sachs, J. D. (2001). The economic burden of malaria. *Am J Trop Med Hyg*, 64(1-2 Suppl):85–96.

Garcia, A., Cot, M., Chippaux, J. P., Ranque, S., Feingold, J., Demenais, F., and Abel, L. (1998a). Genetic control of blood infection levels in human malaria: evidence for a complex genetic model. *Am J Trop Med Hyg*, 58(4):480–8.

Garcia, A., Marquet, S., Bucheton, B., Hillaire, D., Cot, M., Fievet, N., Dessein, A. J., and Abel, L. (1998b). Linkage analysis of blood Plasmodium falciparum levels: interest of the 5q31-q33 chromosome region. *Am J Trop Med Hyg*, 58(6):705–9.

Garrod, A. (1931). The inborn factors in disease. An essay. Oxford University Press, Oxford, U.K.

Giard, D. J., Aaronson, S. A., Todaro, G. J., Arnstein, P., Kersey, J. H., Dosik, H., and Parks, W. P. (1973). In vitro cultivation of human tumors: establishment of cell lines derived from a series of solid tumors. *J Natl Cancer Inst*, 51(5):1417–23.

Glusman, G., Sosinsky, A., Ben-Asher, E., Avidan, N., Sonkin, D., Bahar, A., Rosenthal, A., Clifton, S., Roe, B., Ferraz, C., Demaille, J., and Lancet, D. (2000). Sequence, structure, and evolution of a complete human olfactory receptor gene cluster. *Genomics*, 63(2):227–45.

Goodier, M., Fey, P., Eichmann, K., and Langhorne, J. (1992). Human peripheral blood gamma delta T cells respond to antigens of Plasmodium falciparum. *Int Immunol*, 4(1):33–41.

Goodman, M., Tagle, D. A., Fitch, D. H., Bailey, W., Czelusniak, J., Koop, B. F., Benson, P., and Slightom, J. L. (1990). Primate evolution at the DNA level and a classification of hominoids. *J Mol Evol*, 30(3):260–6.

Grau, G. E., Fajardo, L. F., Piguet, P. F., Allet, B., Lambert, P. H., and Vassalli, P. (1987). Tumor necrosis factor (cachectin) as an essential mediator in murine cerebral malaria. *Science*, 237(4819):1210–2.

Grau, G. E., Heremans, H., Piguet, P. F., Pointaire, P., Lambert, P. H., Billiau, A., and Vassalli, P. (1989a). Monoclonal antibody against interferon gamma can prevent experimental cerebral malaria and its associated overproduction of tumor necrosis factor. *Proc Natl Acad Sci U S A*, 86(14):5572–4.

Grau, G. E., Piguet, P. F., Engers, H. D., Louis, J. A., Vassalli, P., and Lambert, P. H. (1986). L3T4+ T lymphocytes play a major role in the pathogenesis of murine cerebral malaria. *J Immunol*, 137(7):2348–54.

Grau, G. E., Piguet, P. F., Vassalli, P., and Lambert, P. H. (1989b). Tumor-necrosis factor and other cytokines in cerebral malaria: experimental and clinical data. *Immunol Rev*, 112:49–70.

Grau, G. E., Taylor, T. E., Molyneux, M. E., Wirima, J. J., Vassalli, P., Hommel, M., and Lambert, P. H. (1989c). Tumor necrosis factor and disease severity in children with falciparum malaria. *N Engl J Med*, 320(24):1586–91.

Gray, P. W. and Goeddel, D. V. (1982). Structure of the human immune interferon gene. *Nature*, 298(5877):859–63.

Greenwood, B. M., Groenendaal, F., Bradley, A. K., Greenwood, A. M., Shenton, F., Tulloch, S., and Hayes, R. (1987). Ethnic differences in the prevalence of splenomegaly and malaria in The Gambia. *Ann Trop Med Parasitol*, 81(4):345–54.

Grosveld, F., van Assendelft, G. B., Greaves, D. R., and Kollias, G. (1987). Position-independent, high-level expression of the human beta-globin gene in transgenic mice. *Cell*, 51(6):975–85.

Haldane, J. B. S. (1948). The rate of mutation of human genes. Proceedings of the Eighth International Congress of Genetics and Heredity. *Hereditas (Suppl.)*, 35.

Haldane, J. B. S. (1949). Disease and Evolution. *Ric. Sci. (Suppl.)*, 19:68–76.

Hartl, D. L. and Clark, A. G. (1997a). In *Principles of Population Genetics*. Sinauer Associates, Inc., 3 edition.

Hartl, D. L. and Clark, A. G. (1997b). Wright's F Statistics. In *Principles of Population Genetics*, pages 117–120. Sinauer Associates, Inc., 3 edition.

Henri, S., Stefani, F., Parzy, D., Eboumbou, C., Dessein, A., and Chevillard, C. (2002). Description of three new polymorphisms in the intronic and 3'UTR regions of the human interferon gamma gene. *Genes Immun*, 3(1):1–4.

Herndon, C. and Jennings, R. (1951). A twin family study on susceptibility to poliomyelitis. *Am J Hum Genet*, 3:17.

Hill, A. V. (1996). Genetics of infectious disease resistance. *Curr Opin Genet Dev*, 6(3):348–53.

Hill, A. V. (1998). The immunogenetics of human infectious diseases. *Annu Rev Immunol*, 16:593–617.

Hill, A. V. (1999). Genetics and genomics of infectious disease susceptibility. *Br Med Bull*, 55(2):401–13.

Hill, W. and Robertson, A. (1968). Linkage Disequilibrium in finite populations. *Theor Appl Genet*, 38:226–231.

Himmelrich, H., Parra-Lopez, C., Tacchini-Cottier, F., Louis, J. A., and Launois, P. (1998). The IL-4 rapidly produced in BALB/c mice after infection with Leishmania major down-regulates IL-12 receptor beta 2-chain expression on CD4+ T cells resulting in a state of unresponsiveness to IL-12. *J Immunol*, 161(11):6156–63.

Hoffman, S. L., Crutcher, J. M., Puri, S. K., Ansari, A. A., Villinger, F., Franke, E. D., Singh, P. P., Finkelman, F., Gately, M. K., Dutta, G. P., and Sedegah, M. (1997). Sterile protection of monkeys against malaria after administration of interleukin-12. *Nat Med*, 3(1):80–3.

Holland, S. M., Dorman, S. E., Kwon, A., Pitha-Rowe, I. F., Frucht, D. M., Gerstberger, S. M., Noel, G. J., Vesterhus, P., Brown, M. R., and Fleisher, T. A. (1998). Abnormal regulation of interferon-gamma, interleukin-12, and tumor necrosis factor-alpha in human interferon-gamma receptor 1 deficiency. *J Infect Dis*, 178(4):1095–104.

Hollingdale, M. R., McCormick, C. J., Heal, K. G., Taylor-Robinson, A. W., Reeve, P., Boykins, R., and Kazura, J. W. (1998). Biology of malarial liver stages: implications for vaccine design. *Ann Trop Med Parasitol*, 92(4):411–7.

Hudson, R. R., Kreitman, M., and Aguade, M. (1987). A test of neutral molecular evolution based on nucleotide data. *Genetics*, 116(1):153–9.

Hugot, J. P., Chamaillard, M., Zouali, H., Lesage, S., Cezard, J. P., Belaiche, J., Almer, S., Tysk, C., O'Morain, C. A., Gassull, M., Binder, V., Finkel, Y., Cortot, A., Modigliani, R., Laurent-Puig, P., Gower-Rousseau, C., Macry, J., Colombel, J. F., Sahbatou, M., and Thomas, G. (2001). Association of NOD2 leucine-rich repeat variants with susceptibility to Crohn's disease. *Nature*, 411(6837):599–603.

Ikuyama, S., Niller, H. H., Shimura, H., Akamizu, T., and Kohn, L. D. (1992). Characterization of the 5'-flanking region of the rat thyrotropin receptor gene. *Mol Endocrinol*, 6(5):793–804.

Ishii, S., Xu, Y. H., Stratton, R. H., Roe, B. A., Merlino, G. T., and Pastan, I. (1985). Characterization and sequence of the promoter region of the human epidermal growth factor receptor gene. *Proc Natl Acad Sci U S A*, 82(15):4920–4.

Jeffreys, A. J., Kauppi, L., and Neumann, R. (2001). Intensely punctate meiotic recombination in the class II region of the major histocompatibility complex. *Nat Genet*, 29(2):217–22.

Jepson, A. P., Banya, W. A., Sisay-Joof, F., Hassan-King, M., Bennett, S., and Whittle, H. C. (1995). Genetic regulation of fever in Plasmodium falciparum malaria in Gambian twin children. *J Infect Dis*, 172(1):316–9.

Johnson, J. K., Swerlick, R. A., Grady, K. K., Millet, P., and Wick, T. M. (1993). Cytoadherence of Plasmodium falciparum-infected erythrocytes to microvascular endothelium is regulatable by cytokines and phorbol ester. *J Infect Dis*, 167(3):698–703.

Jorde, L. B. (2000). Linkage disequilibrium and the search for complex disease genes. *Genome Res*, 10(10):1435–44.

Jouanguy, E., Altare, F., Lamhamedi, S., Revy, P., Emile, J. F., Newport, M., Levin, M., Blanche, S., Seboun, E., Fischer, A., and Casanova, J. L. (1996). Interferon-gamma-receptor deficiency in an infant with fatal bacille Calmette-Guerin infection. *N Engl J Med*, 335(26):1956–61.

Jouanguy, E., Lamhamedi-Cherradi, S., Altare, F., Fondaneche, M. C., Tuerlinckx, D., Blanche, S., Emile, J. F., Gaillard, J. L., Schreiber, R., Levin, M., Fischer, A., Hivroz, C., and Casanova, J. L. (1997). Partial interferon-gamma receptor 1 deficiency in a child with tuberculoid bacillus Calmette-Guerin infection and a sibling with clinical tuberculosis. *J Clin Invest*, 100(11):2658–64.

Jouanguy, E., Lamhamedi-Cherradi, S., Lammas, D., Dorman, S. E., Fondaneche, M. C., Dupuis, S., Doffinger, R., Altare, F., Girdlestone, J., Emile, J. F., Ducoulombier, H., Edgar, D., Clarke, J., Oxelius, V. A., Brai, M., Novelli, V., Heyne, K., Fischer, A., Holland, S. M., Kumararatne, D. S., Schreiber, R. D., and Casanova, J. L. (1999). A human IFNGR1 small deletion hotspot associated with dominant susceptibility to mycobacterial infection [see comments]. *Nat Genet*, 21(4):370–8.

Jurinke, C., van den Boom, D., Cantor, C. R., and Koster, H. (2002). The use of MassARRAY technology for high throughput genotyping. *Adv Biochem Eng Biotechnol*, 77:57–74.

Kallmann, F. and Reisner, D. (1942). Twin studies on the significance of genetic factors in tuberculosis. *Am. Rev. Respir. Dis.*, 47:549.

Kemper, O. and Wallach, D. (1993). Cloning and partial characterization of the promoter for the human p55 tumor necrosis factor (TNF) receptor. *Gene*, 134(2):209–16.

Khani-Hanjani, A., Lacaille, D., Hoar, D., Chalmers, A., Horsman, D., Anderson, M., Balshaw, R., and Keown, P. A. (2000). Association between dinucleotide repeat in non-coding region of interferon-gamma gene and susceptibility to, and severity of, rheumatoid arthritis. *Lancet*, 356(9232):820–5.

Kimura, M. (1968). Evolutionary rate at the molecular level. *Nature*, 217:624–626.

Kimura, M. (1983). *The neutral theory of molecular evolution*. Cambridge University Press, Cambridge, Massachusetts.

King, M. C. and Wilson, A. C. (1975). Evolution at two levels in humans and chimpanzees. *Science*, 188(4184):107–16.

Kluger, M. J. (1991). Fever: role of pyrogens and cryogens. *Physiol Rev*, 71(1):93–127.

Knappe, A., Hor, S., Wittmann, S., and Fickenscher, H. (2000). Induction of a novel cellular homolog of interleukin-10, AK155, by transformation of T lymphocytes with herpesvirus saimiri. *J Virol*, 74(8):3881–7.

Knight, J. C., Udalova, I., Hill, A. V., Greenwood, B. M., Peshu, N., Marsh, K., and Kwiatkowski, D. (1999). A polymorphism that affects OCT-1 binding to the TNF promoter region is associated with severe malaria [see comments]. *Nat Genet*, 22(2):145–50.

Kobayashi, M., Fitz, L., Ryan, M., Hewick, R. M., Clark, S. C., Chan, S., Loudon, R., Sherman, F., Perussia, B., and Trinchieri, G. (1989). Identification and purification of natural killer cell stimulatory factor (NKSF), a cytokine with multiple biologic effects on human lymphocytes. *J Exp Med*, 170(3):827–45.

Korthauer, U., Graf, D., Mages, H. W., Briere, F., Padayachee, M., Malcolm, S., Ugazio, A. G., Notarangelo, L. D., Levinsky, R. J., and Kroczek, R. A. (1993). Defective expression of T-cell CD40 ligand causes X-linked immunodeficiency with hyper-IgM. *Nature*, 361(6412):539–41.

Kotenko, S. V., Izotova, L. S., Mirochnitchenko, O. V., Esterova, E., Dickensheets, H., Donnelly, R. P., and Pestka, S. (2001). Identification, cloning, and characterization of a novel soluble receptor that binds IL-22 and neutralizes its activity. *J Immunol*, 166(12):7096–103.

Krakauer, T. and Oppenheim, J. J. (1993). IL-1 and tumor necrosis factor-alpha each up-regulate both the expression of IFN-gamma receptors and enhance IFN-gamma-induced HLA-DR expression on human monocytes and a human monocytic cell line (THP-1). *J Immunol*, 150(4):1205–11.

Kremsner, P. G., Neifer, S., Schermuck, S., Chaves, M. F., Sliwa, K., and Bienzle, U. (1991). Interferon-gamma enhances the effect of antimalarial chemotherapy in murine Plasmodium vinckei malaria. *J Infect Dis*, 163(5):1161–3.

Krishna, S. (1997). Science, medicine, and the future. Malaria. *Bmj*, 315(7110):730–2.

Kruglyak, L. (1999). Prospects for whole-genome linkage disequilibrium mapping of common disease genes. *Nat Genet*, 22(2):139–44.

Kruglyak, L. and Nickerson, D. A. (2001). Variation is the spice of life. *Nat Genet*, 27(3):234–6.

Kun, J. F., Klabunde, J., Lell, B., Luckner, D., Alpers, M., May, J., Meyer, C., and Kremsner, P. G. (1999). Association of the ICAM-1Kilifi mutation with protection against severe malaria in Lambarene, Gabon. *Am J Trop Med Hyg*, 61(5):776–9.

Kun, J. F., Mordmuller, B., Lell, B., Lehman, L. G., Luckner, D., and Kremsner, P. G. (1998). Polymorphism in promoter region of inducible nitric oxide synthase gene and protection against malaria. *Lancet*, 351(9098):265–6.

Kwiatkowski, D. (2000). Science, medicine, and the future: susceptibility to infection. *Bmj*, 321(7268):1061–5.

Kwiatkowski, D., Cannon, J. G., Manogue, K. R., Cerami, A., Dinarello, C. A., and Greenwood, B. M. (1989). Tumour necrosis factor production in Falciparum malaria and its association with schizont rupture. *Clin Exp Immunol*, 77(3):361–6.

Kwiatkowski, D., Hill, A. V., Sambou, I., Twumasi, P., Castracane, J., Manogue, K. R., Cerami, A., Brewster, D. R., and Greenwood, B. M. (1990). TNF concentration in fatal cerebral, non-fatal cerebral, and uncomplicated Plasmodium falciparum malaria. *Lancet*, 336(8725):1201–4.

Kwiatkowski, D., Molyneux, M. E., Stephens, S., Curtis, N., Klein, N., Pointaire, P., Smit, M., Allan, R., Brewster, D. R., Grau, G. E., and et al. (1993). Anti-TNF therapy inhibits fever in cerebral malaria. *Q J Med*, 86(2):91–8.

Lander, E. S. (1996). The new genomics: global views of biology. *Science*, 274(5287):536–9.

Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., Funke, R., Gage, D., Harris, K., Heaford, A., Howland, J., Kann, L., Lehoczky, J., LeVine, R., McEwan, P., McKernan, K., Meldrim, J., Mesirov, J. P., Miranda, C., Morris, W., Naylor, J., Raymond, C., Rosetti, M., Santos, R., Sheridan, A., Sougnez, C., Stange-Thomann, N., Stojanovic, N., Subramanian, A.,

Wyman, D., Rogers, J., Sulston, J., Ainscough, R., Beck, S., Bentley, D., Burton, J., Clee, C., Carter, N., Coulson, A., Deadman, R., Deloukas, P., Dunham, A., Dunham, I., Durbin, R., French, L., Grafham, D., Gregory, S., Hubbard, T., Humphray, S., Hunt, A., Jones, M., Lloyd, C., McMurray, A., Matthews, L., Mercer, S., Milne, S., Mullikin, J. C., Mungall, A., Plumb, R., Ross, M., Shownkeen, R., Sims, S., Waterston, R. H., Wilson, R. K., Hillier, L. W., McPherson, J. D., Marra, M. A., Mardis, E. R., Fulton, L. A., Chinwalla, A. T., Pepin, K. H., Gish, W. R., Chissoe, S. L., Wendl, M. C., Delehaunty, K. D., Miner, T. L., Delehaunty, A., Kramer, J. B., Cook, L. L., Fulton, R. S., Johnson, D. L., Minx, P. J., Clifton, S. W., Hawkins, T., Branscomb, E., Predki, P., Richardson, P., Wenning, S., Slezak, T., Doggett, N., Cheng, J. F., Olsen, A., Lucas, S., Elkin, C., Uberbacher, E., Frazier, M., et al. (2001). Initial sequencing and analysis of the human genome. *Nature*, 409(6822):860–921.

Larkin, J., r., Johnson, H. M., and Subramaniam, P. S. (2000). Differential nuclear localization of the IFNGR-1 and IFNGR-2 subunits of the IFN-gamma receptor complex following activation by IFN-gamma. *J Interferon Cytokine Res*, 20(6):565–76.

Lewontin, R. (1964). The interaction of selection and linkage. I. General considerations; heterotic models. *Genetics*, 49:49–67.

Li, W. H. and Sadler, L. A. (1991). Low nucleotide diversity in man. *Genetics*, 129(2):513–23.

Lin, S., Cutler, D. J., Zwick, M. E., and Chakravarti, A. (2002). Haplotype inference in random population samples. *Am J Hum Genet*, 71(5):1129–37.

Lin, T. M., Chen, C. J., Wu, M. M., Yang, C. S., Chen, J. S., Lin, C. C., Kwang, T. Y., Hsu, S. T., Lin, S. Y., and Hsu, L. C. (1989). Hepatitis B virus markers in Chinese twins. *Anticancer Res*, 9(3):737–41.

Lio, D., Marino, V., Serauto, A., Gioia, V., Scola, L., Crivello, A., Forte, G. I., Colonna-Romano, G., Candore, G., and Caruso, C. (2002). Genotype frequencies of the +874T–¿A single nucleotide polymorphism in the first

intron of the interferon-gamma gene in a sample of Sicilian patients affected by tuberculosis. *Eur J Immunogenet*, 29(5):371–4.

Liu, B., Liao, J., Rao, X., Kushner, S. A., Chung, C. D., Chang, D. D., and Shuai, K. (1998). Inhibition of Stat1-mediated gene activation by PIAS1. *Proc Natl Acad Sci U S A*, 95(18):10626–31.

Looareesuwan, S., Merry, A. H., Phillips, R. E., Pleehachinda, R., Wattanagoon, Y., Ho, M., Charoenlarp, P., Warrell, D. A., and Weatherall, D. J. (1987). Reduced erythrocyte survival following clearance of malarial parasitaemia in Thai patients. *Br J Haematol*, 67(4):473–8.

Loots, G. G., Locksley, R. M., Blankespoor, C. M., Wang, Z. E., Miller, W., Rubin, E. M., and Frazer, K. A. (2000). Identification of a coordinate regulator of interleukins 4, 13, and 5 by cross-species sequence comparisons. *Science*, 288(5463):136–40.

Lopez-Maderuelo, D., Arnalich, F., Serantes, R., Gonzalez, A., Codoceo, R., Madero, R., Vazquez, J. J., and Montiel, C. (2003). Interferon Gamma and Interleukin-10 Gene Polymorphisms in Pulmonary Tuberculosis. *Am J Respir Crit Care Med.*

Luzzi, G. A., Merry, A. H., Newbold, C. I., Marsh, K., and Pasvol, G. (1991a). Protection by alpha-thalassaemia against Plasmodium falciparum malaria: modified surface antigen expression rather than impaired growth or cytoadherence. *Immunol Lett*, 30(2):233–40.

Luzzi, G. A., Merry, A. H., Newbold, C. I., Marsh, K., Pasvol, G., and Weatherall, D. J. (1991b). Surface antigen expression on Plasmodium falciparum-infected erythrocytes is modified in alpha- and beta-thalassemia. *J Exp Med*, 173(4):785–91.

MacGregor, R. R., Sheagren, J. N., and Wolff, S. M. (1969). Endotoxin-induced modification of Plasmodium berghei infection in mice. *J Immunol*, 102(1):131–9.

Maheshwari, R. K., Czarniecki, C. W., Dutta, G. P., Puri, S. K., Dhawan, B. N., and Friedman, R. M. (1986). Recombinant human gamma interferon inhibits simian malaria. *Infect Immun*, 53(3):628–30.

Malhotra, K., Salmon, D., Le Bras, J., and Vilde, J. L. (1988). Susceptibility of Plasmodium falciparum to a peroxidase-mediated oxygen-dependent microbicidal system. *Infect Immun*, 56(12):3305–9.

Mao, C., Merlin, G., and Aguet, M. (1990). Differential regulation of the human IFN-gamma receptor expression in Raji and IM9 lymphoblastoid cells versus THP-1 monocytic cells by IFN-gamma and phorbol myristate acetate. *J Immunol*, 144(12):4688–96.

Marsh, K., Forster, D., Waruiru, C., Mwangi, I., Winstanley, M., Marsh, V., Newton, C., Winstanley, P., Warn, P., Peshu, N., and et al. (1995). Indicators of life-threatening malaria in African children. *N Engl J Med*, 332(21):1399–404.

Marth, G., Yeh, R., Minton, M., Donaldson, R., Li, Q., Duan, S., Davenport, R., Miller, R. D., and Kwok, P. Y. (2001). Single-nucleotide polymorphisms in the public domain: how useful are they? *Nat Genet*, 27(4):371–2.

Martin, L. K., Einheber, A., Sadun, E. H., and Wren, R. E. (1967). Effect of bacterial endotoxin on the course of Plasmodium berghei infection. *Exp Parasitol*, 20(2):186–99.

Martin, M. P., Dean, M., Smith, M. W., Winkler, C., Gerrard, B., Michael, N. L., Lee, B., Doms, R. W., Margolick, J., Buchbinder, S., Goedert, J. J., O'Brien, T. R., Hilgartner, M. W., Vlahov, D., O'Brien, S. J., and Carrington, M. (1998). Genetic acceleration of AIDS progression by a promoter variant of CCR5. *Science*, 282(5395):1907–11.

Masternak, K., Peyraud, N., Krawczyk, M., Barras, E., and Reith, W. (2003). Chromatin remodeling and extragenic transcription at the MHC class II locus control region. *Nat Immunol*.

Matsuoka, Y., Moore, G. E., Yagi, Y., and Pressman, D. (1967). Production of free light chains of immunoglobulin by a hematopoietic cell line derived from a patient with multiple myeloma. *Proc Soc Exp Biol Med*, 125(4):1246–50.

McGregor, I. A. (1984). Epidemiology, malaria and pregnancy. *Am J Trop Med Hyg*, 33(4):517–25.

McGuire, W., Hill, A. V., Allsopp, C. E., Greenwood, B. M., and Kwiatkowski, D. (1994). Variation in the TNF-alpha promoter region associated with susceptibility to cerebral malaria. *Nature*, 371(6497):508–10.

McIntosh, N., Helms, P. J., and Smyth, R. L. (2003). *Forfar and Arneil's Textbook of Pediatrics*. Churchill Livingstone, 6th edition.

Merlin, G., van der Leede, B. J., McKune, K., Knezevic, N., Bannwarth, W., Romquin, N., Viegas-Pequignot, E., Kiefer, H., Aguet, M., and Dembic, Z. (1997). The gene for the ligand binding chain of the human interferon gamma receptor. *Immunogenetics*, 45(6):413–21.

Miller, K. L., Silverman, P. H., Kullgren, B., and Mahlmann, L. J. (1989). Tumor necrosis factor alpha and the anemia associated with murine malaria. *Infect Immun*, 57(5):1542–6.

Modiano, D., Petrarca, V., Sirima, B. S., Nebie, I., Diallo, D., Esposito, F., and Coluzzi, M. (1996). Different response to Plasmodium falciparum malaria in west African sympatric ethnic groups. *Proc Natl Acad Sci U S A*, 93(23):13206–11.

Molyneux, M. E., Taylor, T. E., Wirima, J. J., and Borgstein, A. (1989). Clinical features and prognostic indicators in paediatric cerebral malaria: a study of 131 comatose Malawian children [see comments]. *Q J Med*, 71(265):441–59.

Molyneux, M. E., Taylor, T. E., Wirima, J. J., and Grau, G. E. (1991). Tumour necrosis factor, interleukin-6, and malaria. *Lancet*, 337(8749):1098.

Morton, N. E. and Collins, A. (1998). Tests and estimates of allelic association in complex inheritance. *Proc Natl Acad Sci U S A*, 95(19):11389–93.

Mozzato-Chamay, N., Mahdi, O. S., Jallow, O., Mabey, D. C., Bailey, R. L., and Conway, D. J. (2000). Polymorphisms in candidate genes and risk of scarring trachoma in a Chlamydia trachomatis–endemic population. *J Infect Dis*, 182(5):1545–8.

Murphy, P. M. (1993). Molecular mimicry and the generation of host defense protein diversity. *Cell*, 72(6):823–6.

Nagarkatti, R., Rao, C. B., Rishi, J. P., Chetiwal, R., Shandilya, V., Vijayan, V., Kumar, R., Pemde, H. K., Sharma, S. K., Sharma, S., Singh, A. B., Gangal, S. V., and Ghosh, B. (2002). Association of IFNG gene polymorphism with asthma in the Indian population. *J Allergy Clin Immunol*, 110(3):410–2.

Nakagawa, T., Nakagawa, N., Delsing, G. A., Volkman, D., and Kehrl, J. H. (1989). Demonstration and partial characterization of the interferon-gamma receptor on human B lymphocytes. *J Cell Biochem*, 40(4):417–30.

Nakashima, H., Inoue, H., Akahoshi, M., Tanaka, Y., Yamaoka, K., Ogami, E., Nagano, S., Arinobu, Y., Niiro, H., Otsuka, T., and Niho, Y. (1999). The combination of polymorphisms within interferon-gamma receptor 1 and receptor 2 associated with the risk of systemic lupus erythematosus. *FEBS Lett*, 453(1-2):187–90.

Neifer, S., Kremsner, P. G., and Bienzle, U. (1989). Application of anti-TNF to Plasmodium vinckei-infected mice is followed by an increase of parasitaemia. *Acta Trop*, 46(4):273–5.

Newport, M. J., Huxley, C. M., Huston, S., Hawrylowicz, C. M., Oostra, B. A., Williamson, R., and Levin, M. (1996). A mutation in the interferon-gamma-receptor gene and susceptibility to mycobacterial infection. *N Engl J Med*, 335(26):1941–9.

Newton, C. R., Graham, A., Heptinstall, L. E., Powell, S. J., Summers, C., Kalsheker, N., Smith, J. C., and Markham, A. F. (1989). Analysis of any point mutation in DNA. The amplification refractory mutation system (ARMS). *Nucleic Acids Res*, 17(7):2503–16.

Newton, J., Brown, M. A., Milicic, A., Ackerman, H., Darke, C., Wilson, J. N., Wordsworth, B. P., and Kwiatkowski, D. (2003). The effect of HLA-DR on susceptibility to rheumatoid arthritis is influenced by the associated lymphotoxin alpha-tumor necrosis factor haplotype. *Arthritis Rheum*, 48(1):90–6.

Nickerson, D. A., Taylor, S. L., Weiss, K. M., Clark, A. G., Hutchinson, R. G., Stengard, J., Salomaa, V., Vartiainen, E., Boerwinkle, E., and Sing, C. F. (1998). DNA sequence diversity in a 9.7-kb region of the human lipoprotein lipase gene. *Nat Genet*, 19(3):233–40.

Nieters, A., Brems, S., and Becker, N. (2001). Cross-sectional study on cytokine polymorphisms, cytokine production after T-cell stimulation and clinical parameters in a random sample of a German population. *Hum Genet*, 108(3):241–8.

Novelli, F., Bernabei, P., Ozmen, L., Rigamonti, L., Allione, A., Pestka, S., Garotta, G., and Forni, G. (1996). Switching on of the proliferation or apoptosis of activated human T lymphocytes by IFN-gamma is correlated with the differential expression of the alpha- and beta-chains of its receptor. *J Immunol*, 157(5):1935–43.

Novelli, F., D'Elios, M. M., Bernabei, P., Ozmen, L., Rigamonti, L., Almerigogna, F., Forni, G., and Del Prete, G. (1997). Expression and role in apoptosis of the alpha- and beta-chains of the IFN-gamma receptor on human Th1 and Th2 clones. *J Immunol*, 159(1):206–13.

Novelli, F., Di Pierro, F., Francia di Celle, P., Bertini, S., Affaticati, P., Garotta, G., and Forni, G. (1994). Environmental signals influencing expression of the IFN-gamma receptor on human T cells control whether

IFN-gamma promotes proliferation or apoptosis. *J Immunol*, 152(2):496–504.

Ogura, Y., Bonen, D. K., Inohara, N., Nicolae, D. L., Chen, F. F., Ramos, R., Britton, H., Moran, T., Karaliuskas, R., Duerr, R. H., Achkar, J. P., Brant, S. R., Bayless, T. M., Kirschner, B. S., Hanauer, S. B., Nunez, G., and Cho, J. H. (2001). A frameshift mutation in NOD2 associated with susceptibility to Crohn's disease. *Nature*, 411(6837):603–6.

Pain, A., Urban, B. C., Kai, O., Casals-Pascual, C., Shafi, J., Marsh, K., and Roberts, D. J. (2001). A non-sense mutation in Cd36 gene is associated with protection from severe malaria. *The Lancet*, 357(9267):1502–3.

Patil, N., Berno, A. J., Hinds, D. A., Barrett, W. A., Doshi, J. M., Hacker, C. R., Kautzer, C. R., Lee, D. H., Marjoribanks, C., McDonough, D. P., Nguyen, B. T., Norris, M. C., Sheehan, J. B., Shen, N., Stern, D., Stokowski, R. P., Thomas, D. J., Trulson, M. O., Vyas, K. R., Frazer, K. A., Fodor, S. P., and Cox, D. R. (2001). Blocks of limited haplotype diversity revealed by high-resolution scanning of human chromosome 21. *Science*, 294(5547):1719–23.

Penix, L., Weaver, W. M., Pang, Y., Young, H. A., and Wilson, C. B. (1993). Two essential regulatory elements in the human interferon gamma promoter confer activation specific expression in T cells. *J Exp Med*, 178(5):1483–96.

Perlaza, B. L., Arevalo-Herrera, M., Brahimi, K., Quintero, G., Palomino, J. C., Gras-Masse, H., Tartar, A., Druilhe, P., and Herrera, S. (1998). Immunogenicity of four Plasmodium falciparum preerythrocytic antigens in Aotus lemurinus monkeys. *Infect Immun*, 66(7):3423–8.

Phillips, R. E., Looareesuwan, S., Warrell, D. A., Lee, S. H., Karbwang, J., Warrell, M. J., White, N. J., Swasdichai, C., and Weatherall, D. J. (1986). The importance of anaemia in cerebral and uncomplicated falciparum malaria: role of complications, dyserythropoiesis and iron sequestration. *Q J Med*, 58(227):305–23.

Pluzhnikov, A. and Donnelly, P. (1996). Optimal sequencing strategies for surveying molecular genetic diversity. *Genetics*, 144(3):1247–62.

Pokorny, V., McLean, L., McQueen, F., Abu-Maree, M., and Yeoman, S. (2001). Interferon-gamma microsatellite and rheumatoid arthritis. *Lancet*, 358(9276):122–3.

Pravica, V., Asderakis, A., Perrey, C., Hajeer, A., Sinnott, P. J., and Hutchinson, I. V. (1999). In vitro production of IFN-gamma correlates with CA repeat polymorphism in the human IFN-gamma gene. *Eur J Immunogenet*, 26(1):1–3.

Pravica, V., Perrey, C., Stevens, A., Lee, J. H., and Hutchinson, I. V. (2000). A single nucleotide polymorphism in the first intron of the human IFN-gamma gene: absolute correlation with a polymorphic CA microsatellite marker of high IFN-gamma production. *Hum Immunol*, 61(9):863–6.

Pritchard, J. K. (2001). Are rare variants responsible for susceptibility to complex diseases? *Am J Hum Genet*, 69(1):124–37.

Pritchard, J. K. and Cox, N. J. (2002). The allelic architecture of human disease genes: common disease-common variant...or not? *Hum Mol Genet*, 11(20):2417–23.

Pritchard, J. K. and Przeworski, M. (2001). Linkage disequilibrium in humans: models and data. *Am J Hum Genet*, 69(1):1–14.

Przeworski, M. (2002). The signature of positive selection at randomly chosen loci. *Genetics*, 160(3):1179–89.

Puck, J. M. (1994). Molecular and genetic basis of X-linked immunodeficiency disorders. *J Clin Immunol*, 14(2):81–9.

Pulvertaft, R. (1964). Cytology of Burkitt's tumour (African lymphoma). *Lancet*, (1):238–240.

Puri, S. K., Maheshwari, R. K., Dutta, G. P., Friedman, R. M., and Dhar, M. M. (1988). Human interferon-gamma protects rhesus monkeys against sporozoite- induced Plasmodium cynomolgi malaria infection. *J Interferon Res*, 8(2):201–6.

Ramana, C. V., Chatterjee-Kishore, M., Nguyen, H., and Stark, G. R. (2000a). Complex roles of Stat1 in regulating gene expression. *Oncogene*, 19(21):2619–27.

Ramana, C. V., Grammatikakis, N., Chernov, M., Nguyen, H., Goh, K. C., Williams, B. R., and Stark, G. R. (2000b). Regulation of c-myc expression by IFN-gamma through Stat1-dependent and -independent pathways. *Embo J*, 19(2):263–72.

Raval, P., Obici, S., Russell, S. W., and Murphy, W. J. (1995). Characterization of the 5' flanking region and gene encoding the mouse interferon-gamma receptor. *Gene*, 154(2):219–23.

Reich, D. E., Cargill, M., Bolk, S., Ireland, J., Sabeti, P. C., Richter, D. J., Lavery, T., Kouyoumjian, R., Farhadian, S. F., Ward, R., and Lander, E. S. (2001). Linkage disequilibrium in the human genome. *Nature*, 411(6834):199–204.

Reich, D. E. and Lander, E. S. (2001). On the allelic spectrum of human disease. *Trends Genet*, 17(9):502–10.

Rhee, S., Ebensperger, C., Dembic, Z., and Pestka, S. (1996). The structure of the gene for the second chain of the human interferon-gamma receptor. *J Biol Chem*, 271(46):28947–52.

Rihet, P., Traore, Y., Abel, L., Aucan, C., Traore-Leroux, T., and Fumoux, F. (1998). Malaria in humans: Plasmodium falciparum blood infection levels are linked to chromosome 5q31-q33. *Am J Hum Genet*, 63(2):498–505.

Riley, E. M., Jakobsen, P. H., Allen, S. J., Wheeler, J. G., Bennett, S., Jepsen, S., and Greenwood, B. M. (1991). Immune response to soluble exoantigens of Plasmodium falciparum may contribute to both pathogenesis and protection in clinical malaria: evidence from a longitudinal, prospective study of semi-immune African children. *Eur J Immunol*, 21(4):1019–25.

Rioux, J. D., Daly, M. J., Silverberg, M. S., Lindblad, K., Steinhart, H., Cohen, Z., Delmonte, T., Kocher, K., Miller, K., Guschwan, S., Kulbokas, E. J., O'Leary, S., Winchester, E., Dewar, K., Green, T., Stone, V., Chow, C., Cohen, A., Langelier, D., Lapointe, G., Gaudet, D., Faith, J., Branco, N., Bull, S. B., McLeod, R. S., Griffiths, A. M., Bitton, A., Greenberg, G. R., Lander, E. S., Siminovitch, K. A., and Hudson, T. J. (2001). Genetic variation in the 5q31 cytokine gene cluster confers susceptibility to Crohn disease. *Nat Genet*, 29(2):223–8.

Risch, N. and Merikangas, K. (1996). The future of genetic studies of complex human diseases. *Science*, 273(5281):1516–7.

Risch, N. J. (2000). Searching for genetic determinants in the new millennium. *Nature*, 405(6788):847–56.

Roodman, G. D., Bird, A., Hutzler, D., and Montgomery, W. (1987). Tumor necrosis factor-alpha and hematopoietic progenitors: effects of tumor necrosis factor on the growth of erythroid progenitors CFU-E and BFU-E and the hematopoietic cell lines K562, HL60, and HEL cells. *Exp Hematol*, 15(9):928–35.

Rozas, J. and Rozas, R. (1999). DnaSP version 3: an integrated program for molecular population genetics and molecular evolution analysis. *Bioinformatics*, 15(2):174–5.

Rudin, W., Favre, N., Bordmann, G., and Ryffel, B. (1997). Interferon-gamma is essential for the development of cerebral malaria [retracted in Eur J Immunol 1998 Oct;28(10):3398]. *Eur J Immunol*, 27(4):810–5.

Ruwende, C. and Hill, A. (1998). Glucose-6-phosphate dehydrogenase deficiency and malaria. *J Mol Med*, 76(8):581–8.

Ruwende, C., Khoo, S. C., Snow, R. W., Yates, S. N., Kwiatkowski, D., Gupta, S., Warn, P., Allsopp, C. E., Gilbert, S. C., Peschu, N., and et al. (1995). Natural selection of hemi- and heterozygotes for G6PD deficiency in Africa by resistance to severe malaria. *Nature*, 376(6537):246–9.

Sabeti, P. C., Reich, D. E., Higgins, J. M., Levine, H. Z., Richter, D. J., Schaffner, S. F., Gabriel, S. B., Platko, J. V., Patterson, N. J., McDonald, G. J., Ackerman, H. C., Campbell, S. J., Altshuler, D., Cooper, R., Kwiatkowski, D., Ward, R., and Lander, E. S. (2002). Detecting recent positive selection in the human genome from haplotype structure. *Nature*, 419(6909):832–7.

Sachidanandam, R., Weissman, D., Schmidt, S. C., Kakol, J. M., Stein, L. D., Marth, G., Sherry, S., Mullikin, J. C., Mortimore, B. J., Willey, D. L., Hunt, S. E., Cole, C. G., Coggill, P. C., Rice, C. M., Ning, Z., Rogers, J., Bentley, D. R., Kwok, P. Y., Mardis, E. R., Yeh, R. T., Schultz, B., Cook, L., Davenport, R., Dante, M., Fulton, L., Hillier, L., Waterston, R. H., McPherson, J. D., Gilman, B., Schaffner, S., Van Etten, W. J., Reich, D., Higgins, J., Daly, M. J., Blumenstiel, B., Baldwin, J., Stange-Thomann, N., Zody, M. C., Linton, L., Lander, E. S., and Atshuler, D. (2001). A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature*, 409(6822):928–33.

Sachs, J. and Malaney, P. (2002). The economic and social burden of malaria. *Nature*, 415(6872):680–5.

Sakamoto, S., Nie, J., and Taniguchi, T. (1999). Cutting edge: phorbol ester induction of IFN-gamma receptors leads to enhanced DR alpha gene expression. *J Immunol*, 162(8):4381–4.

Sakamoto, S. and Taniguchi, T. (2001). Identification of a phorbol ester-responsive element in the interferon-gamma receptor 1 chain gene. *J Biol Chem*, 276(40):37237–41.

Sanceau, J., Merlin, G., and Wietzerbin, J. (1992). Tumor necrosis factor-alpha and IL-6 up-regulate IFN-gamma receptor gene expression in human monocytic THP-1 cells by transcriptional and post-transcriptional mechanisms. *J Immunol*, 149(5):1671–5.

Schneider, S., Roessli, D., and Excoffier, L. (2000). Arlequin vers.2.000: A software for population genetics data analysis.

Schneider, U., Schwenk, H. U., and Bornkamm, G. (1977). Characterization of EBV-genome negative "null" and "T" cell lines derived from children with acute lymphoblastic leukemia and leukemic transformed non-Hodgkin lymphoma. *Int J Cancer*, 19(5):621–6.

Schofield, L., McConville, M. J., Hansen, D., Campbell, A. S., Fraser-Reid, B., Grusby, M. J., and Tachado, S. D. (1999). CD1d-restricted immunoglobulin G formation to GPI-anchored antigens mediated by NKT cells. *Science*, 283(5399):225–9.

Schofield, L., Villaquiran, J., Ferreira, A., Schellekens, H., Nussenzweig, R., and Nussenzweig, V. (1987). Gamma interferon, CD8+ T cells and antibodies required for immunity to malaria sporozoites. *Nature*, 330(6149):664–6.

Scuderi, P., Sterling, K. E., Lam, K. S., Finley, P. R., Ryan, K. J., Ray, C. G., Petersen, E., Slymen, D. J., and Salmon, S. E. (1986). Raised serum levels of tumour necrosis factor in parasitic infections. *Lancet*, 2(8520):1364–5.

Sedegah, M., Finkelman, F., and Hoffman, S. L. (1994). Interleukin 12 induction of interferon gamma-dependent protection against malaria. *Proc Natl Acad Sci U S A*, 91(22):10700–2.

Sehgal, A., Patil, N., and Chao, M. (1988). A constitutive promoter directs expression of the nerve growth factor receptor gene. *Mol Cell Biol*, 8(8):3160–7.

Seoh, J. Y., Khan, M., Park, S. H., Park, H. K., Shin, M. H., Ha, E. H., Lee, B. E., Yoo, K., Han, H. S., Oh, S., Wi, J. H., Hong, C. K., Oh, C. H., Kim, Y. A., and Park, J. W. (2003). Serum cytokine profiles in patients with Plasmodium vivax malaria: a comparison between those who presented with and without hyperpyrexia. *Am J Trop Med Hyg*, 68(1):102–6.

Shaffer, N., Grau, G. E., Hedberg, K., Davachi, F., Lyamba, B., Hightower, A. W., Breman, J. G., and Phuc, N. D. (1991). Tumor necrosis factor and severe malaria. *J Infect Dis*, 163(1):96–101.

Shear, H. L., Srinivasan, R., Nolan, T., and Ng, C. (1989). Role of IFN-gamma in lethal and nonlethal malaria in susceptible and resistant murine hosts. *J Immunol*, 143(6):2038–44.

Sibley, C. G. and Ahlquist, J. E. (1984). The phylogeny of the hominoid primates, as indicated by DNA-DNA hybridization. *J Mol Evol*, 20(1):2–15.

Sibley, E., Kastelic, T., Kelly, T. J., and Lane, M. D. (1989). Characterization of the mouse insulin receptor gene promoter. *Proc Natl Acad Sci U S A*, 86(24):9732–6.

Sica, A., Tan, T. H., Rice, N., Kretzschmar, M., Ghosh, P., and Young, H. A. (1992). The c-rel protooncogene product c-Rel but not NF-kappa B binds to the intronic region of the human interferon-gamma gene at a site related to an interferon-stimulable response element. *Proc Natl Acad Sci U S A*, 89(5):1740–4.

Sidow, A. (2002). Sequence first. Ask questions later. *Cell*, 111(1):13–6.

Simonsen, K. L., Churchill, G. A., and Aquadro, C. F. (1995). Properties of statistical tests of neutrality for DNA polymorphism data. *Genetics*, 141(1):413–29.

Sorensen, T. I., Nielsen, G. G., Andersen, P. K., and Teasdale, T. W. (1988). Genetic and environmental influences on premature death in adult adoptees. *N Engl J Med*, 318(12):727–32.

Spellberg, B. and Edwards Jr, J. E. (2001). Type 1/Type 2 Immunity in Infectious Diseases. *Clin Infect Dis*, 32(1):76–102.

Spielman, R. S., McGinnis, R. E., and Ewens, W. J. (1993). Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). *Am J Hum Genet*, 52(3):506–16.

Stephens, J. C., Schneider, J. A., Tanguay, D. A., Choi, J., Acharya, T., Stanley, S. E., Jiang, R., Messer, C. J., Chew, A., Han, J. H., Duan, J., Carr, J. L., Lee, M. S., Koshy, B., Kumar, A. M., Zhang, G., Newell, W. R., Windemuth, A., Xu, C., Kalbfleisch, T. S., Shaner, S. L., Arnold, K., Schulz, V., Drysdale, C. M., Nandabalan, K., Judson, R. S., Ruano, G., and Vovis, G. F. (2001a). Haplotype variation and linkage disequilibrium in 313 human genes. *Science*, 293(5529):489–93.

Stephens, M., Smith, N. J., and Donnelly, P. (2001b). A new statistical method for haplotype reconstruction from population data. *Am J Hum Genet*, 68(4):978–89.

Stevenson, M. M. and Ghadirian, E. (1989). Human recombinant tumor necrosis factor alpha protects susceptible A/J mice against lethal Plasmodium chabaudi AS infection. *Infect Immun*, 57(12):3936–9.

Stoute, J. A., Kester, K. E., Krzych, U., Wellde, B. T., Hall, T., White, K., Glenn, G., Ockenhouse, C. F., Garcon, N., Schwenk, R., Lanar, D. E., Sun, P., Momin, P., Wirtz, R. A., Golenda, C., Slaoui, M., Wortmann, G., Holland, C., Dowler, M., Cohen, J., and Ballou, W. R. (1998). Long-term efficacy and immune responses following immunization with the RTS,S malaria vaccine. *J Infect Dis*, 178(4):1139–44.

Taillon-Miller, P., Bauer-Sardina, I., Saccone, N. L., Putzel, J., Laitinen, T., Cao, A., Kere, J., Pilia, G., Rice, J. P., and Kwok, P. Y. (2000). Juxtaposed regions of extensive and minimal linkage disequilibrium in human Xq25 and Xq28. *Nat Genet*, 25(3):324–8.

Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, 123(3):585–95.

Tang, K., Fu, D. J., Julien, D., Braun, A., Cantor, C. R., and Koster, H. (1999). Chip-based genotyping by mass spectrometry. *Proc Natl Acad Sci U S A*, 96(18):10016–20.

Taverne, J., Tavernier, J., Fiers, W., and Playfair, J. H. (1987). Recombinant tumour necrosis factor inhibits malaria parasites in vivo but not in vitro. *Clin Exp Immunol*, 67(1):1–4.

Thompson, E. A., Deeb, S., Walker, D., and Motulsky, A. G. (1988). The detection of linkage disequilibrium between closely linked markers: RFLPs at the AI-CIII apolipoprotein genes. *Am J Hum Genet*, 42(1):113–24.

Tishkoff, S. A., Varkonyi, R., Cahinhinan, N., Abbes, S., Argyropoulos, G., Destro-Bisol, G., Drousiotou, A., Dangerfield, B., Lefranc, G., Loiselet, J., Piro, A., Stoneking, M., Tagarelli, A., Tagarelli, G., Touma, E. H., Williams, S. M., and Clark, A. G. (2001). Haplotype diversity and linkage disequilibrium at human G6PD: recent origin of alleles that confer malarial resistance. *Science*, 293(5529):455–62.

Tomura, M., Yu, W. G., Ahn, H. J., Yamashita, M., Yang, Y. F., Ono, S., Hamaoka, T., Kawano, T., Taniguchi, M., Koezuka, Y., and Fujiwara, H. (1999). A novel function of Valpha14+CD4+NKT cells: stimulation of IL-12 production by antigen-presenting cells in the innate immune system. *J Immunol*, 163(1):93–101.

Valente, G., Ozmen, L., Novelli, F., Geuna, M., Palestro, G., Forni, G., and Garotta, G. (1992). Distribution of interferon-gamma receptor in human tissues. *Eur J Immunol*, 22(9):2403–12.

van der Heyde, H. C., Pepper, B., Batchelder, J., Cigel, F., and Weidanz, W. P. (1997). The time course of selected malarial infections in cytokine-deficient mice. *Exp Parasitol*, 85(2):206–13.

van Hensbroek, M. B., Palmer, A., Onyiorah, E., Schneider, G., Jaffar, S., Dolan, G., Memming, H., Frenkel, J., Enwere, G., Bennett, S., Kwiatkowski, D., and Greenwood, B. (1996). The effect of a monoclonal antibody to tumor necrosis factor on survival from childhood cerebral malaria. *J Infect Dis*, 174(5):1091–7.

Vannucchi, A. M., Grossi, A., Rafanelli, D., Statello, M., Cinotti, S., and Rossi-Ferrini, P. (1994). Inhibition of erythropoietin production in vitro by human interferon gamma. *Br J Haematol*, 87(1):18–23.

Venter, J. C., Adams, M. D., Myers, E. W., Li, P. W., Mural, R. J., Sutton, G. G., Smith, H. O., Yandell, M., Evans, C. A., Holt, R. A., Gocayne, J. D., Amanatides, P., Ballew, R. M., Huson, D. H., Wortman, J. R., Zhang, Q., Kodira, C. D., Zheng, X. H., Chen, L., Skupski, M., Subramanian, G., Thomas, P. D., Zhang, J., Gabor Miklos, G. L., Nelson, C., Broder, S., Clark, A. G., Nadeau, J., McKusick, V. A., Zinder, N., Levine, A. J., Roberts, R. J., Simon, M., Slayman, C., Hunkapiller, M., Bolanos, R., Delcher, A., Dew, I., Fasulo, D., Flanigan, M., Florea, L., Halpern, A., Hannenhalli, S., Kravitz, S., Levy, S., Mobarry, C., Reinert, K., Remington, K., Abu-Threideh, J., Beasley, E., Biddick, K., Bonazzi, V., Brandon, R., Cargill, M., Chandramouliswaran, I., Charlab, R., Chaturvedi, K., Deng, Z., Di Francesco, V., Dunn, P., Eilbeck, K., Evangelista, C., Gabrielian, A. E., Gan, W., Ge, W., Gong, F., Gu, Z., Guan, P., Heiman, T. J., Higgins, M. E., Ji, R. R., Ke, Z., Ketchum, K. A., Lai, Z., Lei, Y., Li, Z., Li, J., Liang, Y., Lin, X., Lu, F., Merkulov, G. V., Milshina, N., Moore, H. M., Naik, A. K., Narayan, V. A., Neelam, B., Nusskern, D., Rusch, D. B., Salzberg, S., Shao, W., Shue, B., Sun, J., Wang, Z., Wang, A., Wang, X., Wang, J., Wei, M., Wides, R., Xiao, C., Yan, C., et al. (2001). The sequence of the human genome. *Science*, 291(5507):1304–51.

Villella, A., Picard, C., Jouanguy, E., Dupuis, S., Popko, S., Abughali, N., Meyerson, H., Casanova, J. L., and Hostoffer, R. W. (2001). Recurrent Mycobacterium avium Osteomyelitis Associated With a Novel Dominant Interferon Gamma Receptor Mutation. *Pediatrics*, 107(4):E47.

Warrell, D., Molyneux, M., and Beales, P. (1990). Severe and complicated malaria. *Trans R Soc Trop Med Hyg*, (84 (supplement 2)):1–65.

Watterson, G. A. (1975). On the number of segregating sites in genetical models without recombination. *Theor Popul Biol*, 7(2):256–76.

Weatherall, D., Clegg, J., and Kwiatkowski, D. (1997). The role of genomics in studying genetic susceptibility to infectious disease. *Genome Res*, 7(10):967–73.

Weatherall, D. J. (2001). Phenotype-genotype relationships in monogenic disease: lessons from the thalassaemias. *Nat Rev Genet*, 2(4):245–55.

Weiss, K. M. and Clark, A. G. (2002). Linkage disequilibrium and the mapping of complex human traits. *Trends Genet*, 18(1):19–24.

WHO (1990). Division of Control of Tropical Diseases: Severe and Complicated Malaria. *Trans R Soc Trop Med Hyg*, 84 (supplement 2):1–65.

WHO (1997). *Weekly Epidemiological Record*, 72(36):269–276.

Wingender, E., Chen, X., Hehl, R., Karas, H., Liebich, I., Matys, V., Meinhardt, T., Pruss, M., Reuter, I., and Schacherer, F. (2000). TRANSFAC: an integrated system for gene expression regulation. *Nucleic Acids Res*, 28(1):316–9.

Wright, S. (1921). Systems of Mating. *Genetics*, 6:111–178.

Wright, S. (1978). Variability within and among Natural Populations. In *Evolution and the Genetics of Populations*, volume 4. University of Chicago Press, Chicago.

Xie, M. H., Aggarwal, S., Ho, W. H., Foster, J., Zhang, Z., Stinson, J., Wood, W. I., Goddard, A. D., and Gurney, A. L. (2000). Interleukin (IL)-22, a novel human cytokine that signals through the interferon receptor-related proteins CRF2-4 and IL-22R. *J Biol Chem*, 275(40):31335–9.

Xu, C. F., Lewis, K., Cantone, K. L., Khan, P., Donnelly, C., White, N., Crocker, N., Boyd, P. R., Zaykin, D. V., and Purvis, I. J. (2002). Effectiveness of computational methods in haplotype prediction. *Hum Genet*, 110(2):148–56.

Xu, W., Presnell, S. R., Parrish-Novak, J., Kindsvogel, W., Jaspers, S., Chen, Z., Dillon, S. R., Gao, Z., Gilbert, T., Madden, K., Schlutsmeyer, S., Yao, L., Whitmore, T. E., Chandrasekher, Y., Grant, F. J., Maurer, M., Jelinek, L., Storey, H., Brender, T., Hammond, A., Topouzis, S., Clegg, C. H., and Foster, D. C. (2001). A soluble class II cytokine receptor, IL-22RA2, is a naturally occurring IL-22 antagonist. *Proc Natl Acad Sci U S A*, 98(17):9511–6.

Ye, K., Dinarello, C. A., and Clark, B. D. (1993). Identification of the promoter region of human interleukin 1 type I receptor gene: multiple initiation sites, high G+C content, and constitutive expression. *Proc Natl Acad Sci U S A*, 90(6):2295–9.

Zhang, K., Deng, M., Chen, T., Waterman, M. S., and Sun, F. (2002). A dynamic programming algorithm for haplotype block partitioning. *Proc Natl Acad Sci U S A*, 99(11):7335–9.

Zhang, L., Cui, X., Schmitt, K., Hubert, R., Navidi, W., and Arnheim, N. (1992). Whole genome amplification from a single cell: implications for genetic analysis. *Proc Natl Acad Sci U S A*, 89(13):5847–51.

Ziegler-Heitbrock, H. W., Thiel, E., Futterer, A., Herzog, V., Wirtz, A., and Riethmuller, G. (1988). Establishment of a human cell line (Mono Mac 6) with characteristics of mature monocytes. *Int J Cancer*, 41(3):456–61.

Zimonjic, D. B., Rezanka, L. J., Evans, C. H., Polymeropoulos, M. H., Trent, J. M., and Popescu, N. C. (1995). Mapping of the immune interferon gamma gene (IFNG) to chromosome band 12q14 by fluorescence in situ hybridization. *Cytogenet Cell Genet*, 71(3):247–8.