

The European CF Twin and Sibling Study; Genetic Susceptibility to Infectious Diseases

Von der Naturwissenschaftlichen Fakultät der
Gottfried Wilhelm Leibniz

Universität Hannover

zur Erlangung des Grades

DOKTOR DER NATURWISSENSCHAFTEN

- Dr. rer. nat. -

genehmigte Dissertation

von

M.F.Sc, VINOD KUMAR MAGADI GOPALIAH

geboren am 01.06.1977 in Bangalore, Indien

Hannover 2007

*This work is dedicated to
my dearest Parents
and to my wonderful brothers*

Die vorliegende Arbeit wurde in der Klinischen Forschergruppe „Molekulare Pathologie der Mukoviszidose“, Zentrum Biochemie und Zentrum Kinderheilkunde der Medizinischen Hochschule Hannover in der Zeit vom 01.10.2003 bis zum 30.09.2006 unter der Leitung von Prof. Dr. Dr. Burkhard Tümmler angefertigt.

Tag der Promotion 24.01.2007

Referent: Prof. Dr. Dr. Burkhard Tümmler
 Klinische Forschergruppe OE6711
 Zentrum Biochemie und Zentrum Kinderheilkunde
 Medizinische Hochschule Hannover

Koreferent Prof. Dr. Gerald-F. Gerlach
 Institut für Mikrobiologie
 Zentrum für Infektionsmedizin
 Stiftung Tierärztliche Hochschule Hannover

A token of gratitude

I would like to express my whole hearted thanks and sincere gratitude to my major advisor, Prof. Dr. Dr. Burkhard Tümmler, Klinische Forschergruppe, Medizinische Hochschule, Hannover, for giving me the scientific freedom and the wonderful opportunity to accomplish this thesis. His pragmatic guidance and constant encouragement throughout the period of my research is gratefully acknowledged.

I would like to thank Prof. Dr. Gerald-F. Gerlach for reviewing my thesis, Prof. Dr. Helmut Holtmann and Prof. Dr. Karl-Heinz Bellgardt for being part of the review committee.

I owe my profound thanks to my supervisor Dr. Frauke Stanke for giving me an unfailing guidance, friendly nature, scholarly advice and the confidence she has showered in me and in my project. Her amazing patience, terrific organizing skills, generous support and deep insights enabled me to succeed in my profession.

I am very much thankful and grateful to,

Frau Silke Jansen and Ms. Stephanie Tamm for their friendly attitude and technical assistance

Dr. Andrea van Barneveld for helping me in setting up the western blot experiments

Dr. Lutz Wiehlmann and Dr. Jens Klockgether for their valuable suggestions and scientific discussions

Dr. Antje Munder for providing me the human cell lines

Ms. Gesa Puls for her help in this thesis preparation

Prof. Christoph Klein for letting me use the FACS and real-time PCR facilities

Dr. Chozhavendan Rathinam and Mr. Giridharan Appaswamy for their help in FACS and real-time PCR experiments

Dr. Kaan Boztug for his help in FACS experiments

Prof. Thomas F. Wienker and Prof. T. Becker for providing me the statistical analysis

Prof. Dave Ussery for performing functional annotation of non-coding variants

The Deutsche Forschungsgemeinschaft (DFG)-sponsored European Graduate College "Pseudomonas: Pathogenicity and Biotechnology" for providing me the fellowship.

I am thankful to Frau Helga Riehn-kopp, co-ordinator of the graduate college, for her kind support and timely help in bureaucratic issues.

My heartfelt thanks to Mr. Dieco Wuerdemann, Ms. Tammy Chang and Dr. Andrea van Barneveld for their friendship and caring, which have supported me in many ways.

I take this opportunity to express my sincere thanks to all the members of “Klinische Forschergruppe” and members of the EGK for their help and necessary support.

My special thanks to my “GURU’S” Dr. Bob Kennedy, Dr. Devaraja and Dr. Mohana Kumara for their moral support, affection, and caring.

I pay my whole hearted thanks to my dear friends Rama, Shanthi, Pal, Kummi, Benki, Umi, Viji, Mona, Giri, Bhanu, Shashi, Anu, Shilpa, Asha, Selvan, Hari, Raghu, Yadhu, Prajeeth and Shivu for their encouragement and cheerful attitude.

Last but not least, I am indebted to my parents and family members, whose love, blessings and sacrifice made it possible for me to accomplish this thesis.

Abstract

Cystic Fibrosis (CF) is the most severe common autosomal recessive congenital disorder in the Caucasian population. It is caused by molecular lesions in the cystic fibrosis transmembrane conductance regulator (*CFTR*) gene located on the long arm of human chromosome 7. While the *CFTR* determines the susceptibility of the airways to the opportunistic pathogen *Pseudomonas aeruginosa* in CF patients, progression and severity of the CF disease can not be predicted by the *CFTR* mutation genotype. Thus, the main objective of this study was to investigate *TLR2*, *TLR4*, *TLR5*, *TLR9*, *SP-D*, *CXCR2*, *PON*, *TNFR1*, *CD14*, and *CD95* as modulators of CF disease severity and susceptibility to *P. aeruginosa* infection. All ten candidate genes were targeted initially by one informative SNP (Single nucleotide polymorphism) to investigate them as CF disease severity modulators in *European CF Twin and Sibling Study* cohort containing 37 families with F508del-*CFTR* homozygous, exhibiting extreme clinical phenotypes. SNPs on *TLR2*, *TLR5* and *TLR9* failed to show the association. Polymorphisms on surfactant protein-D, *CXCR2* and *PON* locus showed only a minor association with CF disease severity. The significant association was found between polymorphisms within *TNFR1*, *TLR4*, *CD14*, *CD95* and CF disease severity. Hence, these four genes were investigated further by typing more SNPs and haplotype analysis. The genomic fragment containing the causative variant was identified by direct comparison of two-marker-haplotype-distributions between sib pair sets displaying a phenotypic contrast. Individuals carrying contrasting haplotypes were subjected to confirmatory sequencing at the outlined genomic fragment and the coding region. Sequencing analysis did not reveal any coding variants. The functional role of non-coding causative variants was explored by in-silico analysis and suitable phenotypic assays. Furthermore, these causative variants were analysed for their role in CF infectious disease in two independent CF cohorts stratified for *P. aeruginosa*-related endophenotypes such as onset of initial and chronic colonisation.

The haplotype analysis showed that the causative haplotype in *TLR4* was located upstream of *TLR4* exon 1 and was associated with CF disease severity but not with *P. aeruginosa* early or late colonisation.

TNFR1 first intron harboured a disease modifying haplotype. In-silico analysis predicted that the alterations in DNase hypersensitive sites, conserved non-coding sequences and inverted local repeats due to intron 1 variants may cause differential transcription. Consistently, western blot analysis showed that the levels of TNFR1 in CF patients' serum correlated with *TNFR1* causative haplotype. The haplotype analysis on *CD95* gene found a causative variant within intron 2. The in-silico analysis predicted a transcription regulatory region within intron 2 and the causative SNP was located within this regulatory region. Thus, it altered the binding site for c-Rel transcription factor. CF patients heterozygous for this intron 2 SNP had significantly lower levels of CD95 mRNA, isolated from rectal suction biopsies of F508del-*CFTR* homozygous patients. This effect was specific to either epithelial cells or to CF context as mRNA quantification by real-time PCR and surface expression by FACS on peripheral blood mononuclear cells from healthy individuals did not reveal any association. *CD95* polymorphisms were not associated with *P. aeruginosa* early or late chronic colonisation.

CD14 promoter polymorphism and 3' UTR polymorphism were associated with CF disease severity and age at onset of *P. aeruginosa* chronic colonisation. The diplotype analysis on *CD14* locus revealed a significant association with age-dependent risk to acquire *P. aeruginosa* colonisation, the level of sCD14 in serum of CF patients and also associated with *P. aeruginosa* O-antigen phenotype. The 3' UTR polymorphism was predicted to alter the binding site for microRNA on CD14 RNA and also binding site for mRNA processing proteins.

In summary, we identified *TLR4*, *TNFR1*, and *CD95* as potential genetic modulators of CF disease severity and *CD14* as both the CF disease modulator as well as the modulator of age dependent risk to acquire *P. aeruginosa* colonisation among CF patients. Furthermore, the importance of non-coding variants in CF disease modulation was illustrated.

Key Words: Cystic Fibrosis, Genetic modulators, Innate immunity

Kurzfassung

Cystische Fibrose ist die häufigste schwere autosomal rezessiv kogenital vererbte Krankheit unter Kaukasiern. Sie wird verursacht durch eine molekulare Veränderung im cystic fibrosis transmembrane conductance regulator- (*CFTR*-) Gen, das sich auf dem langen Arm des menschlichen Chromosoms 7 befindet. Während das *CFTR* die Infektionsanfälligkeit der Atemwege für den Opportunisten *Pseudomonas aeruginosa* in CF- Patienten beeinflusst, können der Verlauf und der Schweregrad der Erkrankung nicht durch den Genotyp der *CFTR*- Mutation vorhergesagt werden. Daher konzentriert sich diese Arbeit hauptsächlich auf die Untersuchung von *TLR2*, *TLR4*, *TLR5*, *TLR9*, *SP-D*, *CXCR2*, *PON*, *TNFR1*, *CD14* und *CD95* als Modulatoren für den Schweregrad der CF- Erkrankung und die Anfälligkeit für *P. aeruginosa*- Infektionen. Für alle zehn Kandidatengene wurde zu Beginn ein informativer single nucleotide polymorphism (SNP) ausgewählt, um sie als Modulatoren des Schweregrades der CF- Erkrankung in der Kohorte der *Europäischen CF Zwillings- und Geschwisterstudie* zu untersuchen, die 37 Familien mit F508del- *CFTR*- Homozygoten mit extremen klinischen Phänotypen umfasste. Die SNPs auf *TLR2*, *TLR5* und *TLR9* zeigten keine Assoziation. Polymorphismen im *SP-D*, *CXCR2* und *PON*- Locus zeigten nur einen geringen Zusammenhang mit dem Schweregrad der CF- Erkrankung. Daher wurden diese sechs Gene von einer weiteren Feinkartierung ausgeschlossen. Ein signifikanter Zusammenhang wurde zwischen dem Schweregrad der CF- Erkrankung und Polymorphismen in *TNFR1*, *TLR4*, *CD14*, *CD95* gefunden. Daher wurden diese vier Gene durch Typisierung weiterer SNPs und Haplotypenanalyse untersucht. Das genomische Fragment, welches die funktionelle Variante trägt, wurde identifiziert durch direkte Vergleiche zwischen Zwei- Marker- Haplotypen- Verteilungen zwischen Geschwisterpaaren, die einen unterschiedlichen Phänotyp zeigten. Individuen, welche verschiedene Haplotypen trugen, wurden in dem ausgewählten genomischen Fragment und der kodierenden Region sequenziert. Die Sequenzanalyse zeigte keine kodierenden funktionellen Varianten. Die Funktion der nicht- kodierenden funktionellen Sequenzen wurde durch in- silico- Analyse und geeignete Phänotypstests untersucht. Darüber hinaus wurden diese funktionellen Varianten auf ihre Rolle in der CF- Erkrankung bei zwei unabhängigen CF- Kohorten, die nach *P. aeruginosa*- assoziierten Endophänotypen, wie z. B. der Zeitpunkt der beginnenden und der chronischen Kolonisation, stratifiziert wurden.

Die Haplotypenanalyse zeigte, dass sich der funktionelle Haplotyp im *TLR4* stromaufwärts des *TLR4*- Exon 1 befand und mit dem Schweregrad der CF- Erkrankung, nicht aber mit der frühen oder späten Kolonisation durch *P. aeruginosa* assoziiert war. Im ersten *TNFR1*- Intron lag der den Schweregrad der Erkrankung modifizierende Haplotyp. Die in- silico- Analyse sagte voraus, dass durch die Intron 1- Varianten Veränderungen in DNase hypersensitiven Stellen, konservierten nicht- kodierenden Sequenzen und invertierten lokalen Repeats entstehen könnten, die eine veränderte Transkription hervorrufen könnten. Entsprechend zeigte die Western Blot- Analyse, dass die *TNFR1*- Mengen im Serum von CF- Patienten mit dem *TNFR1* funktionellen Haplotyp korrelierten.

Die Haplotypenanalyse des *CD95*- Gens zeigte eine funktionelle Variante im Intron 2. Die in- silico- Analyse sagte eine transkriptionsregulatorische Region im Intron 2 voraus und der funktionelle SNP lag in dieser regulatorischen Region. Auf diese Weise veränderte er die Bindungsstelle für den c- Rel- Transkriptionsfaktor. CF- Patienten, die heterozygot für diesen Intron 2- SNP waren, hatten signifikant geringere *CD95*- mRNA- Mengen, die aus Rektumsaugbiopsien von F508del- *CFTR* homozygoten Patienten isoliert wurden. Dieser Effekt ist entweder spezifisch für Epithelzellen oder im Zusammenhang mit CF, denn die Quantifizierung der mRNA durch Real- time Polymerasekettenreaktion (PCR) und der Oberflächenexpression durch Fluoreszenzaktivierte Zellsortierung (FACS) bei peripheren mononukleären Blutzellen aus gesunden Individuen zeigte keinen Zusammenhang. Der *CD95*- Polymorphismus war nicht mit einer frühen oder späten *P. aeruginosa*- Kolonisation von CF- Patienten verknüpft. Der *CD14*- Promoter- Polymorphismus und der 3'UTR- Polymorphismus waren assoziiert mit dem Schweregrad der CF- Erkrankung und dem Alter in dem die chronische Kolonisation mit *P. aeruginosa* begann. Die Analyse der Diplotypen auf dem *CD14*- Locus zeigte einen signifikanten Zusammenhang mit dem altersabhängigen Risiko eine *P. aeruginosa*- Kolonisation zu erwerben, mit der s*CD14*- Menge im Serum von CF- Patienten und war auch mit dem *P. aeruginosa* O- Antigen- Phänotyp verknüpft. Für den 3'UTR- Polymorphismus wurde vorhergesagt, dass er die Bindungsstelle für mikroRNA auf der *CD14*- RNA und auch die Bindungsstelle für mRNA- Prozessierungsproteine verändern würde.

Wir haben *TLR4*, *TNFR1* und *CD95* als mögliche genetische Modulatoren des Schweregrades der CF- Erkrankung und *CD14* sowohl als Modulator der CF- Erkrankung als auch als Modulator des altersabhängigen Risikos eine *P. aeruginosa*- Kolonisation bei CF- Patienten zu erwerben identifiziert. Darüber hinaus wurde die Bedeutung nicht- kodierender Varianten bei der Modulation der CF- Erkrankung veranschaulicht.

Schlüsselbegriffe: Cystische Fibrose, Genetische modulatoren, Angeborene immunität

1.	Introduction	1
1.1.	Cystic Fibrosis (CF)	1
1.2.	The characteristics of <i>Pseudomonas aeruginosa</i> , a successful pathogen	1
1.3.	Host-pathogen interaction in cystic fibrosis	2
1.4.	CF pulmonary hyperinflammatory phenotype	3
1.5.	Innate immunity	5
1.6.	Cystic Fibrosis Genetic Modulators	6
1.7.	Approaches to study CF genetic modulators	7
1.8.	Aim of the thesis	9
2.	Patients and methods	10
2.1.	Patient cohorts	10
2.1.1.	European CF Twin and Siblings	10
2.1.1.1.	Selection of extreme phenotypes	10
2.1.2.	F508del homozygous CF twin and siblings stratified for <i>P. aeruginosa</i> colonisation	11
2.1.3.	F508del homozygous unrelated CF patients with early and late <i>P. aeruginosa</i> chronic colonisation	12
2.1.4.	F508del homozygous unrelated CF patients stratified for birth cohorts	13
2.1.5.	Unrelated F508del homozygous CF patients recruited for global transcriptome analysis	13
2.2.	Methods	14
2.2.1.	Isolation of high molecular weight DNA from blood	14
2.2.2.	DNA isolation from neutrophils	14
2.2.3.	Genotyping	14
2.2.3.1.	Selection of genetic markers	15
2.2.3.2.	Analysis of restriction fragment length polymorphisms (RFLPs)	15
2.2.3.3.	Analysing polymorphic microsatellites by direct blotting process	15
2.2.4.	Genotyping data evaluation	17
2.2.4.1.	Evaluation of genotyping data from CF twin and siblings and from cohorts stratified for <i>P. aeruginosa</i> colonisation	17
2.2.5.	Statistical Analysis	18
2.2.5.1.	Family based evaluation	18
2.2.5.2.	Case-control analysis	18
2.2.5.3.	Correction for sib-pair dependence and multiple testing	18
2.2.5.4.	Hardy-Weinberg equilibrium analysis using FINETTI program	19
2.2.6.	PCR based methods	19
2.2.6.1.	PCR in multiwell plates	19
2.2.6.2.	Long-range PCR	19
2.2.6.3.	Sequencing long-range PCR products	19
2.2.6.4.	Pre-mRNA length determination by PCR	20
2.2.7.	RNA isolation	20
2.2.8.	Real Time PCR	20
2.2.8.1.	First-strand cDNA synthesis using oligo(dT) primers	21

2.2.8.2.	First-strand cDNA synthesis using Random primers	21
2.2.8.3.	Quantification of mRNA on the LightCycler	22
2.2.9.	MicroRNA detection	24
2.2.9.1	MicroRNA isolation	25
2.2.9.2.	Biotin labelling of small RNA fraction (<200 nt)	25
2.2.9.3.	Dot Blotting	25
2.2.9.4.	Prehybridization	25
2.2.9.5.	Hybridization	25
2.2.9.6.	Detection	25
2.2.10.	Western Blotting	26
2.2.10.1.	Sample preparation and separating on a gel	26
2.2.10.2.	Protein blotting	26
2.2.10.3.	Immune detection	26
2.2.11.	ELISA	27
2.2.12.	PBMC isolation	27
2.2.13.	Flow cytometry	28
2.2.13.1.	Staining peripheral blood mononuclear cells	28
2.2.14.	Serotyping of <i>P. aeruginosa</i> CF isolates	31
2.2.15.	Electronic database resources	31
3.	Results and discussion	32
3.1.	Analysis of <i>TLR2</i>, <i>TLR5</i> and <i>TLR9</i> as CF modulators	33
3.1.1.	Toll like receptor (TLR) 2	33
3.1.2.	Toll like receptor (TLR) 5	34
3.1.3.	Toll like receptor (TLR) 9	35
3.1.4.	Role of <i>TLR2</i> , <i>TLR5</i> and <i>TLR9</i> polymorphisms in CF	36
3.2.	Evaluation of <i>SFTPD</i>, <i>IL8RB</i> and <i>PON</i> as CF genetic modulators	38
3.2.1	Surfactant Protein (SP)-D (<i>SFTPD</i>)	38
3.2.1.2.	Role of surfactant protein-D in cystic fibrosis	39
3.2.2.	<i>CXCR2</i> (<i>IL8RB</i> interleukin 8 receptor, beta)	40
3.2.2.1.	No association with <i>P. aeruginosa</i> early or late chronic colonisation among unrelated CF patients	41
3.2.2.2.	Role of <i>CXCR2</i> gene in cystic fibrosis	42
3.2.3.	<i>PON</i> (paraoxonase) gene cluster	42
3.2.3.1.	Role of <i>PON</i> polymorphisms in cystic fibrosis	44
3.3.	Analysis of TNFα receptor <i>TNFRSF1A</i> as a modulator in cystic fibrosis	46
3.3.1.	Rationale for choosing <i>TNFR1</i> as sequencing target	46
3.3.2.	Sequence analysis of <i>TNFR1</i> coding region	47
3.3.3.	Sequencing was focused on the intron 1 of <i>TNFR1</i>	49
3.3.4.	Sequencing analysis of <i>TNFR1</i> intron 1	50
3.3.5.	Functional annotation of <i>TNFRSF1A</i> intron 1	52
3.3.6.	Soluble <i>TNFR1</i> levels in serum of CF patients are associated with D12S889 genotype	53
3.3.7.	Impact of <i>TNFR1</i> defect on innate immunity	54

3.4.	Analysis of TLR4 as a modulator of cystic fibrosis	56
3.4.1.	Association with disease severity in CF on <i>TLR4</i> revealed by single-marker analysis	56
3.4.2.	Association with disease severity in CF on <i>TLR4</i> confirmed by haplotype analysis	57
3.4.3.	Sequence analysis of <i>TLR4</i> gene	58
3.4.4.	<i>TLR4</i> polymorphism is not associated with <i>P. aeruginosa</i> early or late chronic colonisation among unrelated CF patients	58
3.4.5.	<i>TLR4</i> expression analysis by FACS on peripheral blood mononuclear cells	59
3.4.6.	No informative markers within three kb region upstream of rs10759930	60
3.4.7.	Role of <i>TLR4</i> signaling in CF airways	60
3.5.	Analysis of <i>CD14</i> as a modulator of cystic fibrosis	62
3.5.1.	Association with disease discordance in CF on <i>CD14</i> revealed by single-marker analysis	62
3.5.2.	Association with disease discordance in CF on <i>CD14</i> revealed by haplotype analysis	63
3.5.3.	<i>CD14</i> sequencing analysis	64
3.5.4.	Skewed allele distribution on <i>CD14</i> among CF siblings	64
3.5.6.	CF twin and siblings cohort stratified for <i>P. aeruginosa</i> colonisation	65
3.5.7.	F508del homozygous unrelated CF patients with early and late <i>P. aeruginosa</i> chronic colonisation	67
3.5.8.	F508del homozygous unrelated CF patients stratified for birth cohorts	67
3.5.9.	Age dependent risk to acquire <i>P. aeruginosa</i> among CF twin and siblings shown by <i>CD14</i> diplotype analysis	69
3.5.10.	Association between <i>P. aeruginosa</i> O-antigen serotype and <i>CD14</i> diplotype	71
3.5.11.	Influence of <i>CD14</i> diplotype on the sCD14 levels in serum of CF patients	72
3.5.12.	How does <i>CD14</i> 3' UTR polymorphism determine sCD14 levels among CF patients?	74
3.5.13.	<i>CD14</i> 3' UTR polymorphism could affect CD14 mRNA 3' end formation	75
3.5.14.	MicroRNA mediated regulation of CD14 pre-mRNA	76
3.5.14.1.	<i>CD14</i> 3' UTR polymorphism, rs2563298, is located with in a microRNA binding site	77
3.5.14.2.	Biogenesis of hsa-miR-425-5p microRNA	78
3.5.14.3.	CD14 pre-mRNA, not mature mRNA, may be a target for miR-425-5p	79
3.5.14.4.	CD14 pre-mRNA is at least 973bp long from its stop codon	79
3.5.14.5.	MicroRNA binding efficiency is altered by <i>CD14</i> 3' UTR polymorphism rs2563298	80
3.5.14.6.	Standardization of hybridization technique for microRNA detection from different human cell lines	81
3.5.15.	Role of <i>CD14</i> polymorphisms in cystic fibrosis	83

3.6.	Analysis of <i>CD95</i> as a modulator of cystic fibrosis	86
3.6.1.	Skewed allele distribution among CF siblings on <i>CD95</i>	87
3.6.2.	Association between <i>CD95</i> and CF disease severity revealed by single-marker analysis	88
3.6.3.	Haplotype analysis confirmed the association between <i>CD95</i> and CF disease severity	88
3.6.4.	Sequence analysis of <i>CD95</i> coding region and intron 2	90
3.6.5.	Fine-mapping within haplotype block rs2296603-rs7901656	91
3.6.6.	Conserved non-coding sequences (CNS) in <i>CD95</i> gene	92
3.6.7.	Low-helical stability regions within intron 2 of <i>CD95</i> gene	93
3.6.8.	The CNS on intron 2 of <i>CD95</i> gene act as a hot-spot for transcription factor binding	94
3.6.9.	The SNP rs7901656, but not rs1800682, determines the transcriptional activity of <i>CD95</i> among CF patients	94
3.6.10.	Real Time PCR analysis to determine the expression status of the <i>CD95</i> in peripheral blood mononuclear cells	96
3.6.11.	<i>CD95</i> surface expression level on peripheral blood mononuclear cells analysed by FACS	97
3.6.12.	<i>CD95</i> intron SNP (rs7901656) and promoter SNP (rs1800682) are not associated with <i>P. aeruginosa</i> early or late colonisation	98
3.6.13.	The role of <i>CD95/CD95L</i> signaling in cystic fibrosis	98
4.	Conclusions	102
4.1.	Single marker analysis showed no association between <i>TLR2</i> , <i>TLR5</i> and <i>TLR9</i> polymorphisms and CF disease severity	102
4.2.	Surfactant protein-D and CF disease severity	102
4.3.	<i>CXCR2</i> is a modulator of CF disease discordance	103
4.4.	Paraoxonase (<i>PON</i>) gene cluster as a modulator of CF disease severity	103
4.5.	Haplotype block with in <i>TNFR1</i> first intron modifies the CF disease severity	104
4.6.	The <i>TLR4</i> promoter variants modulate CF disease severity but not age at onset of <i>P. aeruginosa</i> colonisation	105
4.7.	The <i>CD14</i> polymorphisms determine both CF disease severity as well as the age at onset of <i>P. aeruginosa</i> chronic colonisation	105
4.8.	<i>CD95</i> is a potential modulator of CF disease severity	107
4.9.	Role of non-coding DNA in determining the susceptibility to infectious diseases	108
5.	Literature	110
6.	Abbreviations	115
7.	Appendices	116

1. Introduction

1.1. Cystic Fibrosis (CF)

CF is known as the most common severe autosomal recessive disease within the Caucasian population, exhibiting an incidence of 1 in 2500 births (Welsh et al, 1995). This monogenic disorder is caused due to mutations in the Cystic Fibrosis Transmembrane conductance Regulator (*CFTR*) gene, which is located on chromosome 7q31.3 (Kerem et al, 1989). The 230 kb of *CFTR* gene encompasses 27 exons which encode for a polypeptide of 1480 amino acids (Riordan et al, 1989, Zielinski et al, 1991). *CFTR* is responsible for the cAMP-activated anionic, with chloride as main substrate, conductance of epithelial cell apical membranes. This anion channel is also involved in cAMP-dependent bicarbonate secretion in airway, intestinal epithelia and exocrine glands (Reddy et al, 2001) and transport of biomolecules like glutathione (GSH) (Kogan et al, 2003) and regulation of other ion channels (Bear et al, 1992, Welsh et al., 1992, 1993, Vankeerberghen et al 2002). More than 1500 disease-associated mutations have been reported to the CF Genetic Analysis Consortium database (www.genet.sickkids.on.ca/cftr/) since the identification of the *CFTR* gene. The majority of which are amino acid substitutions, frameshifts, splice site or nonsense mutations. These mutations can cause disruption of *CFTR* function within epithelial cells in different ways, ranging from complete loss of protein to surface expression with poor chloride conductance (Welsh et al., 1993) depending on the kind of mutation. For example, the most common CF mutation, F508del, a three base pair deletion that codes for phenylalanine at position 508, causes the protein to misfold leading to premature degradation by the ubiquitin proteasome system (Denning et al., 1992). Thus, loss of *CFTR* function at the cell surface leads to mortality in CF patients because of altered hydration of all exocrine epithelia and persistent lung infections (Welsh et al., 1993).

1.2. The characteristics of *Pseudomonas aeruginosa*, a successful pathogen

Pseudomonas aeruginosa is a gram negative, environmental bacterium. It is widely distributed and can grow in almost any aqueous habitat, including soil, surface waters, sewage, plants, and various foods, such as leafy vegetables and fresh fruit juice (Bonten et al., 1999). Although early infection of the CF airways is mostly caused by *Staphylococcus aureus* and *Haemophilus influenza*, chronic infection with *P.*

aeruginosa is of most significance as it is responsible for most of the morbidity and mortality of CF patients (Gibson et al., 2003). The main feature of this bacterium is its very large genome of 6.3 Mbp, which offers a benefit of tremendous adaptability to multiple different environments, including CF airways. The genome of *P. aeruginosa* encodes several cell surface (LPS, flagella, pili, alginate) and secreted (pyocyanin, pyoverdine, protease, elastase, phospholipase, rhamnolipids, exotoxins) virulence factors along with numerous metabolic enzymes. Also CF isolates have been found to have pathogenicity islands, which contain distinct group of genes which directly contribute to disease (He et al., 2004). Furthermore, *P. aeruginosa* isolates have efficient quorum sensing (QS) system. Importantly, the QS regulated genes are virulence genes and genes involved in biofilm formation (Diggle et al., 2002; Whiteley et al., 1999). In *P. aeruginosa*, *las*, *rhl* and PQS (*Pseudomonas* quinolone signal) are the three QS systems. Hierarchical interaction of *las* and *rhl* systems activates autoinducer synthase to produce N-(3-oxododecanoyl)-L-homoserine lactone and N-butyryl-L-homoserine lactone, respectively (Smith and Iglewski, 2003). The ability of *P. aeruginosa* to synchronize the regulation of virulence genes in an entire population explains its increased pathogenicity among QS capable strains compared to QS deficient mutants in some animal models (Tang et al., 1996; Pearson et al., 2000; Wu et al., 2001, Lesprit et al., 2003). Therapy of *P. aeruginosa* pulmonary infection has been problematic, largely due to the high intrinsic resistance of this organism to antimicrobial agents because of the low permeability of the outer membrane (Hancock, 1998), combined with the presence of both beta-lactamases (Philippon et al., 1997) and multidrug efflux pumps (Poole et al., 1996; Köhler et al., 1997).

1.3. Host-pathogen interaction in cystic fibrosis

The molecular mechanism of resistance and susceptibility to a pathogen in humans involves a complex cross talk between host factors and pathogen virulence factors. *P. aeruginosa* cell surface virulence factors are primarily ligands for the pattern recognition/ innate immune receptors on host cells. LPS (Lipopolysaccharide), the major constituent of outer membrane of *P. aeruginosa*, play a major role in inducing immune response mainly by its role in recognition by TLR4/CD14/MD-2 innate receptors (Hajjar et al., 2002; Backhed et al., 2003). In addition, a key component of LPS, the Lipid A, activates multiple pro-inflammatory pathways by its CF specific

modifications that enhance TLR4 activation (Ernst et al., 1999; 2003). During the time of colonisation in CF lung, *P. aeruginosa* isolates switch to quite distinctive phenotypes. They become antibiotic resistant and frequently mucoid characterised by excessive alginate production (Govan and Deretic, 1996). The other phenotypic morphotypes, termed small colony variants (SCV), were shown to be associated with poor lung function (Haussler et al., 1999). Furthermore, *P. aeruginosa* mucoid CF isolates often display rough LPS phenotype due to loss of O-polysaccharide side chains (Hancock et al., 1983) which seems to have a role in LPS signalling. Similarly, flagella and pili, the appendages primarily required for *P. aeruginosa* motility and also facilitate in attachment of *P. aeruginosa* to host cells, play significant roles in inducing inflammation. Binding of *P. aeruginosa* via flagella to asialoGM1 is shown to be sensed by TLR2 and TLR5 (Adamo et al., 2004) and in turn activates Src-Ras-ERK1/2-NF- κ B pathway to release IL (Interleukin) -8 (Lillehoj et al., 2004). Since flagella are very immunogenic, *P. aeruginosa* adapt to the CF lung by selecting aflagellar mutants to evade host response during chronic colonisation (Mahenthiralingam et al., 1994). The quorum sensing (QS) molecules were also reported be immunomodulatory agents and depress host responses (Telford et al., 1998; Smith et al., 2002). On the other hand, most *P. aeruginosa* strains attenuate their virulence after some time during CF lung colonisation. During chronic phase of infection, *P. aeruginosa* isolates secrete less of most common immuno-stimulants such as proteases, exolipase, exotoxin A and hemolysin *in vitro* than their clonal relatives isolated during the initial phase of the infection (Tümmler et al., 1997).

1.4. CF pulmonary hyperinflammatory phenotype

The lung disease of cystic fibrosis is characterized by a cycle of airway obstruction, infection, and inflammation (Gibson et al., 2003). Infection with bacterial pathogens among CF patients followed by intense neutrophilic localisation to the peribronchial and endobronchial spaces (Khan et al., 1995, Muhlebach et al., 1999) leads to airway inflammation with elevated interleukin-8 and neutrophil elastase (Bonfield et al., 1995). As a consequence, CF airways show a prolonged inflammatory response (Gibson et al., 2003). It has been suggested that this inflammatory response remains aggressive by local airway epithelium-pathogen interactions but not by systemic immune response (Chmiel et al., 2002). There is also evidence from *in vivo* experiments that production

of anti-inflammatory cytokines like IL-10 (Bonfield et al., 1999) and lipoxins (Karp et al., 2004) are reduced in CF airway. Further neutrophil influx is mediated and sustained by IL-8, produced by stimulated epithelial cells, macrophages and neutrophils (Chmiel et al., 2002, Tirouvanziam et al., 2000). Thus, inflammation in the CF lung is dominated by neutrophils and their products, including neutrophil elastase. It is reported that the neutrophil elastase can induce up-regulation of NF-kB activation and IL-8 expression (Devaney et al., 2003). Additionally, neutrophil elastase and cathepsin G stimulate airway gland secretion, which removes bacterial pathogens from airway epithelial cells into the airway lumen (Döring and Worlitzsch, 2000). When neutrophil accumulation is exaggerated in CF, it can cause progressive damage to bronchial epithelium significantly mediated by neutrophil elastase (Taggart et al., 2000). Thus, the activated neutrophils are considered to be the primary effector cells for the pathogenesis of CF lung disease. In contrast, a comprehensive study by Aldallal et al. (2002) showed that different inflammatory responses of normal, CF and CFTR-corrected airway epithelial cells were likely due to differences in the cell types and they were unrelated to the presence of CFTR. It is also suggested that changes in volume, ionic composition, and the level of glutathione (GSH) in airway surface liquid due to absence of CFTR may cause increased concentration of secreted products which may be proinflammatory even in the absence of bacterial infections (Machen, 2006). Further, a study from Weber et al. (2001) proposed a model in which mutation in *CFTR* leads to accumulation of excessive amounts of misfolded deltaF508 CFTR in the endoplasmic reticulum (ER) lumen, which may alter the intracellular calcium signalling due to ER stress and in turn activation of NF-kB in the absence of bacterial stimulus. Additionally, Tirouvanziam et al. (2000) demonstrated in human fetal CF airway grafts that before any infection, CF airways are in proinflammatory state characterised by increased intraluminal IL-8 and consistent accumulation of leukocytes in the subepithelial region. Thus, the *CFTR* mutation itself may cause a proinflammatory phenotype of the airways and hence further infection may aggravate mucosal damage due to constant inflammation. In summary, the overproduction of proinflammatory cytokines on one hand and significantly lower levels of the anti-inflammatory cytokine IL-10 on the other hand results in an excessive and persistent inflammation in the CF airways. Consequently, lung function deteriorates more rapidly in *P. aeruginosa* colonised CF patients compared with *P. aeruginosa* negative CF patients (Gibson et al., 2003).

1.5. Innate immunity

Immunity is mediated by humoral and cellular factors elicited by a complex network of innate and adaptive immunity. Innate immunity refers to antigen-nonspecific defense mechanisms that a host use immediately after exposure to antigen. Although the innate immune system was described more than a century ago by Metchnikoff, discovery of adaptive immune system overshadowed it (Silverstein 2003). However recent research has provided considerable insight into the molecules, particularly toll like receptors (TLRs) implicated in innate immune system and their functions. These molecules, termed pattern recognition receptors (PRRs), primarily function in recognising microbial structures referred to as pathogen-associated molecular patterns (PAMPs) and consequently, provide initial protection against microorganisms by stimulating adaptive immune response (West et al., 2006). TLRs are type I transmembrane proteins of the Interleukin-1 receptor (IL-1R) family that possess an N-terminal leucine-rich-repeat (LRR) domain for ligand binding, a single transmembrane domain, and a C-terminal intracellular signalling domain (Bell et al., 2003). Ten human TLRs have been identified in humans and they recognize an array of bacterial, fungal, and viral products, including structural molecules in the microbial cell wall like LPS (Beutler, 2004). For example, TLR2 recognizes gram-positive lipoteichoic acids, TLR4 recognizes gram-negative LPS, TLR5 recognizes flagellin, and TLR9 recognizes unmethylated bacterial DNA. Combinatorial interaction of these TLRs leads to recognition of PAMPs and their signaling initiates with the recruitment of TIR-domain-containing adaptor proteins to the cytoplasmic TIR domain of an activated TLR (Yamamoto et al., 2004). Recruitment of one or more of these adaptors (MyD88, TIRAP/Mal, TRIF/TICAM-1, and RAM/TICAM-2) to a TLR initiates signalling events that activate the NF- κ B, AP-1, and IRF families of transcription factors (Akira and Takeda, 2004). These transcription factors induce the expression of genes involved in host defense from infection.

The innate immune mechanisms defend the airway from the array of microbial products that enter the lung. The lung has a distinctive relationship with the environment and it has developed line of attack to defend itself from microbial attack through evolution. Primary defenses like cough reflex, mucociliary clearance, and antimicrobial properties of the mucosal surface (Strieter et al., 2002) defend the airways from infection. Bacteria and virus particles are carried to the alveolar surface where they

interact with soluble components (IgG, complement, and collectins) and alveolar macrophages, which are sentinel phagocytes of the innate immune system in the lungs. Alveolar fluids also contain high concentrations of lipid binding protein (LBP) and soluble CD14 (sCD14), which are key molecules in the recognition of LPS by alveolar macrophages and other cells in the alveolar environment (Martin et al., 1992, 1997). When bacteria are opsonised by IgG, complement or collectins such as SP-A and SP-D or mannan binding lectin (MBL) in the airspaces, they are ingested by alveolar macrophages and the TLRs in the phagosomal membrane provide discrimination among different microbial products entering the cell (Underhill et. al., 1999). Consequently, macrophages produce proinflammatory cytokines such as IL (interleukin) -8 and CXC chemokines which in turn recruits neutrophils from the lung capillary networks. Furthermore, alveolar macrophages carry microbial antigens into the interstitium and to regional lymph nodes. Specialised antigen presenting cells such as dendritic cells take up these antigens and present them to responding lymphocytes to initiate adaptive immune responses.

1.6. Cystic Fibrosis Genetic Modulators

The heterogeneity of CF disease severity is partly explained by the different mutations of the *CFTR* locus. However, CF patients with the same *CFTR* genotype, displaying significantly variable clinical phenotype strongly indicate the role of factors other than *CFTR* genotype. This notion was confirmed by earlier studies in which significant variation was seen among F508del homozygous patients with respect to their gastrointestinal, hepatobiliary and pulmonary disease manifestation (Johannsen et al., 1991; Kerem et al., 1990). As modulating factors can be genetic and/or environmental, the role of genetic factors in modulating CF disease was clearly dissected by CF twin and siblings study in which monozygous CF twins were more concordant than dizygous CF twins with respect to their disease severity (Mekus et al., 2000). Over the years, many genes such as Angiotensin converting enzyme, Beta 2 Adrenergic receptor, Voltage-gated chloride channel 2 (CICN2), Mannose binding lectin (MBL), Nitric oxide synthase 1, Glutathione-S-transferase-M1, Transforming growth factor beta 1 and *TNF α* were evaluated as CF modulators by several investigators with varying degrees of success (Cutting, 2005; Knowles, 2006). These findings further strengthened the role of genetic component in CF disease modulation. Consistently, the impact of host genetic

factors in *P. aeruginosa* colonisation among CF patients was dissected by comparing the concordance and discordance for *P. aeruginosa* colonisation phenotype among monozygous and dizygous CF twin and siblings. As monozygous twins are genetically identical and dizygous pairs are genetically 50% identical on average, comparing concordance and discordance of *P. aeruginosa* colonisation phenotype between monozygous and dizygous pairs would imply the contribution of genetic factors for the acquisition of *P. aeruginosa*. Thus, comparison of monozygous pairs concordant/discordant for *P. aeruginosa* with dizygous twin and siblings concordant/discordant for *P. aeruginosa* showed a highly significant association in which all monozygous pairs and 78 percent (18 out of 23 pairs) of the dizygous twins were concordant for *P. aeruginosa* infection status (Unpublished observation, Stanke and Tümmler) indicating a strong genetic component involved in modulating the susceptibility to *P. aeruginosa* infection among CF patients. Furthermore, the excessive inflammation in patients with CF (Ramsey et al., 1996) and inflammation in CF infants even without any infection (Khan et al., 1995) suggests that the genes involved in both innate and adaptive immunity as potential genetic modulators in CF. Hence, it is of paramount importance to identify these genetic modulators and to unravel the mechanism of modulation.

1.7. Approaches to study CF genetic modulators

Association-mapping methods attempt to locate disease mutations by detecting association between the incidence of a genetic polymorphism at a gene of interest (the ‘candidate gene’) and that of a disease. Single nucleotide polymorphisms (SNPs) have gained widespread interest as potential molecular markers in disease association studies and linkage disequilibrium mapping (Pritchard and Przeworski, 2001). Further, the power of linkage mapping can be increased by exploiting the fact that polymorphisms in close physical distance occur together in linkage, implying rare recombination events between them. The combinations of adjacent alleles form “haplotypes”, which can be exploited to map causative variants to a haplotype-block (van den Oord and Neale, 2004). Thus, the association-mapping can be used to identify CF genetic modulators by employing either case-control study (association) and/or family-based (linkage) studies (Newton-Cheh et al., 2005; Laird and Lange, 2006). Although, family-based linkage analyses are mostly done with microsatellite markers, majority of the genetic analysis

are moving towards SNP genotyping and analytical strategies based on association and haplotype analysis (Risch, 2000; Schork et al., 2000). The problem of genetic case-control studies has been population stratification due to mixed ethnic groups (Ewens and Spielman, 1995). For this work, to study CF genetic modulators, the problem of population stratification has been tackled by selecting only Caucasians in the study cohort, because CF is very frequent in Caucasians and 70% of the CF alleles are delF508 alleles, and hence both the ethnic background as well as *CFTR* genotype are normalised. On the other hand, as family-based studies circumvents this issue completely, where the parents act as genetic controls for their affected offspring, parents were recruited for most of the cases in this study. Thus, selecting candidate genes based on their biological plausibility and studying its association with disease by employing both case-control and family-based methodologies in combination would be optimal for any specific genetic disorder (Knowles, 2006).

1.8. Aim of the thesis

Cystic fibrosis (CF) patients carrying the identical *CFTR* mutation genotype display significant variability in their disease course and susceptibility to *P. aeruginosa* infection. This evidence suggested that the disease severity variation and differential response towards *P. aeruginosa* infection among CF patients is conferred not only by *CFTR* but also by other genetic factors. Inflammation is thought to contribute significantly to the destruction of the CF lung and pulmonary disease severity. Thus, the fundamental hypothesis guiding this study was that the genes involved in innate immunity and non-specific defense are modulators of *P. aeruginosa* chronic colonisation and CF disease severity. Consequently, the innate immunity genes, primary molecules involved in crosstalk between pathogen and host due to their location at the interface of host and environment, were selected as potential candidate genes.

Combining a candidate-gene based study and systematic haplotype block mapping on European CF Twin and Sibs with extreme clinical phenotypes, this study aims at defining whether or not naturally occurring polymorphisms on toll like receptor-2, toll like receptor-4, toll like receptor-5, toll like receptor-9, CD14, Surfactant protein-D, IL-8 receptor 2 (*CXCR2*), tumour necrosis factor receptor-1 and *TNFRSF6A* (CD95) can be detected as modulators of CF disease severity and susceptibility to *P. aeruginosa* infection.

Additionally, if association occurs, fine-mapping by haplotype analysis and phenotyping of the causative variant(s) by suitable bio-assays are attempted to describe the molecular mechanism.

2. Patients and methods

2.1. Patient cohorts

CF twin and siblings enrolled in the study were recruited in 1996 by a German-Dutch team (core centers are Rotterdam and Hannover). DNA samples from twin and sibs were obtained in 1996-1999. Serum and microbiological samples were collected in 1997-1999. Details of patients and mode of selection are described in chapters 2.1.1 and 2.1.1.1.

2.1.1. European CF Twin and Siblings

The “*European Cystic Fibrosis Twin and Sibling Study*” consists of 442 CF twin and siblings, representing different *CFTR* mutation genotypes. They were recruited from 158 CF centers from 14 different European countries during 1995 and 1996 (Mekus et al., 2000) to assess the role of genetic and environmental factors on CF disease severity. The patient panel studied in this project is a subgroup of the European CF Twin and sibling study (Table 1). 277 sib pairs, 12 pairs of dizygous twins (DZ) and 29 pairs of monozygous twins (MZ) were included in the original study, 114 of whom were reported to be homozygous for F508del. The 98 patients belonged to 43 pairs and 4 trios. The trios were grouped into three separate sib pairs for the sake of data comparison, resulting in a total of 55 pairs. 13 of these were MZ, 5 were DZ and 37 were siblings.

Table 1: Number of twin and sib pairs recruited in the European cystic Fibrosis Twin and Sibling Study cohort to investigate the CF genetic modulators.

	Sib pairs	Dizygous twin pairs	Monozygous pairs
Number of pairs recruited	277	12	29
Number of F508del homozygous pairs	37	5	13

2.1.1.1. Selection of extreme phenotypes

Two most sensitive clinical parameters to CF course and prognosis (Corey and Farewell, 1996; Corey et al, 1997; Lai et al, 1999), forced expiratory volume (FEV_{Perc}) and weight for height (wfh%), were considered to assess the disease severity among twin and siblings (Mekus et al., 2000). First, lung function was evaluated from predicted

values of forced expiratory volume in one second (FEV_1) expressed as a predicted value ($FEV_1\%pred$) based on the data by Knudson et al, 1993. While in CF, $FEV_1\%pred$ declines with age (Corey M et al. 1996), age and gender-dependent percentiles (FEVPerc) for the CF population were employed to correct for this disease specific effect by taking data from the European CF registry report of 1995. Second, nutritional status was assessed by weight, expressed as predicted weight for height percentage (wfh%), based on percentiles for weight and height corrected for age and gender (Prader et al., 1989). All the patients were assigned a rank number based on their FEVPerc and wfh%. Disease

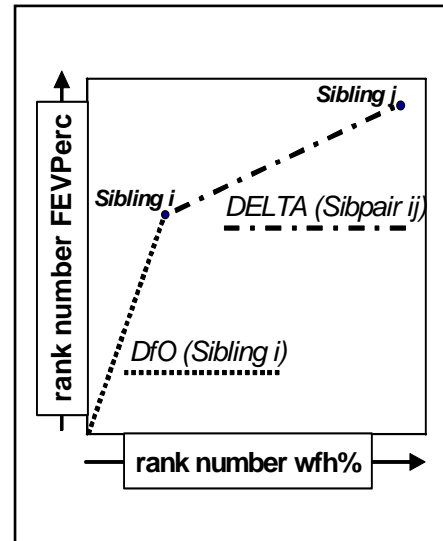


Figure 1: Definition of DfO (disease severity) and DELTA (discordance), which are both determined by the rank numbers for FEVPerc and wfh% for sibling i and sibling j of a pair.

severity was chosen by plotting FEVPerc ranks on Y-axis and wfh% ranks on X-axis for every patient (Fig.1). The overall disease severity of a patient i is determined by the distance from origin (DfO) in the plot resulting from the rank numbers x_i and y_i . Intrapair discordance was defined by the distance (DELTA) between two data points representing two siblings i and j. By this approach, both sibs in a pair displaying highest rank numbers were grouped as concordant mild (CON+), and if both sibs were displaying lowest rank numbers, they were grouped as concordant severe (CON-). If sibs within a pair were discordant for rank numbers, then the pair was grouped as discordant pair (DIS).

2.1.2. F508del homozygous CF twin and siblings stratified for *P. aeruginosa* colonisation

A set of Caucasian F508del homozygous patient pairs selected from the sib pair population, were recruited from the European Cystic Fibrosis Twin and Sibling Study (Mekus et al., 2000), whereby extreme and intermediate phenotypes were enrolled to represent the CF patient pair community. The manifestation of the CF basic defect of this cohort was previously described (Bronsveld et al., 2000). Monozygous and dizygous pairs were comparable with respect to clinical characteristics such as lung function, nutritional status, age at chronic colonisation and gender distribution although

the median of age at the day of investigation was lower for monozygous pairs ($P < 0.1$). 41 dizygous and 12 monozygous pairs were included in this study to assess the role of genetic factors on *P. aeruginosa* acquisition (Table 2). *P. aeruginosa* status of these patients, reported by physician in 1995-1996, was confirmed again during 1997-1998 by sputum bacteriology and serology. Data for age at onset of chronic *P. aeruginosa* acquisition, reported by the patient or/and physician, was available for these pairs.

Table 2: Phenotypic characteristics of F508del homozygous dizygous and monozygous pairs recruited for determination of *P. aeruginosa* colonisation status

	Dizygous pairs	Monozygous pairs
Number of pairs	35	12
Number of trios	2 (form 6 pairs)	-
Total number of pairs	41	12
Gender		
Female (n)	41	16
Male (n)	35	8
Age in years, Median (inner quartiles; range)	13 (10.6-19.4; 5-39)	11 (9.9-18.8; 9-32)
Age of onset of chronic colonisation, Median (inner quartiles; range)	8 (4-14; 1-35)	10.7 (10-12; 1-22)

2.1.3. F508del homozygous unrelated CF patients with early and late *P. aeruginosa* chronic colonisation

To evaluate the role of genetic modulators on the onset of *P. aeruginosa* chronic colonisation among CF patients, a CF cohort, containing 21 CF patients, stratified for early or late colonisation of *P. aeruginosa*, was recruited. The 21 unrelated CF patients were enrolled from the CF clinic Hannover as described elsewhere (Derichs et al. 2003). Briefly, all patients were homozygous for the F508del mutation, matched for gender and had longitudinal data on *P. aeruginosa* colonisation. Hence, these CF patients are grouped according to their age at first *P. aeruginosa* colonisation. The first group of 13 patients (PA-early) had an early chronic *P. aeruginosa* colonisation before the age of 7 years whereas the second group of 8 patients (PA-late) showed *P. aeruginosa* colonisation at the age of 14 years or later. The infection status was verified by routine sputum bacteriology in the microbiology department of the MHH.

2.1.4. F508del homozygous unrelated CF patients stratified for birth cohorts

This cohort consists of unrelated F508del homozygous CF patients from Hannover CF clinic who were born in 1959 to 1975. These patients were recruited initially for CFTR mutation genotype analysis during 1989-1992 and the age at the day of investigation (blood taking) ranges from 34 to 14 years. We categorized these patients into two groups in which 22 patients who were born in 1959 to 1967 as Early-born group and 23 patients who were born in 1973 to 1975 as Late-born group. In order to test the hypothesis that the changing environmental factors (treatment) and improvement of survival have an impact on selection of favourable alleles and genotypes among CF population, Early-born group and Late-born groups were typed for CD14 polymorphisms and the distribution of genotypes were compared with genotype distribution among all other cohorts investigated.

2.1.5. Unrelated F508del homozygous CF patients recruited for global transcriptome analysis

Fifteen unrelated patients, homozygous for F508del CFTR mutation, were selected from a cohort containing CF patients, who were recruited initially to study the impact of residual CFTR on global transcriptome. For transcriptome analysis, the total RNA was isolated from rectal biopsies using Qiagen RNeasy protocol (Qiagen Corp, Hilden, Germany) and the quality was determined by gel electrophoresis. cDNA was synthesized from total RNA and hybridized on the affymetrix chips (GeneChip® Human Genome U133 Plus 2.0 Array, Affymetrix, Santa Clara, CA) according to the protocol from Affymetrix manual (Version 700217 rev 3). The expression data was normalized for chips and evaluated using Affymetrix Microarray Suite v5.1 software. (The transcriptome analysis was done in collaboration with Larissa Pusch and Torsten Kroll, University of Jena). Thus, these patients had global transcriptome data. Hence, these 15 patients were typed for candidate genes and mRNA levels were compared against genotypes at different polymorphisms.

2.2. Methods

2.2.1. Isolation of high molecular weight DNA from blood

The DNA isolation method from human blood described by Gross-Bellard et al. (1973) was followed. 5-10 ml K⁺-EDTA blood (fresh or thawed on ice after storage at -20°C) was incubated with 40 ml lysis-buffer for 30 minutes on ice. Later on, the nuclei of the lymphocytes were separated from the cell debris by centrifugation for 15 minutes at 600 x g. The lysed pellet was then incubated with 5 ml proteinase K (0.25 mg/ml), SDS (0.5% w/v) in STE for 8 to 14 hours at 56°C in a shaking water bath. Subsequently, digested proteins and DNA were separated by phenol/chloroform extraction. 3 ml chloroform/ isoamylalcohol (29:1) and 3 ml phenol were added to the digest, mixed carefully for 15 minutes and centrifuged for 10 minutes at 600 x g. The organic phase, found at the bottom of the tube, was carefully removed, and the extraction step was repeated. Finally, an extraction step with 6 ml chloroform was performed to remove remaining phenol from the aqueous phase. Precipitation of DNA was done on ice by adding 1/10 of the volume of the aqueous phase of 3M Na₂CO₃ aq (pH 5.5) and 30 - 40 ml of 70% EtOH (cooled down to -20°C). DNA was then transferred to an Eppendorf tube containing 1 ml 70% EtOH and centrifuged for 3 minutes at 12,000 x g. The supernatant was discarded and the pellet washed twice with 70% EtOH. DNA was dissolved in 300 µl of TE at 4°C for one week. Finally, a 1:25 dilution of the stock solution was prepared and DNA concentration was determined by measuring the OD at 260 nm.

2.2.2. DNA isolation from neutrophils

Neutrophils were collected as a pellet during peripheral blood mononuclear cells isolation. The pellet was washed in 10ml of PBS and centrifuged at 8000 rpm to remove red blood cells (washing is repeated depending on the amount of RBCs). The washed pellet was subjected to DNA isolation using Peqlab DNA isolation kit (PEQLAB Biotechnologie GmbH).

2.2.3. Genotyping

Single nucleotide polymorphisms (SNPs) and microsatellites were selected for genotyping the candidate genes. The suitable markers were selected as it is described in chapter 2.2.3.1. SNPs were analysed by PCR-RFLP and microsatellites were analysed

by direct blotting electrophoresis as described by Mekus et al., 1995. Analysis of SNPs by RFLP and analysis of microsatellites by direct blotting are described in chapters 2.2.3.2 and 2.2.3.3, respectively.

2.2.3.1. Selection of genetic markers

All SNPs were selected from the NCBI database based on their location within the region of interest and also based on their heterozygosity value as an informative measure. SNPs were typed on 9 different controls (DNA samples from healthy volunteers), to estimate the allele frequencies, by PCR-RFLP method. Preferably, SNPs displaying a 50% frequency of each allele were selected for further genotyping of twin and sibling samples (A list of all the SNPs chosen from NCBI data base is given in appendix). Microsatellite markers were selected based on their proximity to the candidate genes by screening the genomic sequences for repeat motifs. Polymorphic information content (PIC) of a microsatellite marker was estimated based on the allele frequencies among 9 control samples and a marker displaying a PIC of more than 0.5 was typed on CF twin and siblings.

2.2.3.2. Analysis of restriction fragment length polymorphisms (RFLPs)

RFLP typing was performed by using commercially available restriction-endonucleases. Briefly, 15µl of PCR amplified product was incubated with 10µl of restriction mix containing 3µl of restriction buffer and 2 to 10 U of suitable restriction enzyme (NE Biolabs) at its optimum temperature for overnight restriction digestion. Digested samples were then analysed on a 2 to 4% agarose gel.

2.2.3.3. Analysing polymorphic microsatellites by direct blotting process

The direct blotting process uses a denaturing polyacrylamide (PAA) gel and high voltage electrophoresis for high resolution separation of small DNA fragments. The apparatus (GATC, Konstanz) consists of a vertical electrophoresis gel with a conveyor belt directly beneath the lower edge of the acrylamide gel. Reaching the end of the vertical gel, the products are directly transferred on a positively charged nylon membrane that is transported down the gel by a conveyor belt running through the lower buffer reservoir. Resolution of separation depends on the matrix of the gel and conveyor belt speed.

2.2.3.3.1. Gel preparation

For microsatellite analysis, a gel of 4% PAA and 8.3 M urea in 1 x TBE buffer with a thickness of 0.19 cm was used. After polymerisation, the gel was mounted into the blotting chamber and electrophoresis was performed at 1500V for at least 30 min in 1x TBE for equilibration of the gel.

2.2.3.3.2. Sample preparation

For analysis of microsatellites, the selected markers were amplified using an asymmetric primer ratio of 5 pmol biotin labelled primer and 25 pmol of unlabelled primer in multiwell plates. PCR was carried out in a final volume of 15 µl [1 µM MgCl₂, 2 µM dNTP's and 0.25 U Invitak polymerase (Invitek)] with an oil overlay. Once PCR was done, 8 µl of PCR products were transferred to another multiwell plate and dried at 37°C for overnight. Dried products were dissolved in 10 µl of formamide containing 0.2% w/v bromophenol blue and xylene cyanol. Before loading, samples were denatured for 5 minutes at 95°C in a heating block and immediately cooled at -20°C. 0.5 µl of each sample was then loaded onto the polyacrylamide gel and pre-run into the gel by electrophoresis at 500 V for 3 minutes.

2.2.3.3.3. Electrophoresis

Electrophoresis was performed at 1900 V. Transport of the membrane was started shortly after the first front of bromophenol blue had reached the membrane, the delay between the first contact of the bromophenol blue marker with the membrane and the start of transport depended on the size of the tested product. The speed of the conveyor belt was kept constant at 17cm/h. 30 cm of Hybond N⁺ nylon membrane were used for single markers. After electrophoresis, the blue fronts of the various runs were marked at the side of the membrane and membranes were stored at room temperature until developing.

2.2.3.3.4. Membrane-Developing and detection

Biotinylated products were detected by chemiluminescence. Membrane was incubated for 1 hour with 75ml of 1.5% of blocking reagent (Boehringer), which was dissolved in 150 ml of buffer 1 (100 mM Tris pH 7.5, 150 mM NaCl). Next, membranes were rinsed

for another 1 - 2 hours in fresh 75ml of the blocking solution containing 6µl of anti-biotin-alkaline-phosphatase fab-fragments (0.75 U/ µl). Unbound antibody and blocking agent were removed by washing the membranes 3 times for 8 minutes each in buffer 1 mixed with 1% Triton-X 100. The membrane was equilibrated for 20 minutes in developing buffer 3 (100 mM Tris pH 9.5, 100 mM NaCl, 50 mM MgCl₂). The chemiluminescence reaction was induced by incubating the membranes in 50 ml of buffer 3 containing 10% v/v Sapphire II (Tropix) and 300 µl of CDPstar (Tropix) for 5 min. Later, the solution was retained and the membrane was rinsed with 50 ml of buffer 3 containing 1% v/v Sapphire II and 30 µl CDPstar. Membranes were sealed in plastic foil and chemiluminescence was detected with help of Kodak X-o-mat X ray films. Exposure time varied between 5 sec and 1 hour depending on the efficiency of the PCR.

2.2.4. Genotyping data evaluation

The genotyping data obtained for all candidate genes, from CF twin and sibling cohort and from cohorts stratified for *P. aeruginosa* phenotype, were subjected to statistical analysis. Association between CF disease severity and susceptibility to *P. aeruginosa* with polymorphisms on all the candidate genes was evaluated with appropriate statistical tests as described in chapter 2.2.4.1.

2.2.4.1. Evaluation of genotyping data from CF twin and siblings and from cohorts stratified for *P. aeruginosa* colonisation

Three different hypotheses were tested in CF twin and sibling cohort. Firstly, preferential transmission of certain alleles (or haplotypes) in CF offsprings from their parents, was tested by comparing the transmitted alleles (or haplotypes) vs. non-transmitted ones by considering all nuclear families irrespective of their disease severity categories. Secondly, allele and genotype distributions were compared in a case-control fashion. Briefly, the comparison was done between mildly (CON+) vs severely (CON-) affected pairs, employing the phenotypic contrast to identify CF modulators. Thirdly, both CON+ and CON- were grouped as CONC (Concordant) and compared against DIS to analyse the role of modulators in “trans” (encoded elsewhere in the genome). The polymorphisms on candidate genes characterised in CF twin and siblings are analysed in other cohorts stratified for *P. aeruginosa* colonisation. The allele and genotype distributions were compared between both *P. aeruginosa* early colonised and *P.*

aeruginosa late colonised groups as well as between *P. aeruginosa* positive and *P. aeruginosa* negative groups to determine the role of polymorphisms on candidate genes in susceptibility to *P. aeruginosa* colonisation.

2.2.5. Statistical Analysis

2.2.5.1. Family based evaluation

Preferential transmission of alleles or haplotypes, at tested loci among CF offspring from their parents, was tested by analysing nuclear families with the Monte Carlo simulation based association test (Knapp and Becker, 2003) that can be viewed as an extension of the transmission-disequilibrium test (Spielman et al., 1993).

2.2.5.2. Case-control analysis

Case-control based evaluation was done by comparing allele/genotype/haplotype frequencies between two groups (For example: CON+ vs. CON- or DIS vs. CONC or *P. aeruginosa* colonised vs. *P. aeruginosa* non-colonised) using CLUMP software package. CLUMP is a program that was designed for use in genetic case-control association studies, and especially for distributions with a large number of categories (Sham and Curtis, 1995). Thus, CLUMP calculates Chi-square statistic for k x 2 contingency tables and significance of association is determined by Monte Carlo simulations.

2.2.5.3. Correction for sib-pair dependence and multiple testing

Dependence of genotypes of the individuals within each sib-pair was corrected by accounting the affection status in each permutation replicate and was simultaneously permuted or not permuted with equal probability for both sibs (Sib-pair correction was kindly carried out by T. Becker, IMBIE). Further, the haplotype counts were obtained by using likelihood-weighted haplotype explanations for each individual as described by Becker et al. (2005). As several markers on a candidate gene are tested individually and also in combination for the same null hypothesis in this study, all P-values obtained for CON- versus CON+ and CONC versus DIS comparisons are corrected for multiple testing by applying Monte Carlo simulation based strategy.

Corrected P-value was obtained for single marker analysis and also for multi-marker-combination analysis as described by Becker and Knapp (2004b).

2.2.5.4. Hardy-Weinberg equilibrium analysis using FINETTI program

Deviations from Hardy-Weinberg equilibrium (HWE) for polymorphisms in CD14 were analysed using the program FINETTI (Kindly provided by Prof. Thomas F. Wienker; unpublished data). The De Finetti diagram includes a parabola indicating genotype distributions consistent with HWE. The distance between the parabolic curve and the genotypes indicates the extent of deviation from HWE and the significance is calculated by Chi-square test.

2.2.6. PCR based methods

2.2.6.1. PCR in multiwell plates

The inner 60 wells of a 96 well plate (Greiner) were coated with 5 μ l (50ng) of DNA. To each well, 25 μ l reaction mixture, containing 3 μ l of 5 μ M primer, 3 μ l of 2mM dNTPs, 3 μ l of 10X buffer (InviTek), 1 to 3 μ l of 50mM MgCl₂, 0.05 μ l of *Taq*-polymerase (5U/ μ l) and 12 to 10 μ l of H₂O, was added. The outer rim of the plate was sealed with agarose to avoid evaporation. PCR was carried out at optimum annealing temperature in a Hybaid thermocycler with a heated lid.

2.2.6.2. Long-range PCR

For PCR amplification of sequencing targets, the Failsafe™ PCR System (EPICENTRE Technologies, WI USA) was employed according to the manufacturer's instructions. Briefly, for 50 μ l amplification reaction, 25 μ l of Failsafe PreMix, 10 μ l of 5 μ M primers and 4 μ l of H₂O was added to a 100 μ l reaction tube containing 1 μ l (100ng) of DNA template. Later, 0.2 μ l of Failsafe PCR Enzyme Mix was added and PCR was carried out at suitable annealing temperature. The above combination was tried with all 12 different Failsafe Premixes on control samples and the optimal combination was chosen for further applications.

2.2.6.3. Sequencing long-range PCR products

As good sequencing results can be obtained for a maximum fragment size of 500-700bp, chromosome walking strategy was followed for sequencing >1kb fragments.

Briefly, the sequencing region (1kb to 3kb) was targeted by terminal primers and amplified using Failsafe™ PCR System (EPICENTRE Technologies, WI USA). Later on semi-overlapping internal primers were designed using Primer3 for every 500 to 600bp of the target for sequencing. Internal primers were designed in such a way that the second product should include at least 30bp of the first product. By this approach, sequencing results of the primer binding regions were confirmed. Sequencing was performed at QIAGEN (Hilden). Sequencing results were analysed by aligning both 5' primer read as well as 3' primer read of a fragment against reference sequence using the software "CodonCode Aligner" (CodonCode Corporation). Sequence variations found on both 5' and 3' reads were accounted for further genotyping. Variations found only on one read were confirmed by re-sequencing the fragment of interest with a new set of primers.

2.2.6.4. Pre-mRNA length determination by PCR

In order to find CD14 pre-mRNA length, total RNA was isolated from 293T and T84 cell lines (chapter 2.2.6). cDNA was synthesised using random primers (chapter 2.2.7.2). Primers were designed to amplify products of different lengths. For this, single forward primer (5'-GGGCTTTGCCTAAGATCCAA) was used in combination with four different reverse primers (Primer 1: 5'-CCATTATGTCGGGGAGTGAC, primer 2: 5'-GTGAGGCAGGCATCTAGCTC, primer 3: 5'-TATAGGCATGAGCCACCACA, primer 4: 5'-GTGGTGGTGCATGCCTATAA), located on different regions to amplify 697bp, 986bp, 1604bp and 2075bp products respectively. Products were amplified using Failsafe™ PCR System (EPICENTRE Technologies, WI USA). Two genomic DNA samples, isolated from blood, were used as controls.

2.2.7. RNA isolation

Total RNA from adherent cells and PBMCs were isolated with "Absolutely RNA miniprep kit" (Stratagene, La Jolla, CA). The concentration of RNA was determined by measuring the OD at 260nm and the quality was assessed by taking ratio of ODs at 260 and 280nm.

2.2.8. Real Time PCR

Total RNA isolated with "Absolutely RNA miniprep kit" (Stratagene, La Jolla, CA) was used for cDNA synthesis. cDNA was synthesised by using oligo dT primer and

expand reverse transcriptase (Roche). GAPDH specific primers were used as internal controls (forward primer 5'-GTCAGTGGTGGACCTGACC; reverse primer 5'-TGAGCTTGACAAAGTGGTCG). CD95 specific primers to give product size of 226bp were used to analyse the CD95 mRNA level (forward primer 5'-TTTCACTTCGGAGGATTGCT; reverse primer 5'-ACCTGGAGGACAGGGCTTAT). The PCR reaction was performed in duplicates using a LightCycler–FastStart DNA Master SYBR Green I kit (Roche) according to the manufacturer's instructions. Real-Time PCR was done by using Roche Applied Science's "The LightCycler System" (ROCHE). PCR reactions were monitored using SYBR Green I Dye which is a double stranded (ds) DNA binding dye. It is thought to bind in the minor groove of dsDNA and upon binding increases in fluorescence over 100 fold. The fluorescence intensity depends on the amount of dsDNA which in turn depends on the initial amount of cDNA template put for the PCR reaction.

2.2.8.1. First-strand cDNA synthesis using oligo(dT) primers

First strand cDNA was synthesised from 10µg total RNA by extension with oligo(dT)₁₅ and expand reverse transcriptase enzyme (Roche). Briefly, 10µl of RNA solution was mixed with 1µl of 4µM oligo(dT)₁₅. This mixture was incubated at 65°C for 10 min and immediately cooled on ice. Subsequently 9µl reaction mixture (4µl of 5X RT-Buffer (Roche), 1µl DTT (100mM) (Roche), 1µl of 10mM dNTPs, 2.5µl of H₂O and 0.5µl of 50U/µl expand reverse transcriptase enzyme) was added and reverse transcribed at 42°C for 1 hour followed by an inactivation step of 95 °C /5 minutes, and brief centrifugation. Samples were stored at -80°C.

2.2.8.2. First-strand cDNA synthesis using Random primers

For 20µl of reaction volume, 10µl of RNA solution was mixed with Random primers at a concentration of 200ng primer/µg RNA. This mixture was incubated at room temperature for 10min. Subsequently 9µl reaction mixture (4µl of 5X RT-Buffer (Roche), 1µl DTT (100mM) (Roche), 1µl of 10mM dNTPs, 2.5µl of H₂O and 0.5µl of 50U/µl expand reverse transcriptase enzyme) was added and reverse transcribed at 42°C for 1 hour followed by brief centrifugation. Samples were stored at -80°C.

2.2.8.3. Quantification of mRNA on the LightCycler

A PCR reaction profile in a LightCycler can be divided into three segments: an early background phase, an exponential phase (or log phase) and a plateau. The background phase lasts until the signal from the PCR product is greater than the background signal of the system. The exponential phase begins when sufficient product has accumulated to be detected above background, and ends when the reaction efficiency falls as the reaction enters the plateau. Typical real-time PCR curves monitored in on the LightCycler are shown in Fig. 2a. The cycle at which each reaction first rises above background is dependent on the amount of target present at the beginning of the reaction.

Checking efficiency of amplification

For relative quantification of an unknown target, consistent standard (predefined template concentration) was used with a minimum of three dilutions (1 point per log of concentration) and standard curve was plotted to calculate the efficiency of amplification. This was done by plotting Cp (crossing point defined as cycle at which the signal rises above the background) vs. the log of the initial template concentration as shown in Fig. 2a. The linear plot indicates that the efficiency was constant over the concentration range analysed. Based on this efficiency the tool calculates the concentration of target template.

Relative quantification

All the reactions were amplified in duplicates and the final results were reported as the ratio of number of mRNA copies of a target gene relative to GAPDH mRNA (house keeping gene) from the same cDNA sample. This was calculated as shown in the Fig. 2b. Thus, four normalised ratios were obtained for a sample and standard deviation is plotted for all the four values. As a standardising experiment, real-time PCR was done on 293T cell line and T-84 cell line cDNA to analyse the CD95 mRNA level. Two different targets (FAS-1 and FAS-2) for the same gene i.e. CD95, were amplified using two set of primers FAS-1 (forward primer 5'- TTTCACTTCGGAGGATTGCT; reverse primer 5'- ACCTGGAGGACAGGGCTTAT gives 226bp product) and FAS-2 (forward primer 5'- GGAAAGCTAGGGACTGCACA; reverse primer 5'- TGTTACATTTGGTGCAAGG gives 250bp product). Results are shown in Figure 2d.

As this method showed that the technique is specific and consistent for CD95 mRNA analysis, only FAS-1 target was analysed on all other samples.

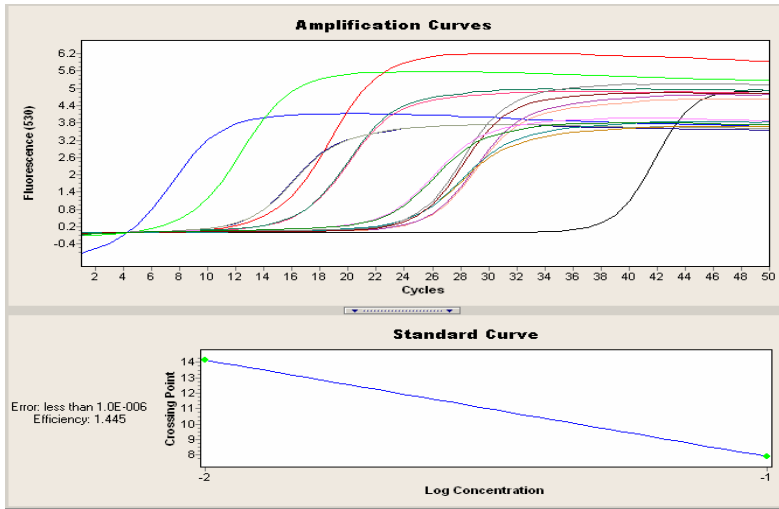


Figure 2a: Efficiency of the amplification was determined by plotting standard curve (cp vs. log concentration)

Samples			Results			Calls		
Include	Color	Pos Name	CP	Concentration	Standard	Combined	Target	Control
1	Blue	Standard 1			1.00E0	Success	Positive	Success
2	Green	Standard 2	7.89	1.00E-1	1.00E-1	Success	Positive	Success
3	Red	Standard 3	14.14	1.00E-2	1.00E-2	Success	Positive	Success
4	Black	No Template	37.62	1.76E-6		Success	Positive	Success
5	Pink	293T GAPDH	15.61	[5.83E-3]		Positive	Positive	Success
6	Light Green	Repl. of 293T GAPDH	15.63	[5.78E-3]		Positive	Positive	Success
7	Dark Green	T84 GAPDH	11.72	2.44E-2		Positive	Positive	Success
8	Light Blue	Repl. of T84 GAPDH	11.80	2.37E-2		Positive	Positive	Success
9	Orange	293T FAS - I	24.67	[2.07E-4]		Positive	Positive	Success
10	Purple	Repl. of 293T FAS - I	24.62	[2.11E-4]		Positive	Positive	Success
11	Yellow	T84 FAS - I	23.49	[3.20E-4]		Positive	Positive	Success
12	Teal	Repl. of T84 FAS - I	23.95	[3.12E-4]		Positive	Positive	Success
13	Brown	293T FAS - 2	23.77	[2.89E-4]		Positive	Positive	Success
14	Pink	Repl. of 293T FAS - 2	23.73	[2.93E-4]		Positive	Positive	Success
15	Grey	T84 FAS - 2	21.66	[6.27E-4]		Positive	Positive	Success
16	Light Green	Repl. of T84 FAS - 2	21.70	[6.19E-4]		Positive	Positive	Success

Figure 2b: Relative quantification by taking normalised ratio of house keeping gene to target

$$\text{Normalised Ratio} = \left. \begin{matrix} C/A \\ D/A \\ C/B \\ D/B \end{matrix} \right\} \text{Standard deviation}$$

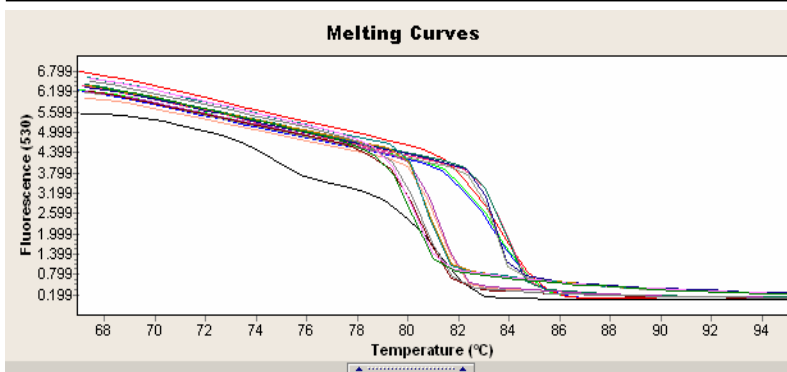
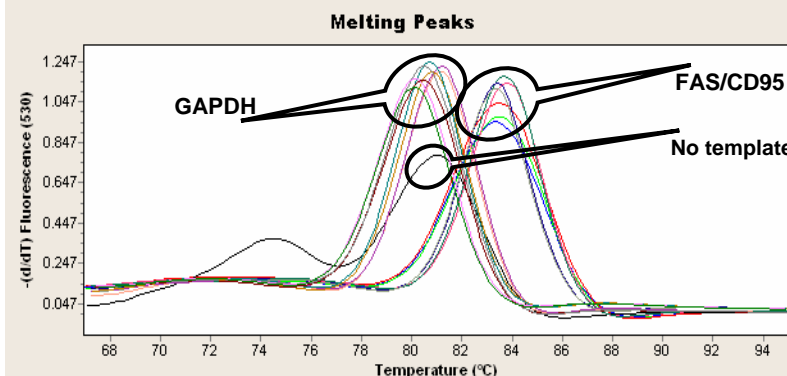
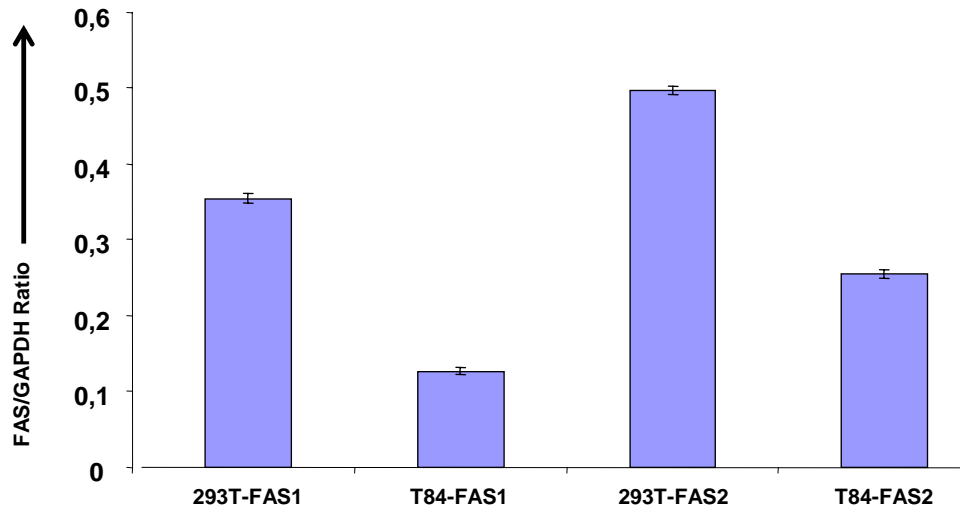


Figure 2c: Specificity of the amplified products is checked by analysing melting curves and melting peaks



Separate and specific melting peaks for both GAPDH and CD95 indicate specificity of the PCR amplification.

Figure 2d: CD95 mRNA quantification in 293T and T84 cell lines using real-time PCR



The mRNA copies are quantified by taking the ratio of target gene to house keeping gene (FAS/GAPDH) mRNA copies and comparing the ratios between two samples.

2.2.9. MicroRNA detection

MicroRNAs are small (≈ 21 to 24nt) non-coding RNAs that serve as posttranscriptional regulators of gene expression in higher eukaryotes. Their widespread and important role in animals is highlighted by recent estimates that 20%–30% of all genes are microRNA targets. In this study, CD14 3' UTR polymorphism was found to be located on the microRNA binding site, which raised the possibility of altering the binding efficiency of a microRNA on CD14 and in turn differential gene regulation. Hence, in an effort to detect the targeted microRNA, dot blot and hybridization techniques were standardised.

2.2.9.1 MicroRNA isolation

For isolating micro RNAs from all the cell types, *mirVana*TM miRNA Isolation kit from Ambion® was employed. Essentially manufacturer's instructions were followed to isolate small RNAs from total RNA fraction and quantity and quality was determined by taking OD measurements at 260 and 280nm.

2.2.9.2. Biotin labelling of small RNA fraction (<200 nt)

2 µg of small RNA isolated using *mirVana*TM miRNA Isolation kit from Ambion® was biotin-labelled according to manufacturer's instructions using Biotin-Chem-Link kit (Roche Diagnostics GmbH).

2.2.9.3. Dot Blotting

One µl of each RNA samples were coated manually on Hybond N⁺ nylon membrane (8x7cm in size) with serial dilutions (400ng/µl to 4ng/µl). Coated membranes were dried at room temperature for 10min. Membranes were incubated at 70°C for 10 min followed by UV cross linking for 3 minutes. These membranes were stored at 4°C until use.

2.2.9.4. Prehybridization

Blotted membranes were washed twice in 3xSSC/0.1% SDS at room temperature and prehybridized for 1 hour at 37°C in 10ml of prehybridization solution (6xSSC, 5x Denhardt's solution, 0.05% sodium pyrophosphate, 100µg/ml boiled herring sperm DNA, 0.5% SDS). (Volume of both prehybridization and hybridization solution depends on the size of the membrane).

2.2.9.5. Hybridization

Membranes were removed from the prehybridization solution and immersed in 5ml of hybridization solution (6xSSC, 1x Denhardt's solution, 100µg/ml yeast tRNA, 0.05% sodium pyrophosphate) within a falcon tube. Later, 10ng /ml of biotin-labelled probe (Biotin-labelled microRNA) was added to the hybridization solution and hybridized at 48°C for 14 to 18 hours.

2.2.9.6. Detection

Hybridized membranes were washed for 10 min, 3 times, in 6xSSC/0.05% pyrophosphate at room temperature on a shaker. Further, membranes were washed for 45 min at 60°C in 6xSSC/0.05% pyrophosphate which was prewarmed to 60°C. The temperature of the washing solution was measured using thermometer during the washing procedure and once the temperature of the membrane reached 60°C, washing

was continued for another 20 minutes. After this step, membranes were subjected to the detection protocol as described in chapter 2.2.2.3.4.

2.2.10. Western Blotting

The quantity of both 55kDa and 28kDa fractions of TNFR1 in serum samples was measured by western blotting as described by Hawari et al. (2004) with some modifications in the protocol to deplete excess immunoglobulin.

2.2.10.1. Sample preparation and separating on a gel

Western blotting was done to quantify target protein in human serum samples. Initially, the serum samples were pre-cleared with Protein A (10 μ l /100 μ l serum) and Protein G (10 μ l /100 μ l serum) beads by incubating at 4°C on a shaker for 1hr. Then, samples were centrifuged at 6500 rpm/1 min at 4°C. The supernatant was taken and total protein was measured using Bradford's protein quantification method. 60 μ g of total protein was mixed in sample loading buffer and heated at 90°C-95°C for 10 min and protein were separated on a 10%-12% SDS-polyacrylamide gel at 80 volts until the bands reached the separating gel and later on at 140-160 volts.

2.2.10.2. Protein blotting

PVDF membrane and blotting papers were cut according to the gel size. The membrane was soaked in methanol for 30sec and immediately transferred to transfer buffer and incubated for more than 5min. Blotting papers were also incubated in transfer buffer along with the membrane. The gel was removed from the gel-running apparatus and transferred to transfer buffer and incubated for 5min. Then, the membrane was placed on the wet blotting papers. The gel was placed on the membrane and blotting papers were kept on top of the gel. Blotting was done at 320Amp for 40 to 60min depending on the size of the protein.

2.2.10.3. Immune detection

The membrane was blocked using blocking buffer (I-Block, TROPIX) overnight at 4°C or for 2hr at RT on an orbital shaker. Later the membrane was rinsed with wash buffer twice for 2 min and incubated with primary antibody (1:100 to 1:500 in blocking buffer) in a sealed plastic bag for 1 to 2hrs at RT. After that the membrane was rinsed briefly

with two changes of wash buffer and washed for 15min with wash buffer at room temperature. Again with fresh changes of wash buffer, the membrane was washed 3 times for 5min at RT. HRP-labelled secondary antibody was diluted 1: 10,000 in blocking buffer (dilution depends on the secondary antibody and the sample used). The membrane was incubated in diluted secondary antibody solution for 2hrs at RT (light sensitive HRP-labelled antibody was protected from light by covering the membrane with aluminium foil). The membrane was rinsed briefly with wash buffer and washed for 15min at RT. Another time, membrane was washed for 15min with fresh changes of wash buffer at every 5min. For detection of the signals on the membrane, ECL chemiluminescence detection kit from Amersham was used. Briefly, both detection solutions, A and B, are mixed (in 1:1 ratio). The membrane was placed on a clean surface with blot-protein upside and mixed detection solution was poured on the membrane and incubated at RT for 5min. Later, the excess detection solution was drained using blotting paper. Membranes were sealed in plastic foil and chemiluminescence was detected with help of Kodak X-o-mat X ray films. Exposure time varied between 5 sec and 1min depending on the amount of protein.

2.2.11. ELISA

Soluble CD14 (sCD14) concentrations in serum was determined by Human sCD14 Quantikine ELISA Kit (R & D Systems, DC140) according to the manufacturer's instructions.

2.2.12. PBMC isolation

Venous blood (15-20ml) was collected into K⁺-EDTA or heparinized tubes and immediately processed. Peripheral blood mononuclear cells (PBMC) were isolated using gradient centrifugation in Ficoll-Paque solution (Pharmacia). The blood was mixed with equal volume of PBS (137mM NaCl, 2.7mM KCl, 8.1mM Na₂HPO₄, and 1.5mM KH₂PO₄, pH 7.2). 15 ml of Ficoll-Paque Plus was filled into a 50 ml Falcon tube and blood/PBS solution was layered over it carefully. Tubes were centrifuged at 1600rpm for 30min without brake.

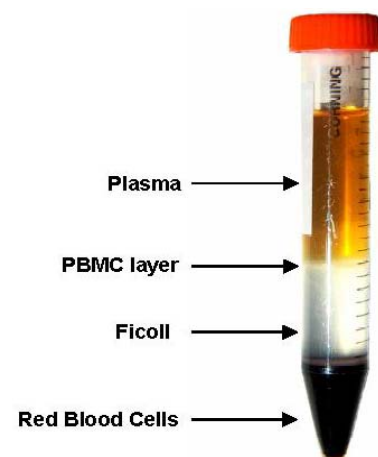


Figure 3: Ficoll-gradient separation of PBMCs

After centrifugation, four fractions were separated in four different layers (Fig. 3). The upper phase contains the plasma/PBS fraction. Below this phase a small buffy coat contains the PBMC fraction. Third phase was ficoll and then follows the neutrophil/erythrocyte fraction. With a 10ml pipette, the PBMC fraction was sucked out carefully and transferred to a new 50ml falcon tube. Later, this fraction was washed twice with PBS and centrifuged at 4°C, 1300rpm for 8min. Then, the pellet was resuspended in 5ml of RPMI (containing 10% FCS, 100U/ml penicillin, 0.1mg/ml streptomycin, 0.3mg/ml glutamine, and 10 μ M 2-Mercaptoethanol) and cells were counted. Cells were centrifuged again and resuspended in RPMI + 10%FCS + 10%DMSO at 5 X 10⁶ cells/ml. Then the tubes containing PBMCs were stored at -80°C.

2.2.13. Flow cytometry

PBMCs were analysed by flow cytometry using a FACS-Canto (BD Biosciences) and FlowJo software (Tree Star, Inc). The following monoclonal antibodies (all from BD Pharmingen, San Diego, CA except noted otherwise) were used: CD95-APC (Cat. No. 558814, DX2 clone), CD3 ϵ -FITC, CD4-FITC & -PE, CD8 α -FITC, -PE & -PerCP, CD14-FITC, CD19-PE and TLR4-PECy7 (eBioscience, Cat. No. 25-9917). In all experiments, cells were also stained with corresponding isotype-matched monoclonal antibodies. (FACS was done in co-operation with Kaan Boztug, MHH)

2.2.13.1. Staining peripheral blood mononuclear cells

Frozen peripheral blood mononuclear cells were revived by thawing them in 37°C water bath. Immediately after thawing, cells in 1ml (\approx 4 x 10⁶ cells/ml) media were washed twice with 10 to 15ml of PBS (containing 50% FCS). Cells were pelleted by centrifugation, resuspended in 1ml of ice cold PBS (containing 5% FCS) and kept on ice for blocking non-specific receptors for 15 to 20min. Antibody cocktail was prepared by mixing chosen antibodies and staining was performed according to the manufacturer's instructions. Expression of CD95 was analysed on different cell types of peripheral blood mononuclear cells (Table 3). FACS analysis and interpretation of data is described in detail in Fig. 4a to 4c.

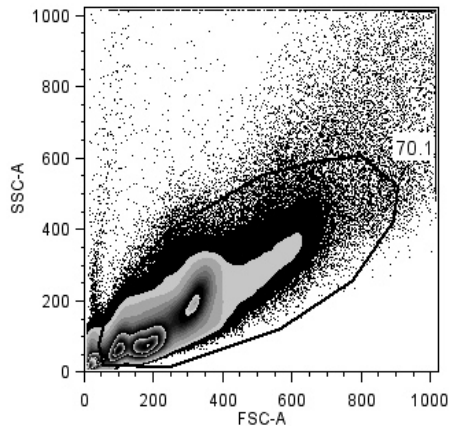
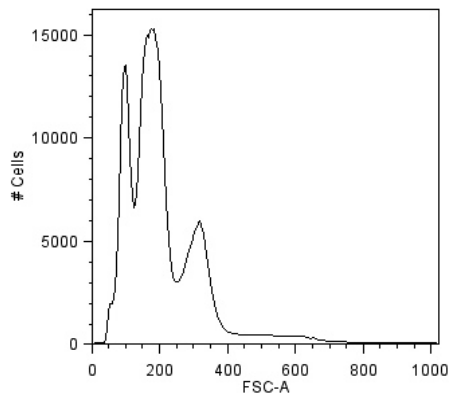


Figure 4a: The typical scatter property of total PBMCs

PBMCs are mixed population of cells and thus we see different light scatter due to light scatter properties of cells

The forward scatter denoted FSC measures the size, and the sideward scatter indicated SSC Measure the granularity of the cells. The black line enclosing 70.1% of cells indicates gating for typical total PBMCs.



The gated population in the above scatter is represented as histogram, showing different peaks with forward scatter in the x axis indicating different populations

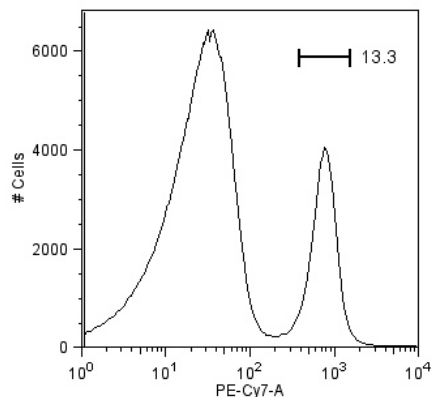
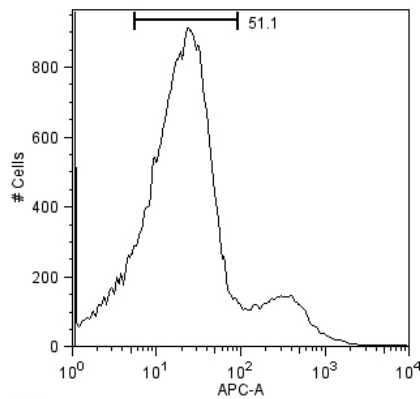


Figure 4b: Gating CD19 positive cells to analyse CD95 expression

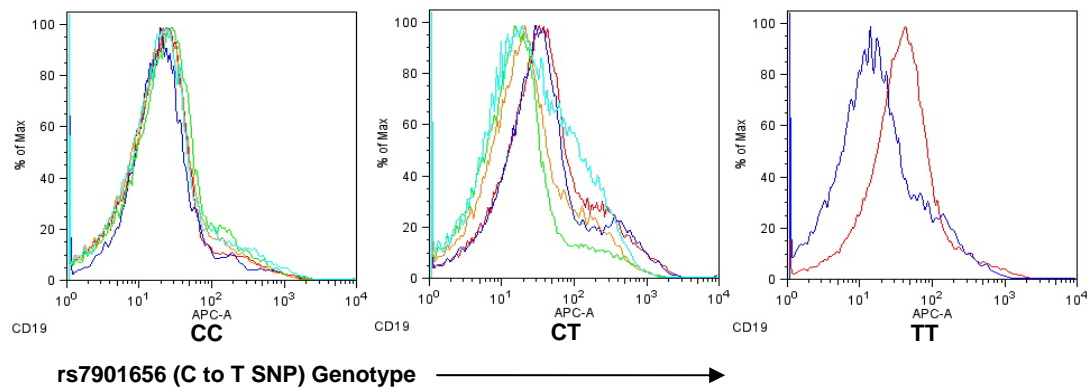
Staining for surface markers using flurochrome conjugated antibodies and looking for respective colour on the x-axis gives the cell population positive for the targeted surface marker.

For example, the histogram indicates that 13.3% of total PBMCs are CD19 positive and hence they are gated and looked for CD95 expression.



The expression of CD95 was analysed on CD19 positive cells. CD95 antibody is conjugated with APC, and thus the histogram indicate that 51.1% of total CD19+ cells are CD95 positive.

Figure 4c: CD95 expression on CD19+ (B-cells) cells



As an example, the expression of CD95 on CD19+ cells is grouped according to the rs7901656 genotypes.

Figure 4a to 4c: The acquired cells are gated for CD19+ cells (B-cells) and these B-cells are gated again for CD95 expressing cells. Later, CD19+CD95+ cells are subjected to statistics to obtain geometric mean of CD95 expression on total B-cells. Similar approach is followed to analyse the CD95 or CD14 or TLR4 expression on different cell types.

Table 3: Markers for surface antigens on different cell types of peripheral blood mononuclear cells

Surface markers	Cell types	Functions
CD14+	Monocytes	Pattern recognition receptor that detects antigenic molecules on the surface of bacteria (lipoteichoic acid on gram positive, LPS on gram negative), myobacteria (glycolipids), fungi (mannans), as part of the innate immune system
CD19+	B-cells	Earliest B cell antigen in fetal tissue. Regulates B cell development, activation and differentiation. May define intrinsic and antigen receptor-induced signaling thresholds critical for clonal expansion of the B cell pool and humoral immunity
CD3e+	Total T-cells	Complex of delta, epsilon, gamma, zeta and eta chains of integral membrane glycoproteins that associates with T cell antigen receptor (TCR), and is required for TCR cell surface expression and signal transduction
CD4+	T-helper cells	Non-polymorphous glycoproteins belonging to immunoglobulin superfamily. Expressed on surface of T helper cells; serves as co-receptor in MHC class II-restricted antigen induced T cell activation
CD8+	Cytotoxic T-cells	Heterodimer of an alpha and a beta chain linked by two disulfide bonds; heterodimer on thymocytes and homodimer on peripheral blood T cells. Can kill target cells by recognizing peptide-MHC complexes on them or by secreting cytokines capable of signaling through death receptors on target cell surface

2.2.14. Serotyping of *P. aeruginosa* CF isolates

Serotypes of *P. aeruginosa* clinical isolates were tested according to the manufacturer's instructions using *Pseudomonas aeruginosa* serotyping kit (ERFA Canada Inc.). Briefly, bacteria were grown on Luria-Bertani agar plate overnight for agglutination test. Small amount of live bacteria from the culture was mixed with 10 µl antibody solution (17 serotype specific monoclonal antibodies) or 10 µl of sterile saline on a clean glass slide. Agglutination was compared with standard positive (ATCC 33348 - ATCC 33364) and negative control (sterile saline) for the respective serotypes. Strains agglutinated with only one antibody were considered as typable or O-antigen intact strains (smooth LPS) where as strains which showed agglutination with more than one type of antibodies (polyagglutinable) or no agglutination with any antibodies were included under non-typable strains or O-antigen deficient strains (Rough LPS).

2.2.15. Electronic database resources

Several electronic database resources were used in this study to understand the mechanism of CF disease modulation via non-coding variants.

The Vista Genome Browser

For functional annotation of non-coding variants, VISTA (Visualisation Tools for Alignments) tool was employed. The Vista Genome Browser, available at <http://genome.lbl.gov/vista/index.shtml>, aligns genome sequences from different organisms and identifies conserved non-coding sequences.

TRANSFAC database

The non-coding sequences were analysed for probable transcription factor binding sites using TRANSFAC (<http://www.gene-regulation.com/pub/databases.html>). TRANSFAC is the database on eukaryotic transcription factors, their binding sites and DNA binding profiles.

miRBase (MicroRNA database)

MicroRNA database (<http://microrna.sanger.ac.uk/sequences/>) is a searchable database of published miRNA sequences and annotation and also provides predicted miRNA targets in different animals.

3. Results and discussion

We hypothesised that genes which are involved in innate immune and non-specific defense will play a key role in determining the rate and duration of *P. aeruginosa* colonisation among CF patients, and consequently CF disease severity. Thus, twelve candidate genes were chosen to evaluate as modulators, namely; toll like receptor-2 (*TLR2*), toll like receptor-4 (*TLR4*), toll like receptor-5 (*TLR5*), toll like receptor-9 (*TLR-9*), *CD14*, *CD95* (*TNFRSF6*), surfactant protein-D (*SFTPD*), *CXCR2* (*IL8RB*), TNF α receptor 1 (*TNFRSF1A*), Paraoxonase (*PON*), Lipid binding protein (*LBP*) and myeloid differentiation primary response gene 88 (*MYD88*) based on their functional importance in recognition of bacterial ligands and in imparting immune response. Initial mapping by SNP-RFLP or microsatellite typing on these candidate genes led us to focus on four candidate genes out of twelve for further fine mapping. For *LBP* and *MYD88*, no informative markers could be established.

Hence, firstly, I will summarise the results of three genes *TLR2*, *TLR5*, and *TLR9*, for which no clear evidence of association was found with the disease phenotype, very briefly in chapter 3.1.

Secondly, results of the three genes *SFTPD*, *CXCR2*, and *PON*, for which a minor association was found with single marker analysis, in chapter 3.2.

Finally, results from four major modulators *TNFRSF1A*, *TLR4*, *CD14* and *CD95* will be described in detail in chapters 3.3, 3.4, 3.5 and 3.6, respectively.

3.1. Analysis of *TLR2*, *TLR5* and *TLR9* as CF modulators

3.1.1. Toll like receptor (TLR) 2

TLR2 was the first human TLR to be described and it is found to interact with a series of bacterial ligands such as lipopeptides (Hoshino et al., 1999), peptidoglycon (Lien et al., 1999), and lipoteichoic acid of gram positive bacteria (Kobe and Deisenhofer, 1995). It is suggested that the activation of epithelial proinflammatory responses, stimulated with *P. aeruginosa*, is predominantly handled by TLR2 presented at the surfaces of airway cells within the context of an asialoganglioside/lipid raft microdomain or caveolae (Soong G et al., 2004). Furthermore, TLR2 expression and response were reported to be strongly enhanced in human CF bronchial epithelial cell lines possibly due to hypomethylation of TLR2 gene promoter (Shuto T et al., 2006).

TLR2, chromosome 4q32

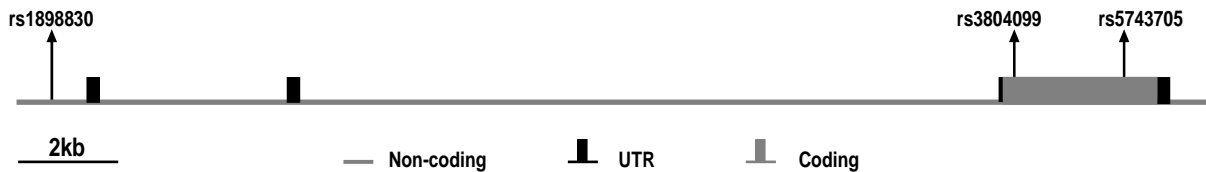


Figure 5: Schematic representation of *TLR2* gene structure; *TLR2*, located on chromosome 4q32, is 21.8kb long and has one exon. Polymorphisms shown on the *TLR2* gene are targeted for mapping analysis. SNPs are indicated by rs followed by their unique numbers (Location and size are based on NCBI, NC_000004.10 Reference assembly: Build 36.2)

To investigate *TLR2* as a modulator in CF, rs3804099, a C to T SNP located in the exon of *TLR2* was targeted (Fig. 5). This SNP, a synonymous polymorphism, has an allele frequency of 52% and 48% for T and C alleles, respectively (NCBI SNP data base). Hence, this SNP was typed on 9 pairs of mildly affected (CON+), 9 pairs of severely affected (CON-) and 16 pairs of discordant (DIS) CF twin and sibling samples. The allele and genotype (Table 4) distributions were compared among CON+, CON- and DIS pairs. A small trend of association was observed between mildly and severely affected pairs at rs3804099 allele distribution, in which allele C was over represented (63%) in mildly affected pairs compared to severely affected pairs. However, the association was not statistically significant ($P = 0.15$).

Table 4: Comparison of genotype and allele distribution among F508del homozygous CF twin and siblings in *TLR2* at rs3804099

Genotypes	CON-	CON+	DIS	CONC*
CC	3	7	10	10
CT	10	9	16	19
TT	5	2	6	7
Total	18	18	32	36
Alleles				
C	16	23	36	39
T	20	13	28	33
Total	36	36	64	72
Freq. of allele C	0.44	0.63	0.56	0.54
Freq. of allele T	0.56	0.37	0.44	0.46

* Both CON- and CON+ pairs are grouped as CONC, concordant pairs, for disease severity

3.1.2. Toll like receptor (TLR) 5

Flagellin is a protein component of gram negative bacterial flagella and TLR5 is found to be receptor for it (Hayashi et al., 2001). Amino acids 386-407 of TLR5 extracellular domain was shown to be responsible for specific binding of flagellin (Mizel et al., 2003). It has been suggested that TLR5 is expressed at the basal surface of epithelial cells, while it has been shown that the addition of *P. aeruginosa* to the apical surface of human nasal CF epithelial cells elicited a considerable amount of gene upregulation (Hybiske et al., 2004). Furthermore, it was shown that human airway epithelial cells sense *P. aeruginosa* infection via recognition of flagellin by TLR5 and play an important role in the initiation of inflammatory response against its invasion (Zhang et al., 2005).

TLR5, chromosome 1q41.q42

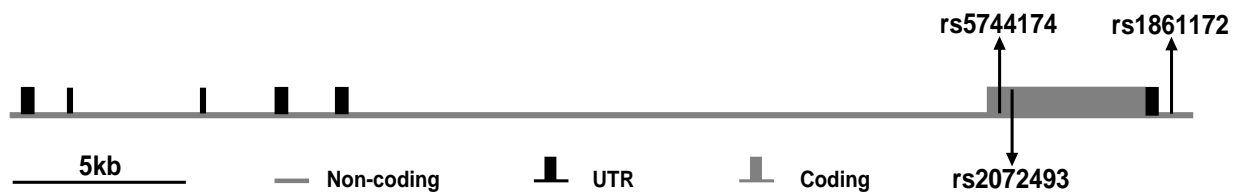


Figure 6: Schematic representation of TLR5 gene structure; TLR5, located on chromosome 1q41.q42, is 32.97kb long and has one exon. Polymorphisms shown on the TLR5 gene are targeted for fine mapping. SNPs are indicated by rs followed by their unique numbers (Location and size are based on NCBI, NC_000001.9 Reference assembly: Build 36.2)

Three SNPs, rs 5744174, rs2072493 and rs1861172 were selected for optimisation. Among these three SNPs, A to G SNP, rs1861172, was found to be informative marker with allele frequencies of 53% and 47% for alleles A and G, respectively. The SNP rs1861172 is approximately 400bp distant to the 3' end of the *TLR5* gene (Fig. 6). The comparison of allele and genotype distributions between eight mildly affected pairs, eight severely affected pairs and 14 pairs of DIS pairs at this locus did not show any association with disease severity (Table 5).

Table 5: Comparison of genotype and allele distribution among F508del homozygous CF twin and siblings in *TLR5* at rs1861172

Genotypes	CON-	CON+	DIS	CONC*
AA	6	2	9	8
AG	6	9	12	15
GG	4	5	7	9
Total	16	16	28	32
Alleles				
A	18	13	30	31
G	14	19	26	33
Total	32	32	56	64
Freq. of allele A	0.56	0.40	0.53	0.48
Freq. of allele G	0.44	0.60	0.47	0.52

* Both CON- and CON+ pairs are grouped as CONC, concordant pairs, for disease severity

3.1.3. Toll like receptor (TLR) 9

TLR9 recognises bacterial DNA containing unmethylated CpG motifs and TLR9 deficient mice are not responsive to CpG DNA challenge (Hemmi et al., 2000). It has previously been shown that unmethylated CpG DNA isolated from CF sputum induces lower respiratory tract inflammation in an animal model (Schwartz et al., 1997). Additionally, signalling through TLR9 was shown to be important in *P. aeruginosa* keratitis (Huang et al., 2005). Two SNPs were found to be informative after initial analysis on control samples. A synonymous C to T SNP, rs352140, located on exon 2 of the *TLR9* gene and a promoter SNP rs187084 (Fig. 7) were targeted to evaluate the role of *TLR9* in CF disease modulation by typing them on twin and sibling samples.

TLR9, chromosome 3p21.3

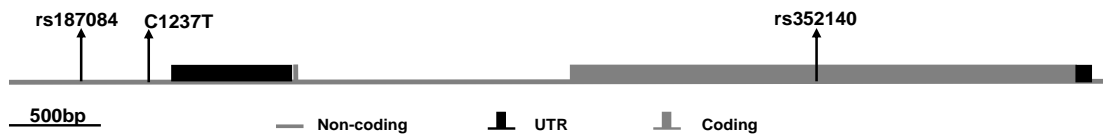


Figure 7: Schematic representation of TLR9 gene structure; TLR9, located on chromosome 3p21.3, is 5.08kb long and has two exons. Polymorphisms shown on the TLR9 gene are targeted for fine mapping. SNPs are indicated by rs followed by their unique numbers. C1237T is a SNP chosen from Lazarus et al., 2003. (Location and size are based on NCBI, NC_000003.10 Reference assembly: Build 36.2)

At first, the allele and genotype distributions at these two SNPs were compared individually for CON+, CON- and DIS pairs. Later on, construction of 2-marker haplotype (rs187084-rs352140) and its distribution among CF twin and sibling pairs was compared. Both the single marker analysis as well as two-marker haplotype analysis (Table 6) did not reveal any association with CF disease modulation.

Table 6: TLR9 two-marker haplotype (rs187084-rs352140) distribution among F508del homozygous CF twin and siblings

TLR9 rs187084- rs352140 haplotype	CON-	CON+	DIS
1-1	8	6	7
1-2	10	10	13
2-1	8	16	18
2-2	2	0	0

CON-/CON+: P = 0.18, CON-/DIS: P = 0.16, CON+/DIS: P = 1.0

3.1.4. Role of TLR2, TLR5 and TLR9 polymorphisms in CF

This is the first study to analyse TLR2, TLR5 and TLR9 as cystic fibrosis modulators. Earlier investigations on polymorphisms in innate immunity genes have suggested that these polymorphisms may play a significant role in determining the susceptibility to infection in general (Lazarus et al., 2002). Consistent with this hypothesis, polymorphisms on TLR2 are reported to be associated with inflammatory bowel disease (Pierik et al., 2006), susceptibility to asthma and allergies in children of farmers (Eder et al., 2004), and risk of developing tuberculosis (Ogus et al., 2004). Similarly, TLR5 stop codon polymorphism was shown to be associated with susceptibility to legionnaires' disease (Hawn et al., 2003), resistance to systemic lupus erythematosus (Hawn et al.,

2005), and negatively associated with Crohn's disease (Gewirtz et al., 2006). *TLR9* polymorphism was reported to be associated with Crohn's disease (Torok et al., 2004), and risk of low birth weight among *Plasmodium falciparum*-infected women (Mockenhaupt et al., 2006). However, no studies have reported the role of *TLR2*, *TLR5* and *TLR9* polymorphisms in CF and susceptibility to *P. aeruginosa* infection. On the other hand, the importance of TLRs, particularly TLR2 signalling in the pulmonary host response to *P. aeruginosa* was clearly shown in which TLR2 deficient mice exhibited augmented cytokine response and delayed bacterial killing (Skerrett et al., 2006), and mucoid *P. aeruginosa* preferentially induced lipoprotein genes which in turn triggered the proinflammatory response in TLR2 dependent manner (Firoved et al., 2004). Furthermore, TLR2 expression was slightly reduced in bronchial epithelium of CF patients (Hauber et al., 2005).

Table 7: P-values obtained for comparison of genotype and allele distribution among F508del homozygous CF twin and siblings in *TLR2*, *TLR5* and *TLR9*

	Disease phenotype	TLR2 rs3804099	TLR5 rs1861172	TLR9 rs187084	TLR9 rs352140
Genotypes	CON-/CON+	0.3	0.3	0.8	0.3
Genotypes	CONC/DIS	0.9	0.8	0.9	0.8
Alleles	CON-/CON+	0.1	0.3	1.0	0.5
Alleles	CONC/DIS	0.8	0.5	0.8	0.5

However, there was no significant association found in our study between *TLR2*, *TLR5* and *TLR9* polymorphisms and CF disease severity (Table 7), which may be due to two possibilities. Firstly, polymorphisms on *TLR2*, *TLR5* and *TLR9* may not play any role in CF disease modulation among CF twin and sibling cohort. Secondly, the number of polymorphisms tested on *TLR2*, *TLR5* and *TLR9* are not sufficient to detect the causative variant, if any, on these genes due to recombination events between selected markers and actual causative variant(s). Thus, it is necessary to perform the systematic haplotype analysis to examine both possibilities.

3.2. Evaluation of *SFTPD*, *IL8RB* and *PON* as CF genetic modulators

3.2.1 Surfactant Protein (SP)-D (*SFTPD*)

SP-D belongs to the collectin family of calcium-dependent carbohydrate binding proteins, which includes SP-A and the serum collectins, mannan-binding lectins and bovine conglutinin (Hoppe and Reid, 1994). SP-A and SP-D are considered to be molecules of the innate immune system, involved in first-line defense of mucosal surfaces, especially the lung, against bacterial, viral, fungal, or allergen challenge (Crouch and Wright., 2001; Clark et al., 2000). In CF scenario, it is shown that *P. aeruginosa* elastase degrades SP-A and SP-D (Mariencheck et al., 2003) and the levels of SP-D in the CF airway are insufficient for suppression of inflammation and endotoxin clearance, which may lead to bronchiectasis (Noah et al., 2003). Furthermore, it has been suggested that recombinant SP-D could play a role in therapeutic strategies for cystic fibrosis (Clark and Reid, 2003). Thus, SP-D was analysed as a CF genetic modulator among CF twin and sibling cohort.

SFTPD, chromosome 10q22.2-q23.1

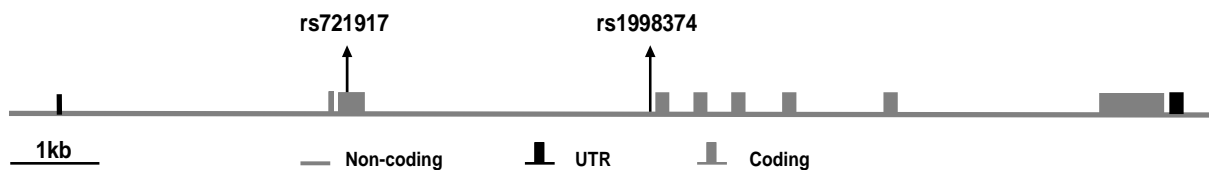


Figure 8: Schematic representation of SP-D (SFTPD) gene structure; SP-D, located on chromosome 10q22.2-q23.1, is 11.36kb long and has eight exons. Polymorphisms shown on the SP-D gene are targeted for fine mapping. SNPs are indicated by rs followed by their unique numbers (Location and size are based on NCBI, NC_000010.9 Reference assembly: Build 36.2)

Initial typing of two polymorphisms on control samples showed rs721917 as an informative marker. The rs721917 is a non-synonymous C to T polymorphism located on exon 2 of the gene (Fig. 8) and changes threonine to methionine at the 31st aminoacid position. Nine pairs of CON+, nine pairs of CON- and 15 DIS pairs of CF twin and siblings were genotyped at rs721917. Comparison of allelic and genotypic distribution at this locus among CF twin and siblings showed an allelic imbalance between mildly and severely affected pairs in which allele C was underrepresented (34%) among mildly affected pairs compared to 56% among severely affected pairs (Table 8). However, the association failed to reach statistical significance ($P = 0.06$).

Table 8: Comparison of genotype and allele distribution among F508del homozygous CF twin and siblings in SP-D at rs721917

Genotypes	CON-	CON+	DIS	CONC*
CC	7	1	7	8
CT	6	10	11	16
TT	5	7	12	12
Total	18	18	30	36
Alleles				
C	20	12	25	32
T	16	24	35	40
Total	36	36	60	72
Freq. of allele C	0.56	0.34	0.42	0.45
Freq. of allele T	0.44	0.66	0.58	0.55

* Both CON- and CON+ pairs are grouped as CONC, concordant pairs, for disease severity

3.2.1.2. Role of surfactant protein-D in cystic fibrosis

Surfactant protein-D plays an important role in pulmonary diseases (Hartl and Griese, 2006). It is encoded by a single gene located on the long arm of Chromosome 10q22.2-23.1, 335kb downstream of surfactant protein-A (SP-A) another important collectin. Polymorphisms on SP-D are shown to be associated with infectious pulmonary diseases, in which SNP rs721917 was associated with susceptibility to tuberculosis (Floros et al., 2000) and severe respiratory syncytial virus bronchiolitis in infants (Lahti et al., 2000). Further, a specific haplotype on SP-D was associated with decreased serum levels of SP-D (Heidinger et al., 2005). Mannose binding lectin (MBL), belongs to the collectin family, was successfully shown to be a modulator of cystic fibrosis by many investigations (Garred et al., 1999; Gabolde et al., 1999) in which variant alleles of MBL are associated with severity of pulmonary disease and survival of CF patients. In contrast, there are no associations between CF and SP-D polymorphisms have been identified up to this study. In this study, a minor association ($P = 0.06$) was found between a frequent polymorphism, rs721917, and CF disease severity. Although it was not statistically significant, it is interesting to note that the same polymorphism was shown to be associated with susceptibility to tuberculosis (Floros et al., 2000) and severe respiratory syncytial virus bronchiolitis in infants (Lahti et al., 2000). However, SP-A is also located close to SP-D on the genetic locus which necessitates the discrimination between two adjacent, equally plausible genes. Thus, it is essential to

fine map the region by constructing haplotypes with more number of informative markers and thereby delineating the actual causative variant(s).

3.2.2. CXCR2 (IL8RB interleukin 8 receptor, beta)

Neutrophil recruitment to the sites of infection is primarily dependent upon chemokines such as IL-8 (CXCL8). Peripheral blood neutrophils express two main chemokine receptors on cell surface, namely CXCR1 and CXCR2 (Chuntharapai et al., 1995). Among them, CXCR2 is thought to be predominantly responsible for recruitment of neutrophils in response to IL-8 (Sabroe I et al., 2002). Additionally, Tsai et al. (2000) reported that CXCR2 is essential for protective innate host response in murine *P. aeruginosa* pneumonia. Hence, the chemokine receptor CXCR2 was evaluated as a CF disease modulator.

IL8RB, chromosome 2q35



Figure 9: Schematic representation of CXCR2 (IL8RB) gene structure; CXCR2, located on chromosome 2q35, is 11.23kb long and has one exon. SNP rs2230054 shown on the CXCR2 gene is targeted for evaluation of association. (Location and size are based on NCBI, NC_000002.10 Reference assembly: Build 36.2)

The SNP rs2230054 is a synonymous SNP located on exon 3 of CXCR2 (Fig. 9) with allele frequencies of 57% and 43% for C and T alleles respectively (NCBI database). This polymorphism was typed on 10 pairs of severely affected pairs, 12 pairs of mildly affected pairs and 17 pairs of DIS pairs. Interestingly, the allele distribution between CONC and DIS groups were significantly different ($P = 0.04$). Allele T was overrepresented with 66% among CONC compared to 50% among DIS pairs (Table 9) indicating the modulation via trans-acting factors as described for *CFTR* hitchhiking genes (Mekus et al., 2003).

Table 9: Comparison of genotype and allele distribution among F508del homozygous CF twin and siblings in CXCR2 at rs2230054

Genotypes	CON-	CON+	DIS	CONC*
TT	9	11	10	20
TC	10	8	14	18
CC	1	5	10	6
Total	20	24	34	44
Alleles				
T	28	30	34	58
C	12	18	34	30
Total	40	48	68	88
Freq. of allele T	0.70	0.63	0.50	0.66
Freq. of allele C	0.30	0.37	0.50	0.34

Both CON- and CON+ pairs are grouped as CONC, concordant pairs, for disease severity

3.2.2.1. No association with *P. aeruginosa* early or late chronic colonisation among unrelated CF patients

Single-marker analysis at rs2230054 showed an association with CF disease discordance. Thus, rs2230054 was further evaluated for its association with *P. aeruginosa* chronic colonisation. The cohort containing 21 unrelated F508del homozygous CF patients, stratified for early and late *P. aeruginosa* chronic colonisation, was typed at rs2230054. Genotype distributions at rs2230054 were compared between 13 patients belonging to *P. aeruginosa* early colonised and eight patients of *P. aeruginosa* late colonised group. The genotype distributions ($P = 0.72$) and allele distributions ($P = 0.37$) in both the groups were similar (Table 10). No association was found between CXCR2 SNP, rs2230054 and early or late colonisation of *P. aeruginosa*.

Table 10: Comparison of allele and genotype distribution among *P. aeruginosa* early and late colonised delF508 homozygous CF patients at rs2230054

	<i>P. aeruginosa</i> early	<i>P. aeruginosa</i> late
Genotypes		
TT	1	4
CT	6	4
CC	3	3
Alleles		
T	8	12
C	12	10

3.2.2.2. Role of CXCR2 gene in cystic fibrosis

CXCR2 is an important chemokine receptor for mediating cellular activity of neutrophils by attracting IL-8 (Sabroe et al., 2002) and CXCR2 deficient mice had defects in neutrophil migration (Cacalano et al., 1994). Furthermore, neutralization of CXCR2 in mice increased the mortality due to reduced accumulation of neutrophils and impaired *P. aeruginosa* clearance from the lung (Tsai et al., 2000). Consistently, in CF, the activated neutrophils are considered to be the primary effector cells for the pathogenesis of CF lung disease as neutrophil influx is mediated and sustained by IL-8 (Chmiel et al., 2002, Tirouvanziam et al., 2000). However, no association studies of CXCR2 and CF have been reported. On the other hand, the CXCR2 SNP, rs2230054, was shown to be associated with chronic obstructive pulmonary disease (COPD), in which patients carrying T allele had a better lung function (Matheson et al., 2006), supporting a role for this gene in the inflammatory processes in the airways. In this study, evaluation of rs2230054 for association with cystic fibrosis disease showed a significant association ($P = 0.04$) with disease discordance. This finding strongly suggested that gene(s) encoding elsewhere in the genome (trans-acting factors) interact differentially with CXCR2 locus to modify CF disease severity. Hence, fine-mapping of CXCR2 is needed to locate the causative variant. Additionally, analysis of the same polymorphism for *P. aeruginosa* early or late colonisation phenotype did not show any association. Nevertheless, it is interesting to note that individuals carrying TT genotype were underrepresented in *P. aeruginosa* early colonised group, indicating a possible role of CXCR2 in modulating susceptibility to *P. aeruginosa* colonisation among CF patients. Clearly, further evaluation of CXCR2 polymorphism on larger cohort, stratified for *P. aeruginosa* colonisation, is necessary to confirm this finding.

3.2.3. PON (paraoxonase) gene cluster

The *PON* gene cluster (Fig. 10) contains at least three members, including *PON1*, *PON2* and *PON3*, located on chromosome 7q21.3-22.1. These three paraoxonases are a family of mammalian lactone hydrolases with high sequence similarity (Li et al., 2003). However, they show distinct substrate specificities and expression patterns (Primo-Parmo., 1996). The PON, in general, hydrolyses many substrates including aromatic carboxylic acid and organophosphates such as paraoxon. These enzymes can use acyl-HSL molecules as their substrates (Draganov et al., 2005) and can degrade N-(3-

oxododecanoyl)-L-homoserine lactone (3OC12-HSL) from *P. aeruginosa* by hydrolysing the lactone ring (Ozer et al., 2005). As *P. aeruginosa* uses acyl-HSL quorum sensing molecules to regulate virulence genes and biofilm formation, degradation of this key molecule by *PON* may modulate the quorum sensing. With this rationale, we targeted the *PON* cluster as a modulator of CF and evaluated by typing a tri-nucleotide repeat microsatellite among CF twin and sibling cohort.

PON Cluster, chromosome 7q21.3

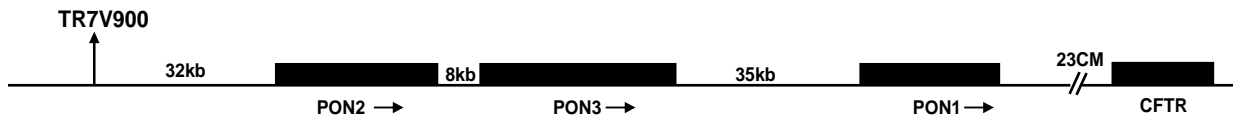


Figure 10: Schematic representation of *PON* gene cluster *PON2*, *PON3* and *PON1*. *TR7V900* is a trinucleotide repeat polymorphism located 32kb upstream of *PON* cluster. (Location and size are based on NCBI, NC_000007.12 Reference assembly: Build 36.2)

Initial optimisation of 6 microsatellites on this locus, for informative markers, indicated that *TR7V900* as a suitable marker. *TR7V900* is a (tgg)₁₀ repeat polymorphism, located 32 kb upstream to *PON2* and has three different alleles (allele 9, 10 and 11; Fig. 11).

Figure 7: Genotypes of a family at *TR7V900*

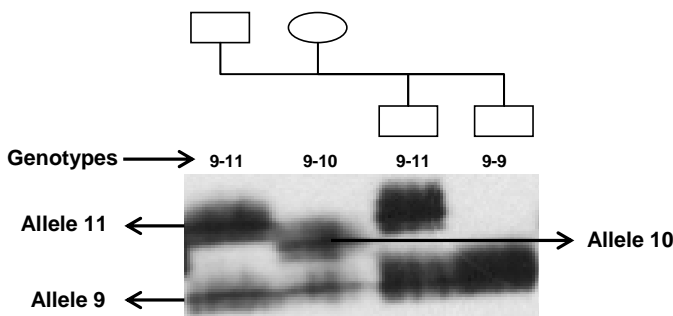


Figure 11: Three different alleles (9, 10, and 11) observed at *TR7V900* with in a family. Alleles are named by assigning lowest band as allele 9. (9-11); genotype of father, (9-10); genotype of mother, (9-11); genotype of 1st sib, (9-9); genotype of 2nd sib

Ten severely affected pairs, twelve mildly affected pairs and 19 DIS pairs were genotyped at *TR7V900* (Table 11). Comparison of allele and genotype distribution between mildly affected and severely affected pairs showed a minor association (P = 0.07) with disease severity in which allele 11 was over represented among mildly affected pairs (54%) compared to severely affected pairs (37%).

Table 11: Comparison of genotype and allele distribution among F508del homozygous CF twin and siblings at TR7V900

Genotype	CON-	CON+	DIS	CONC*
11-11	2	9	9	11
10-10	3	5	7	8
9-9	0	0	0	0
11-10	8	8	19	16
11-9	3	0	3	3
10-9	4	2	4	6
Total	20	24	38	44
Alleles				
11	15	26	40	41
10	18	20	33	38
9	7	2	3	9
Total	40	48	76	88
Freq. of allele 11	0.37	0.55	0.53	0.46
Freq. of allele 10	0.45	0.41	0.43	0.43
Freq. of allele 9	0.18	0.04	0.04	0.11

* Both CON- and CON+ pairs are grouped as CONC, concordant pairs, for disease severity

3.2.3.1. Role of PON polymorphisms in cystic fibrosis

PONs, including PON1, PON2, and PON3, are a group of enzymes that play a key role in organophosphate detoxification and in prevention of atherosclerosis (Harel et al., 2004). PON1 and PON3 can hydrolyse a broad range of esters and lactones (Yang et al., 2005) and interestingly recombinant human PON2 could effectively hydrolyse several AHL compounds (Draganov et al., 2005). Considering all these points, PON cluster most likely play a role in AHL-inactivation in *P. aeruginosa* infected CF airways.

Table 12: P-values obtained for comparison of genotype and allele distribution among F508del homozygous CF twin and siblings at TR7V900

	Phenotype	TR7V900
Genotypes	CON-/CON+	0.07
Genotypes	CONC/DIS	0.94
Alleles	CON-/CON+	0.07
Alleles	CONC/DIS	0.28

In this study, targeted repeat polymorphism, TR7V900, showed a minor association ($P = 0.07$) of this locus with CF disease severity. However, all the patients analysed in this study were homozygous for deltaF508 mutation on the *CFTR* gene and *PON* cluster,

located 23 CM distant from *CFTR* on chromosome 7, is linked to *CFTR* (Schmiegelow et al., 1986). Previous study on genetic modulators in European CF twin and siblings has identified genes in the vicinity of *CFTR* as modulators of CF (Mekus et al., 2003). Therefore, fine-mapping of this locus with more informative markers is obligatory to delineate recombination breakpoints (haplotype blocks) and thereby differentiating the association found between CF disease and *CFTR* hitchhiking genes with association between CF disease severity and *PON* locus.

3.3. Analysis of TNF α receptor *TNFRSF1A* as a modulator in cystic fibrosis

TNFR1 is constitutively expressed in most tissues and is found to be the key mediator of the pleiotropic actions of TNF α proinflammatory cytokine on its target cells (Wajant et al., 2003). TNFR1, both membrane bound and soluble form, exerts proinflammatory response upon binding to TNF α by activating NF- κ B and AP-1 transcription factors (Wajant et al., 2003) and can also initiate apoptosis when NF- κ B signaling is blocked (Varfolomeev and Ashkenazi, 2004). Several studies in TNFR1 deficient mice established the requirement of this receptor in host defense against microorganisms (Wajant et al., 2003; Steinshamn et al., 1996; Deckert-Schluter et al., 1998; Castanos-Velez et al., 1998; O'Brien et al., 1999) and in the induction and maintenance of various chronic inflammatory pathologies (Kollias et al., 1999). Similarly, TNFR1 role in humans was highlighted by mutations in *TNFR1* gene causing autoinflammatory diseases, termed TNFR1-associated periodic syndromes (TRAPS) due to reduced levels of TNFR1 shedding (McDermott et al., 1999).

3.3.1. Rationale for choosing *TNFR1* as sequencing target

Previously in our lab, alpha, beta and gamma subunits of ENaC (Epithelial sodium channel) were targeted as modulators of CF. The comparison of transmitted and non-transmitted alleles and haplotypes from all nuclear families of 37 F508del-CFTR homozygous sibs on alpha subunit of ENaC (*SCNN1A*, maps to 12p13), revealed a skewed haplotype distribution. However, the haplotype in question was located exactly within its neighbouring gene *TNFR1* (*TNFRSF1A*), separated by 5kb of intergenic sequence (Fig. 12).

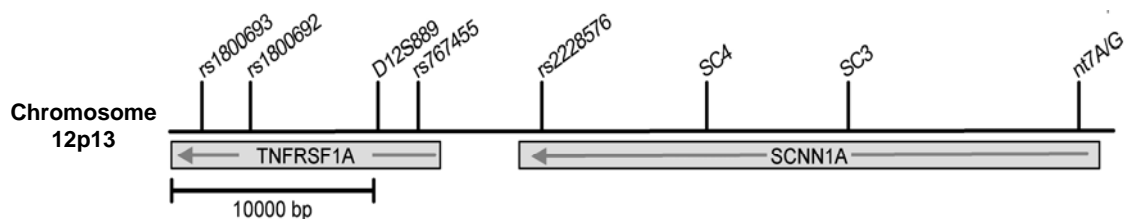


Figure 12: Schematic representation of *TNFR1* (*TNFRSF1A*) and *SCNN1A* location on chromosome 12p13; polymorphisms shown on the map are targeted for fine mapping. D12S889 is a di-nucleotide repeat polymorphism; SC4 and SC3 are tetra- and tri-nucleotide polymorphisms- the others are SNPs

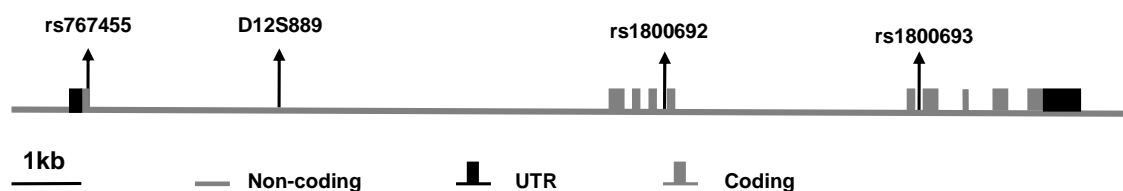
Frequency distribution of the haplotype block rs740842-rs1800693-rs1800692-D12S889-rs767455-rs2228576, encompassing the complete *TNFR1* gene, was significantly skewed ($P = 0.008$). Furthermore, the haplotype block D12S889-rs767455 (intron 1 to codon 12 in exon 1 of *TNFR1*) was highly significantly skewed, in which the haplotype 10-G was preferentially transmitted to CF offspring while the haplotype 13-A was more frequent among non-transmitted parental haplotypes (Table 13), indicating that chromosomes carrying haplotype 10-G harbour a benign variant. These results implied that the sequence variant(s) are located within *TNFR1* and hence the complete coding region of *TNFR1* was sequenced.

Table 13: Distribution of D12S889-rs767455 haplotype block on *TNFR1* among CF twin and siblings

TNFR1 D12S889-rs767455 haplotype	Transmitted	Non- transmitted	CON-	CON+
13-A	0.20	0.41	0.397	0.216
14-A	0.15	0.04	0.068	0.104
8-G	0.07	0.10	0.091	0.027
10-G	0.58	0.28	0.250	0.500
13-G	0.0	0.15	0.080	0.054
Other			0.115	0.099

3.3.2. Sequence analysis of *TNFR1* coding region

In order to find the exact causative variant(s) in *TNFR1*, genomic DNA samples from individuals harbouring contrasting haplotypes at rs1800693-rs1800692-D12S889-rs767455 in homozygous state were sequenced.



*Figure 13: Schematic representation of *TNFR1* (*TNFRSF1A*) gene structure; *TNFR1*, located on chromosome 12p13.2, is 14kb long and has ten exons. Polymorphisms shown on the *TNFR1* gene are targeted for fine mapping. SNPs are indicated by rs followed by their unique numbers. D12S889 is a di-nucleotide repeat polymorphism (Location and size are based on NCBI, NC_000012.10 Reference assembly: Build 36.2)*

The genomic sequence of *TNFR1* was determined by direct sequencing of four unrelated F508del homozygotes (eight chromosomes). Two individuals carried G-C-10-

G haplotype in homozygous state at rs1800693-rs1800692-D12S889-rs767455. Three chromosomes carried the contrasting haplotype A-T-13-A. The sequenced fragment of *TNFR1* coding regions composed three targets. The target 1 of size 780bp covered the near promoter, the 5' UTR and exon 1. The second target of size 1690bp covered the exon 2 to exon 5 region and the target 3 of size 2665bp covered the exon 6 to 3' UTR (Fig. 13). The primers employed for sequencing the coding regions of *TNFR1* are tabulated in table 14.

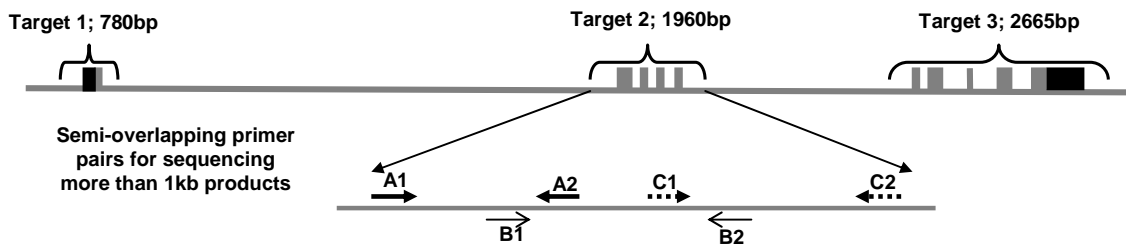


Figure 14: Schematic representation of *TNFR1* (*TNFRSF1A*) gene structure and its sequenced regions. Sequencing strategy for products more than 1kb in size within *TNFR1* was shown for target 2. A1 and A2 primer pair amplifies 486bp product; B1 and B2 primer pair amplifies 597bp product; C1 and C2 primer pair amplifies 470bp product. B1 primer overlaps approximately 30bp of A1A2 product.

All the fragments were amplified and sequencing was performed at Qiagen. For sequencing target 2 and target 3, semi-overlapping internal primers were designed to amplify smaller products from already amplified target 2 and target 3 fragments (Fig. 14).

Table 14: Primers employed for *TNFR1* coding region sequencing

Fragment/internal primer	PCR product size	Forward primer	Reverse primer
Target 1; promoter, exon 1	780 bp	5' ccactctgggaaaggctagt	5' ctgaccggaggacagaat
Target 2; exon 2 to exon 5	1690 bp	5' gccgtagtcccagctattca	5' acagagctgccaattcaacc
Target 2; internal primer set 1	486 bp	5' agattgtatggcccaactg	5' gccgattccctgaagtctct
Target 2; internal primer set 2	597 bp	5' cctgggctgggattttct	5' ctgcaattgaagcactggaa
Target 2; internal primer set 3	470 bp	5' ggctgcaggaagaaccagt	5' gctaattggtcccaccagtc
Target 3; exon 6 to 3' UTR	2665 bp	5' ttcaagatccctgccaatc	5' tgagccccacagaaagtgt
Target 3; internal primer set 1	561 bp	5' ttcaagatccctgccaatc	5' ctcctctccctccaaag
Target 3; internal primer set 2	500 bp	5' tctgaccaaacactgctttg	5' tgggagtaactctctcattcatc
Target 3; internal primer set 3	581 bp	5' gaggcatgtcaccacaagtc	5' cacttctgaaggggttgg
Target 3; internal primer set 4	671 bp	5' agaggtggcaccaccctatc	5' aggaccctccttccaga
confirmatory sequencing of intron 2	491 bp	5' ccctcaggggtattggact	5' ctgtgcactcacccttcc
confirmatory sequencing of exon 4	573 bp	5' tgtgggtcctgtctatgtg	5' gtgcacacgggttctgttt
confirmatory sequencing of intron 5	572 bp	5' tccccctctgtattctgtg	5' acagagctgccaattcaacc
confirmatory sequencing of intron 7	428 bp	5' catcccaccatccatcta	5' gaatggtcaggacatttgg
confirmatory sequencing of exon 9	565 bp	5' ctgcgccaccttctctt	5' gatctctgtgtcgtcag
confirmatory sequencing of 3' UTR	618 bp	5' gtacgcgtggtggagaac	5' aaaaaggctcaggacgaac

Briefly, target 2 was split into 3 products of size 486bp, 597bp and 470bp. Similarly, target 3 was split into four fragments of size 561bp, 500bp, 581bp and 671bp. Both strands of each targeted fragment were sequenced and the sequenced results were aligned against reference sequence using the software CodonCode aligner and scanned for sequence variants. Additionally, small PCR products were amplified for confirming the ambiguous results obtained within intron 2, exon 4, intron 5, intron 7, exon 9 and the 3' UTR. The sequencing analysis confirmed already typed polymorphisms rs1800693 in intron 6 and rs1800692 in intron 4. The entire coding sequence of *TNFR1*, the 5' UTR and near promoter sequence was unaltered in all sequenced samples. A rare SNP was found within intron 4; however, further analysis of this SNP on CF twin and siblings showed that the SNP was not informative (polymorphism information content of the rare SNP: 0.11).

3.3.3. Sequencing was focused on the intron 1 of *TNFR1*

To test if preferentially transmitted haplotype is associated with disease severity, haplotype distribution was compared between both severely and mildly affected patient pairs as well as discordant and concordant pairs. Discordant and concordant pairs were comparable at all tested loci on 12p13. The distribution of haplotype block D12S889-rs767455 was significantly different between mildly and severely affected pairs ($P = 0.02$). The haplotype distributions at D12S889-rs767455 reflected the results of the family-based analysis, in which the haplotype 10-G was overrepresented among mildly affected patient pairs and the haplotype 13-A was overrepresented among severely affected pairs (Table 13), supporting the finding that haplotype 10-G at D12S889-rs767455 is carrying a benign variant and is located within intron 1. Furthermore, family based haplotype analysis showed that the chromosome segment carrying allele 10 at D12S889 was shared preferentially among all CF twin and siblings and the same segment was enriched among mildly affected pairs compared to severely affected pairs (Table 13). Accordingly, reconstruction of haplotypes and analysis for recombination break points showed a 10kb segment, defined by markers rs1800692 and D12S889 (intron 4 to intron 1 of *TNFR1*), was shared differently among mildly and severely affected pairs, in which majority of the mildly affected sibs shared C-10 haplotype. This finding suggested that the causative variants are located within intron 1 of *TNFR1*.

3.3.4. Sequencing analysis of *TNFR1* intron 1

The fact that the haplotype of interest, D12S889-rs767455, was located within intron 1 and absence of causative coding variants prompted us to screen the entire 7.5kb intron 1 by sequencing. Similar strategy, as shown for the *TNFR1* coding region sequencing (Fig. 14), was followed to sequence the complete intron 1 of *TNFR1* (Fig. 15). The same eight chromosomes, which were sequenced for *TNFR1* coding region, were also sequenced at *TNFR1* intron 1. The primers employed for sequencing intron 1 are displayed in table 15.



Figure 15: TNFR1 first intron sequencing strategy; the two targets of size 3481bp and 3800bp, covering the complete intron 1, were PCR-amplified. Further, target 1 and target 2 of intron 1 were targeted by six and seven sets of semi-overlapping internal primers respectively to obtain the complete intron 1 sequence.

Intron 1 sequencing identified seven previously reported polymorphisms (Fig. 16). Interestingly, 7.5kb sequence of *TNFR1* intron 1 differed by haplotype at these seven SNPs. Comparing the two contrasting haplotypes G-C-10-G and A-T-13-A at rs1800693-rs1800692-D12S889-rs767455, four chromosomes of G-C-10-G carried the respective seven-marker haplotype G-G-T-G-T-T-10-A and three chromosomes of A-T-13-A carried C-T-C-A-C-C-13-G haplotype at rs2284344-rs887477-rs1860545-rs4149581-rs4149580-rs4149577-D12S889-rs4149576 in intron 1 of *TNFR1*. Furthermore, all chromosomes carrying allele 10 at D12S889 shared the haplotype C-10 at rs1800692-D12S889 in our CF families.

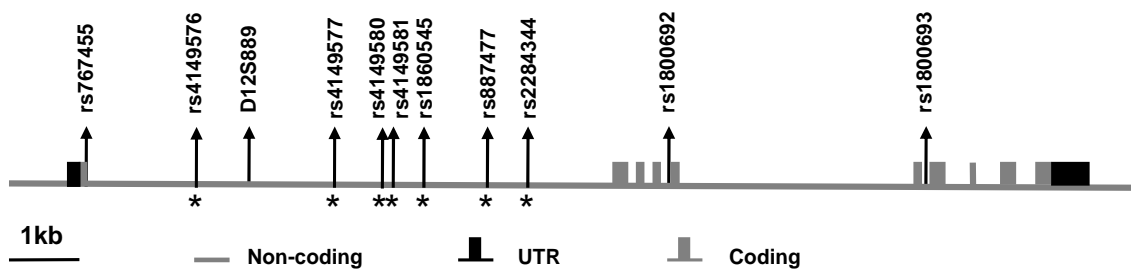


Figure 16: *SNPs identified by sequencing within intron 1 of *TNFR1* gene are shown. SNPs are indicated by rs followed by their unique numbers. D12S889 is a di-nucleotide repeat polymorphism (Location and size are based on NCBI, NC_000012.10 Reference assembly: Build 36.2)

Mapping of haplotype fragment within intron 1 of *TNFR1* and absence of disease causing sequence variants in the coding region indicated that non-coding variants located within intron 1 might play a functional role in determining the CF disease severity. The mapped chromosome segment carrying allele 10 at D12S889, however, differed at seven SNPs within the 7.5kb of intron 1 of *TNFR1*. Thus, it is plausible that either one of the intron 1 SNPs or a haplotype composed of several allelic variants within intron 1 alters the functionality of TNFR1.

Table 15: Primers employed for *TNFR1* intron 1 sequencing

Fragment/internal primer	Size of PCR product	Forward primer	Reverse primer
TNFR1 intron 1 – target 1	3481 bp	5' ctgcaggtcctaacctcagc	5' ttccctgacccttaacatgc
Target 1: internal primer set 1	591 bp	5' gaagattagtctcggggagt	5' agacgtggaggacgatcaag
Target 1: internal primer set 2	561 bp	5' gcctcagccttgttcaatg	5' tgtgacaccatttggaga
Target 1: internal primer set 3	589 bp	5' ttccatgcctctgttcttt	5' gcttggccacagtctctct
Target 1: internal primer set 4	584 bp	5' gagaagtgggtaggggagt	5' accctgacaggcagtgactt
Target 1: internal primer set 5	597 bp	5' gaaaagtggatgtggagctga	5' ggtgtggtcagggctcact
Target 1: internal primer set 6	581 bp	5' agactgagctcgggaggac	5' ttccctgacccttaacatgc
TNFR1 intron 1 – target 2	3800 bp	5' gatggatgaattccgtctgg	5' gggccatacaatctgatgct
Target 2: internal primer set 1	557 bp	5' gatggatgaattccgtctgg	5' tgaactcgaagcacgtgaac
Target 2: internal primer set 2	593 bp	5' ttggggagatcatatctgca	5' cctcagactccccaccact
Target 2: internal primer set 3	598 bp	5' ctgctgaggactgggcttac	5' gaagggaagtctccctcgac
Target 2: internal primer set 4	596 bp	5' agggcacagacctgatgg	5' agaagccactggagcaccta
Target 2: internal primer set 5	562 bp	5' ccttggatgggattcaaagt	5' ttgatctctgccagcttca
Target 2: internal primer set 6	499 bp	5' aaatgaagcccagggaaac	5' tggggagtggacagaagaag
Target 2: internal primer set 7	567 bp	5' ggaggcagattgaggttga	5' cccacctagcctctgtaac

3.3.5. Functional annotation of *TNFRSF1A* intron 1

The finding of an association between the non-coding sequence variants necessitates the identification of functional elements within *TNFR1* intron 1. For this purpose, the complete intron 1 sequence of *TNFR1* was analysed for conserved non-coding sequences (CNS) using AltaVista Genome Browser. In addition, the Genome Atlas software was employed to analyse the primary sequences of two contrasting haplotypes, encompassing the complete intron 1 of *TNFR1*, to identify DNase hypersensitive sites and other functional elements (Hallin and Ussery, 2004).

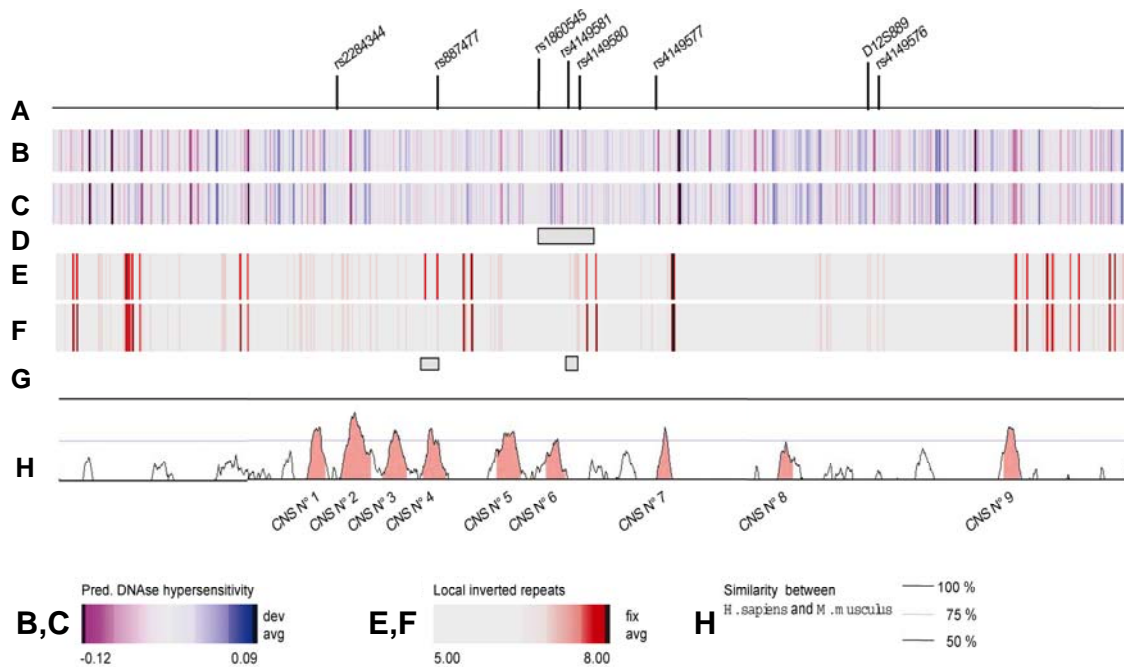


Fig. 17: Functional annotation of *TNFR1* intron 1. (A); the figure shows the entire 7.5kb of *TNFR1* intron 1 with allocated genetic markers. (B, C, E, F); Putative functional elements were visualised using GenomeAtlas software. Primary sequences for 8-marker haplotypes at rs2284344-rs887477-rs1860545-rs4149581-rs4149580-rs4149577-D12S889-rs4149576 were compared with respect to predicted DNase hypersensitive sites based on the trinucleotide model of Brukner et al. (195a, 1995b). Results for haplotype G-G-T-G-T-T-10-A are shown in (B) and for C-T-C-A-C-C-13-G are shown in (C). Comparing the two haplotypes, patterns for predicted DNase hypersensitive sites were different near rs1860545, rs4149581 and rs4149580 as indicated by gray boxes in (D). Primary sequences were monitored for local inverted repeats in (E) for haplotype G-G-T-G-T-T-10-A and (F) for haplotype C-T-C-A-C-C-13-G, whereby differences between the two haplotypes are indicated in gray boxes in (G) near rs887477 and rs4149581-rs4149580. (H): VISTA plot displaying the percentage identity between mouse and human in the *TNFR1* intron 1.

The AltaVista Genome Browser identified nine CNSs within intron 1 of *TNFR1* and interestingly SNP rs887477 was located on CNS 4 (Fig. 17). Similarly, altered patterns for predicted DNaseI hypersensitive sites were observed near rs1860545-rs4149581-rs4149580. Functional significances of these alterations due to polymorphisms can be explained by considering the DNA structural properties such as bendability of the DNA helix near DNase hypersensitive sites (Brukner et al., 1995a, b) or nucleosome positioning signals (Fitzgerald et al., 1994; Baldi et al., 1996), which are determined by the primary sequence. Furthermore, it is plausible that the two haplotypes within intron 1 might differ with respect to their accessibility of and responsiveness to transcriptional regulators. Alternatively or additionally, causative variants might act through the pre-mRNA as splicing requires a specific tertiary structure of the pre-mRNA and its interaction with RNA-binding proteins such as hnRNPs (hetero nuclear ribo-nucleo proteins). All these mechanisms may ultimately determine the transcriptional efficiency of *TNFR1* and thereby affecting the TNF α signalling.

3.3.6. Soluble TNFR1 levels in serum of CF patients are associated with D12S889 genotype

The haplotype mapping indicated that the marker D12S889 is strongly linked to the causative haplotype associated with the disease. Hence, the levels of sTNFR1 in serum of CF twin and siblings were determined by western blotting and compared against their genotype at D12S889 in order to validate the hypothesis that variants within intron 1 of *TNFR1* are causative variants. It is shown that the *TNFR1* exists both in 55kDa full length and 28kDa cleaved ectodomain forms and among these two fractions, 55-kDa full length form is the predominant fraction in human serum and lung epithelial lining fluid (Hawari et al., 2004). Accordingly, 55kDa fraction was found to be a major form in most of the samples tested. However, few samples showed higher levels of 28kDa cleaved *TNFR1* ectodomain form. Strikingly, both forms of the sTNFR1 levels were associated with D12S889 genotypes, in which individuals carrying genotype 13-13 were showing higher level of sTNFR1 compared to individuals carrying 10-10 genotypes (Fig. 18).

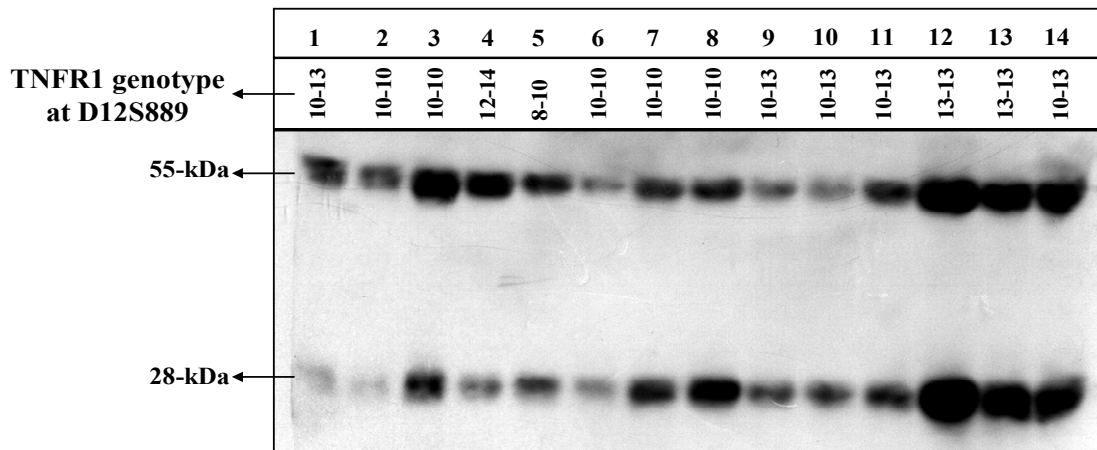


Fig. 18 Soluble TNFR1 levels in CF patients carrying different D12S889 genotypes. Homozygous genotypes 10-10 at D12S889 are shown in lanes 2, 3, 6 and 7. Homozygous genotypes 13-13 at D12S889 are shown in lanes 12 and 13. Samples in lanes 1, 2, 3, and 4 were from four severely affected (CON-) individuals; samples in lanes 5& 6, and 7&8, and 9&10 were from three mildly affected (CON+) sib pairs. Samples in lanes 11, 12, 13, and 14 were from four discordant (DIS) individuals.

Furthermore, 13-13 genotypes showed considerably higher levels of 28kDa form compared to all other genotypes. Thus, the D12S889 genotype was associated with both quantity of TNFR1 protein as well as the ratio of two fractions of TNFR1. The fact that the TNFR1 protein levels were associated with D12S889 genotype strongly implies that the sequence variants within the intron 1 haplotype are influencing transcriptional activity of *TNFR1*.

3.3.7. Impact of *TNFR1* defect on innate immunity

In this study, polymorphisms within intron 1 of *TNFR1* were associated with CF disease severity (Table 13). In-silico analysis of non-coding intron 1 sequence predicted several possibilities such as shift in DNaseI hypersensitivity sites and location of SNP within a conserved non-coding sequence may interfere with gene transcription (Fig. 17). This possibility was further strengthened by the finding that the levels of TNFR1 protein in CF patient serum were correlated with the intron 1 haplotype (Fig. 18). The levels of TNFR1 in circulation appear to be critical in regulating TNF α induced innate inflammatory responses. In general, the release of membrane bound TNFR1 from cells into circulation involves two distinct pathways. First pathway, by the proteolytic cleavage and shedding of TNFR1 ectodomain by the receptor sheddases (Wajant et al., 2003) and second pathway is by release of full length, 55-kDa TNFR1 within the membranes of exosome-like vesicles (Hawari et al., 2004). Furthermore, shedding of

TNFR1 was shown to be an important process in immune response as TNFR1 shedding may attenuate TNF α activity either by rapidly down regulating the cell surface expression of TNF receptors or by acting as competitive antagonists for soluble TNF α and thereby reducing the sensitivity of a target cell to soluble TNF α (Aderka, 1996). Additionally, rapid shedding of TNFR1 from endothelial cells after histamine stimulation rendered endothelial cells refractory to the proinflammatory effects of TNF α (Wang et al., 2003). The role of TNFR1 shedding in innate immune response was further strengthened by the study in mutant mice expressing nonshedddable TNFR1 in which defective TNFR1 shedding caused innate immune hyperresponsiveness and increased susceptibility to chronic inflammatory conditions (Xanthoulea et al., 2004).

In CF scenario, TNF α plays a significant role in mediating innate immune response against opportunistic bacterial pathogens that colonise the CF airways. Upon binding of TNF α , TNFR1 signaling triggers several intracellular signaling pathways which control gene expression through NF-kB and AP-1 transcription factors. However, these actions are beneficial for the healthy immuno-competent host, but in case of CF they reinforce chronic airway inflammation: The bacterial pathogens evade the host response and sustain a vicious cycle of fooled self-destructive host defense characterised by a pronounced imbalance between pro-and anti-inflammatory cytokines (Tümmler and Kiewitz, 1999). Furthermore, the accumulation of misfolded F508del-CFTR in ER (endoplasmic reticulum) may cause the proinflammatory ER overload response that ultimately results in activation of NF-kB (Knorre et al., 2002). Thus, the transcription factor NF-kB that plays an important integrating role in the intracellular regulation of immune response and inflammation, is induced in F508del-*CFTR* homozygotes by two major stimuli, the mutant CFTR protein and the TNF α signaling cascade. Taken together, the variability in TNFR1 protein levels may cause imbalance in the ratio of TNFR1 to TNF α and thus affecting the major pro-inflammatory signaling pathway. Hence, it is of vital importance to investigate how the polymorphisms within intron 1 of TNFR1 determine the levels of both 55kDa and 28kDa forms (Fig. 7) and its consequences on CF pathology.

3.4. Analysis of TLR (Toll like receptor)-4 as a modulator of cystic fibrosis

TLR4 recognizes gram negative bacterial LPS and transduces the signals necessary for production of inflammatory mediators (Beutler, 2000, Lien et al., 2000). Additionally, TLR4 increases the efficiency of LPS recognition by forming complex with CD14, MD-2 and LBP. *P. aeruginosa* is known to synthesize hexa-acylated LPS instead of penta-acylated LPS during adaptation to the CF airway. Interestingly, human TLR4 has been shown to specifically recognize hexa-acylated LPS and induce robust proinflammatory signals (Hajjar et al., 2002).

TLR4, chromosome 9q32-q33

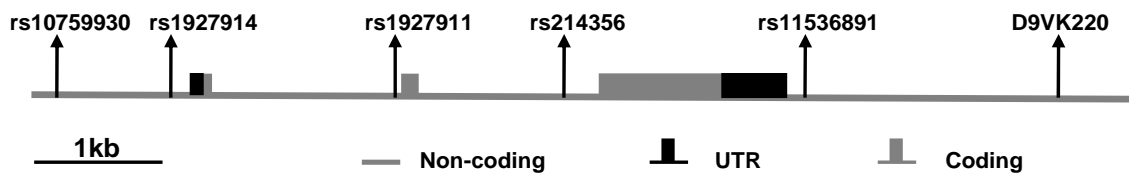


Figure 19: Schematic representation of *TLR4* gene structure; *TLR4*, located on chromosome 9q32-q33, is 3.95kb long and has three exons. Polymorphisms shown on the *TLR4* gene are targeted for fine mapping. SNPs are indicated by rs followed by their unique numbers and D9VK220 is a microsatellite marker. (Location and size are based on NCBI, NC_000009.10 Reference assembly: Build 36.2)

3.4.1. Association with disease severity in CF on *TLR4* revealed by single-marker analysis

In order to investigate *TLR4* as a modulator in CF, the locus was analysed by typing 5 single nucleotide polymorphisms and one CA₍₂₂₎ repeat polymorphism, D9VK220, on CF twin and siblings (Fig. 19). Case-control analysis was done by comparing the allele distribution in both mildly (CON+) and severely (CON-) affected patient pairs as well as discordant (DIS) with concordant (CONC) pairs. Allele and genotype distributions at all loci were comparable between discordant and concordant pairs. Interestingly, the comparison revealed an allelic imbalance between mildly and severely affected patient pairs at D9VK220 (P = 0.008), rs1927914 (P = 0.03), rs1927911 (P = 0.03) and at rs214356 (P = 0.05). However, at rs10759930 (P = 1.0) and rs11536891 (P = 1.0) allele distributions were unaltered. This finding indicated that *TLR4* locus is a modulator of CF disease severity. Hence, to detect the exact causative variant, haplotype distribution was evaluated among CF twin and siblings.

3.4.2. Association with disease severity in CF on *TLR4* confirmed by haplotype analysis

Two-marker haplotypes for adjacent markers were constructed step by step. Distribution of each two-marker haplotype block was compared in both mildly and severely affected patient pairs as well as discordant and concordant pairs.

Fig. 20 Haplotype block rs10759930-rs1927914 on *TLR4* is associated with CF disease severity among CF twin and siblings.

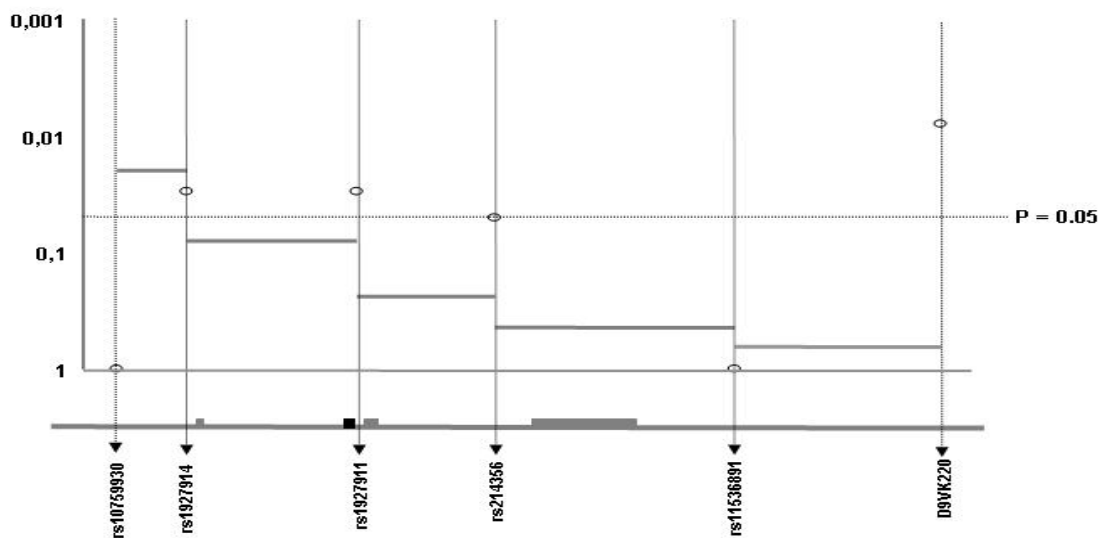


Fig. 20 Case-control analysis comparing mildly (CON+) and severely (CON-) affected patient pairs. P-values are shown for single markers (open circles) and haplotypes of adjacent markers (lines). The x-axis depicts the physical distance between the markers and the y-axis denotes the P-value. P = 0.02 was observed for the haplotype block rs10759930-rs1927914

Haplotype distributions between DIS and CONC pairs at all loci were unaltered while, as it is shown in figure 20, the distribution of haplotype block rs10759930-rs1927914, located in the *TLR4* promoter, was highly dissimilar between mildly and severely affected patient pairs ($P = 0.02$), in which haplotype C-T was overrepresented among mildly affected pairs (75%) compared to 38% in severely affected pairs. The haplotype analysis also revealed that rs10759930 and rs1927914 are linked without any recombination events between them as only three haplotype combinations were observed (C-T, C-C and T-T). This observation indicates that the haplotype may extend upstream of rs10759930. Thus, to identify sequence variants within this haplotype,

genomic DNA from individuals harbouring all the three observed haplotypes at rs10759930- rs1927914, were sequenced.

3.4.3. Sequence analysis of *TLR4* gene

The 500bp promoter region, exon 1, exon 2 with its complete 5' UTR and flanking intron sequences were analysed by sequencing eight unrelated F508del homozygotes. Among eight individuals, four individuals were carrying C-T haplotype, 3 individuals were carrying C-C haplotype and one individual was carrying T-T haplotype at rs10759930-rs1927914. Sequencing analysis did not reveal any novel sequence variants. However, previously reported polymorphisms in the 5' UTR were confirmed (Table 16). As the haplotype analysis revealed that the causative variant(s) is/are located upstream of rs1927914 and the 5' UTR is downstream of rs1927914 (Fig. 19), the polymorphisms in the 5' UTR were not evaluated for association.

Table 16: *TLR4* sequencing confirmed previously reported polymorphisms

Region	NCBI SNP ID
5' UTR	rs760361
5' UTR	rs2727192
5' UTR	rs760363
5' UTR	rs1413088
5' UTR	rs10759933

3.4.4. *TLR4* polymorphism is not associated with *P. aeruginosa* early or late chronic colonisation among unrelated CF patients

Both single-marker analysis as well as haplotype analysis showed that rs1927914 was significantly associated with CF disease severity among CF twin and siblings. Thus, rs1927914 was selected to evaluate *TLR4* locus and its association with *P. aeruginosa* chronic colonisation. The cohort containing 22 unrelated F508del homozygous CF patients, stratified for early and late *P. aeruginosa* chronic colonisation, was typed at rs1927914. Allele and genotype distributions at rs1927914 were compared between 13 patients belonging to *P. aeruginosa* early colonised and nine patients of *P. aeruginosa* late colonised group. The allele ($P = 1.0$) and genotype ($P = 0.36$) distributions in both the groups were similar (Fig. 21) and no association was observed between rs1927914 SNP and early or late colonisation of *P. aeruginosa*.

Fig. 21 Comparison of allele and genotype distribution at rs1927914 between *P. aeruginosa* early and *P. aeruginosa* late colonised groups

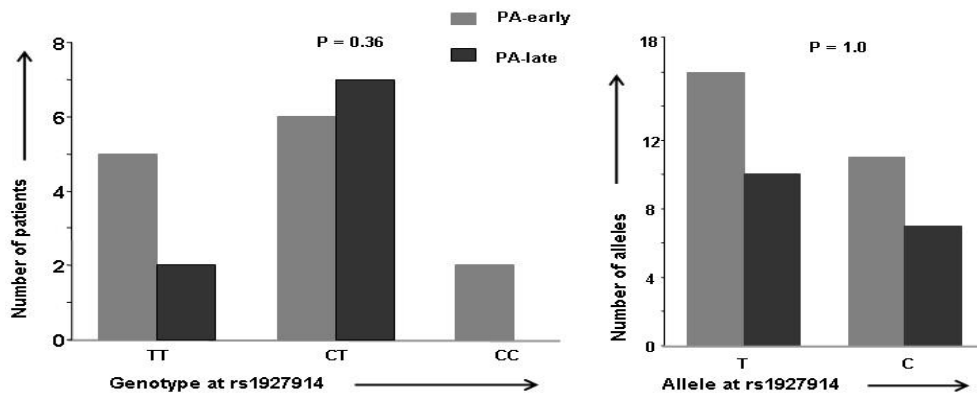


Figure 21: C to T SNP rs1927914 genotype and alleles are denoted on x-axis. PA-early: Patients chronically colonised within seven years of their age. PA-late: Patients chronically colonised after 15 years of their age. No significant association was found

3.4.5. TLR4 expression analysis by FACS on peripheral blood mononuclear cells

In an effort to investigate the role of rs1927914 polymorphism in TLR4 expression, peripheral blood mononuclear cells from 10 healthy volunteers were collected and surface expression level of TLR4 was studied by FACS. The expression levels were compared with genotype at rs1927914. Six individuals were homozygous for allele T, three individuals were heterozygous and only one individual was homozygous for allele C at rs1927914.

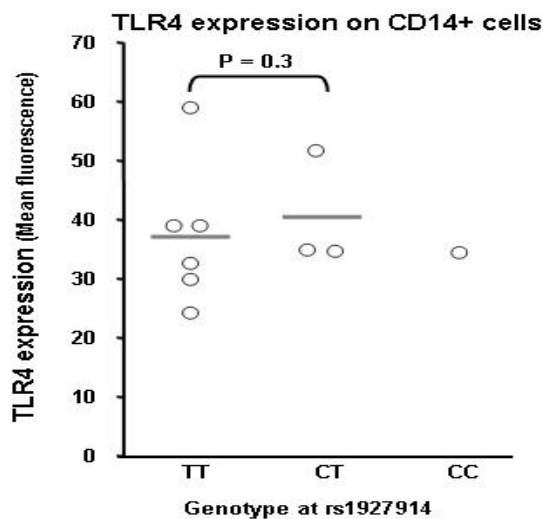


Figure 22: TLR4 surface expression level vs. rs1927914 genotypes; P-values are derived by comparing CC and CT groups by Mann-Whitney U test. $P \leq 0.05$ was considered significant. These data are representative of two independent experiments.

The comparison of TLR4 surface expression on CD14+ cells (monocytes) was done between TT and CT individuals only. No association was observed between genotypes at rs1927914 and TLR4 surface expression level ($P = 0.3$; Fig. 22), although there was a small trend of higher expression among heterozygotes.

3.4.6. No informative markers within three kb region upstream of rs10759930

As haplotype block rs10759930-rs1927914 indicated that the causative haplotype may extend upstream of rs10759930, six SNPs (Table 17), which are reported in the NCBI database, were typed on nine control samples to analyse the information content. However, none of the tested polymorphisms can be employed to construct the haplotype in CF twin and siblings due to their low information content.

Table 17: Polymorphisms with low polymorphism information content on *TLR4* locus upstream of rs10759930

SNP NCBI ID	PIC [†]	Position (bp) [‡]
rs7875169	0.11	119500162
rs7855597	0.19	119490033
rs10983754	0.11	119498428
rs10983753	0.24	119495836
rs7873159	-	119490834
rs4837495	0.11	119490586

[†]*Polymorphism information content determined by accounting allele frequencies from nine unrelated controls*

[‡]*Position of the polymorphisms on chromosome 9q32-q33*

3.4.7. Role of TLR4 signaling in CF airways

The innate immune system is of vital importance to the early control of infection in CF. TLR4, the pathogen recognition receptor, plays an important role in signaling the innate immune response against bacterial infection by recognising LPS. Recognition of LPS occurs by the CD14/MD-2/TLR4 complex which is present on many cell types including macrophages and dendritic cells (Shimazu et al., 1999; Hoshino et al., 1999). Multiple signal transduction pathways are activated by this complex in macrophages upon LPS stimulation, including the activation of transcription factor NF- κ B and increased transcription of pro-inflammatory cytokines such as IL-1 β , TNF α , neutrophil recruiting chemokines, and metalloproteinases (Medzhitov, 2001). Furthermore, it has been shown that human TLR4 recognised the hexa-acylated lipid A of *P. aeruginosa*,

synthesised during adaptation to the CF lung, and activated NF- κ B 100 times greater than with the penta-acylated lipid A (Hajjar et al., 2002).

Several DNA polymorphisms in genes linked with the inflammatory response have been associated with an enhanced or suppressed inflammatory response (Casanova and Abel, 2005). *TLR4* coding variants, D299G and T399I, are shown to be associated with attenuated response to inhaled LPS in humans (Arbour et al., 2000). Translating these findings to CF, Urquhart et al. (2006) analysed *TLR4* D299G polymorphism in 100 CF children stratified for clinical phenotype and reported *TLR4* as a non-modulator in CF since they found no association with *TLR4* D299G polymorphism. By contrast, in our CF twin and sibling cohort, the haplotype analysis and the sequencing clearly demonstrated that the CF modifying variant(s) is/are located within the haplotype fragment on *TLR4* promoter or further upstream (Fig. 20), and not any coding variants.

It is plausible that sequence variants in promoter regions would cause differential gene transcription, mostly mediated by modification of transcription factor binding sites. However, sequencing of 500bp *TLR4* promoter region did not detect any sequence variants. On the other hand, since the haplotype block involved in CF disease modulation is not completely mapped, it is possible that the causative variants may be located very distant towards 5' region of *TLR4*. Hence, it is necessary to fine map the region to identify the true causative variants. Furthermore, analysis of *TLR4* promoter polymorphism genotype and allele distributions in CF patients stratified for *P. aeruginosa* early or late colonization did not show any association (Fig. 21). This would imply that TLR4 signaling is not directly determining the age at onset of *P. aeruginosa* chronic colonization even though it is capable of inducing effective NF- κ B response against hexa-acylated lipid A of *P. aeruginosa* (Hajjar et al., 2002). On the other hand, this overwhelming TLR4 signaling may induce continued neutrophil infiltration in the airways. Consistent with this notion, it is shown that the TLR4 activation could cause down-regulation of the G protein receptor kinases, which are involved in chemokine receptor desensitisation and consequently, enhance neutrophil migration (Fan and Malik, 2003). Thus in CF airways, these neutrophils are not able to clear bacterial infections and instead further damage the airway and lung tissue (Döring and Worlitzsch, 2000). In this regard, it is extremely important to find out the *TLR4* promoter variants, responsible for CF disease modulation.

pairs (CON-), discordant pairs (DIS) and concordant pairs (CONC) were unaltered at both the loci (Table 18 and 19). However, a minor allelic imbalance between CONC and DIS pairs was observed at CD14 3' UTR polymorphism, rs2563298 (P = 0.08).

Table 18: Comparison of genotype and allele distribution among F508del homozygous CF twin and siblings in CD14 at promoter polymorphism rs2569190

Genotypes	CON-	CON+	DIS	CONC*
CC	5	8	14	13
CT	10	14	22	24
TT	5	6	6	11
Total	20	28	42	48
Alleles				
C	20	30	50	50
T	20	28	34	46
Total	40	58	84	96
Freq. of allele C	0.50	0.52	0.60	0.52
Freq. of allele T	0.50	0.48	0.40	0.48

* Both CON- and CON+ pairs are grouped as CONC, concordant pairs, for disease severity

Table 19: Comparison of genotype and allele distribution among F508del homozygous CF twin and siblings in CD14 at 3' UTR polymorphism rs2563298

Genotypes	CON-	CON+	DIS	CONC*
TT	0	1	4	1
TG	7	11	19	18
GG	13	12	15	25
Total	20	24	38	44
Alleles				
T	7	13	27	20
G	33	35	49	68
Total	40	48	76	88
Freq. of allele T	0.18	0.27	0.35	0.22
Freq. of allele G	0.82	0.73	0.64	0.78

* Both CON- and CON+ pairs are grouped as CONC, concordant pairs, for disease severity

3.5.2. Association with disease discordance in CF on CD14 revealed by haplotype analysis

To confirm the association found between CONC and DIS pairs and also to find out the exact causative variant, the two-marker haplotype (rs2569190-rs2563298) was constructed on CD14 and its distribution was analysed both in mildly and severely affected pairs as well as DIS and CONC pairs. In striking contrast to single-marker analysis, a highly significant distortion in rs2569190-rs2563298 haplotype distribution

was observed between CONC and DIS pairs ($P=0.007$), while the distribution of haplotype was unaltered between CON+ and CON- (Table 20). This finding indicated that the locus may have an impact on CF disease by modulating the binding efficiency of factors encoded elsewhere in the genome. Thus, the complete rs2569190-rs2563298 haplotype fragment was screened for other sequence variants, if any, by sequencing.

Table 20: *CD14* two-marker haplotype (rs2569190-rs2563298) frequency distribution among CF twin and siblings

<i>CD14</i> rs2569190-rs2563298 haplotype [†]	CON-	CON+	‡DIS	‡CONC*
C-T	0.16	0.18	0.39	0.17
C-G	0.28	0.27	0.30	0.28
T-G	0.54	0.54	0.30	0.54

‡CONC and DIS comparison, $P = 0.007$

* Both CON- and CON+ pairs are grouped as CONC, concordant pairs, for disease severity

† Haplotype T-T was not observed among CF twin and siblings

(The haplotype counts were obtained by using likelihood-weighted haplotype explanations for each individual – kindly performed by T. Becker, Bonn.)

3.5.3. *CD14* sequencing analysis

2600bp fragment, encompassing the complete *CD14* gene with its 678bp of promoter from major transcriptional start site, 5'UTR, exon 1, intron 1, exon 2 and 3' UTR, was sequenced by selecting 2 unrelated CF F508del homozygous patients, where one individual was carrying C-G haplotype and the other was carrying T-C haplotype at rs2569190-rs2563298. Sequencing of complete *CD14* gene did not reveal any novel sequence variants. However, the results confirmed C-159T promoter polymorphism and rs2563298. Thus, sequencing confirmed that there is no recombination between rs2569190 and rs2563298 and both markers in strong linkage. As preferential transmission of certain alleles from parents to CF offspring point towards a modifying locus, both polymorphisms on *CD14* locus were analysed for transmission-disequilibrium to find out the mechanism.

3.5.4. Skewed allele distribution on *CD14* among CF siblings

All CF sib pair families, irrespective of their disease status were analyzed with the Monte Carlo simulation based association test (Becker and Knapp, 2004) for

investigating transmission-disequilibrium at *CD14* polymorphisms. Family-based analysis showed evidence for disproportionate transmission of the haplotype block rs2569190-rs2563298, encompassing the complete *CD14* gene, in which the haplotype C-T at rs2569190-rs2563298 was overrepresented in CF offspring while haplotype T-G was observed more frequently among non-transmitted parental haplotypes. ($P = 0.07$; Table 21). However, the transmission distortion was highly significant at rs2563298 ($P = 0.019$), While transmission at rs2569190 was unaltered ($P = 0.39$). Furthermore, T-T haplotype was not observed in any of the individuals studied, implying that both the markers are in absolute linkage without any recombination between them. All other haplotypes (C-T, C-G and T-G) were transmitted equally in both CON- and CON+ pairs (Table 20).

Table 21: Frequency distribution of rs2569190-rs2563298 haplotype on CD14 among CF offspring

CD14 rs2569190-rs2563298 haplotype	‡Transmitted	‡Non- transmitted
C-T	0.38	0.17
C-G	0.30	0.40
T-G	0.32	0.43

‡comparison for transmitted and non transmitted haplotypes, $P = 0.07$; (Comparison of at single locus $P = 0.39$ for rs2569190 and $P = 0.019$ at rs2563298)
(The haplotype counts were obtained by using likelihood-weighted haplotype explanations for each individual – kindly performed by T. Becker, Bonn.)

3.5.6. CF twin and siblings cohort stratified for *P. aeruginosa* colonisation

CD14 promoter and 3' UTR polymorphisms were analysed in CF twin and sibling cohort stratified for *P. aeruginosa* colonisation. For this purpose, all CF twin and sib pairs were categorised as *P. aeruginosa* colonised (PA-yes) and *P. aeruginosa* non-colonised (PA-nil). Monozygous and dizygous pairs were analysed separately. As genotypes of the individuals in each sib pair were interdependent, the transmission of genotypes in combination within each sib pair was considered for the analysis. Parental genotype frequencies were derived from their allele frequencies and expected genotype combinations at both the loci were calculated according to HWE law.

Strikingly, the distribution of genotype combinations among dizygous pairs was significantly different from the distribution of expected genotype combinations among parents at rs2569190 ($P = 0.02$), in which the genotype combination CC/CC (both sibs

carrying CC genotype) was over represented and TT/CT combination (one sib carrying TT and the other sib carrying CT genotype) was under represented among dizygous pairs. The distortion of the genotype distribution from the expectancy values was strongest in the subgroup of siblings who were both colonized with *P. aeruginosa* ($P = 0.11$; Table 22). Similarly, the distribution of genotype combinations among dizygous pairs at rs2563298 was different (Table 23) but not statistically significant ($P = 0.11$).

Table 22: *CD14* genotype at rs2569190 and *P. aeruginosa* status of F508del homozygous twins and sib pairs recruited for determination of *P. aeruginosa* colonization status.

Genotype of sib or twin A Genotype of sib or twin B	CC CC	CT CC	TT CC	CT CT	TT CT	TT TT	*P-value
Dizygous/YY (23)	7	6	0	7	1	2	0.11
Dizygous/NN (12)	1	3	0	6	2	0	0.77
Dizygous/YN(6)	1	1	0	3	0	1	0.77
All dizygous (41)	9	10	0	16	3	3	0.02
Monozygous/YY (9)	1	0	0	4	0	4	0.72
Monozygous/NN (3)	1	0	0	2	0	0	1.0
All monozygous (12)	2	0	0	6	0	4	1.0

Table 23: *CD14* genotype at rs2563298 and *P. aeruginosa* status of F508del homozygous twins and sib pairs recruited for determination of *P. aeruginosa* colonization status.

Genotype of sib or twin A Genotype of sib or twin B	TT TT	TG TT	GG TT	TG TG	GG TG	GG GG	*P-value
Dizygous/YY (22)	1	2	0	8	7	4	0.36
Dizygous/NN (10)	2	0	0	3	3	2	0.50
Dizygous/YN(6)	0	1	0	13	0	3	0.74
All dizygous (38)	3	3	0	10	10	9	0.11
Monozygous/YY (9)	0	0	0	3	0	6	0.72
Monozygous/NN (3)	0	0	0	3	0	0	1.0
All monozygous (12)	0	0	0	6	0	6	0.42

Table 22 and 23: Data are represented for monozygous and dizygous pairs separately. YY; both sibs are colonized in a pair. NN; both sibs are not colonized in a pair. YN; one sib is colonised and the other one in not in a pair. * The parental genotype frequencies were derived from their allele frequencies and the expectancy frequencies of 6 combinations of genotypes were calculated according to Hardy-Weinberg equilibrium (HWE) law. From these parental genotype expectancy frequencies, the expected numbers of 6 combinations of genotypes among twin and sibling pairs (for each pair-sub groups from the above table) were derived and compared with observed numbers of 6 combinations of genotypes among pairs.

In contrast, the distributions in monozygous pairs were similar to the expected distribution at both the loci. As this finding suggested that the *CD14* locus is involved in differential susceptibility to *P. aeruginosa* among CF twin and siblings, both

polymorphisms were analysed in an independent cohort stratified for age at onset of *P. aeruginosa* chronic colonisation.

3.5.7. F508del homozygous unrelated CF patients with early and late *P. aeruginosa* chronic colonisation

Comparison of rs2569190 genotype distribution among unrelated CF patient cohort, displaying extreme phenotype for the onset of *P. aeruginosa* chronic colonisation (Chapter 2.1.3), showed significant association ($P = 0.008$; Fig. 24a) in which heterozygotes are underrepresented in group who were chronically colonised before their seven years of age. Distortion of genotype frequencies in early chronically colonised group from Hardy-Weinberg equilibrium (HWE) was also seen in the De Finetti diagram (Fig. 25) while no distortion from HWE was observed among the patients colonised at an age of 15 years or later. In contrast, both allele and genotype distributions were similar at rs2563298 (Fig. 24b).

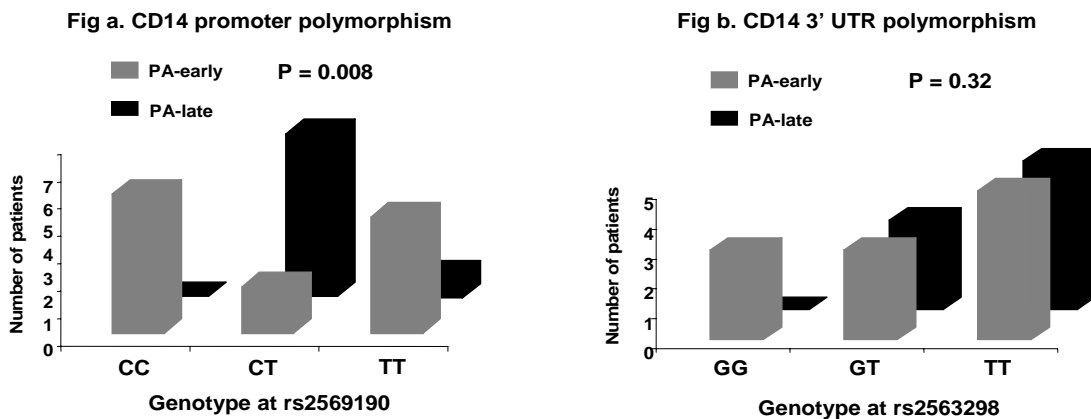


Fig. 24 CD14 promoter polymorphism is associated with *P. aeruginosa* early or late colonization. Genotypes at CD14 polymorphisms are compared against PA-early (chronically colonized with in seven years of age) and PA-late (chronically colonized after 15 years of age).

3.5.8. F508del homozygous unrelated CF patients stratified for birth cohorts

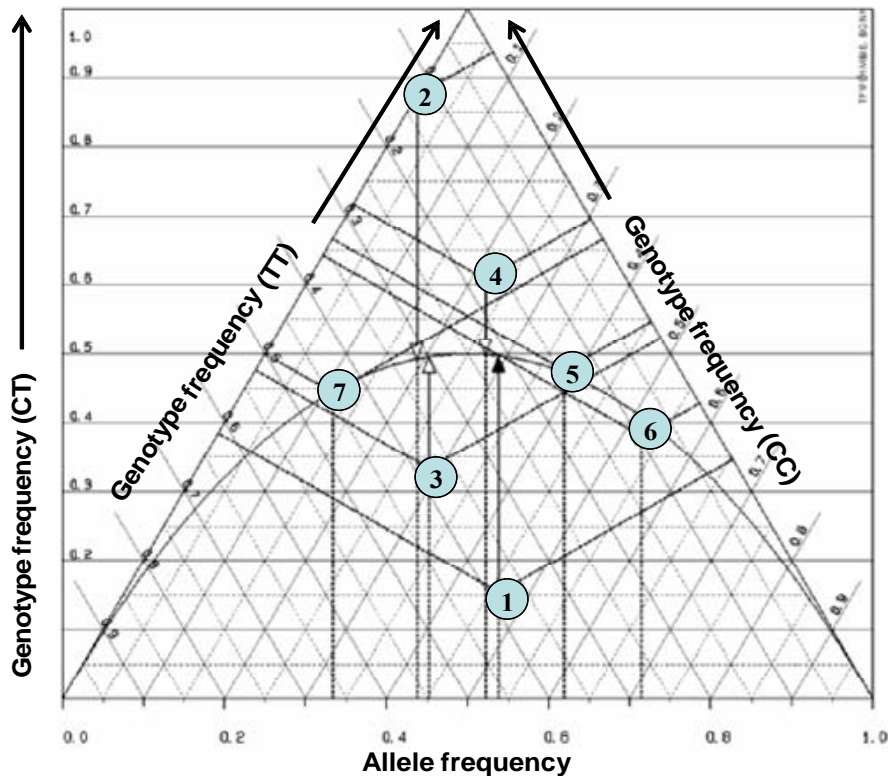
In order to test the hypothesis that the changing environmental factors and improvement of survival have an impact on selection of favourable alleles and genotypes among CF population, we have genotyped 22 patients who were born in 1959 to 1967 from Early-born group and 23 patients who were born in 1973 to 1975 from Late-born group (Chapter 2.1.4).

Table 24: Genotype distributions of all the cohorts tested on *CD14* promoter polymorphism

	1	2	3	4	5	6	7
<i>CD14</i> rs2569190	PA-Early [†]	PA-late [‡]	Birth cohort-Early born ^a	Birth cohort-Late born ^b	Dizygous first born sib (PA+)	Dizygous second born sib (PA+)	Monozygous Pairs (PA+)
CC	6	0	6	5	8	11	1
CT	2	7	7	14	10	8	4
TT	5	1	8	4	3	2	4

[†]Colonized by *P. aeruginosa* before 7 years of their age; [‡]Colonized by *P. aeruginosa* after 15 years of their age; ^aEarly-born group: Individuals born between 1959 - 1967
^bLate-born group: Individuals born between 1973 -1975; (PA+): Positive for *P. aeruginosa* colonisation

Figure 25: De Finetti Diagram with Hardy-Weinberg equilibrium (HWE) parabola for patient panels investigated on *CD14* promoter polymorphism



*Fig. 25 Base of the triangle denotes the allele frequencies and 2-sides of a triangle and y-axis denote genotype frequencies at rs2569190. The parabolic curve shows HWE, the distance between a circle and the curve corresponds to the deviation from HWE. 1: PA-early (n = 13), 2: PA-late (n = 8), 3: Birth cohort-early born (n = 21), 4: Birth cohort-late born (n = 23), 5: Dizygous first born sibs positive for *P. aeruginosa* (n = 21), 6: Dizygous second born sibs positive for *P. aeruginosa*, 7: Monozygous sibs positive for *P. aeruginosa*.*

Thus, as visualized in figure 25, patients who were chronically colonized in their early age are not in HWE ($P = 0.02$) indicating the loss of heterozygotes at rs2569190. Further distortion from HWE was seen for late colonized group ($P = 0.13$; Fig. 25, group 2). Both monozygous and dizygous pairs were in HWE. No difference in genotype distribution was observed between *P. aeruginosa* positive first born dizygous sib and second born dizygous sib implying that there was no influence of age difference within the dizygous sibs. However, comparison of allele distribution of monozygous pairs to either *P. aeruginosa* positive dizygous first born sibs ($P = 0.051$) or second born sibs ($P = 0.008$), both showed a significant difference. In addition, the distribution of genotypes among patients born during 1959 to 1967 was similar to genotype distribution in *P. aeruginosa* early chronically colonized group, sharing an underrepresentation of heterozygotes in comparison to patients born during 1973 to 1975. This finding, however, highlighted the concept that heterozygotes at *CD14* promoter polymorphism are selected as a favorable genotype which leads to an accumulation of carriers of heterozygous genotypes at rs2569190. To validate this finding, a cohort containing CF twin and siblings, chronically colonized with *P. aeruginosa* was genotyped at *CD14* polymorphisms.

3.5.9. Age dependent risk to acquire *P. aeruginosa* among CF twin and siblings shown by *CD14* diplotype analysis

The *CD14* 3' UTR was associated with CF disease discordance but not with *P. aeruginosa* early or late colonisation phenotype. In contrast, the *CD14* promoter polymorphism was associated with *P. aeruginosa* early or late colonisation but not with disease discordance. Furthermore, the haplotype information strongly suggested that both the markers are in linkage disequilibrium and associated with disease discordance and hence these two markers might interactively modulate the phenotype.

Therefore, interactive effect of these two markers were analysed by constructing diplotypes (combination of haplotypes). As a diplotype can be homozygous or heterozygous, there are sixteen possible diplotype pairs. However, because of linkage disequilibrium between the two *CD14* markers, only six diplotypes were observed. The distribution of diplotypes from CF twin and siblings were compared against age at onset of *P. aeruginosa* chronic colonisation (Fig. 26). Interestingly, diplotype CG-CG was associated with early acquisition of *P. aeruginosa* chronic colonisation, as two pairs

homozygous for CG were colonised at age 10 years or earlier (pairs 17 & 18 in Fig. 26). In contrast, diplotype CT-CT was associated with late acquisition of *P. aeruginosa* chronic colonisation as one pair (pair 10 in Fig. 26) and two unrelated patients (number 21 & 22 in Fig. 26) were colonised at age 15 years or later. Thus, diplotype analysis suggests that diplotype CG-CG was a risk factor for early onset of *P. aeruginosa* chronic colonisation, while diplotype CT-CT was found to be beneficial. Sibs carrying both risk as well as beneficial diplotype, CT-CG, clustered as early colonised patients indicating that the risk variant CG determines the phenotype in dominant fashion (pairs 1, 11, 12, 13, and unrelated patients 23, 25, 31, 32 in Fig. 26). Heterozygotes of the mild CT and the non-classified TG alleles show later onset of colonisation (Fig. 26) showing that CT is dominant over TG. Homozygous for haplotype TG were observed predominantly among monozygote twins as 4 out of 9 pairs carried diplotype TG-TG while only two out of 11 dizygous pairs with shared haplotype were TG homozygotes.

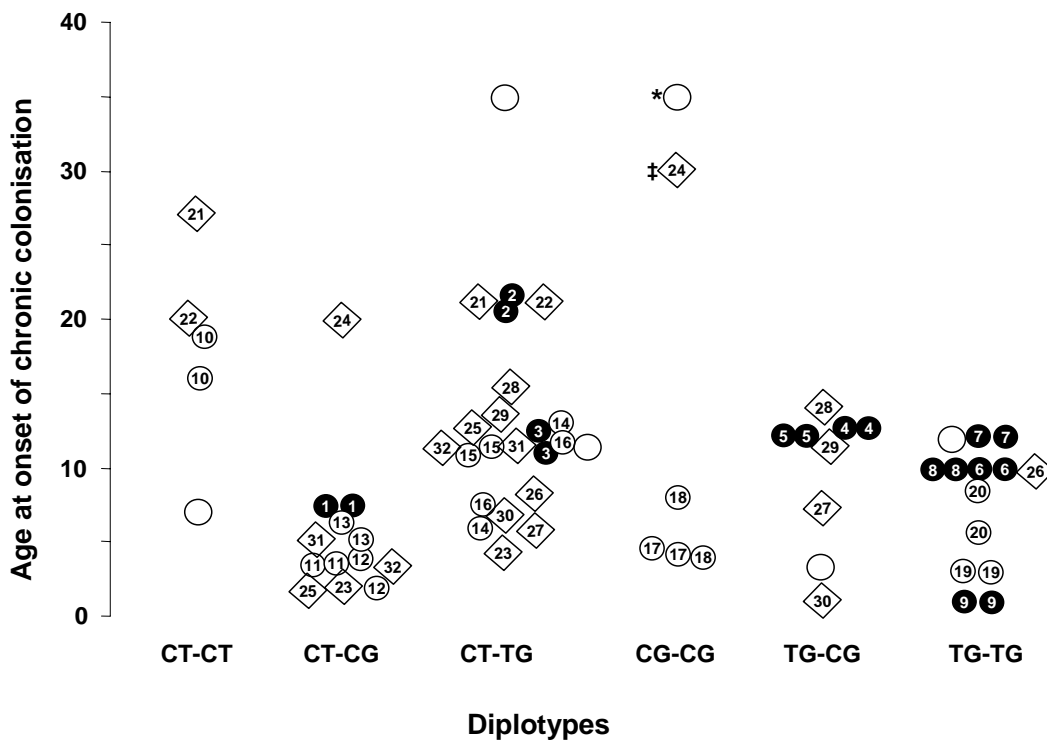


Figure 26: Age dependent risk to acquire *P. aeruginosa* chronic colonisation among CF twin and siblings is determined by diplotype at CD14 polymorphisms; Monozygous pairs are numbered from 1 to 9 in black circles; Dizygous pairs concordant for diplotype-pair are numbered from 10 to 20 in white circles; Dizygous pairs discordant for diplotype-pair are numbered from 21 to 32 in blank diamonds; Dizygous sibs discordant for *P. aeruginosa* colonisation (one sib in a pair is colonised) are shown in blank circles. * Patient had unusual clinical history with excessive humoral immune response. ‡Discordant for clinical phenotype.

3.5.10. Association between *P. aeruginosa* O-antigen serotype and *CD14* diplotype

P. aeruginosa isolates become non-typeable or poly-typeable, once they chronically colonise the CF lung, due to lack of or variability in the number of O-antigen polymer. The *CD14* has been described as the receptor for complexes of LPS with lipid binding protein. Further, it has been shown that the LPS molecules containing typical O-antigen, long-chain carbohydrates, and two phosphorylated Kdo moieties, signals via *CD14*-dependent pathway since the affinity of *CD14* for these LPS structures is very high (Gangloff SC et al., 1999). In an effort to address the hypothesis that the O-antigen phenotype among CF *P. aeruginosa* isolates is dependent on the *CD14* genotype, 60 *P. aeruginosa* strains, isolated from CF twin and siblings, were serotyped for their O-antigen. The number of typeable (O-antigen intact/smooth LPS) and non-typeable (O-antigen deficient and poly-agglutinable/rough LPS) strains was compared against the *CD14* diplotypes among CF monozygous and dizygous twin and siblings (Table 25).

Table 25: Distribution of typeable and non-typeable *P. aeruginosa* isolates among CF twin and siblings according to *CD14* diplotype pairs

Monozygous pairs							
	N (%)	CT-CT	CT-CG	CT-TG	CG-CG	TG-CG	TG-TG
NT [†]	14 (87%)	-	2	4	-	2	6
T [‡]	2 (12%)	-	-	-	-	2	-
Dizygous pairs							
NT [†]	33 (75%)	4	8	10	6	2	3
T [‡]	11 (25%)	1	-	5	-	2	3

[†] Number of non-typeable strains (both poly-typeable and non-typeable strains) including O-antigen deficient and poly-agglutinable strains

[‡] Number of typeable strains (O-antigen intact)

The proportion of non-typeable strains was higher among monozygous pairs (14 strains out of 16) compared to dizygous pairs (Table 25), suggesting that the *P. aeruginosa* strains undergo faster adaptability in monozygous pairs due to identical host genetic background. Furthermore, patients carrying diplotype CG-CG and CT-CG were found to harbour only non-typeable strains, while patients carrying other diplotypes were colonised with both typeable and non-typeable strains. Interestingly, CG-CG diplotype was associated with early chronic colonisation of *P. aeruginosa* among CF twin and siblings (Fig. 26) and the same diplotype was found to be prominently susceptible for faster adaptability of *P. aeruginosa* isolates to the CF lung. Diplotype association with *P. aeruginosa* early or late colonisation and O-antigen serotype of *P. aeruginosa*

suggests that both promoter polymorphism and 3' UTR polymorphism on *CD14* might interactively affect the phenotype.

3.5.11. Influence of *CD14* diplotype on the sCD14 levels in serum of CF patients

In order to test whether *CD14* diplotypes determine CD14 protein levels, ELISA was performed to quantify the sCD14 levels in serum of F508del homozygous CF twin and siblings. Serum samples from 47 CF individuals, who were carrying homozygous *CD14* diplotypes (CT-CT, CG-CG, and TG-TG), were available. AbProtF (Antibody against *P. aeruginosa* outer membrane protein F) values measured by ELISA in 1990, as a measure of *P. aeruginosa* infection, were available for at least one time point from all these 47 patients. 34 serum samples were available from 24 CF patients, in which 8 patients had two serum samples collected at two different time points and one patient had three samples collected at three different time points. The patient's age varied from 8.9 to 26.7 years.

Table 26: Age categories of 24 CF patients (34 serum samples) to test the role of *CD14* diplotypes in determining age-dependent sCD14 levels

Age group	Number of patients	Mean age (Years)	Minimum age (Years)	Maximum age (Years)	Mean AbProtF values†
I	12	10.4	8.9	11.9	8.3
II	10	15.0	13.6	16.4	11.6
III	12	22.5	19.9	26.7	16.9

† Antibody titre against *P. aeruginosa* outer membrane protein-F was measured by ELISA (kindly performed by Peter Kubesch and Jutta Boßhammer)

As the risk to acquire *P. aeruginosa* increase with patient's age, these 24 patients were grouped into three different age periods (Table 26) and the sCD14 levels were compared against *CD14* diplotype to test whether levels of sCD14 is also determined by *CD14* diplotype in an age-dependent manner (Table 27). Mean AbProtF values were found to be positively correlated with age of patients and this association was found to be prominent among patients carrying CT-CT diplotype (Table 27).

Table 27: The soluble CD14 levels for different homozygous *CD14* diplotypes

Age Group	CD14 diplotype	Number of Patients	Mean sCD14 ($\mu\text{g/ml}$)	Mean AbProtF values
I	CT-CT	5	1711	13.1
	CG-CG	3	1461	9.7
	TG-TG	4	1621	3.6
II	CT-CT	4	1534	13.5
	CG-CG	2	1231	7.0
	TG-TG	4	1536	12.1
III	CT-CT	4	2087	28.1
	CG-CG	2	1575	14.3
	TG-TG	6	1871	12.1

Mean sCD14 levels were compared against *CD14* diplotypes in all three age groups. The patients carrying the CG-CG diplotype had lower levels of sCD14 compared to the other patients irrespective of age groups implying that *CD14* diplotype determines the sCD14 levels in the serum of CF patients, albeit the rareness of the haplotype CG did not allow a statistical comparison.

Therefore, to analyse whether *CD14* variants influence the course of *P. aeruginosa* infection as visualised by AbProtF titers, AbProtF data from 47 patients were compared against their age according to their *CD14* diplotypes (Fig. 27).

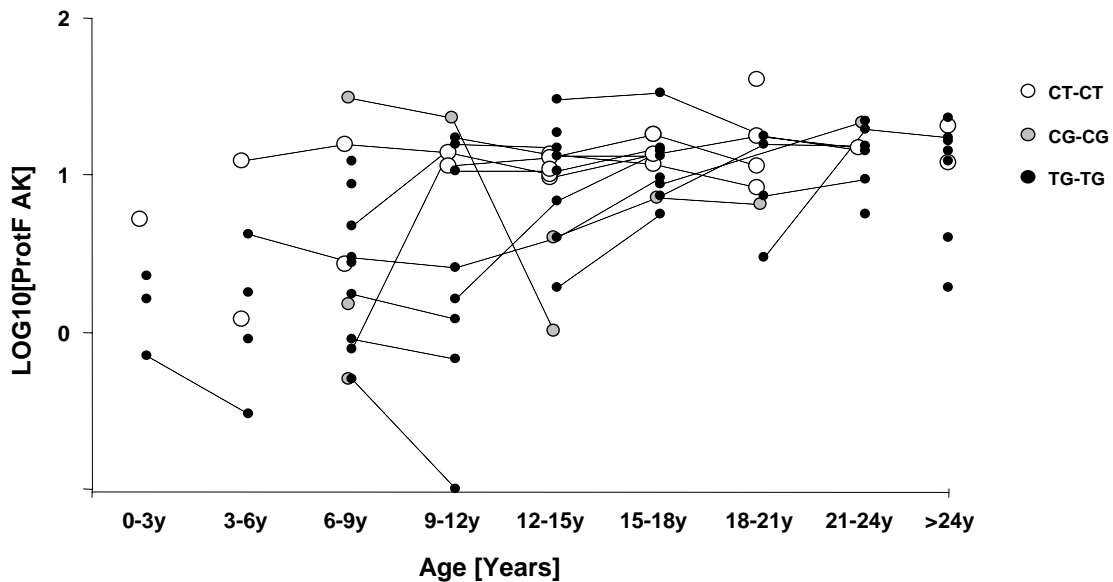


Fig. 27 AbProtF values plotted against age and CD14 diplotypes of CF patients. X-axis denotes age groups of 3 year interval; y-axis denotes mean AbProtF values in log scale; for each patient, all AbProtF values for one age period were represented as average value.

All patients who were 15 years or older had high AbProtF titers irrespective of *CD14* diplotype. Four out of seven patients carrying TG-TG during 9 to 12 years of their age had lower AbProtF titers, while both patients carrying CT-CT diplotype had higher AbProtF titers. Patients carrying CG-CG diplotype had lower AbProtF titers. An unusual course of AbProtF titers was observed for one patient carrying CG-CG diplotype who showed a decline of AbProtF titers from his 6 to 15 years of age. In conclusion the data show that individuals carrying CT-CT diplotype were able mount a higher antibody response against *P. aeruginosa* in their early age compared to other individuals carrying CG-CG diplotypes. These observations were in accordance with the earlier findings, in which CG-CG diplotype was associated with early colonisation of *P. aeruginosa* followed by faster adaptability to the CF lung, as shown by O-antigen deficiency, and low sCD14 levels in the serum. Similarly CT-CT diplotype was associated with late colonisation of *P. aeruginosa* and higher sCD14 levels. Taken together, these findings show consistently that *CD14* diplotypes determines the levels sCD14 and thereby mounting the specific antibody response against *P. aeruginosa*, which means, CF patients carrying higher sCD14 levels in serum are capable of eliminating *P. aeruginosa* from their lung early in their life.

3.5.12. How does *CD14* 3' UTR polymorphism determine sCD14 levels among CF patients?

The *CD14* promoter polymorphism was shown to alter the binding site for SP1/SP3 transcription factors and thereby determining the sCD14 level (Tricia et al., 2001). However, the single marker analysis as well as the haplotype analysis in this study strongly suggested that the *CD14* 3' UTR polymorphism as a major causative variant in CF scenario. Furthermore, the *CD14* diplotypes re-constructed by genotyping the *CD14* promoter polymorphism and the *CD14* 3' UTR polymorphism were associated with *P. aeruginosa* related endophenotypes. Hence, this finding constitutes the first example of a causative relationship between the CD14 phenotype and a polymorphism in the 3' UTR of *CD14*.

Several studies have shown that the sequence variants within coding region or in the promoter region or in the 5' UTR exert their effects on transcription or splicing or translation leading to variability in protein levels (Pickering and Willis, 2005; Cartegni et al., 2002). The 3' UTRs are shown to be involved in gene expression regulation at

multiple levels such as mRNA 3' end formation and polyadenylation at pre-mRNA level and also at mature mRNA level by regulating mRNA stability/degradation, nuclear export, subcellular localisation and translation efficiency (Conne et al., 2000; Mignone et al., 2002; Chabanon et al., 2004). In support of these findings, several 3' regulatory polymorphisms, which are disease causing have been reported (Chen et al., 2006). Thus, the impact of regulatory polymorphisms within the 3' UTRs can be envisaged at different level of mRNA formation. Accordingly, we attempted to study the possible role of polymorphism within the *CD14* 3' UTR in determining the different levels of sCD14.

3.5.13. CD14 3' UTR polymorphism could affect CD14 mRNA 3' end formation

In general, 3' end processing of mRNA involves interaction of multiprotein complex with cis-acting sequence elements on the pre-mRNA (Zhao et al., 1999): Such as polyadenylation signal AAUAAA interacts with multimeric cleavage/polyadenylation specificity factor (CPSF), a high density of G/U or U residues, located downstream of cleavage site, interacts with the 64-kDa subunit of the heterotrimeric Cleavage-stimulating factor (CstF). Interactions between CPSF and CstF, as well as polyA polymerase (PAP) and Cleavage factors I and II (CFI and CFII respectively) are minimal essential requirements for efficient cleavage and for *in vitro* polyadenylation (Colgan and Manley, 1997). This RNA-protein interaction determines the site of cleavage 10-30 nt downstream of polyadenylation signal, preferentially immediately 3' of a CA dinucleotide (Chen et al., 1995). Thus, it is conceivable that variations within these cis-acting elements may alter binding efficiency of these proteins and hence may cause poor mRNA processing. However, the *CD14* 3' UTR polymorphism was located 14 nt downstream to polyadenylation signal AUUAAA and 3 nt upstream of GU/U element (Fig. 28) excluding the possibility of altering the binding efficiency of both CPSF and CstF proteins. On the other hand, it is interesting to note that the polymorphism was located 2 nt downstream of the wild-type CD14 mRNA cleavage site indicating a probable role in cleavage site selection. Consistently, heterogeneity in the CD14 3' end cleavage sites were observed and interestingly the CD14 cDNA isolated from insulinoma tissue (NCBI: BQ477440) was found to include the *CD14* 3' UTR polymorphism as a site for 3' end cleavage (Fig. 28). However, it is not clear

whether or not this heterogeneity in the CD14 3' cleavage site is due to 3' UTR polymorphism.



Fig. 28 BLAST search against CD14 cDNA in human expressed sequence tag (EST) data base identified variable lengths of CD14 3' UTR that are isolated from different human cell lines; X13334:THP1 macrophage cell line; Wild-type (NM_000591): general CD14 transcript deposited in NCBI; CD366124: Alveolar macrophage; BQ477440: Human insulinoma tissue. * denotes cleavage and polyadenylation site. TATTA is polyadenylation signal. T is CD14 3' UTR G to T SNP. Probable binding site, GTGT, for Cleavage-stimulating factor (CstF) is shown in dotted line.

Nevertheless, a convincing consensus for the downstream regulatory elements is not achieved. Thus, it is reasonable to suppose that the presence of sequence variant exactly within the putative cleavage site might alter the overall efficiency of cleavage complex assembly and in turn the efficiency of the 3' end formation.

3.5.14. MicroRNA mediated regulation of CD14 pre-mRNA

MicroRNA's are short, often phylogenetically conserved, non-protein-coding RNA molecules (Bartel, 2004). They bind to near perfect complementary sequences in the 3' UTR of a target gene and mediate post-transcriptional gene repression either by degradation of target transcripts or by inhibiting protein translation. In order to test the possibility of microRNA mediated CD14 regulation, microRNAs which are predicted to be able to bind CD14 3' UTR were explored in miRBase (<http://microrna.sanger.ac.uk/>). The "miRBase" is a database for all known microRNAs from different organisms and it manages the nomenclature and annotation of microRNAs from all species with the "miR" designation followed by a unique number (Ambros et al., 2003; Griffiths-Jones, 2004). The miRBase has reported three human microRNAs, hsa-miR-610, hsa-miR-377, and hsa-miR-369-5p (hsa denotes *Homo sapiens*) to be able bind human CD14 pre-mRNA. However, the binding sites for all these microRNAs are located more than 700bp downstream of 3' UTR (Fig. 29).



Fig. 29: MicroRNAs and their binding sites on *CD14* pre-mRNA predicted by miRBase; Black arrows indicate the position of microRNA binding sites on *CD14*; Nucleotide numbering starts from the *CD14* mRNA stop codon; *hsa-miR* denotes *Homo sapiens*-microRNA. Perfect pairing was shown by vertical lines and G to U pairing, which is allowed in RNA base pairing, is shown by dots.

3.5.14.1. *CD14* 3' UTR polymorphism, rs2563298, is located with in a microRNA binding site

It has recently been shown that a G to A transition in the 3' UTR of *GDF8* gene (myostatin) in sheep created binding site for microRNAs, mir1 and mir206, and hence translation inhibition of *GDF8* gene expression, which in turn caused muscular hypertrophy (Clop et al., 2006). Similarly, Abelson et al. (2005) reported that a sequence variant in the binding site for microRNA has-miR-189 on *SLITRK1* gene was associated with Tourette's syndrome. To test if the *CD14* 3' UTR polymorphism, rs2563298, can modify the binding of any microRNA's, *CD14* 3' UTR and its 2 kb downstream sequence was screened for possible binding sites for microRNA's by searching in miRBase (Release 9; October 2006). Interestingly, microRNA from bovine (*Bos taurus*), named bta-miR-425-5p, was predicted to be able to bind bovine and rabbit *CD14* 3' UTR (Fig. 30a). Although, it is not reported in the miRBase that bta-miR-425-5p can bind human *CD14* 3' UTR, two findings strengthened the possibility of miR-425-5p binding to human *CD14*. Firstly, miR-425-5p binding site exactly overlapped the cleavage and polyadenylation site of human *CD14* 3' UTR. Secondly, the existence of same microRNA (hsa-miR-425-5p) in humans was experimentally shown (Altuvia et al., 2005). Consistent with this finding, miR-425-5p binding site is evolutionarily conserved in human, bovine and rabbit (Fig. 30b) and the eight nucleotides in the 5' proximal end of microRNA, 'seed' region, (positions 2-8 nucleotides) allowed perfect base pairing to the target sequence without any mismatches. Surprisingly, the *CD14* polymorphism rs2563298 was located within in the hsa-miR-425-5p binding site (Figure 30c). These findings raised the possibility that the polymorphism may cause

differential binding of hsa-miR-425-5p to *CD14* 3' UTR and in turn may play a role in modifying the CD14 transcript level.

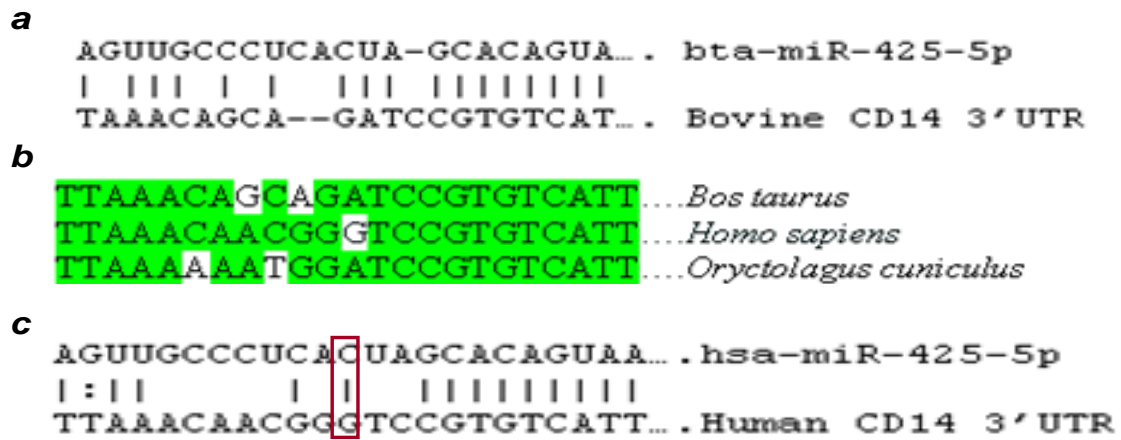


Fig. 30a Alignment of bovine microRNA (*bta-miR-425-5p*) sequence and its binding site on bovine *CD14* 3' UTR; Fig. 30b Alignment of miR-425-5p binding sites on *CD14* in bovine (*Bos taurus*), human (*Homo sapiens*) and rabbit (*Oryctolagus cuniculus*). Nucleotides conserved at least in two species are shown in grey background; Fig.30c Alignment of human microRNA (*hsa-miR-425-5p*) sequence and its binding site on human *CD14* 3' UTR. Location of G to T polymorphism is shown in rectangle. Perfect pairing is shown by vertical lines and G to U pairing, which is allowed in RNA base pairing, is shown by dots.

3.5.14.2. Biogenesis of hsa-miR-425-5p microRNA

The microRNA miR-425-5p is 23 nucleotides in length and is derived from its precursor pre-microRNA, miR-425. The miR-425 microRNA is encoded by gene *MIRN425*, located on chromosome 3p21.31. General microRNA biogenesis pathway involves processing of pri-microRNAs by group of RNase III enzymes to yield mature microRNA (Lee et al., 2003; Denli et al., 2004; Grishok et al., 2001; Hutvagner et al., 2001). Similarly, the *MIRN425* transcript of length 86bp, a stem-loop structure, is processed by dicer, RNase III like nuclease, to yield two mature microRNAs from both arm of the stem loop namely hsa-miR-425-5p from 5' arm and hsa-miR-425-3p from 3' arm (Fig. 31).

hsa-mir-425 stem-loop

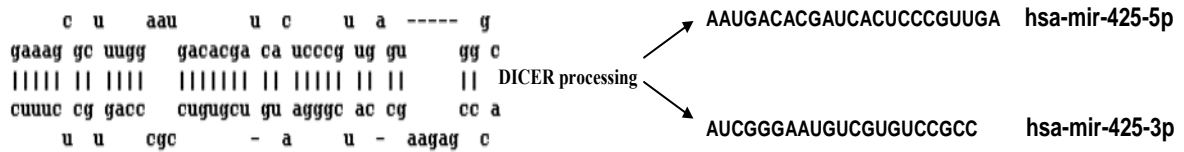


Figure 31: Pri-microRNA miR-425 is processed by enzyme DICER to yield two mature microRNAs. hsa-miR-425-5p was predicted to bind human CD14 3' UTR in this study.

3.5.14.3. CD14 pre-mRNA, not mature mRNA, may be a target for miR-425-5p

It is shown that the “seed” region, approximately 8 to 10 nt from 5' end, of microRNA provides most of the pairing specificity to its target (Tomari and Zamore, 2005). Taken this evidence, hsa-miR-425-5p can not bind CD14 mRNA due to the fact that the CD14 mature mRNA completely excludes the binding site for hsa-miR-425-5p “seed” region from its 3' UTR. On the other hand, microRNA 425-5p can bind CD14 pre-mRNA prior to CD14 mRNA processing. However, there is no data available on the length of CD14 pre-mRNA made prior to mRNA processing. Thus, in order to test whether CD14 pre-mRNA (prior to mRNA processing) is extended enough to possess the binding site for hsa-miR-425-5p binding site, the CD14 pre-mRNA length was determined.

3.5.14.4. CD14 pre-mRNA is at least 973bp long from its stop codon

The length of CD14 pre-mRNA was determined by PCR-amplifying the CD14 3' end from cDNA synthesized using random primers. Random primers are supposed to amplify all the transcripts irrespective of polyadenylation. Hence, if CD14 pre-RNA is longer than the polyadenylated CD14 mRNA, synthesizing cDNA using random primers should amplify them. With this rationale, cDNA synthesized using random primers were used as templates to determine the length of CD14 pre-mRNA. The methodology is explained in detail in chapter 2.2.5.4. Briefly, one constant forward primer (enclosing the stop codon of CD14 mRNA), in combination with different reverse primers, was employed to amplify 697bp, 986bp, 1604bp and 2075bp fragments from genomic DNA and from cDNA (Fig. 32).

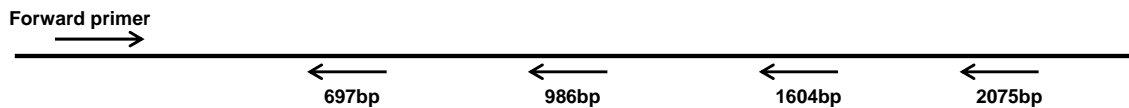


Figure 32: Map showing the location primers employed to amplify CD14 pre-mRNA

All the fragments were amplified from genomic DNA, while only 697bp and 986bp fragments were amplified from cDNA (Fig. 33) indicating that CD14 pre-mRNA extends at least 973bp long from its stop codon. Thus, it is quite reasonable to speculate that the miR-425-5p may bind CD14 pre-mRNA to mediate its regulatory effects.

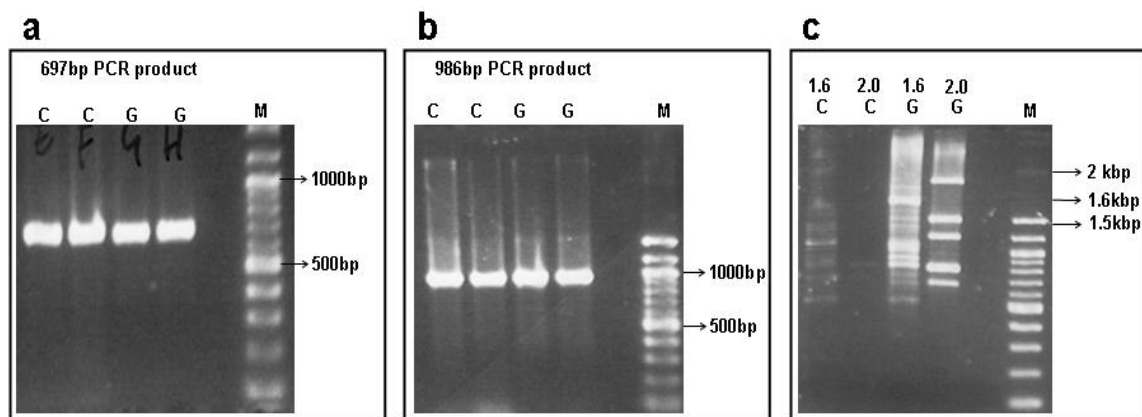


Figure 33: CD14 pre-mRNA length determination; lanes C: cDNA was used as template; lanes G: genomic DNA was used as template; lanes M: 100bp marker; Fig a. 697bp product amplification; Fig b. 986bp product amplification; Fig c. 1.6kb and 2.0kb products were amplified from genomic DNA but not from cDNA. (cDNA was tested for genomic DNA contamination: Intron spanning primers for TNFRSF6, were employed to amplify products of size 250bp and 600bp from both cDNA as well as genomic DNA, respectively. Amplifying only 250bp product but not 600bp from cDNA indicated that cDNA is not contaminated with genomic DNA)

3.5.14.5. MicroRNA binding efficiency is altered by CD14 3' UTR polymorphism rs2563298

As a next step, the consequence of CD14 3' UTR polymorphism rs2563298 on miR-425-5p binding efficiency was determined by using a web tool 'RNAhybrid' (Rehmsmeier et al., 2004). RNAhybrid is primarily designed to predict microRNA targets and it also calculates the minimum free energy (mfe) of hybridization of a long and a short RNA molecules. The binding efficiencies of T allele and G allele at CD14 3' UTR polymorphism, rs2563298, with hsa-miR-425-5p microRNA were estimated (Fig. 34).

The T allele showed a higher mfe value (-24.3 kcal/mol) for hybridizing with microRNA compared to G allele (-28.5 kcal/mol). This finding suggested that hsa-mir-425-5p microRNA can bind CD14 G allele at rs2563298 more efficiently than binding CD14 T allele at rs2563298. Thus, as a next step, the expression of mir-425-5p in different cell lines was tested.

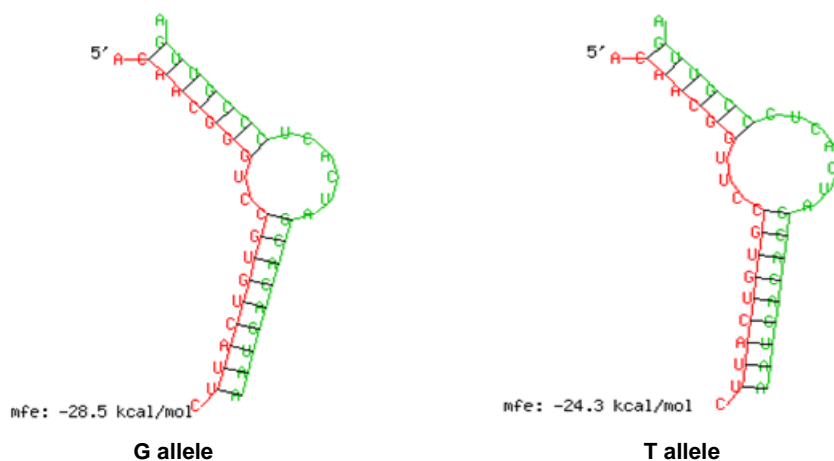


Figure 34: hsa-miR-425-5p binding efficiencies at CD14 rs2563298 G and T alleles; microRNA sequence is shown in green colour and CD14 3' UTR sequence is shown in red colour. G allele shows the lower energy requirement (-28.5 kcal/mol) for hybridization.

3.5.14.6. Standardization of hybridization technique for microRNA detection from different human cell lines

In an effort to detect microRNA hsa-miR-425-5p from different cell types, hybridization protocol was standardized by using artificial hsa-miR-425-5p microRNA as a probe and its complimentary sequence as a target. The miR-425-5p probe was labeled with biotin (Fig. 35a). The miR-425-5p complimentary oligo, in different concentrations (300ng, 100ng, and 50ng), was coated on the membrane and hybridized at 48°C for overnight with 50ng of biotin-labeled miR-425-5p probe (Fig. 35b). The specificity of hybridization was confirmed by testing hybridization conditions on randomly chosen complementary sequences of known microRNAs (Fig. 35c) in which only 425-5p showed a signal. 10ng of probe on the membrane and 50ng of biotin-labeled miR-425-5p in the hybridization solution was enough to detect hybridization signal. Thus, it was possible to discriminate different microRNAs using this method.

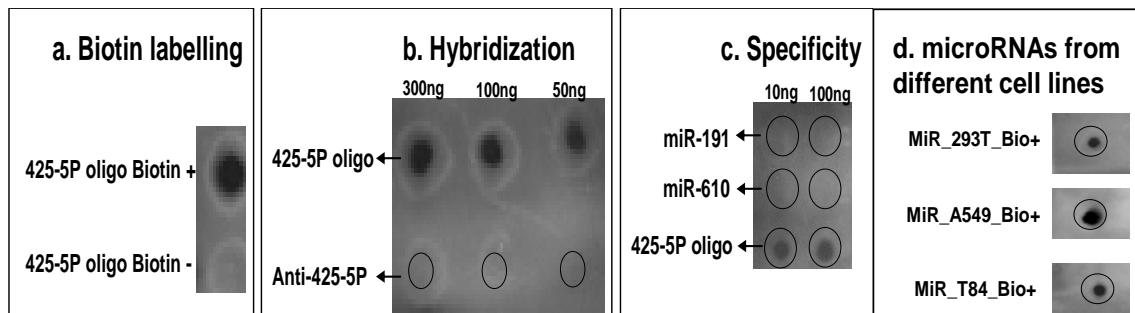


Figure 35a: Confirmation of biotin labelling; artificial microRNA 425-5p oligo was biotin labeled, efficiency of labeling was tested by coating them on membrane and signal was detected (Biotin + is labeled with biotin and 425-5p biotin- is not labeled).

Fig. 35b: 425-5p oligo, artificial complementary sequence for miR-425-5p microRNA, employed as a positive control and Anti-425-5p is a negative control. Both oligos were coated on the membrane and hybridized with 425-5p oligo Biotin+.

Fig. 35c: miR-191 and miR-610 complementary sequences were employed to test the specificity of miR-425-5p hybridization. 425-5P oligo Biotin+ hybridized only with its complementary oligo.

Fig. 35d: Total microRNAs were isolated from 293T, A549, and T84 cell lines and biotin labeled. The labeling efficiency was tested by coating them on the membrane followed by detection.

Hence, the total microRNAs were isolated from 293T, T84, and A549 cell lines and 2 μ g of microRNAs were labeled with biotin. Biotin labeling of total microRNAs, isolated from different cell lines, was successful (Fig. 35d). Complementary oligos, both for hsa-miR-425-5p and other 15 different randomly chosen human microRNAs, were coated on the membrane in two different concentrations (50ng and 10ng) and hybridized with 10ng of biotin-labeled total microRNAs isolated from different cell lines.

However, hybridization of total microRNAs failed to detect any signal. As the hybridization was optimized for single microRNA (50ng) with its complementary oligo (10ng), total microRNAs may not contain such a high concentration of targeted microRNA species. Moreover, the microRNA detection technology is developing currently. The optimized protocols for microRNA quantification, especially from cell lines which I used in this study, are still not established. Thus, hybridization with labeled total microRNAs still needs to be optimized with different concentrations of microRNA and its complementary sequence. The hybridization parameters have to be standardized further prior to making any conclusions on miR-425-5p microRNA mediated CD14 regulation.

3.5.15. Role of CD14 polymorphisms in cystic fibrosis

The persistent, severe neutrophilic inflammation and inability of innate immune response to clear chronic bacterial infection due to pronounced imbalance between pro- and anti-inflammatory cytokines is the characteristic of CF airways (Tümmler and Kiewitz, 1999). Macrophages and polymorphonuclear neutrophils are the phagocytic cells which play a central role in innate immunity through expression of pathogen recognition receptors such as CD14 and TLRs. CD14 is the LPS binding receptor that presents LPS to its signaling receptor complex, MD-2/TLR4 and can participate in activation of NF- κ B to induce inflammatory response (Hajjar et al., 2002). The soluble CD14 readily binds LPS monomers and transfers to m (membrane bound) CD14 (Hailman et al., 1996) or directly to the MD-2/TLR4 complex on cells that do not express mCD14 (Frey et al., 1992; Pugin et al., 1993). Additionally, CD14 was shown to bind other bacterial ligands such as peptidoglycans and lipoteichoic acid (Pugin et al., 1994). Thus, for efficient recognition of bacterial invasion and effective induction of down stream signal via TLR4 or TLR2, CD14 protein levels are very critical.

In this study *CD14* promoter and 3' UTR polymorphisms were investigated to determine the role of *CD14* genetic variants in CF. The haplotype analysis showed a significant association with CF disease discordance (Table 20) indicating that the trans-acting factors, which are encoded elsewhere in the genome, may bind CD14 alleles differentially and in turn modulate the CF disease. This finding was further confirmed by transmission disequilibrium test (TDT) at CD14, in which a significant transmission distortion of haplotype was observed (Table 21). Sequencing analysis of the complete *CD14* gene revealed no sequence variants but promoter polymorphism and 3' UTR polymorphism. Further analysis of *CD14* genotypes at both polymorphisms among *P. aeruginosa* stratified cohorts identified an association between *CD14* diplotypes and early or late acquisition of *P. aeruginosa* (Fig. 26). Furthermore, the O-antigen phenotype of *P. aeruginosa* strains and level of antibodies against *P. aeruginosa* outer membrane protein F (AbProtF) was found to be associated with the *CD14* diplotypes (Table 25, 26, and 27). Thus it is possible that the higher level of CD14 may effectively interfere with *P. aeruginosa* adaptation and chronic colonization in early ages of CF patients.

A recent association study in CF children showed that the CF patients having higher sCD14 levels remained *P. aeruginosa* free and CC genotype at *CD14* promoter polymorphism had an increased risk of early infection with *P. aeruginosa* (Martin et al., 2005). However, they failed to show an association between the *CD14* genotypes and sCD14 levels in plasma of CF children. By contrast, in this study the *CD14* diplotype, not genotype of any single polymorphism, was associated with both the levels of sCD14 as well as the early acquisition of *P. aeruginosa* among CF patients. This discrepancy could be due to dissimilar selection of phenotypes for the study and evaluation approach. Moreover, the genotype analysis at single marker can detect the true association with the haplotype due linkage which is prominent in *CD14*. Accordingly, in this study, the association between *CD14* promoter polymorphism and *P. aeruginosa* early acquisition was significant among unrelated CF patients in which CT genotypes were associated with late colonization of *P. aeruginosa* (Fig. 24). However, the diplotype analysis among CF twin and siblings clearly indicated that the association between *P. aeruginosa* related endophenotype and the *CD14* polymorphisms is indeed due to interaction of both loci. Both polymorphisms may have a functional impact on CD14 levels, as shown by published data and in-silico work carried out in this thesis.

It is shown by Tricia DL et al. (2001) that *CD14* promoter polymorphism alters the binding site for Sp1/Sp2/Sp3 transcription regulators and further they hypothesized that the enhanced binding of Sp3 to C allele at *CD14* promoter polymorphism could result in transcriptional repression of CD14. Although, several studies have shown that the variations within 3' UTRs of different genes can cause diseases in humans (Chen et al., 2006), no studies have investigated the *CD14* 3' UTR polymorphism to explain its role in determining sCD14 levels. In-silico analysis in this study predicted two potential possibilities of how 3' UTR polymorphism may regulate sCD14 levels. One possibility could be by altering the binding efficiency of CD14 RNA to the proteins involved in mRNA 3' end formation (Fig. 28). This might lead to reduced efficiency of mRNA processing and in turn reduced level of CD14 transcripts. The other possibility identified in this study is that the polymorphism in CD14 pre-mRNA may alter the binding efficiency of hsa-miR-425-5p microRNA (Fig. 30-34), which might cause defect in mRNA processing or complete degradation of CD14 RNA molecules.

However, further experiments are necessary to detect the microRNA hsa-miR-425-5p from different cell lines, including peripheral blood mononuclear cells and to investigate its role in CD14 mRNA regulation. If the polymorphism alters the efficiency of microRNA mediated CD14 mRNA quantity then, the quantity of microRNA and CD14 mRNA both isolated from the same sample should be inversely proportional if the sample is homozygous for G at *CD14* 3' UTR polymorphism (Fig. 11). This can be tested by isolating both microRNA fractions as well as CD14 mRNA from peripheral blood mononuclear cells from individuals harboring different genotypes at *CD14* 3' UTR polymorphism. The CD14 mRNA and microRNA hsa-miR-425-5p can be quantified by real-time PCR and hybridization experiments, respectively. By this approach, the association between the microRNA, the CD14 mRNA and the CD14 3' UTR polymorphism can be explored.

In summary, we identified *CD14* promoter polymorphism and *CD14* 3' UTR polymorphism as causative variants of both CF disease modulation as well as determinants of age at onset of *P. aeruginosa* chronic colonization in CF patients. Furthermore, by phenotyping assays and in-silico analysis we highlighted the mechanistic role of both polymorphisms in regulating the level of sCD14, and thereby affecting the CF pathology.

3.6. Analysis of CD95 as a modulator of cystic fibrosis

Apoptosis is an essential physiological process for the homeostasis of epithelial and inflammatory cells. This process is regulated by many factors such as oxidative stress (Buttke, T.M. et al., 1994), extracellular matrix proteins (Meredith et al., 1993), and CD95 ligation (Nagata and Golstein, 1995). CD95 is a cell-surface receptor belonging to the tumour necrosis factor (TNF) receptor family of apoptosis signalling molecules. Cloning of CD95 molecule showed it to be a 319-aa type 1 transmembrane glycoprotein and found to be expressed in several tissues, including thymus, spleen, ovary and heart, and on a number of cell types, including activated T- and B-lymphocytes (Itoh et al., 1991; Watanabe-Fukunaga et al., 1992). CD95 and CD95 (L) ligand are particularly essential for the elimination of haematopoietic cells, including lymphocytes, monocytes and macrophages, as these cells are increased in CD95-deficient *lpr* and CD95L-deficient *gld* mice (Cohen and Eisenberg, 1991). Similarly, mutations in human CD95 gene were shown to cause lymphoproliferation and autoimmunity (Fisher et al., 1995). Furthermore, in *P. aeruginosa* pneumonia model, it was shown that *P. aeruginosa* infection induces apoptosis of lung epithelial cells by activation of endogenous CD95/CD95 ligand system and deficiency of either CD95 or CD95 ligand caused increased mortality in mice (Grassme et al., 2000).

TNFRSF6, chromosome 10q24.1

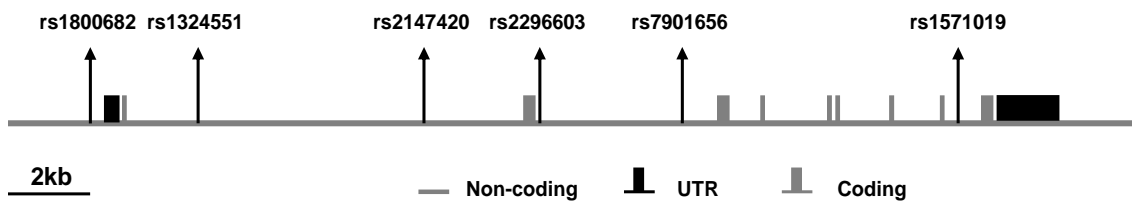


Figure 36: Schematic representation of CD95 (TNFRSF6) gene structure; CD95, located on chromosome 10q24.1, is 25.2kb long and has nine exons. Polymorphisms shown on the CD95 gene are targeted for fine mapping. SNPs are indicated by rs followed by their unique numbers (Location and size are based on NCBI, NC_000010.9 Reference assembly: Build 36.2)

3.6.1. Skewed allele distribution among CF siblings on *CD95*

All families of 37 F508del homozygous sib pairs were subjected to the Monte Carlo simulation based test and the analysis revealed disproportionate transmission of alleles on *CD95*. The comparison of transmitted and non-transmitted alleles and haplotypes showed transmission distortion, in which frequency distribution was significantly skewed for the haplotype block rs1324551-rs7901656 ($P = 0.03$, Fig. 37), which enclose intron 1 and intron 2. However, single-marker analysis did not reach the statistical significance after correction for multiple testing ($P = 0.16$). The haplotype A-C at rs1324551-rs7901656 was over transmitted to CF offspring while the haplotype G-C was frequently observed among non-transmitted parental haplotypes (Table 28). Consistently, haplotype A-C was found to be risk haplotype and was overrepresented among severely affected pairs. While haplotype G-T was underrepresented among severely affected (CON-) pairs compared to mildly affected pairs (Table 28).

Figure 37: Family based analysis of haplotype blocks on *CD95* among CF twin and siblings

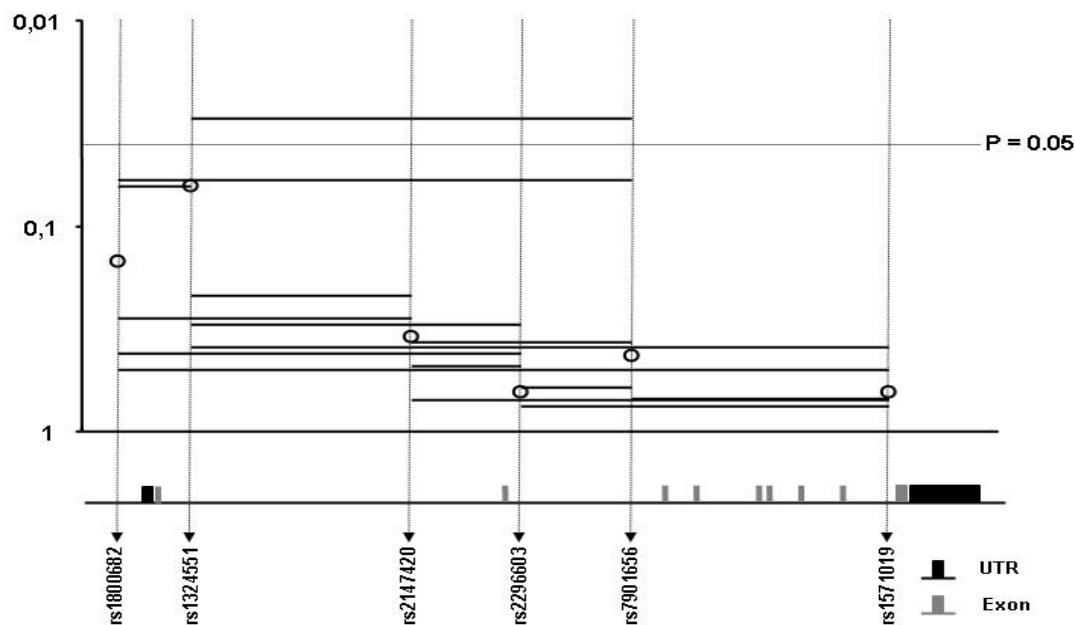


Fig. 37: Family based analysis of haplotype blocks. P-values are shown for single markers (open circles) and haplotypes of adjacent markers (lines). The x-axis depicts the physical distance between the markers and the y-axis denotes the P-value. The corrected P-value for haplotype block rs1324551-rs7901656 is $P = 0.03$. (Data is corrected for sib-pair dependence in co-operation with T. Becker, Bonn).

Table 28: Frequency distribution of rs1324551-rs7901656 haplotype block on CD95 among CF twin and siblings

CD95 rs1324551-rs7901656 haplotype	Transmitted	Non- transmitted	CON-	CON+	DIS
G-C	0.26	0.38	0.14	0.14	0.21
G-T	0.30	0.35	0.19	0.40	0.28
A-C	0.43	0.21	0.67	0.46	0.50
A-T	0	0.05	0.0	0.0	0.0

3.6.2. Association between CD95 and CF disease severity revealed by single-marker analysis

CD95 was evaluated as a modulator in CF by initially typing six SNPs located on CD95 (Fig. 36). Allele and genotype distributions were compared between both mildly affected and severely affected patient pairs as well as CONC and DIS pairs. The allele distributions were similar at all tested loci in both the comparison. However, genotype distributions were significantly different between severely and mildly affected pairs at all loci but rs2296603 ($P = 0.26$) and rs1571019 ($P = 0.75$) revealed by uncorrected single-marker P-values, in which heterozygotes were overrepresented among CON-group. To identify the exact sequence block involved in disease variation, haplotype blocks for adjacent markers were constructed and their distributions were analysed in CF twin and sibling cohort.

3.6.3. Haplotype analysis confirmed the association between CD95 and CF disease severity

Two-marker haplotypes were constructed for all available families and distribution was systematically evaluated among CF twin and siblings. As it is shown in Figure 38, distribution of haplotype block rs2296603-rs7901656 was significantly different comparing CON- and CON+ ($P = 0.04$), while all other blocks were similarly distributed among all the tested groups. The haplotype distributions were corrected for sib-pair dependence and multiple testing (Fig. 39). Indeed, the same haplotype block (rs2296603-rs7901656) was found to be associated with disease severity ($P = 0.06$), assuring that the two-marker haplotype distribution evaluation as a robust and accurate method.

Figure 38: Direct comparisons of CD95 haplotype distributions among CF twin and siblings using Monte-Carlo simulation of k x 2 tables

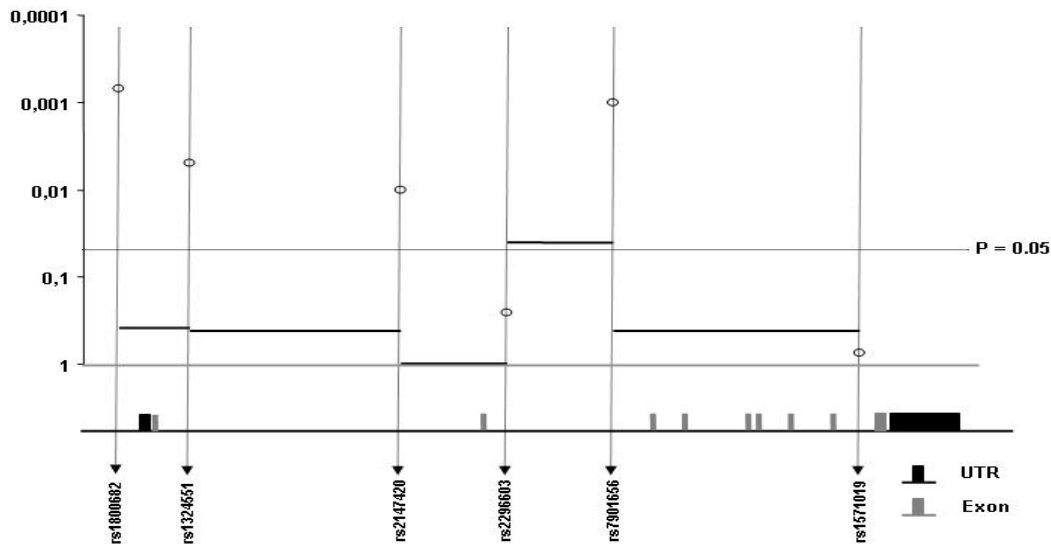


Fig. 38 Case-control analysis comparing mildly (CON+) and severely (CON-) affected patient pairs. P-values are shown for genotype distribution at single markers (open circles) and haplotypes of adjacent markers (lines). The x-axis depicts the physical distance between the markers and the y-axis denotes the P-value. P = 0.04 was observed for the haplotype block rs2296603-rs7901656

Figure 39: CD95 haplotype distributions among CF twin and siblings analysed by correction for multiple testing and sib-pair dependence

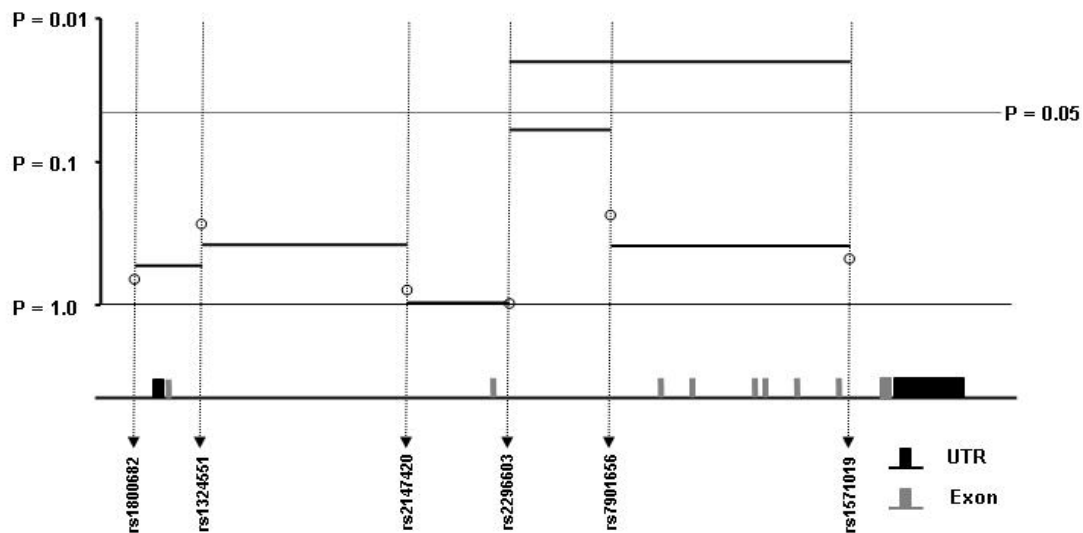


Fig. 39 Case-control analysis comparing mildly (CON+) and severely (CON-) affected patient pairs. P-values are shown for single markers (open circles) and haplotypes of adjacent markers (lines). The x-axis depicts the physical distance between the markers and the y-axis denotes the P-value. P = 0.06 (corrected for multiple testing) was observed for the haplotype block rs2296603-rs7901656

Furthermore, three-marker haplotype block (rs2296603-rs7901656-rs1571019) showed differences in haplotype distribution comparing CON+ and CON- pairs ($P = 0.02$). However, two out of the 3 row comparing this haplotype did not reveal altered genotype distribution comparing CON+ and CON- (rs1571019; $P = 0.81$, rs7901656-rs1571019 ; $P = 0.39$). This finding strongly suggested that the sequence variant is located within the haplotype block rs2296603-rs7901656. Thus, the entire haplotype block rs2296603-rs7901656, located within intron 2 of *CD95* gene, was sequenced. Furthermore, to exclude the possibility of any coding variants for the association, the complete coding region from exon 1 to exon 8 of *CD95* gene was sequenced.

3.6.4. Sequence analysis of *CD95* coding region and intron 2

Genomic DNA samples from nine individuals, in which four individuals were harbouring haplotype A-C at rs2296603-rs7901656, 3 individuals were carrying haplotype G-T at rs2296603-rs7901656 and two individuals were carrying haplotype A-C at rs2296603-rs7901656, were selected for sequence analysis. The sequenced region contained part of promoter, exons 1 to 8, including intron-exon boundaries and entire intron 2.

Table 29: Sequencing of *CD95* coding and promoter region confirmed the previously reported polymorphisms

Region	NCBI SNP ID	Heterozygosity†	PIC‡
Promoter	rs2234768	0.156	-
Promoter	rs1800682*	0.500	0.37
Intron 2	rs2296603*	0.488	0.37
Intron 2	rs2031611	0.270	-
Intron 2	rs9658741	0.491	-
Intron 2	rs9658742*	0.500	0.37
Intron 2	rs9658748	0.119	-
Intron 2	rs7901656*	0.457	0.35
Intron 2	rs9658750	0.124	-
Intron 2	rs2031613	0.272	-
Intron 2	rs2031612	0.488	-
Exon 3	rs3218612 – Synonymous	0.116	-
Intron 5	rs7911226	0.476	-
Intron 6	rs2296600	0.484	-
Exon 7	rs2234978 – Synonymous	0.333	-

* Polymorphisms tested in this study; † Heterozygosity values reported in NCBI; ‡ Polymorphism information content determined by accounting parental allele frequencies

Targeted fragments were PCR-amplified and sequenced. Sequencing of the coding region and intron 2 did not reveal any novel sequence variants; however it confirmed previously reported polymorphisms (Table 29). Sequencing of *CD95* intron 2 fragment, encompassing the targeted haplotype block rs2296603-rs7901656, showed three previously reported polymorphisms rs9658741, rs9658742 and rs9658748. The genotype distribution at these three loci indicated the transmission of alleles irrespective of genotypes at rs2296603 and rs7901656 (Table 30).

Table 30: Sequencing results from the *CD95* rs2296603-rs7901656 haplotype block

	rs2296603*	rs9658741	rs9658742*	rs9658748	rs7901656*
1	AA	AA	C/G	TT	CC
2	GG	GG	CC	TT	TT
3	A/G	GG	CC	G/T	C/T
4	AA	GG	CC	TT	CC

* *Polymorphisms tested in this study*

3.6.5. Fine-mapping within haplotype block rs2296603-rs7901656

To test whether any of the SNPs lying within the haplotype block rs2296603-rs7901656 show an association with the disease severity, rs9658742 was typed by PCR-RFLP on CF twin and siblings. Single-marker analysis at rs9658742 showed an allelic imbalance between CON+ and CON- pairs ($P = 0.06$). Further, subsequently breaking the rs2296603-rs7901656 haplotype fragment with rs9658742 data did not show any association with the disease severity and haplotype distribution (Fig. 40). However, haplotype structure indicated that the causative variants are located between rs9658742 and rs7901656 as haplotype block rs9658742-rs7901656 showed a trend of dissimilar distribution between mildly and severely affected patient pairs.

Figure 40: Fine-mapping the targeted haplotype rs2296603-rs7901656 with a SNP rs9658742

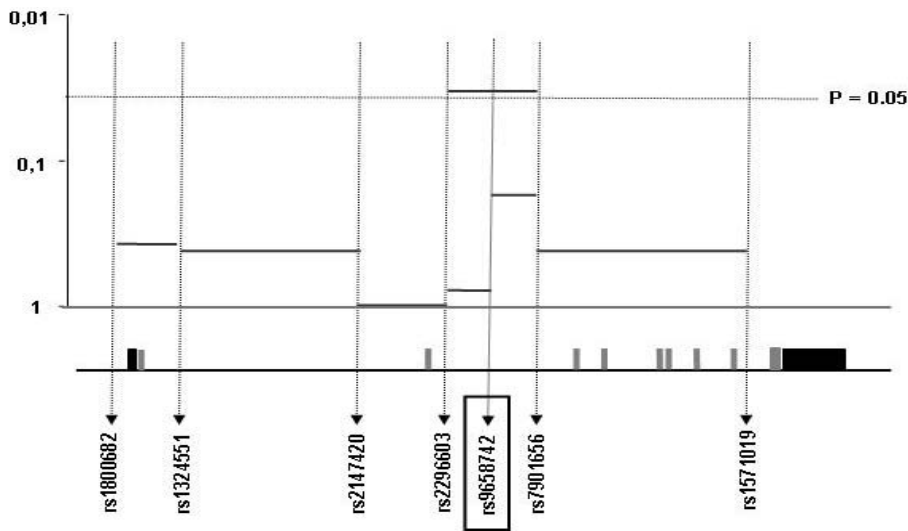


Fig. 5 Case-control analysis comparing mildly (CON+) and severely (CON-) affected patient pairs. P-values are shown for haplotypes of adjacent markers (lines). The x-axis depicts the physical distance between the markers and the y-axis denotes the P-value. P = 0.1 was observed for the haplotype block rs9658742-rs2296603

3.6.6. Conserved non-coding sequences (CNS) in *CD95* gene

Often regulatory sequences are only very short (only a few base-pairs long) and difficult to find. One way to discover them is by searching conserved areas, sequences that have persisted for millions of years of evolution among different organisms. Thus, comparison of genomic sequences by sequence aligning tools such as VISTA predicts the conserved non-coding sequences among different organisms (Mayor et al., 2000; Frazer et al., 2004). Analysis of the entire *CD95* genomic sequence by aligning against mouse *CD95* genomic sequence by VISTA (Visualisation Tools for Alignments) tool identified 14 blocks of conserved non-coding sequences (CNSs) within *CD95* gene (Fig. 41). Surprisingly, three CNSs were located within intron 2 of *CD95* gene. Screening the haplotype block rs9658742-rs7901656 block for CNSs revealed that there were actually two CNSs blocks of size 100bp each, positioned within rs9658742-rs7901656 fragment separated by only 260bp from one another. SNP rs7901656 is exactly located within the second CNS block of intron 2 (Fig. 42). Stretch of 500bp sequence encompassing two CNSs blocks and location of probable causative SNP on one of these blocks raise the possibility that this region is involved in regulation of gene function in an unknown way.

3.6.8. The CNS on intron 2 of *CD95* gene act as a hot-spot for transcription factor binding

The P-Match tool (Chemenev et al., 2005) combines both pattern matching and weight matrix approaches thus providing higher accuracy of recognition of transcription factor binding sites. It is closely interconnected with the TRANSFAC database (Chapter 2.2.14) and hence P-Match tool was used in this study to find out the possible transcription factors and their binding sites in intron 2 of *CD95* gene. Fig. 43 shows the transcription factors and their binding sites found on intron 2 of *CD95* gene. Strikingly, all the four transcription factor binding sites were located exclusively on second CNS of intron 2. Furthermore, the SNP rs7901656 altered the binding site, in which allele C creates a binding site for c-Rel transcription factor. Thus we assumed that the SNP rs7901656 is a causative variant and it is involved in differential transcription activity of the gene.

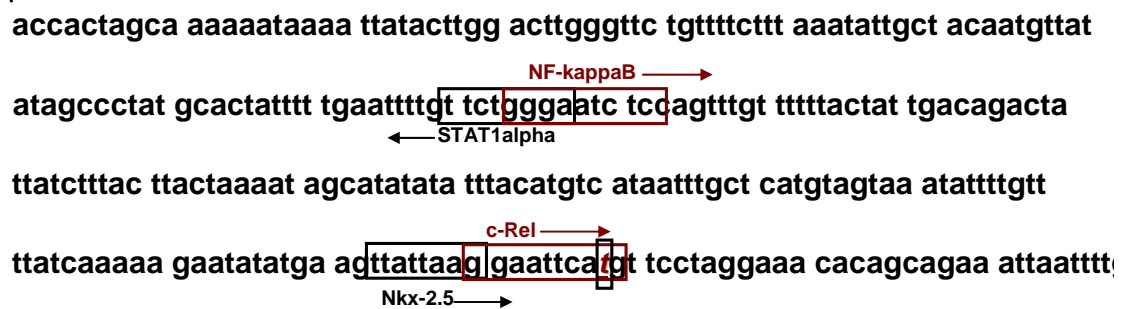


Fig. 43: Composite sequence elements for transcription factors with in intron 2 of *CD95*. The T to C SNP (rs7901656), shown in small rectangle, alters the binding site for c-Rel transcription factor.

3.6.9. The SNP rs7901656, but not rs1800682, determines the transcriptional activity of *CD95* among CF patients

To investigate the role of rs7901656 polymorphism in CF, a cohort, containing 15 unrelated F508del homozygous CF patients, was analysed. These 15 patients were initially recruited, as part of another study, to investigate the residual activity of CFTR and its effect on global transcriptome (Chapter 2.1.5).

Fig. a CD95 transcriptome level against intron polymorphism

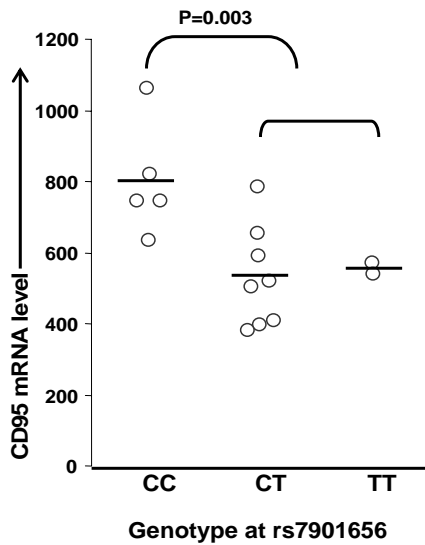


Fig. b CD95 transcriptome level against promoter polymorphism

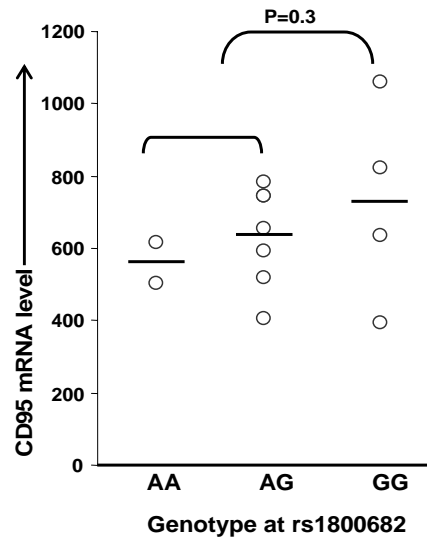


Fig. 44 CD95 intron polymorphism is associated with mRNA levels; Genotypes at rs7901656 (Fig. a) and at rs1800682 (Fig. b) were plotted against CD95 mRNA data from the affymetrix chips. Dark horizontal lines denote mean values. P-values are derived from Mann-Whitney U test.

Total RNA was isolated from rectal suction biopsies from these patients and mRNA was quantified by affymetrix chips. Thus global transcriptome data was available for these patients and interestingly *CD95* gene was found to be differentially expressed among this cohort (Unpublished data). Hence, these 15 patients were genotyped at rs7901656 and rs1800682. CD95 expression data was compared against the genotypes (Fig. 44). The polymorphism rs1800682 is located at nucleotide position -670 in the enhancer region of the *CD95* promoter and it modifies a potential binding site for GAS transcriptional element (Huang et al., 1997; Shuai, 1994). However, this polymorphism failed to show any association with the CD95 mRNA quantity ($P = 0.3$). Surprisingly, a clear association between genotype at rs7901656 and CD95 mRNA level was observed. The CC individuals had a significantly higher level of CD95 mRNA compared to CT and TT individuals at rs7901656 ($P=0.003$). Thus, the SNP rs7901656 is presumably of biological significance since variables associated strongly with transcriptional activity.

3.6.10. Real Time PCR analysis to determine the expression status of the CD95 in peripheral blood mononuclear cells

In order to investigate whether the CD95 polymorphism rs7901656 also play a role in peripheral blood mononuclear cells, the defense effector cells, FACS and real time PCR analysis were carried out on 12 samples of peripheral blood mononuclear cells, isolated from 12 different healthy individuals. The CD95 mRNA level was analysed by real time PCR and compared against their genotypes at CD95 intron 2 SNP, rs7901656.

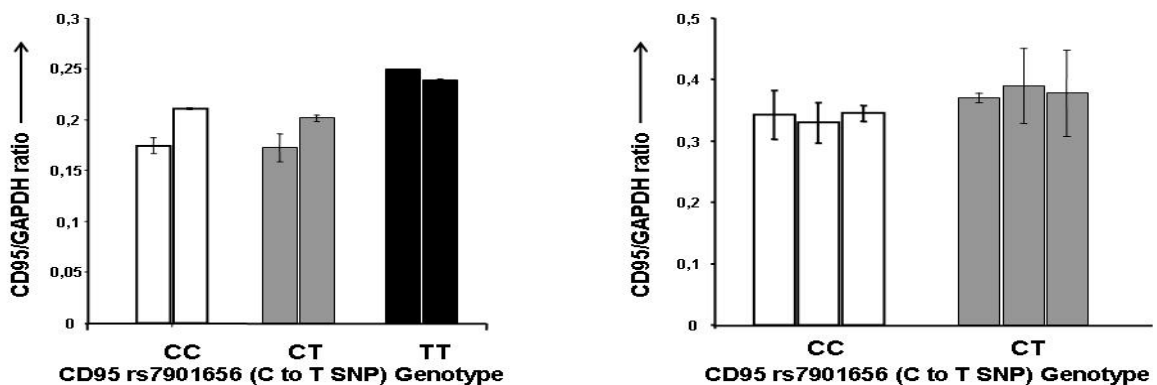


Figure 45: Quantification of CD95 expression on PBMCs from 12 healthy volunteers. Left panel: comparison of CD95 to GAPDH mRNA ratio between first 6 unrelated non-CF individuals [two CC (white bars), two CT (grey bars) and two TT (black bars) genotypes at rs7901656]. Right panel: Comparison of CD95 to GAPDH mRNA ratio between second 6 individuals [three CC (white bars) and three CT (grey bars) genotypes at rs7901656]

Among 12 volunteers, five individuals were homozygous for C allele, five individuals were heterozygous (CT) and only two individuals were homozygous for T allele. The mRNA level was expressed as CD95 to GAPDH ratio. Two set of experiments were done to analyse the mRNA ratio on 12 different samples of peripheral blood mononuclear cells, in which six samples were analysed in each experiment. In contrast to the association seen with mRNA data from CF patients, in whom heterozygotes are associated with lower level of mRNA, the real time PCR data on peripheral blood mononuclear cells from healthy individuals showed no significant association with the genotype (Fig. 45). Thus, CD95 intron 2 SNP was not associated with CD95 mRNA levels in peripheral blood mononuclear cells from healthy volunteers.

3.6.11. CD95 surface expression level on peripheral blood mononuclear cells analysed by FACS

The CD95 surface expression levels were assessed by FACS to test the role of *CD95* polymorphism in peripheral blood mononuclear cells from healthy volunteers. The surface expression levels of CD95 on peripheral blood mononuclear cells were compared against their genotype at rs7901656 (Fig. 46).

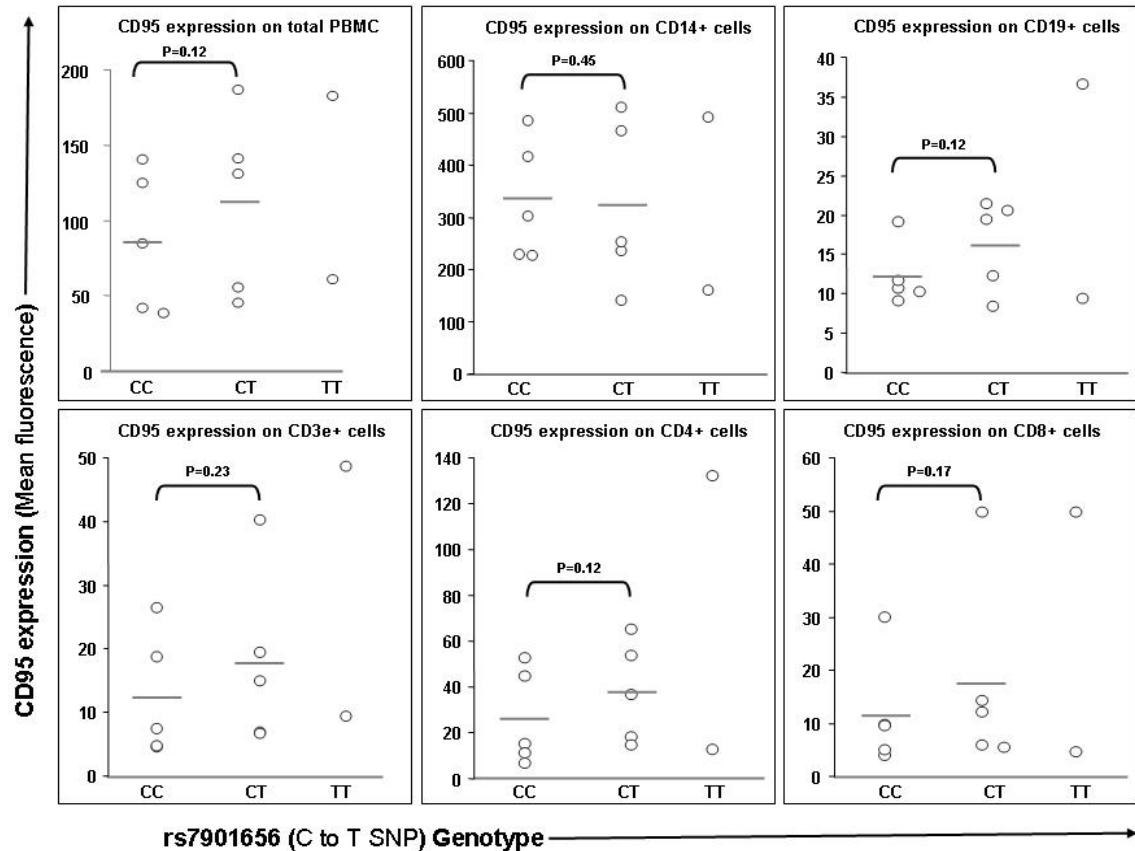


Figure 46: CD95 surface expression level vs. rs7901656 genotypes; P-values are derived by comparing CC and CT groups by Mann-Whitney U test. $P \leq 0.05$ was considered significant. These data are representative of two independent experiments.

The expression level of CD95 was analysed on different cell populations of peripheral blood mononuclear cells such as total peripheral blood mononuclear cells, CD14+ cells (monocytes), CD19+ cells (B-cells), CD3e+ cells (total T-cells), CD4+ cells (T-helper cells) and CD8+ cells (cytotoxic T-cells) to investigate whether the CD95 intron 2 SNP play any role in cell specific expression of CD95. No significant association was observed between genotype at CD95 intron 2 SNP and CD95 surface expression levels on different cell types of peripheral blood mononuclear cells. This finding suggested that the CD95 polymorphism is not determining the CD95 expression levels in

peripheral blood mononuclear cells of healthy volunteers and hence intron SNP function could be specific to either or both epithelial cells and to CF context.

3.6.12. CD95 intron SNP (rs7901656) and promoter SNP (rs1800682) are not associated with *P. aeruginosa* early or late colonisation

Both rs7901656 and rs1800682 were tested to evaluate CD95 locus for its association with *P. aeruginosa* chronic colonisation. The cohort containing 22 unrelated F508del homozygous CF patients, stratified for early and late *P. aeruginosa* chronic colonisation, was typed at rs7901656 and rs1800682.

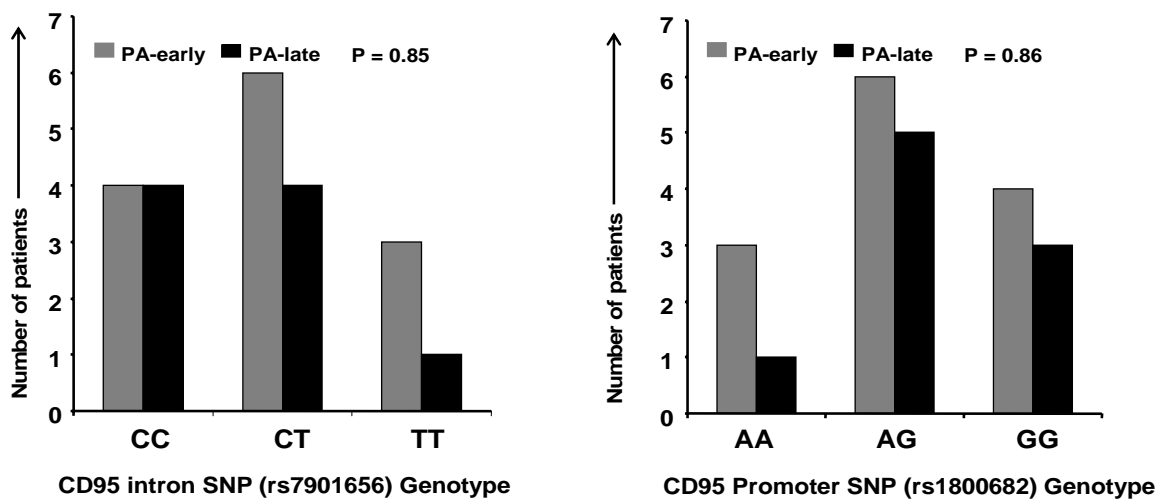


Figure 47: C to T SNP rs7901656 genotype and A to G SNP rs1800682 are denoted on x-axis. PA-early: Patients chronically colonised within seven years of their age. PA-late: Patients chronically colonised after 15 years of their age.

Genotype distributions at rs7901656 and rs1800682 were compared between 13 patients belonging to *P. aeruginosa* early colonised and nine patients of *P. aeruginosa* late colonised groups. The genotype distributions (Fig. 47) in both *P. aeruginosa* early colonised and *P. aeruginosa* late colonised patients were similar at rs7901656 ($P = 0.85$) and at rs1800682 ($P = 0.86$).

3.6.13. The role of CD95/CD95L signaling in cystic fibrosis

A precise balance between cellular apoptosis and cellular survival is essential for the proper functioning of the immune system. Apoptosis mediated by CD95/CD95L pathway is one of the mechanisms used to control immune responses. CD95 (FAS/APO-1), a member of tumor necrosis factor receptor superfamily, is a cell surface

receptor interacts with its ligand CD95L to regulate programmed cell death in different lymphoid cells (Ashkenazi and Dixit, 1998). Furthermore, it is shown that *P. aeruginosa* binds to CD95 and trigger apoptosis by activating the acid sphingomyelinase enzyme that releases ceramide in membrane rafts (Grassme et al., 2000, 2001) which in turn promotes the fusion of small rafts to a large ceramide-enriched membrane platform that serves to cluster and signal CD95 (Gulbins et al., 2004). Indeed, clustering of CD95 as a general mechanism involved in signaling via this receptor has been shown to occur in different cell types, including lymphocytes, hepatocytes, fibroblasts, phagocytes and thymocytes (Fanzo et al., 2003; Hueber et al., 2002). The clustering of CD95 and its ligand leads to recruitment of Fas Associated Death Domain (FADD) to the receptor death domain and initiates programmed cell death, mediated by activation of a family of proteases termed caspases (Salvesen and Dixit, 1997).

In this study, CD95 receptor was analysed as a modulator of CF by genotyping *CD95* polymorphisms in CF twin and sibling cohort stratified for disease severity and *P. aeruginosa* colonisation. The family based analysis to identify transmission disequilibrium among CF siblings at *CD95* locus revealed a significant transmission distortion of haplotype block encompassing entire intron 1 and intron 2 of *CD95* (Fig. 37) indicating *CD95* locus as a modulator in CF. Further haplotype analysis and fine mapping showed that the non-coding variant located within intron 2 of *CD95* was significantly associated with disease severity (Fig. 38 and 39) and also associated with differential expression of CD95 mRNA in rectal tissues of CF patients (Fig. 44), in which the low expressing genotypes were associated with severe CF disease. However, CD95 mRNA quantification by RT-PCR (Fig. 45) and CD95 surface expression analysis by FACS (Fig. 46) in peripheral blood mononuclear cells from healthy volunteers failed to show the similar association with the intron 2 SNP genotype. This outcome suggested that the intron SNP function is specific to cell type (only in epithelial cells) or specific to CF phenotype (hyperinflammatory). This finding is of high significance, as in CF scenario, it has been reported that the CD95 and CD95L expression was increased in lung sections obtained from CF patients colonised with *Staphylococcus aureus* or *P. aeruginosa* compared with uninfected control subjects (Durieu et al., 1999). Similarly, Cannon et al., (2003) reported that the expression of CD95/CD95L after infection of *P. aeruginosa* was CFTR dependent, in which *P.*

aeruginosa induced a delayed apoptotic response in cultured cells expressing mutant CFTR compared with cells expressing wild-type CFTR. The increased expression of CD95 and CD95 ligand on epithelial cells after infection by *P. aeruginosa* with active type III secretion system has also been shown by several other studies (Grassme et al., 2000; Jendrossek et al., 2001; 2003). These observations indicate the significance of CD95/CD95L mediated apoptosis as a host defense mechanism in epithelial cells against *P. aeruginosa* among CF patients.

The analysis of CD95 polymorphisms in a cohort stratified for *P. aeruginosa* colonisation did not show any evidence of association (Fig. 47). This would imply that the level of CD95 expression in CF patients is not solely determined by *P. aeruginosa* infection but the overall CF phenotype. It is shown that inflammation occurs very early in CF airways before any infection (Tirouvanziam et al., 2000). Hence, it is plausible that the proinflammatory cytokines which are specifically induced due to CF basic defect may stimulate the CD95 expression by activating tissue specific transcriptional regulators, for instance NF- κ B. In this regard, it is important to note that the polymorphism within intron 2 was located within a conserved non-coding sequence (CNS) and altered the binding site for NF- κ B family member c-Rel (Fig. 43). Additionally, the CNS within intron 2 also harboured composite sequence elements for several transcription factors, including a perfect consensus binding site “GGGRNNYYCC” for NF- κ B (Fig. 43). It is known that the NF- κ B members dimerize to form homo- or heterodimers, which are associated with specific responses to different stimuli and differential effects of transcription (Ghosh et al., 1998). Furthermore, experiments with cell lines deficient for single or multiple NF- κ B subunits revealed that the in vivo specificity of cellular gene activation does not only lie within the sequence of the κ B DNA site, but is also likely to be greatly influenced by combinatorial protein-protein interactions with other promoter-bound factors (Hoffmann et al., 2003). Accordingly, *CD95* was shown to be one of the several κ B-dependent genes as it harbours NF- κ B binding sites in its promoter (Hoffmann et al., 2003). Thus, it is reasonable to speculate that the interaction of NF- κ B members bound both in intronic enhancer of *CD95* and the promoter might determine the expression level.

Hence, it is necessary to validate the in-silico finding that c-Rel binding site with in enhancer region of *CD95* intron 2 determines the expression level of CD95 in CF patients. This issue can be resolved by performing experiments such as electrophoretic mobility shift assay (EMSA) or chromatin immuno-precipitation (ChIP) techniques to find out whether or not the transcription factor c-Rel binds to *CD95* intron. Furthermore, the question of cell specificity and CF specificity of the polymorphism can be answered by performing CD95 expression analysis (Real-time PCR and FACS) on both epithelial cells and peripheral blood mononuclear cells isolated from the CF patients as well as healthy individuals harbouring different genotypes at *CD95* intron SNP. Taken together, a combination of human genetic, comparative genomic and functional analyses cumulatively demonstrates that the *CD95* non-coding variants modulate the CF disease severity via variable CD95 gene expression.

4. Conclusions

The main objectives of this study were to identify the role of innate immunity genes such as toll like receptor-2, toll like receptor-4, toll like receptor-5, toll like receptor-9, *CD14*, Surfactant protein-D, IL-8 receptor 2 (*CXCR2*), tumour necrosis factor receptor-1 and *TNFRSF6A* (CD95) in CF disease severity modulation and modulation of susceptibility to *P. aeruginosa* infection in CF patients. In order to evaluate them as CF modulators, informative single nucleotide polymorphisms and DNA microsatellites within or in the vicinity to these genes were typed in European CF Twins and Sibs with extreme clinical phenotypes and unrelated CF patients stratified for *P. aeruginosa* early or late colonisation.

4.1. Single marker analysis showed no association between *TLR2*, *TLR5* and *TLR9* polymorphisms and CF disease severity

TLR2, *TLR5* and *TLR9* are receptors for bacterial lipoproteins, flagellin and unmethylated CpG motifs, respectively. These innate immunity genes were investigated as CF disease severity modulators by genotyping informative single nucleotide polymorphisms in CF twin and siblings. The allele and genotype distributions were compared between mildly affected and severely affected CF patient pairs. The single marker analysis on *TLR2*, *TLR5* and haplotype analysis on *TLR9* did not reveal any association with CF disease severity among CF twin and siblings at all tested loci in these three genes (Chapter 3.1). To conclude that these genes do not modulate CF more markers have to be analysed in the gene, as a gene can not be excluded as a modulator by single locus analysis using one SNP.

A minor association was found between surfactant protein-D, *CXCR2*, *PON* locus polymorphisms and CF disease severity

4.2. Surfactant protein-D and CF disease severity

Surfactant protein-D (SP-D) is a serum collectin and plays an important role in first-line defense of mucosal surfaces. SP-D is located on long arm of chromosome 10, 335kb downstream of another important innate immune collectin, surfactant protein-A. The allele and genotype distributions at a non-synonymous polymorphism within second exon of SP-D were compared between mildly and severely affected CF patient pairs. A

non-statistical significant association was found between disease severity and SP-D polymorphism (chapter 3.2.1). This finding points to a potential involvement of the SP-D locus in CF disease modulation. However, SP-A is also located close to SP-D. Hence, to discriminate between SP-A and SP-D, two adjacent equally plausible genes, more markers have to be typed.

4.3. *CXCR2* is a modulator of CF disease discordance

CXCR2 is a chemokine receptor expressed mainly on neutrophil surface. This receptor is responsible for the recruitment of neutrophils in response to IL-8. The comparison of allele and genotype distributions at a synonymous SNP, located within exon 3 of *CXCR2*, between mildly and severely affected patient pairs did not show any association, while the allele distribution between concordant and discordant pairs was significantly different (chapter 3.2.2). This finding indicated that trans-factors encoded elsewhere in the genome may bind differentially to cis-regulatory elements on *CXCR2* locus and in turn cause CF disease modulation. Further analysis of this polymorphism on CF cohort stratified for early or late colonisation of *P. aeruginosa* showed similar distribution of allele and genotypes between both groups. In order to identify the causative variant and to unravel the molecular mechanism, fine-mapping of the *CXCR2* locus is necessary.

4.4. Paraoxonase (PON) gene cluster as a modulator of CF disease severity

The *PON* gene cluster, *PON1*, *PON2* and *PON3*, is located 23cM distant from *CFTR* gene on chromosome 7. These genes codes for paraoxanase enzymes that are reported to be capable of degrading *P. aeruginosa* quorum sensing molecules termed homoserine lactones. The *PON* locus was investigated as a modulator of CF by genotyping a tri-nucleotide repeat polymorphism located 32kb upstream of *PON* locus. The allele distribution was found to be dissimilar between mildly and severely affected CF pairs (chapter 3.2.3). This finding indicates that *PON* genes modulate CF disease severity, albeit a denser map is necessary to differentiate the effect caused due to hitchhiking with the *CFTR* gene and due to the *PON* locus itself.

4.5. Haplotype block with in TNFR1 first intron modifies the CF disease severity

TNFR1 receptor binds TNF α proinflammatory cytokine and exerts proinflammatory response by activating NF- κ B and AP-1 transcription factors. The previous work from our lab identified *TNFR1* locus as a CF modulator by serendipity. The significant transmission distortion of the 6-marker haplotype block (rs740842-rs1800693-rs1800692-D12S889-rs767455-rs2228576), encompassing complete *TNFR1* gene, among CF offspring indicated that the *TNFR1* locus is a CF modulator. Hence, initially the complete *TNFR1* coding region was screened to identify the causative variants by sequencing the genomic DNA samples from individuals carrying contrasting *TNFR1* haplotypes. The sequencing results did not reveal any sequence variants within the coding region. However, fine-mapping of *TNFR1* locus revealed that the two-marker haplotype (D12S889-rs767455) distribution was significantly different between mildly and severely affected patient pairs. Furthermore, this two-marker haplotype (D12S889-rs767455) was located within intron 1 of *TNFR1* gene. Thus, the complete 7.5kb of intron 1 was analysed by sequencing to identify the causative variants. Sequencing analysis identified seven naturally occurring SNPs within intron 1. Interestingly, 7.5kb sequence of *TNFR1* intron 1 differed by haplotype at these seven SNPs. Thus, it was concluded that either one of the intron 1 SNPs or a haplotype composed of several allelic variants within intron 1 alters the functionality of TNFR1 (chapter 3.3.4).

This hypothesis was further strengthened by in-silico analysis. The functionality of the intron 1 variants at the haplotype level was predicted by the AltaVista Genome Browser and Genome Atlas software. Comparison of two contrasting haplotypes by these two tools revealed that the alterations in DNase hypersensitive sites, local inverted repeats and conserved non-coding sequences due to intron 1 variants might cause differential transcription of *TNFR1*. This notion was validated by comparing the soluble TNFR1 levels (both 55kDa- and 28kDa-forms) in serum from CF patients to their intron 1 haplotype. Thus, the western blot analysis confirmed the association between intron 1 haplotype and sTNFR1 levels in CF patients' serum, in which the levels of both 55-kDa as well as 28kDa forms of TNFR1 correlated with *TNFR1* intron 1 haplotype. Hence, it is necessary to investigate how the intron 1 variants of *TNFR1* determine the levels of soluble TNFR1 in serum. In conclusion, the *TNFR1* gene was identified as a modulator

of CF disease severity and the non-coding variants within intron 1 may modulate CF disease severity by determining the sTNFR1 levels in CF patients.

4.6. The *TLR4* promoter variants modulate CF disease severity but not age at onset of *P. aeruginosa* colonisation

TLR4, an innate immunity gene, transduces signals necessary for the production of inflammatory mediators by recognising the gram negative bacterial LPS. The *TLR4* locus was analysed as CF modulator by typing five SNPs and a di-nucleotide polymorphism in CF twin and siblings (chapter 3.4). The single marker analysis showed a significant difference in allele distribution between mildly and severely affected pairs. Sequencing of *TLR4* coding region did not reveal any sequence variants. Further two-marker haplotype analysis revealed a significant difference in the distribution of two-marker haplotype block located on the promoter of the *TLR4* gene between mildly and severely affected CF patient pairs. This finding suggested that the *TLR4* locus is a modulator of cystic fibrosis disease severity. The TLR4 expression on monocytes, isolated from healthy volunteers, was compared with their *TLR4* promoter genotype in an attempt to disclose the functionality of the SNP. We found no association with the TLR4 promoter SNP genotype and the expression level TLR4 on monocytes from healthy individuals. Next, the *TLR4* promoter polymorphism was tested in CF unrelated patients stratified for *P. aeruginosa* early or late chronic colonisation to investigate its role in determining the susceptibility to *P. aeruginosa* among CF patients. The allele and genotype distributions of the *TLR4* promoter polymorphism were similar in both *P. aeruginosa* early or late colonised pairs. Thus, it was concluded that the haplotype block on *TLR4* promoter plays a role in CF disease severity modulation but not age at onset of *P. aeruginosa* chronic colonisation.

4.7. The *CD14* polymorphisms determine both CF disease severity as well as the age at onset of *P. aeruginosa* chronic colonisation

CD14 is a receptor for several bacterial ligands including LPS. The CD14-LPS complex is recognised by TLR4/MD-2 complex to activate NF- κ B and induce inflammatory response. The *CD14* gene was evaluated as CF modulator by testing the *CD14* promoter polymorphism, the *CD14* 3' UTR polymorphism and the haplotype of these two

polymorphisms (chapter 3.5). The analysis of haplotype distribution revealed a significant difference between concordant and discordant CF patient pairs. The significance was stronger at the 3' UTR polymorphism. Thus it was concluded that the trans-factors, encoded elsewhere in the genome, bind *CD14* locus differentially and in turn modulate the CF disease severity. The sequencing of the complete *CD14* gene confirmed the promoter SNP and 3' UTR SNP but did not reveal any other sequence variants, indicating that either one or both of these SNPs were causative variants. This notion was further confirmed by the family based analysis in which the transmission disequilibrium was observed at these two loci between concordant and discordant CF patient pairs. However, the stronger distortion was seen at the *CD14* 3' UTR polymorphism. Thus, it was concluded that the *CD14* 3' UTR polymorphism mediates CF disease severity modulation.

Furthermore, the role of the *CD14* polymorphisms in determining the susceptibility to *P. aeruginosa* infection among CF patients was tested by comparing the transmission of genotype combinations among *P. aeruginosa* colonised patient pairs and *P. aeruginosa* non-colonised patient pairs. The transmission of genotype combination at the *CD14* promoter polymorphism was significantly different between *P. aeruginosa* colonised and non-colonised CF dizygous pairs. Testing these polymorphisms in a CF cohort stratified for the different age of birth indicated that heterozygotes at *CD14* promoter polymorphism were selected as favourable genotypes among late born CF patients. Additionally, the genotype distribution at *CD14* promoter polymorphism between *P. aeruginosa* early colonised and late colonised unrelated CF patients was also significantly different in which heterozygotes were overrepresented among late-colonised CF patients. This finding confirmed the role of *CD14* promoter polymorphism in determining the age dependent susceptibility to *P. aeruginosa* infection among CF patients. However, *CD14* promoter polymorphism and *CD14* 3' UTR polymorphism were in tight linkage. Thus, analysis of *CD14* diplotype for the age dependent risk to acquire *P. aeruginosa* among CF twin and siblings revealed that the combination of *CD14* polymorphisms determines the susceptibility to *P. aeruginosa*. Moreover, the early colonised *CD14* diplotypes were associated with lower amount of soluble CD14 levels in serum.

The *CD14* promoter polymorphism was shown to alter the binding site for Sp transcription regulators and thus regulate the level of *CD14* transcription, while nothing was known about the *CD14* 3' UTR polymorphism. In this study, two possibilities of

how the *CD14* 3' UTR polymorphism may regulate sCD14 levels were proposed. As *CD14* 3' UTR polymorphism was located within the mRNA cleavage and polyadenylation site, the first possibility was that the *CD14* 3' UTR polymorphism alters the binding efficiency of CD14 RNA to the proteins involved in mRNA processing. Secondly, the 3' UTR polymorphism was located within a microRNA binding site and thus the *CD14* 3' UTR polymorphism may vary the microRNA mediated CD14 regulation. In an effort to validate the microRNA mediated regulation, artificial microRNA detection protocol was standardised. In conclusion, the combinatorial effect of both the *CD14* 3' UTR polymorphism as well as the *CD14* promoter polymorphism determine the CF disease severity and age dependent risk to acquire *P. aeruginosa* colonisation among CF patients.

4.8. *CD95* is a potential modulator of CF disease severity

As *P. aeruginosa* infection is shown to trigger apoptosis via CD95 signaling, this apoptosis receptor was evaluated as CF modulator in this study (chapter 3.6). Genotyping the *CD95* SNPs showed a significant difference in genotype distribution between mildly and severely affected CF patient pairs, in which heterozygotes were overrepresented among severely affected pairs. Sequencing the complete coding region of *CD95* did not reveal any sequence variants. The transmission disequilibrium test showed a significant transmission distortion of haplotype block encompassing the first and second introns of *CD95*. Further fine-mapping by haplotype analysis showed a strong association of intron 2 SNP with disease severity. Sequencing analysis of complete haplotype sequence within intron 2 confirmed previously reported polymorphisms.

The *CD95* intron 2 SNP genotypes were compared against the quantity of *CD95* mRNA, isolated from rectal suction biopsies of CF patients and assessed by affymetrix chip. The intron 2 SNP was strongly associated with *CD95* mRNA in which, heterozygotes were expressing very low levels of *CD95*. On the other hand the *CD95* intron 2 SNP was not associated with the *CD95* expression level on peripheral blood mononuclear cells from healthy volunteers as tested by real-time PCR and FACS analysis. Hence, it was concluded that the role of intron 2 SNP is specific to epithelial cells or modulates *CD95* expression under CF context (pro-inflammatory phenotype). On the other hand, the *CD95* polymorphisms were not associated with *P. aeruginosa*

early or late colonisation among CF patients implying that CD95 modulates CF disease severity but not early or late colonisation of *P. aeruginosa* among CF patients.

Thus, to understand the functional role of non-coding variant within *CD95* intron 2 in determining the CD95 expression level, the intron 2 sequence was assessed by in-silico examination. The intron 2 SNP was located within a 100bp conserved non-coding sequence as revealed by AltaVista Genome Browser. This conserved non-coding sequence also found to be low-helical stability region indicating this sequence as a hot-spot for transcription factor binding. Consistently, the 100bp sequence within intron 2 of *CD95* had composite sequence elements for several transcription factors including NF- κ B and the intron 2 SNP altered the binding site for NF- κ B family member c-Rel transcription factor. In conclusion, CD95 is CF disease severity modulator and is mediated by differential expression of CD95 due to intron 2 SNP, possibly regulated by altered interaction of NF- κ B transcription factors. Further experiments such as electrophoretic mobility-shift assay or chromatin immunoprecipitation are necessary to validate the in-silico finding.

In summary, this study illustrated several important points. Firstly it demonstrated that the non-coding, naturally occurring polymorphisms on innate immunity and non-specific defense genes modulate the CF disease severity and susceptibility to *P. aeruginosa* infection. Secondly, the modulation via these genes is not exclusively due to *P. aeruginosa* infection but due to basic CF defect (pro-inflammatory phenotype). Finally, it indicated that relatively common genetic variation, with little or no overall phenotypic effect on the general population, can have significant effect in the context of CF.

4.9. Role of non-coding DNA in determining the susceptibility to infectious diseases

Susceptibility to infection arises from complex interplay between environment and host genetic factors. In general, many genetic loci contribute to human disease susceptibility and most of the focus in the field has been on identifying variants within the coding sequence as causative variants. The implications of a coding variant may be readily apparent as a change in the amino acid directly affects the protein. By contrast, the role of non-coding variations has been largely neglected either by considering them as “Junk DNA” or due to difficulties in assaying the functional effect of these polymorphisms. In this regard, the findings from this study greatly emphasize the importance of variations

within the non-coding sequence in determining the susceptibility to infectious disease. All four major CF modulators, characterised in this study, harboured their causative variants within their non-coding region. The functional consequences of these non-coding sequence variants were at different levels of gene regulation depending on their location, starting from the promoter variant to the 3' UTR variant. The fine-tuning of expression levels of these genes can be envisaged to result from; firstly, alterations in the transcription factor binding sites. Secondly, alterations in DNA conformations which are required for efficient splicing. Thirdly, alterations in the binding sites for proteins required for efficient splicing and mRNA processing and finally, alterations of microRNA target sites mediating gene regulation. However, irrespective of their location, the non-coding variants regulate the level of mRNA and in turn the quantity of protein made. Hence, this knowledge points to a target for therapy as the manipulations of regulatory polymorphisms that affect expression levels is easier than correcting the effects of abnormal protein. In this context, the findings reported upon in this thesis have added significant input for a clearer understanding of the functional significance of non-coding variants in determining the susceptibility to infectious diseases.

5. Literature

- Abelson et al., *Science* 310: 317-320, 2005
- Adamo et al., *Am J Respir Cell Mol Biol* 30: 627-634, 2004
- Aderka, *Cytokine Growth Factor Rev* 7: 231-240, 1996
- Akira and Takeda, *Nat Rev Immunol* 4: 499-511, 2004
- Altuvia et al., *Nucleic Acids Res* 33: 2697-2706, 2005
- Ambros et al., *RNA* 9: 277-279, 2003
- Anonymous, Cystic Fibrosis Genetic Analysis Consortium (CFGAC). *Am J Hum Genet* 47: 354-359, 1990
- Ashkenazi and Dixit, *Science* 281: 1305-1308, 1998
- Backhed et al., *Microbes Infect* 5: 1057-1063, 2003
- Baldi et al., *J Mol Biol* 263: 503-510, 1996
- Bartel, *Cell* 116: 281-297, 2004
- Bear et al., *Cell* 68: 809-818, 1992
- Becker and Knapp, *Am J Hum Genet* 74: 589-591, 2004
- Becker and Knapp, *Am J Hum Genet* 75: 561-570, 2004
- Becker and Knapp, *Genet Epidemiol* 27: 21-32, 2004
- Becker et al., *Ann Hum Genet* 69: 747-756, 2005
- Bell et al., *Trends Immunol* 24: 528-533, 2003
- Beutler, *Curr Opin Immunol* 12: 20-26, 2000
- Beutler, *Nature* 430: 257-263, 2004
- Bonfield et al., *Am J Respir Crit Care Med* 152: 2111-2118, 1995
- Bonfield et al., *J Allergy Clin Immunol* 104: 72-78, 1999
- Bonten et al., *Am J Respir Crit Care Med* 160: 1212-1219, 1999
- Bronsveld et al., *Gastroenterology* 119: 32-40, 2000
- Brukner et al., *EMBO J* 14: 1812-1818, 1995
- Brukner et al., *J Biomol Struct Dyn* 132: 309-317, 1995
- Buttke, et al., *Immunol Today* 15: 7-10, 1994
- Cacalano et al., *Science* 265: 682-684, 1994
- Cannon et al., *Am J Respir Cell Mol Biol* 29: 188-197, 2003
- Cartegni et al., *Nat Rev Genet* 3: 285-298, 2002
- Casanova and Abel, *J Exp Med* 202: 197-201, 2005
- Castanos-Velez et al., *Infect Immun* 66: 2960-2968, 1998
- Chabanon et al., *Brief Funct Genomic Proteomic* 3: 240-256, 2004
- Chemenev et al., *Nucl Acids Res* 33: W432-W437, 2005
- Chen et al., *Hum Genet* 120: 1-21, 2006
- Chen et al., *Nucleic Acids Res* 23: 2614-2620, 1995
- Chmiel et al., *Clin Rev Allergy Immunol* 23: 5-27, 2002
- Chuntharapai et al., *J Immunol* 153: 5682-5688, 1994
- Clark and Reid, *Arch Dis Child* 88: 981-984, 2003
- Clark et al., *Microbes Infect* 2: 273, 2000
- Cobb et al., *J Immunol* 173: 5659-5670, 2004
- Cohen and Eisenberg, *Annu Rev Immunol* 9: 243-269, 1991
- Colgan and Manley, *Genes and Development* 11: 2755-2766, 1997
- Conne et al., *Nat Med* 6: 637-641, 2000
- Corey et al., *Am J Epidemiol* 143 : 1007-1017, 1996
- Crouch and Wright, *Annu Rev Physiol* 63: 521, 2001
- Cutting, *Annu Rev Genomics Hum Genet* 6: 237-260, 2005
- Deckert-Schluter et al., *J Immunol* 160: 3427-3436, 1998

Denli et al., *Nature* 432: 231-235, 2004
 Denning et al., *Nature* 358: 761-764, 1992
 Derichs et al., *Pediatr Res* 55: 69-75, 2004
 Devaney et al., *FEBS Letters* 544: 129-132, 2003
 Diggle et al., *J Bacteriol* 184: 2576, 2002
 Durieu et al., *Thorax* 54: 1093-1098, 1999
 Döring and Worlitzsch, *Paediatr Respir Rev* 1: 101-106, 2000
 Eder et al., *J Allergy Clin Immunol* 113: 482-488, 2004
 Emerson et al., *Pediatr Pulmonol* 34: 91-100, 2002
 Ernst et al., *J Endotoxin Res* 9: 395-400, 2003
 Ernst et al., *Science* 286: 1561-1565, 1999
 Ewens and Spielman, *Am J Hum Genet* 57: 455-464, 1995
 Fan and Malik, *Nat Med* 9: 315-321, 2003
 Fanzo et al., *Cancer Biol Ther* 2: 392-395, 2003
 Firoved et al., *Infect Immun* 72: 5012-5018, 2004
 Fisher et al., *Cell* 81: 935-946, 1995
 Fitzgerald et al., *J Biol Chem* 269: 21303-21314, 1994
 Floros et al., *J Infect Dis* 182: 1473-1478, 2000
 Frazer et al., *Nucleic Acids Res* 32: W273-W279, 2004
 Frey et al., *J Exp Med* 176: 1665-1671, 1992
 Fuchs et al., *Genomics* 13: 219-224, 1992
 Fukunaga, et al., *J Immunol* 148: 1274, 1992
 Gehring et al., *Nat Genet* 28: 389-392, 2001
 Gewirtz et al., *Am J Physiol gastrointest Liver Physiol* 290: G1157-G1163, 2006
 Ghosh et al., *Annu Rev Immunol* 16: 225-260, 1998
 Gibson et al., *Am J Respir Crit Care Med* 168: 918-951, 2003
 Govan and Deretic, *Microbiol Rev* 60: 539-574, 1996
 Grassme et al., *J Biol Chem* 276: 20589-20596, 2001
 Grassme et al., *Science* 290: 527-530, 2000
 Griffiths-Jones, *Nucleic Acids Res* 32: D109-D111, 2004
 Grishok et al., *Cell* 106: 23-34, 2001
 Gulbins et al., *J Mol Med* 82: 357-363, 2004
 Hailman et al., *J Immunol* 156: 4384-4390, 1996
 Hajjar et al., *Nat Immunol* 3: 354-359, 2002
 Hallin and Ussery, *Bioinformatics* 20: 3682-3686, 2004
 Hancock, *Clin Infect Dis* 1: S93-S99, 1998
 Harel et al., *Nat Struct Mol Biol* 11: 412-419, 2004
 Hartl and Griese, *Eur J Clin Invest* 36: 423-435, 2006
 Hauber et al., *Can Respir J* 12: 13-18, 2005
 Haussler et al., *Clin Infect Dis* 29: 621-625, 1999
 Hawari et al., *Proc Natl Acad Sci USA* 101: 1297-1302, 2004
 Hawn et al., *J Exp Med* 198: 1563-1572, 2003
 Hawn et al., *Proc Natl Acad Sci USA* 102: 10593-10597, 2005
 Hayashi et al., *Nature* 410: 99-103, 2001
 He et al., *Proc Natl Acad Sci USA* 101: 2530-2535, 2004
 Heidinger et al., *Immunogenetics* 57: 1-7, 2005
 Hemmi et al., *Nature* 408: 740-745, 2000
 Hoffmann et al., *EMBO J* 22: 5530-5539, 2003
 Hoppe and Reid, *Structure* 2:1129, 1994
 Hoshino et al., *J Immunol* 162: 3749-3752, 1999

Huang and Kowalski, *Nucl Acids Res* 31: 3819-3821, 2003
Huang et al., *Invest Ophthalmol Vis Sci* 46: 4209-4216, 2005
Huang et al., *Mol Immunol* 34: 577-582, 1997
Hueber et al., *EMBO Rep* 3: 190-196, 2002
Hutvagner et al., *Science* 293: 834-838, 2001
Hybiske et al., *Cell Microbiol* 6: 49-63, 2004
Itoh et al., *Cell* 66: 233-243, 1991
Jendrossek et al., *Infect Immun* 69: 2675-2683, 2001
Jendrossek et al., *Infect Immun* 71: 2665-2673, 2003
Johannsen et al., *Lancet* 337: 631-634, 1991
Karp et al., *Nat Immun* 5: 388-392, 2004
Kerem et al., *Science* 245 : 1073-1080, 1989
Kerem et al., *N Engl J Med* 323: 1517-1522, 1990
Khan et al., *Am J Respir Crit Care Med* 151: 1075-1082, 1995
Knapp and Becker, *Hum Hered* 56: 2-9, 2003
Knorre, *Biol Chem* 383: 271-282, 2002
Knowles, *Curr Opin Pulm Med* 12: 416-421, 2006
Kobe and Deisenhofer, *Nature* 374: 183-86, 1995
Koch and Høiby, *Lancet* 341: 1065-1069, 1993
Kogan et al., *EMBO J* 22: 1981-1989, 2003
Kollias et al., *Immunol Rev* 169: 175-194, 1999
Köhler et al., *Mol Microbiol* 23: 345-354, 1997
Lahti et al., *Pediatr Res* 51: 696-699, 2000
Laird and Lange, *Nat Rev Genet* 7: 385-394, 2006
Lazarus et al., *Immunol Rev* 190: 9-25, 2002
Lee et al., *Nature* 425: 415-419, 2003
Lesprit, *Am J Respir Crit Care Med* 167: 1478-1482, 2003
Li et al., *J Mol Med* 81: 766-779, 2003
Lien et al., *J Biol Chem* 274: 33419-33425, 1999
Lien et al., *J Clin Invest* 105: 497-504, 2000
Lillehoj et al., *Am J Physiol Lung Cell Mol Physiol* 287: L809-L815, 2004
Mahenthalingam et al., *Infect Immun* 62: 596-605, 1994
Makris et al., *Thromb Haemost* 78: 1426-1429, 1997
Mariencheck et al., *Am J Respir Cell Mol Biol* 28: 528-537, 2003
Martin et al., *Am J Respir Crit Care Med* 155: 937-944, 1999
Martin et al., *J Clin Invest* 90: 2209-2219, 1992
Matheson et al., *J Hum Genet* 51: 196-203, 2006
Mayor et al., *Bioinformatics* 16: 1046, 2000
McDermott et al., *Cell* 97: 133-144, 1999
Medzhitov, *Nat Rev Immunol* 1: 135-145, 2001
Mekus et al., *Electrophoresis* 16: 1886-1888, 1995
Mekus et al., *Hum Genet* 112: 1-11, 2003
Mekus et al., *Twin Res* 3: 277-293, 2000
Meredith et al., *Mol Biol Cell* 4: 953-961, 1993
Mignone et al., *Genome Biol* 3: 4, 2002
Miller et al., *Nat Rev Microbiol* 3: 36-46, 2005
Mizel et al., *J Biol Chem* 278: 23624-23629, 2003
Mockenhaupt et al., *J Infect Dis* 194: 184-188, 2006
Muhlebach et al., *Am J Respir Crit Care Med* 160: 186-191, 1999
Nagata and Golstein, *Science* 267: 1449-1456, 1995

Newton-Cheh et al., *Mutat Res* 573: 54-69, 2005
 Noah et al., *Am J Respir Crit Care Med* 168: 685-691, 2003
 O'Brien et al., *Infect Immun* 67: 595-601, 1999
 Ogun et al., *Eur Respir J* 23: 219-223, 2004
 Ozer et al., *FEMS Microbiol Lett* 253: 29-37, 2005
 Pearson, *Infect Immun* 68: 4331-4334, 2000
 Philippon et al., *Antimicrob Agents Chemother* 41: 2188-2195, 1997
 Pickering and Willis, *Semin Cell Dev Biol* 16: 39-47, 2005
 Pierik et al., *Inflamm Bowel Dis* 12: 1-8, 2006
 Poole et al., *Mol Microbiol* 21: 713-724, 1996
 Poort et al., *Blood* 88: 3698-3703, 1996
 Prader et al., *Helv Paediatr Acta Suppl* 52: 1-125, 1989
 Primo-Parmo, *Genomics* 33: 498-507, 1996
 Pritchard and Przeworski, *Am J Hum Genet* 69: 1-14, 2001
 Pugin et al., *Proc Natl Acad Sci USA* 90: 2744-2748, 1993
 Ran et al., *Proc Natl Acad Sci USA* 100: 14315-14320, 2003
 Reddy, Quinton PM *J. Pancreas* 2: 212-218, 2001
 Rehmsmeier et al., *RNA* 10: 1507-1517, 2004
 Riordan et al., *Science* 245 : 1066-1073, 1989
 Risch, *Nature* 405: 847-856, 2000
 Sabroe et al., *Eur Respir J* 19: 350-355, 2002
 Salvesen and Dixit, *Cell* 14: 443-446, 1997
 Schmiegelow et al., *Clin Genet* 29: 374-377, 1986
 Schork, et al., *Clin Genet* 58: 250-264, 2000
 Schwartz et al., *J Clin Invest* 100: 68, 1997
 Sham and Curtis, *Ann Hum Genet* 59: 97-105, 1995
 Shimazu et al., *J Exp Med* 189: 1777-1782, 1999
 Shuai, *Curr Opin Cell Biol* 6: 253-259, 1994
 Shuto et al., *The FASEB Journal express article*, 2006
 Silverstein, *Nat Immunol* 4: 3-6, 2003
 Skerrett et al., *Am J Physiol Lung Cell Mol Physiol*, 2006
 Smith and Iglewski, *Curr Opin Microbiol* 6: 56-60, 2003
 Smith et al., *J Bacteriol* 184: 1132-1139, 2002
 Smith et al., *J Immunol* 167: 366-374, 2001
 Soong et al., *J Clin Invest* 113: 1482-1489, 2004
 Spielman et al., *Am J Hum Genet* 52: 506-516, 1993
 Steinshamn et al., *J Immunol* 157: 2155-2159, 1996
 Strieter et al., *J Clin Invest* 109: 699-705, 2002
 Taggart et al., *Am J Physiol Lung Cell Mol Physiol* 278: L33-L41, 2000
 Tang, *Infect Immun* 64: 37-43, 1996
 Telford, *Infect Immun* 66: 36-42, 1998
 Tirouvanziam et al., *Am J Respir Cell Mol Biol* 23: 121-127, 2000
 Tomari and Zamore, *Genes Dev* 19: 517-529, 2005
 Tomashefski Jr et al., *Am J Clin Pathol* 91: 522-530, 1989
 Tricia et al., *J Immunol* 167: 5838-5844, 2001
 Tsai et al., *Infect Immun* 68: 4289-96, 2000
 Tümmler and Kiewitz, *Mol Med Today* 5: 351-358, 1999
 Tümmler et al., *Behring Inst Mitt* 98: 249-255, 1997
 Underhill et al., *Nature* 401: 811-815, 1999
 van den Oord and Neale, *Mol Psychiatry* 9: 227-36, 2004

Vankeerberghen et al., *J Cystic Fibrosis* 1: 13-29, 2002
Wajant et al., *Cell Death Differ* 10: 45-65, 2003
Wang et al., *J Biol Chem* 278: 21751-21760, 2003
Watanabe-Fukunaga et al., *Nature* 356: 314-317, 1992
Welsh et al., *Cell* 73 : 1251-1254, 1993
Welsh et al., *McGraw-Hill, New York* 3799-3876, 1995
Welsh et al., *Neuron* 8: 821-829, 1992
West et al., *Annu Rev Cell Dev Biol* 22: 409-37, 2006
Whiteley et al., *Proc Natl Acad Sci USA* 96: 13904, 1999
Wu, *Microbiol* 147: 1105-1113, 2001
Xanthoulea et al., *J Exp Med* 200: 367-376, 2004
Yamamoto et al., *Mol Immunol* 40: 861-868, 2004
Yang et al., *FEBS Lett* 579: 3713-3717, 2005
Zhang et al., *Infect Immun* 73: 7151-7160, 2005
Zhao et al., *Microbiol Mol Biol Rev* 63: 405-445, 1999
Zielinski et al., *Genomics* 10: 241-248, 1991

6. Abbreviations

bp	Base pair
CF	cystic fibrosis
CFTR	cystic fibrosis transmembrane conductance regulator
CON ⁻	concordant severe CF phenotype
CON ⁺	concordant mild CF phenotype
DIS	discordant CF phenotype
DZ	dizygous twins
MeOH	Methanol
MZ	monozygous twins
ND	nondiscordant CF phenotype
nt	Nucleotides
PCR	polymerase chain reaction
PIC	polymorphism information content
RFLP	restriction fragment length polymorphism
SNP	single nucleotide polymorphism
TEMED	N,N,N',N'-tetramethyl-ethylene-diamine

7. Appendices

Material and Equipment

Membranes, separation equipment, X-ray films

Amicon Microcon YM-100 centrifugal filter device	Millipore, Billerica, MA, USA
Filters and blotting paper	Schleicher & Schuell, Einbeck, D
Hybond N ⁺ Nylon membrane	Amersham, Freiburg, D
Kodak X-o-Mat AR X-ray films	Eastman Kodak, New Haven, CT, USA

Technical equipment

Autoclave	Tuttnauer, Breda, NL
Centrifuges:	
- Beckman GS-15R	Beckman Coulter, Fullerton, CA, USA
- Beckman Microfuge R	Beckman Coulter, Fullerton, CA, USA
- Eppendorff Mini Spin Plus	Eppendorff, Hamburg, D
Direct blotter:	
- GATC1500 direct blotting electrophoresis sequencer	GATC, Konstanz, D
Gel chambers for agarose electrophoresis	Forschungswerkstätten, Medical School, Hannover
Gel documentation system:	
Gel-Print 2000	Biophotonics, Ann Arbor, MI, USA
Mechanical pipettes	Eppendorff, Hamburg, D Gilson, Villier le Bel, F SLG, Gauting, D
Multiwell pipettes	Biohit, Köln, D Biozym, Hess. Oldendorf, D Eppendorff, Hamburg, D Hitachi, Tokyo, J
Photometer	BioRad, Hercules, CA, USA
Power supplies	Sartorius, Göttingen, D
Scales	
Thermocycler:	
- Hybaid Omnigene	Hybaid, Teddington, UK
- Landgraf	Landgraf, Laatzen, D
Thermomixer	Eppendorff, Hamburg, D
UV transilluminator	Bachofer, Reutlingen, D

Consumables

Filter pipette tips (sterile)
Latex exam gloves (powder free)
Multiwell plates
Pipette tips

Reaction vials

Greiner Bio-One, Frickenhausen, D
Kimberley Clark, Zaventem, B
Greiner-Bio one
Eppendorff, Hamburg, D
Sarstedt, Nümbrecht, D
SLG, Gauting, D
Eppendorff, Hamburg, D
Sarstedt, Nümbrecht, D

Chemicals

General chemicals

Acrylamide:

- AccuGel 19:1
USA

- Rotiphorese Gel 40

Agarose ultra pure

Ammoniumperoxodisulfate (APS)

Boric acid, crystalline(H₃BO₃)

Bromphenolblue

Diethyl-pyrocabonate (DEPC)

DTT

Chloroform p.a.

Acetic acid, 96%

Ethanol

Formamide

Glycerol

Isoamylalcohol

Isopropanol

KCl, p.a.

MgCl₂ x 6 H₂O, p.a.

NaCl, p.a.

NaOH, p.a.

Na₂-EDTA

Paraffin

Phenol:

- Roti- Phenol pH 7.5 - 8

- peq- Gold Trifast

TEMED

Tris ultra pure

Triton X 100

Urea, pearls, purest

Xylenecyanol

National Diagnostics Atlanta, Georgia,

Roth, Karlsruhe, D

Life Technologies, Paisley, Scotland, UK

Roth, Karlsruhe, D

Merck, Darmstadt, D

Serva, Heidelberg, D

Sigma-Aldrich, Steinheim, D

ICN, Aurora, OH, USA

Roth, Karlsruhe, D

Merck, Darmstadt, D

JT Baker, Deventer, NL

Sigma-Aldrich, Steinheim, D

Fluka, Buchs, CH

Roth, Karlsruhe, D

Fluka, Buchs, CH

Merck, Darmstadt Germany

Merck, Darmstadt, D

Merck, Darmstadt, D

Merck, Darmstadt, D

Merck, Darmstadt, D

Fluka, Buchs, CH

Roth, Karlsruhe

peqLab, Erlangen

Serva, Heidelberg, D

ICN, Aurora, Ohio, USA

Serva, Heidelberg, D

Merck, Darmstadt, D

Sigma, St. Louis, MO, USA

Buffers and standard solutions for direct electrophoresis

Developing buffer I:

100 mM	Tris-HCl	pH 7.5
150 mM	NaCl	

Developing buffer III:

100 mM	Tris-HCl	pH 9.5
100 mM	NaCl	
50 mM	MgCl ₂	

Loading buffer I (agarose gels):

0.05% (w/v) xylene cyanol, 0.05% (w/v) bromphenolblue, 40% (w/v) glycerol

Loading buffer II (direct blotter):

0.2% (w/v) xylene cyanol, 0.2% (w/v) bromphenolblue in formamide

Lysis buffer:

50 mM	Tris-HCl	pH 7.5
109.5 g/l	saccharose	
1% (w/v)	Triton X- 100	

STE:

50 mM	Tris-HCl	pH 7.5
100 mM	NaCl	
1 mM	Na ₂ -EDTA	

10 x TBE:

1.275 M	Tris-HCl	pH 9.0
0.42 M	H ₃ BO ₃	
0.024 M	Na ₂ -EDTA	

TE:

10 mM	Tris-HCl	pH 8.0
1 mM	Na ₂ -EDTA	

Buffers for western blotting

Separating gel buffer

1.5M	Tris-HCl	pH 8.8
------	----------	--------

Stacking gel buffer

1M	Tris-HCl	pH 6.8
----	----------	--------

Separating gel (10%)

Water	12.3ml
Acrylamide (30%)	10.3ml
1.5M Tris	7.7ml
10% SDS	310 μ l
TEMED	23 μ l
APS	310 μ l

Stacking gel

Water	7.8ml
Acrylamide (30%)	1.95ml
Tris (pH 6.8)	1.45ml
10% SDS	115 μ l
TEMED	11.5 μ l
APS	310 μ l

3X loading buffer

SDS	6%
Glycerin	30%
Tris-HCl	150mM
Bromophenol blue	0.02%

Running buffer (10X; pH 8.1-8.4)

Tris	30g
Glycine	144g
SDS	10g
Water	Make up to 1 lit.

Transfer buffer (1X)

Tris	25mM
Glycine	192mM
MeOH	10%

Biochemicals, enzymes and enzyme-related buffers, photo-chemicals

AMV reverse transcriptase	Stratagene, La Jolla CA, USA
Anti- Biotin- AP, Fab- fragments	Roche Diagnostics, Mannheim, D
Blocking reagent	Roche diagnostics Mannheim, D
CDP Star	Tropix, Bedford, MA, USA
Developing solution for X-ray films	Agfa-Gevaert, Mortsels, B
Desoxynukleotides (dNTPs)	Roth, Karlsruhe, D
Express fixative salt	Tetenal, Norderstedt, D
Luminescence-enhancer Sapphire II	Tropix, Bedford, MA, USA
Oligonucleotides	Invitrogen, Karlsruhe, D
	Life Technologies, Paisley, UK
	MWG-Biotech, Ebersberg, D
	Roth, Karlsruhe, D
Restriction enzymes	New England Biolabs, Frankfurt/ Main, Germany
Germany	
RNAasin Ribonuclease inhibitor	Promega, Madison, WI, USA
Taq polymerase	InvitTek, Berlin, D

dNTPs:

100 mg	dNTP (A, G, C or T)
+ 100 μ l	Tris-HCl [1M] pH 7.0
+ 900 μ l	H ₂ O _{STE} → check pH, add more Tris-HCl if necessary
ad 1.1 ml	H ₂ O _{STE}

2 μ M of each dNTP was put together and filled up to 1 ml with H₂O_{STE}, resulting in a dNTP pre-mix with a concentration of 2mM of each dNTP

PCR 10x NH₄⁺ reaction buffer (Invittek):

500 mM	Tris-HCl pH 8.8
160 mM	(NH ₄) ₂ SO ₄
0.1 %	Tween® 20

Recipe for testing TLR2 polymorphisms:

SNP NCBI ID	Primer sequence (5' to 3')	Restriction enzyme	Informative and tested on Twin and sibs?	Product size: Fragment size (bp)
rs3804099	TGCTGGACTTACCTTCCTTGA CAAACATTCCACGGAAGTTG	HpyCH IV	Yes	172: 87+85
rs5743705	ACTTCATTCTGGCAAGTGG GAATGAGAATGGCAGCATCA	Hpy 188 III	No	174: 89+85
rs1898830	GAATAGTAAAATAAATCCCCA TGTCTTGCCAGAGGTTTCATC	Bsl I	No	147: 122+25

Recipe for testing TLR5 polymorphisms:

SNP NCBI ID	Primer sequence (5' to 3')	Restriction enzyme	Informative and tested on Twin and sibs?	Product size: Fragment size (bp)
rs2072493	GGAACCAGCTCCTAGCTCCT AAACCCAGAGAACGAGTCA	MfeI	No	189: 104+85
rs5744174	ACAGCCCAGAGACTGGTGTT GATAGCATCCTGGATATTGG	BclI	No	228: 193 + 35
rs1861172	AGATAAGAGGTGGCCCCAAA GCTGAAAAGGTAGGTTGGTGA	Hpy CH4 111	Yes	196: 138 + 58

Recipe for testing TLR9 polymorphisms:

SNP NCBI ID	Primer sequence (5' to 3')	Restriction enzyme	Informative and tested on Twin and sibs?	Product size: Fragment size (bp)
rs352140	TTGGCTGTGGATGTTGTTGT AAGCTGGACCTCTACCACGA	Bst U I	Yes	177: 135+42
rs187084	TGTACTGGATCCTGGGGATG GAGCTCCTTTGCCTGGTCTA	Hpy188 III	Yes	168: 126+42

Recipe for testing Surfactant protein-D polymorphisms:

SNP NCBI ID	Primer sequence (5' to 3')	Restriction enzyme	Informative and tested on Twin and sibs?	Product size: Fragment size (bp)
rs721917	GAAGACCTACTCCCACAGAGCAA TTGGGAGGAAGAAACACGTC	BsrD I	Yes	248: 221+27
rs1998374	CCATTCCTGGATCACCCT TTGGTCCAGGTTCTCCAAC	Hpy188 III	No	250: 186+64

Recipe for testing CXCR2 polymorphism:

SNP NCBI ID	Primer sequence (5' to 3')	Restriction enzyme	Informative and tested on Twin and sibs?	Product size: Fragment size (bp)
rs2230054	GCTGTCGTCCTCATCTTCCCG AGTCCATGGCGAAACTTCTG	BsrBI	Yes	212: 191 + 21

Recipe for testing PON polymorphisms:

Repeat ID	Primer sequence (5' to 3')	Informative
TR7V899	GCCTGGCTGACAGAGTAAGA CCTGTTGAAAACCTGTGCTT	NO
TR7V900	TGCAAATAAGATTAAGTACCCTTTG CAGTACTCAGGCAGGTGAA	Yes
TE7V901	AATTAGCCCAGTGTGGTGGT CCCTTCTTCCCTTCTTTCG	NO
TE7V902	CCAGCTTGCCAAACAATACA GGAGGCCTAAGCCAGAAGAT	NO

Recipe for testing TLR4 polymorphisms:

SNP NCBI ID	Primer sequence (5' to 3')	Informative and tested on Twin and sibs?	Restriction enzyme	Product size: Fragment size (bp)
rs5030718	CATCTTCAATGGCTTGCCA ACCTGGAGGGAGTTCAGACA	No	Taq I	249: 225+24
rs5030713	GAGAACTTCCCATTGGACA TCATAGGGTTCAGGGACAGG	No	BstB I	223: 138+85
rs1927914	TGGGATTAATGAACTGGCATT ACAAAATGGTCCCTCACAGC	Yes	SphI	202: 116+86
rs11536891	GGGTGTGTTCCATGTCTCA GCATAAGGGATAAGGGGAGA	Yes	Hha I	246: 43+203
rs10759930	GGGTGCACTCACTCACCTCT CCTTGGACACCCATTACCAG	Yes	Dra I	255 : 165 + 90
rs2149356	TGACTGGTAAATATCCATTTCAGAGA TTTCCACAAAACCTCGCTCCT	Yes	Tsp509 I	163: 130+33
rs1927911	TCCATATCATTGGGGAGACTG TGGAATCCATGCACTCTAAA	Yes	BsaJ I	157: 87 + 70

Oligonucleotides employed for sequencing the TLR4 gene

Region	Primer sequence (5' to 3')	Fragment size
Promoter and Exon 1	GCAGCCCCAGCAAATAAT AGTTCTGGGCAGAAGTGAGG	611 bp
Exon 2	TGTGTGTCATCCTTGTGCAG CTCCCAACTCCCCTCCTA	499 bp

Recipe for testing CD14 polymorphisms

SNP NCBI ID	Primer sequence (5' to 3')	Restriction enzyme	Informative and tested on Twin and sibs?	Product size: Fragment size (bp)
rs2569190	CCTCCCCACCTCTCTTCCT CACCCACCAGAGAAGGCTTA	AVaII	Yes	206 : 78 + 128
rs2563298	CCGCAGTCTTTTCTTGAGG CGTCAGGACGTTGAGGACTT	Sau96I	Yes	217: 145+72bp

Oligonucleotides employed for sequencing the complete CD14 gene

Targets	Category	Sequence (5' to 3')	Product size (bp)
Target 1	Terminal primer left	ATTTGGTGGCAGGAGATCAA	2111
	Terminal primer right	ATCTGCTCAGAAAGCCCTGA	
	1 internal primer left	AAGAGAGGTGGGGAGGTGAT	599
	1 internal primer right	TAGCTGAGCAGGAACCTGTG	
	2 internal primer left	GTATGCTGACACGGTCAAGG	570
	2 internal primer right	CTCGGAGCGCTAGGGTTTAC	
	3 internal primer left	GCAACACAGGAATGGAGACG	700
	3 internal primer right	TTCTTGAGGAGGACAGATAGGG	
Target 2	Primer left	AAAAGGAAGGGGGAATTTTTC	596
	Primer right	GCTTCCAGGCTTCACACTTG	

Recipe for testing CD95 polymorphisms:

SNP NCBI ID	Primer sequence (5' to 3')	Informative and tested on Twin and sibs?	Restriction enzyme	Product size: Fragment size (bp)
rs1800682	ATATAGCTGGGGCTATGCGA CTCAGAGAAAGACTTGCGGG	Yes	ScrF I	216: 141+75
rs1324551	ATCTGCAAGCTGGCATTCT ACTCCCATCGTGATTTCTGC	Yes	Hph I	145: 100+45
rs2147420	TTCTGTCTCTGATGAAATCTGG ACAGCGCAATGAGATCCTAAA	Yes	HpyCH4 IV	180: 121+59
rs2296603	AAATTTATCCATAACCACATCAAAT TTTTACAGTTTTTGGTTCCCCTA	Yes	Nla III	127: 95+32
rs7901656	TCTGGGAATCTCCAGTTTGTTT GCTCTGCTCACCTATACAGCAA	Yes	HpyCH4 IV	199: 148+51
rs1571019	GATCTTTT TAGGCAGGAGTTCTGT ACCTGCTCAGCATAAAGCAT	Yes	Rsa I	164: 111+53

Oligonucleotides employed for sequencing the CD95 gene

Region	Primer sequence (5' to 3')	Fragment size
Exon 1	ACGAACCCTGACTCCTTCCT CCTATCCCCGGGACTAAGAC	471 bp
Exon 2	GTGGAGCCCTCACATTGTCT AACCACATCAAATAAGCGTGA	485 bp
Exon 3	GCTTTTGTCTTGGGAGACTTTC CAGTAGTTAGCTCGGCACCTG	505 bp
Exon 4	AAATGATTGCTGGCCATTTC GCTTTCCTTGACTGTCTGTGC	566 bp
Exon 5 and 6	ATTTTTCATCCAGCCATCCA GAATGAGGCAAATCTTTGTGAA	542 bp
Exon 7	GTTCCAAAATCAGCGGTCTC TGGGCTATGGAGCAAGACTC	381 bp
Exon 8	TTGCTTAGTTTCTGGCAAGG AATGCTTTATGCTGAGCAGGT	387 bp

Oligonucleotides employed for sequencing the CD95 intron 2, fragment 1

	Primer sequence (5' to 3')	Fragment size
Target 1; terminal primers	AAGCCCTGTCTCCAGGTAT TGCCGTTATTTTGTTCATT	2422 bp
Internal primers set 1	CAAGTGACATGCACCTCTGAA GGGTGGGTTATAGGGGAGA	639 bp
Internal primers set 2	AGCACATGGGTTTCATGTGATAG GCAAGTATAAACAGTGGGTTGG	695 bp
Internal primers set 3	GGGTACACATGCACAATCTG TCACATCCATAATGCATCACC	884 bp

Oligonucleotides employed for sequencing the CD95 intron 2, fragment 2

	Primer sequence (5' to 3')	Fragment size
Target 2; terminal primers	GGTGATGCATTATGGATGTGA TGGAAGAAAAATGGGCTTTG	2543 bp
Internal primers set 1	CCTCCCCACATCATCATATTC CAGGTTCTTTCGTTCTCCA	677 bp
Internal primers set 2	CCTGCCTTTTCCAACATTC ACTTCACCTTCTTGGGATGC	759 bp
Internal primers set 3	CATCTAGCTTTCCCATAGCA TTGGAGAAAGGGGAGAGGTT	745 bp

Curriculum Vitae

Vinod Kumar Magadi Gopalaiah

Klinische Forschergruppe
Medizinische Hochschule Hannover
Carl-Neuberg Str.2
Hannover-30625
Germany
Phone: +49-511-5326722
vinodkumarmg31@yahoo.co.in, kumar.vinod@mh-hannover.de

Geburtsdatum: 1 st June 1977	Geschlecht: Männlich	Nationalität: Indisch
--	-----------------------------	------------------------------

Ausbildung :

- **1999-2002 M.F.Sc. Master of Fishery Sciences (Microbiology), University of Agricultural Sciences, College of fisheries, Mangalore, India.** Dissertation project on Luminescent Vibriosis and its control using bacteriophage as a Biocontrol tool.
- **1995-1999 B.F.Sc. Bachelor of Fishery Sciences, University of Agricultural Sciences, College of Fisheries, Mangalore, India.**
- **1993-1995 PUC. Pre-University. Sree Siddaganga College of Arts, Science and Commerce, Tumkur, India.**
- **1990-1993 SSLC. Secondary School Leaving Certificate. Government Pre-University College, Magadi, Bangalore (R) Dist. India.**
- **1983-1990 Primary School. Government Higher-Primary School, Hosapete, Magadi, Bangalore (R) Dist. India.**

Publikationen :

- Stanke F, Becker T, Cuppens H, **Kumar V**, Cassiman JJ, Jansen S, Radojkovic D, Siebert B, Yarden J, Ussery DW, Wienker TF, Tummeler B. The TNFalpha receptor *TNFRSF1A* and genes encoding the amiloride-sensitive sodium channel *ENaC* as modulators in cystic fibrosis. *Hum Genet.*, 2006 Apr;119(3):331-43.
- **M.G. Vinod**, M.M. Shivu, K.R. Umesha, B.C. Rajeeva, G. Krohne, Indrani Karunasagar, and Iddya Karunasagar. Isolation of *Vibrio harveyi* bacteriophage with a potential for biocontrol of luminous vibriosis in hatchery environments. *Aquaculture*. 2006 (255): 117-124.
- Karunasagar, I., **M.G.Vinod**, Bob Kennedy, M.D., Atnur Vijay, A Deepanjali, K.R. Umesha and Iddya Karunasagar. Biocontrol of bacterial pathogens in aquaculture with emphasis on phage therapy. In: P.Walker, R. Lester and M.G. Bondad-Reantaso (eds.). *Diseases in Asian Aquaculture V*, Fish Health Section, Asian Fisheries Society, Manila. 2005, pp. 535-542.