Genetic variants of surfactant proteins A, B, C, and D in bronchopulmonary dysplasia

J. Pavlovic[a], C. Papagaroufalis[b], M. Xanthou[b], W. Liu[c], R. Fan[d], N.J. Thomas[c,e], I. Apostolidou[b],
E. Papathoma[f], E. Megaloyianni[g], S. DiAngelo[a] and J. Floros[a,e,h,*]
[a]*Department of Cellular and Molecular Physiology, PSU, Hershey, PA, USA*
[b]*Aghia Sophia, Hospital, Athens, Greece*
[c]*Department of Health Evaluation Sciences, PSU, Hershey, PA, USA*
[d]*Department of Statistics, Texas A&M University, College Station TX, USA*
[e]*Department of Pediatrics, PSU, Hershey, PA, USA*
[f]*Alexandra Hospital, Athens, Greece*
[g]*A. Kyriakou Hospital, Athens, Greece*
[h]*Department of Obstetrics and Gynecology, PSU, Hershey, PA, USA*

Abstract. BPD 28D (O_2 dependency at 28 days of life) and BPD 36W (O_2 dependency at 36 wks post-menstrual age) are diseases of prematurely born infants exposed to mechanical ventilation and/or oxygen supplementation. In order to determine whether genetic variants of surfactant proteins (SPs-A, B, C, and D) and SP-B-linked microsatellite markers are risk factors in BPD, we performed a family based association study using a Greek study group of 71 neonates (<30 wks gestational age) from 60 families with, 52 BPD 28D and 19 BPD 36W, affected infants. Genotyping was performed using newly designed pyrosequencing assays and previously published methods. Associations between genetic variants of SPs and BPD subgroups were determined using Transmission Disequilibrium Test (TDT) and Family Based Association Test (FBAT). Significant associations ($p \leqslant 0.01$) were observed for alleles of SP-B and SP-B-linked microsatellite markers, and haplotypes of SP-A, SP-D, and SP-B. Specifically, allele B-18 C associated with susceptibility in BPD 36W. Microsatellite marker AAGG 6 associated with susceptibility in BPD 28D/36W group. Haplotype analysis revealed ten susceptibility and one protective haplotypes for SP-B and SP-B-linked microsatellite markers and two SP-A-SP-D protective haplotypes. The data indicate that SP loci are linked to BPD. Studies in different study groups and/or of larger sample size are warranted to confirm these observations and delineate genetic background of BPD subgroups.

Keywords: BPD, surfactant proteins, alleles/haplotypes, TDT, FBAT, pyrosequencing

1. Introduction

Bronchopulmonary dysplasia (BPD) is the most common cause of morbidity in prematurely born infants who require prolonged mechanical ventilation [30]. BPD is almost exclusively a disease of severely premature (24–28 weeks of gestation), extremely low birth weight (ELBW, less than 1000 gr) infants who have been exposed to high airway pressures and/or high inspired oxygen concentrations [26,38]. Despite the dramatic improvements in neonatal care the incidence of BPD is on the rise due to the increased survival of the smallest infants [2]. BPD infants have structurally and biochemically immature lungs, the development of which is disrupted due to a number of harmful stimuli [29,30]. These processes lead to an impaired alveolar and capillary growth and overall an abnormal lung structural development that can have long term consequences [3,14].

*Corresponding author: Joanna Floros, Ph.D., Evan Pugh Professor of Cellular and Molecular Physiology, Pediatrics, and Obstetrics and Gynecology, The Pennsylvania State University College of Medicine, 500 University Drive; Hershey, PA 17033, USA. E-mail: jfloros@psu.edu.

It has been noted that genetic factors may contribute to BPD [27]. Gene expression profile studies of an animal model of BPD implicated several genes in the pathogenesis of this disease [57]. Candidate gene and linkage analysis are approaches used to identify genes associated with multifactorial diseases such as BPD. Although there have been very few studies on the genetic background of BPD in humans, preliminary reports have shown an association between BPD and surfactant proteins [39,62]. A recent case control study has demonstrated that the frequency of SP-B intron 4 deletion variant alleles is increased in BPD versus control infants even when essential external confounding factors, such as birth order, are included in the analysis [50].

Surfactant proteins (SPs) are good candidate genes of neonatal disease due to their role in lung development and maturation [4,15,19,63]. These proteins are components of pulmonary surfactant, a lipoprotein complex necessary for lung function [15]. Surfactant protein (SP-) B and SP-C play important roles in surfactant structure and surface tension lowering properties [61], while SP-A and SP-D play a role in local host defense and regulation of inflammatory processes in the lung [6,46,65]. Deficiency of surfactant can result in Respiratory Distress Syndrome (RDS) in premature infants. The lungs of these infants have been identified to have low SP levels [9,10]. Association of genetic variants of SP-A and SP-B with RDS has been demonstrated by both case control and family-based linkage studies [16–19,21,22,41,42]. Since BPD is commonly preceded by RDS, overlapping underlying mechanisms regulated by the same genetic factors may play a role in the etiology of both diseases.

In this report, we sought to perform a family based association study in order to identify whether alleles and/or haplotypes of surfactant protein genes (SP-A, SP-B, SP-C, and SP-D) and SP-B-linked microsatellite markers are susceptibility or protective factors in BPD. Associations between single nucleotide polymorphisms (SNPs) of surfactant proteins and BPD were tested using Transmission Disequilibrium Test (TDT) [55] and Family Based Association Test (FBAT) [24,34,47]. The extended TDT (ETDT) [7, 52] and multi-allelic FBAT [25] were used to test for markers with multiple alleles.

To accelerate genotype analysis necessary for this study, we adapted our previously described PCR-cRFLP method [11] to the recently introduced pyrosequencing method. Pyrosequencing is a primer extension based method that has been used for a variety

Table 1
Characteristics of population used in the study

Number of children per family	BPD at 28 days	BPD at 36 weeks
1	$35^{¥} + 6^{*}$	$11^{€} + 4^{*}$
2	4	2
3	1	0
Total number of families	46	17

*These families have a second child that was diagnosed with either RDS or the other form of BPD
¥ Two families have only mothers and no fathers, all others have both mother and father.
€ One family has only mother and no father, all others have both mother and father.

of biological applications, including SNP genotyping, SNP discovery, haplotyping, insertion/deletion studies, methylation studies and allele frequency studies [12, 49]. This flexible, high-throughput method of genotyping improves on our previously published PCR-cRFLP method [11] by reducing the time and effort necessary for data acquisition.

2. Materials and methods

2.1. Study population

The study population consisted of 60 families from Greece with 71 affected infants, as outlined in Table 1. All infants required supplemental oxygen at 28 days of life. All infants were intubated and required $FiO_2 > 0.30$ on day 1 of life. Sixty four of them had RDS and were treated with surfactant. Out of 71 affected infants, 52 fulfilled the criteria for BPD at 28 days as described by Bancalari et al. [1]. All infants with BPD at the 28 days category met the Bancalari et al. criteria of supplemental oxygen requirement 28 days after birth, persistent abnormalities in the chest radiograph, and tachypnea in the presence of rales or retractions. This subgroup of infants is referred to here as BPD 28D. Nineteen infants went on to be oxygen dependent at 36 weeks of post-menstrual age [53]. This subgroup of infants is referred to here as BPD 36W. When the two subgroups are combined for analysis, they are referred to as BPD 28D/36W. The study protocol was approved by institutional committees of the participating hospitals and written parental consent obtained from each family.

2.2. Isolation of DNA

DNA was extracted from 200 μl of blood using the QIAamp DNA Blood Midi kit (Qiagen, Valencia, CA)

Sequence to analyze
AA140: CT**A/G**GAGAAGACCTT
Dispensation order: CTCGACGAGAGACT

Sequence to analyze
AA223: **G/T**CTCTTTTCCCCGA
Dispensation order: ATGACTCTCGA

Sequence to analyze
AA140: CT**A/G**GAGAAGACCTT AA223:
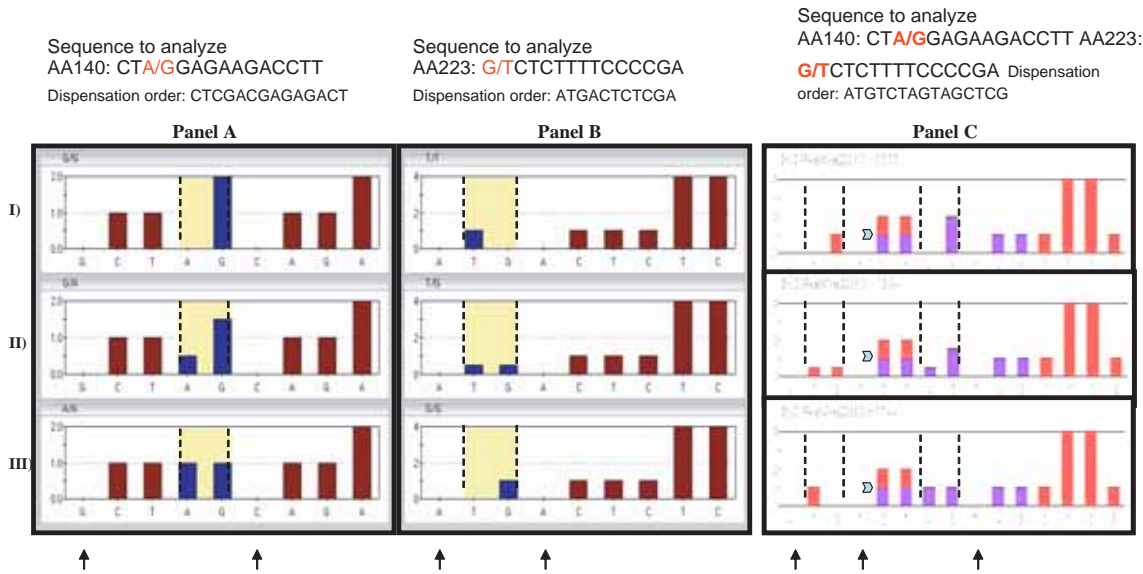**G/T**CTCTTTTCCCCGA Dispensation
order: ATGTCTAGTAGCTCG

Fig. 1. Simplex and multiplex pyrosequencing for SP-A (AA140 and AA223): Expected patterns from simplex pyrosequencing analysis (panels A and B) and multiplex analysis (panel C) of two SNPs (AA140 and AA223) are presented as histograms. There are three possible patterns of nucleotide distribution in simplex pyrosequencing for each SNP (panel A for AA140 SNP; panel B for AA223 SNP), while nine combinations are possible for multiplex pyrosequencing (only three are represented in panel C). The positions of the SNPs of interest are highlighted in yellow and outlined with a dashed line. **Panel B:** The sequencing primer for the AA223 SNP (G/T) is right in front of the SNP, and this makes it the most straight forward type of sequencing reaction (as compared to that in panel A). Since the AA223 sequence (**G/T**CTCTTTTCCCC, Panel B), does not contain an A nucleotide, this can serve as negative control. The nucleotides serving as negative controls are identified by an arrow at the bottom of each panel. Whenever possible, the negative control nucleotide is dispensed first. For example, in panel B, the nucleotide dispensation order is as follows: the negative control (A) is followed by a T and then a G nucleotide (constituting the SNP). These are followed by another negative control A nucleotide (arrow). The dispensation order then continues to sequence through a short segment of the surrounding DNA sequence, in this case CTCTTTTCCCC (panel B), in order to ensure that the SNP analyzed is within the DNA segment of interest. Note the difference in peak height when a single T or C is dispensed versus four Ts or four Cs are dispensed in sequence. Panel portions I and III depict homozygous SNP patterns (G and T, respectively in panel B), while the II portion of each panel depicts a heterozygous SNP pattern (G/T in panel B). **Panel A:** For AA140 SNP, the sequencing primer is not right in front of the SNP site, as it is for the AA223 SNP (panel B). In this case two nucleotides (C and T) precede the SNP (A/G). Therefore, in the sequencing reaction, a G nucleotide (negative control, arrow) is dispensed. This is followed by dispensation of a C and a T that sequence the CT preceding the A/G SNP. Due to the fact that another G nucleotide is immediately adjacent to the A/G SNP, the specific SNP signal for G overlaps with the flanking G, as represented by the double peak height in panel A.I. A negative control C nucleotide (arrow) is dispensed following the A/GG sequence, and sequencing of surrounding bases continues as in panel B. **Panel C:** Both SNPs (AA140 and AA223) are sequenced at once. Due to the dispensation order set-up AA223 SNP is sequenced first (red bars), as in panel B, followed by a negative control T. The dispensation of the following C and T nucleotides, sequences through both AA223 and AA140 sequence at once (block arrow, red and blue bars overlap). Dispensation of the A and G nucleotides, sequences through the AA140 SNP site (blue bars). This SNP site is followed by a negative control (T nucleotide, arrow). The remaining stretch of nucleotides is the surrounding sequence for both AA140 (AG, blue bar) and AA223 (CTTTTCCCCG, red bar).

according to manufacturer's instructions. DNA was eluted in the final step using 200 μl of DNase free dH$_2$O (Ambion). Upon extraction DNA was stored at $-20\,^{\circ}$C until further handling.

2.3. Pyrosequencing: General considerations, primer design, template preparation and plate set-up

a) Background: Pyrosequencing is a primer based DNA sequencing method, where DNA synthesis is monitored by four different enzymes and detected through luminescence in the form of a light pulse. It is performed on single stranded DNA templates pro-

duced by PCR and separation of the biotinylated single strands (see below). Then a sequencing primer hybridization to the complementary bases on the single stranded PCR template initiates nucleotide incorporation by DNA polymerase. As each new nucleotide is incorporated, a pyrophosphate (PPi) group is released in an equimolar proportion to the amount of the incorporated nucleotide. This pyrophosphate is converted to ATP by the ATP sulfurylase, which in turn drives the conversion of luciferin to oxyluciferin by luciferase. The light produced as a result of this reaction is detected by a charge-coupled device (CCD) camera. Each light signal is proportional to the number of nucleotides in-

Table 2
SNPs of surfactant protein genes studied

SP-A1	SP-A2	micosatellite markers	SP-B	SP-C	SP-D
AA19(C/T)	AA9(A/C)	D2S388	B-18(A/C)	CA138(A/C)	DA11(C/T)
AA50(C/G)	AA91(C/G)	D2S2232	B1013(A/C)	CA186(A/G)	DA160(A/G)
AA62(A/G)	AA140(C/T)	AAGG	B1580(C/T)		
AA133(A/G)	AA223(A/C)		B9306(A/G)		
AA219(C/T)					

SP-A1, SP-A2, SP-C, and SP-D SNP positions are based on the amino acid number of the amino acid sequence deduced from the cDNA (for SP-D the amino acid numbering starts after cleavage of signal peptide, while for SP-A and SP-C starts prior to the signal peptide cleavage). The first letter denotes the gene (A for SP-A, C for SP-C, D for SP-D), and the second letter (A) stands for amino acid. For SP-B SNPs, the numbers refer to the nucleotide position of the polymorphism [36]. The polymorphic variant of each SNP is given in parenthesis. For SP-A, haplotypes consisting of SNPs within the coding sequence that may or may not change the encoded amino acid are denoted as 6A or 1A for SP-A1 and SP-A2, respectively. For example, the SP-A2 $1A^2$ haplotype referred to in Fig. 3 consists of SNPs encoding amino acids 9, 91, 140, and 223 (AA9 C/AA91 G/AA140 C/AA223 C) [11]. The SP-D locus is in physical proximity with the SP-A locus [23] and in Fig. 3 the co-transmission of the SP-D SNPs with the $1A^2$ haplotype are shown.

corporated onto the DNA strand. As each nucleotide is dispensed, the complementary strand is synthesized and its signals generate a Pyrogram^{TM}. Each light signal corresponds to a single peak in the pyrogram. Pyrograms are scored by pattern-recognition software that compares the predicted SNP pattern (histogram) to the observed pattern (pyrogram) (Pyrosequencing AB, Uppsala, Sweden). The ability to quantitate each peak of the pyrogram provides for numerous advantages. For example, efficiency of the PCR amplification can be determined from the peak height. Moreover, specificity of the template or any non-specific allelic amplification can be monitored by ensuring appropriate heights of specific peaks, and lack of incorporation of negative control nucleotides.

b) Primer design for pyrosequencing: PCR primers were designed using commercially available software DNAStar (www.dnastar.com) or Pro-oligo (www.changbioscience.com) (Table 3). Multiplex PCR and/or multiplex pyrosequencing primers were designed to ensure that no homo- and/or heterodimers occurred, and that there were no false priming sites for either of the primers on either template. Biotin label was added to the 5' end of one of the primers in each primer pair. The biotin labeling of the primers is currently carried out by Biomers (www.biomers.com), although other companies have been used in the past. Sequencing primers for the pyrosequencing reaction were designed using the SNP Primer Design Software (technical support web site of Pyrosequencing AB, Uppsala, Sweden). These primers anneal directly in front of the SNP position (or several nucleotides before it) on the reverse strand of the PCR template to be analyzed. Depending on the position of the sequence primer docking, the dispensation order of nucleotides is determined

by the Pyrosequencing SNP analysis software (Fig. 1). Nucleotides that are not present in the sequence, can serve as negative controls, and are dispensed, if possible, at the beginning of the sequencing reaction as well as in positions flanking the SNP position (Fig. 1). The negative control ensures the specificity of the priming and sequencing reaction. The DNA strand to be analyzed is elongated through the SNP site.

c) Template preparation and plate set-up for pyrosequencing: Twenty μl of biotinylated PCR product was immobilized to streptavidin-Sepharose HP beads (Amersham Biosciences) and incubated with binding buffer, 10 mmol/L Tris-HCl, 2 mol/L NaCl, 1 mmol/L EDTA, 1 mL/L Tween 207 in a 96-well microtiter plate for 10 min at 25°C with mixing at 1400 rpm. The 96-well plate was then processed through vacuum filtration using the Vacuum Prep Tool (Pyrosequencing AB), where the bead-bound PCR products were transferred to the filter at the end of each 96-Filter probe, and the remaining liquid was removed by vacuum filtration. The Vacuum Prep Tool was passed through a series of solutions including the denaturation solution (0.2 mol/L NaOH) that removed non-biotinylated strands. Sepharose bead-bound biotinylated strands were retained on the filter and rinsed in washing buffer (10 mmol/L Tris-acetate, pH 7.6). The vacuum was removed to release the beads from the filter. Bead-bound single-stranded DNA templates for each SNP were resuspended in annealing buffer (20 mMTris-acetate, pH 7.6; 2 mM magnesium acetate) containing 20 pmol of each respective sequencing primer, in wells of a PSQ 96 plate (Pyrosequencing AB). This mixture was then heated at 80°C for 2 min on a compact heat block and cooled to room temperature for 10 min to facilitate annealing of the Pyrosequencing primers to templates.

Table 3
Primers used in the study

SNP	Gene	F #	F primer	R #	R primer	Seq # S primer	Direction*	Simplex/ multiplex
AA19(C/T)	SP-A1	1327A	cgactggacccagagccatgtc	1328	B-ctggagggccagggtctcctt	1331 cagcctctggtgctg	Forward	S and M with AA50
AA50(C/G)	SP-A1	1327A	cgactggacccagagccatgtc	1328	B-ctggagggccagggtctcctt	1332 cagggacgggagag	Forward	S and M with AA19
AA62(A/G)	SP-A1	1344	B-tagctccaaccagttgtgggtgac	1345	tgatttccaggaggacatggca	1346 catggcatttctctccag	Reverse	S only
AA133(A/G)	SP-A1	1233	B-agtctgcagggctccataatga	28A	accctcagtcaggcctacat	1338 ggagaagaccttctctcc	Reverse	S and M with AA219
AA219(C/T)	SP-A1	1321	aagtacaacacatatgcctat	494	B-tcagaactcacagatggtca	1339 aggggagcccgca	Forward	S and M with AA133
AA9(A/C)	SP-A2	1327A	cgactggacccagagccatgtc	1328	B-ctggagggccagggtctcctt	1333 gccctctggccctca	Forward	S only
AA91(C/G)	SP-A2	1397	tgggaataatgggctgcctg	1398	B-ctgacttcagggtcgctgtg	1399 gagagaagggggag	Forward	S only
AA140(C/T)	SP-A2	1233	B-agtctgcagggctccataatga	28A	accctcagtcaggcctacat	1336 tggactgcccattg	Reverse	S and M with AA223
AA223(A/C)	SP-A2	1321	B-aagtacaacacatatgcctat	494	tcagaactcacagatggtca	1337 tgtgtacatctccacacact	Reverse	S and M with AA140
B1580(T/C)	SP-B	1313	tcccaggacacgatgaggaag	1390	B-gcaggcaggaggtgagc	1341 tacttccagaacca	Forward	S only
CA138(A/C)	SP-C	1393	B-tccagagagcatccccagtc	1394	gtagtagagcggcacctcgc	1395 tggaagttgtggacttt	Reverse	S only
CA186(A/G)	SP-C	1318	gggctgcacatgggatagaaactc	1319	B-ccctgggatcacacctgctcacc	1320 ggcatggccgtga	Forward	S only
DA11(C/T)	SP-D	1305	gaagcagaaatgaagacctactcc	1301	B-gccactctccactgagctacaca	1302 aagacctactcccacagaa	Forward	S and M with AA160
DA160(A/G)	SP-D	1306	gaggtgtccctggtgagcgt	1307	B-gtgggtggaggggggtgta	1308 gagtccctggaaac	Forward	S and M with AA11

All primer sequences are 5' to 3'. Forward (F) primers are in sense and reverse (R) primers in anti-sense orientation. For forward and reverse sequencing, the sequencing (Seq) primers are in sense and antisense orientation, respectively. *The forward and reverse direction of the sequencing assay are noted. In a multiplex reaction two SNPs can be sequenced from a single template (as for AA19/AA50) or from two different templates, as is the case in all other multiplex reactions showed. Primer numbers are arbitrary laboratory numbers. B stands for the biotin label at the 5' end of the primer. S denotes simplex reaction, M denotes multiplex reaction.

The plate was placed into the PSQ HS 96MA System (Pyrosequencing AB) for analysis. Simplex pyrosequencing was performed for genotyping SNPs of all four genes. Where multiplexing could be achieved, duplexing of two SNPs of the same gene was designed, with the sequencing reaction taking place from either one or two PCR templates.

2.4. Genotyping of SP SNPs and SP-B microsatellite markers

Table 2 shows the SNPs studied for all four surfactant protein genes and SP-B linked microsatellite markers, and Table 3 the primers used.

I. **SP-A analysis**

a) **SP-A PCR:**

i) General: All DNA samples were individually amplified from non-quantified genomic DNA by PCR using Taq DNA polymerase (Roche Diagnostics, Indianapolis, IN). Two μl of individual DNA samples were combined with 1 μl of 1.25 mM deoxynucleotide triphosphate (dNTP) mix, various amounts of each PCR primer (25 ng/μl to 100 ng/μl), 2.5 μl l0x Buffer 1 and

2.5 μl of l0x Buffer 2 (Roche) in 50 μl overall volume (this was done for all the SNPs of all four SP genes). Cycling conditions for SP-A (gene specific reaction) were as follows: 95°C for 2 min followed by 35 cycles of 95°C for 30', 58°C for 1 min, and 72°C for 1 min. This was followed by one cycle of 72°C for 5 min. PCR profiles for SP-A nested reaction were as follows: 95°C for 2 min followed by 35 cycles of 95°C for 30', 60°C for 30', and 72°C for 30'. This was followed by one cycle of 72°C for 5 min. An aliquot of the PCR reaction was run on 8% PAGE gel to verify size and quantity of PCR product, and most importantly to ensure the presence of a single PCR band in simplex PCR or two bands in multiplex PCR. The outline of the SP-A SNP genotyping approach is presented in Fig. 2.

ii) SP-A1: primer pair 1327A/293 was used to generate a 549 bp long PCR product, which was used as template for nested reactions for AA19, AA50, and AA62. Due to the spatial proximity of AA19 and AA50, these SNPs were amplified on the single DNA strand of a 187 bases product generated with 1327A/1328 primer pair. The 549 bp PCR product was also used as a template in a nested reaction with primers 1344/1345 to generate a fragment for analysis of AA62. The second gene spe-
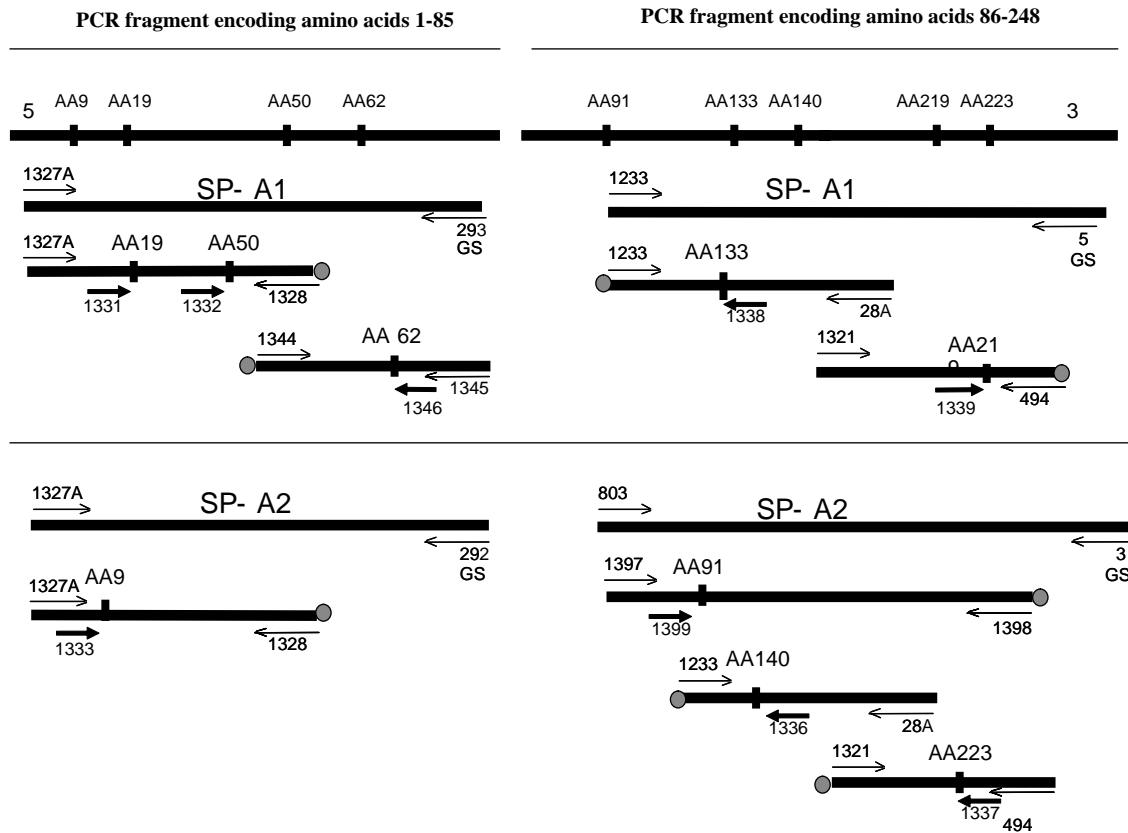
PCR fragment encoding amino acids 1-85   **PCR fragment encoding amino acids 86-248**

Fig. 2. Schematic representation of SP-A PCR and pyrosequencing **for PCR fragments encoding amino acids 1-248**: Forward and reverse primers are presented with thin arrows, and GS (gene specific) stands for gene specific primers (SP-A1 or SP-A2 specific). Grey circles represent biotin label at the primer end. Sequencing primers are presented with thick arrows.

cific fragment of SP-A1 was amplified using primers 1233/5 to generate a 422 bp fragment. Two nested reactions were performed on this template; a 200 bp fragment for AA133 with primer pair 1233/28A and a 209 bp fragment for AA219 with primer pair 1321/494.

iii) SP-A2: A 541 bp long PCR product was amplified with primer pair 1327A/292 and used as template in nested PCR with primer pair 1327A/1328 to generate a 186 bp fragment for AA9 analysis. The second gene specific fragment of SP-A2 was amplified using primers 803/3 to generate a 1700 bp fragment. This fragment was used as a template in three nested PCR reactions: 1) with primers 1397/1398 to generate a 129 bp fragment for AA91 pyrosequencing, 2) with primers 1233/28A to generate a 200 bp fragment for AA140 pyrosequencing, 3) with primers 1321/494 to generate a 223 bp fragment for AA223 pyrosequencing. Due to the cost of biotinylated primers, the objective was to have as many as possible overlapping primers in the nested reactions for the two genes. Therefore, the same primer pairs were used for AA9, AA19, and

AA50; for AA133 and AA140; and for AA223 and AA219.

b) **SP-A pyrosequencing:** All pyrosequencing reactions were first performed in simplex, where a single SNP was analyzed at a time, using 20 μl of the PCR reaction. Where possible, cost effective multiplex pyrosequencing assays were designed. For SP-A1, AA19 and AA50 were sequenced in duplex using 20 μl of a singular template and both SNPs were analyzed in forward assays. A reverse AA133 assay was performed in multiplex with a forward AA219 assay, using 15 μl of AA133 PCR template and 25 μl of AA219 template, to achieve optimal intensity in the ratio of the signal from each template. For SP-A2, 20 μl of AA140 PCR product was pyrosequenced in multiplex with 20 μl of AA223 PCR product. Both SNPs were pyrosequenced either in simplex or multiplex assays as schematically presented in the Fig. 1. Overall, complete genotyping of both SP-A genes requires 12 PCR reactions and 6 pyrosequencing reactions.

c) **SP-A molecular haplotypes:** The advantage of being able to pyrosequence AA19 and AA50 on a single DNA strand, allows for direct molecular haplotyping of these two SNPs. In SP-A1, the AA19(C)AA50(G) haplotype was not observed in 152 individual DNA samples genotyped by pyrosequencing and in 100 samples genotyped by PCR-cRFLP [11]. FBAT was used to estimate frequencies of SP-A1, AA19(C/T)AA50(C/G), haplotypes in 1534 individuals. The frequency of AA19(T)AA50(G) haplotype was 0.532, the frequency of AA19(T)AA50(C) was 0.383, the frequency of AA19(C)AA50(C) was 0.083, while the frequency of AA19(C)AA50(G) was 0.001. These observations together indicate that the SP-A1 haplotype AA19(C)AA50(G) is extremely rare (approximately 1 in 1000).

II. **SP-B analysis:**

a) **SP-B SNP genotyping:** Genotyping for the SP-B SNPs B-18(A/C), B1013(A/C), and B9306(A/G) was performed for all samples using the PCR based converted Restriction Fragment Length Polymorphisms (cRFLP) method, as described before [36,51]. Genotyping for SP-B marker B1580(C/T) was done by both PCR-cRFLP and pyrosequencing. PCR for marker B1580(T/C) was performed using oligos 1313/1390 and cycling conditions were as follows: 95°C for 2 min followed by 35 cycles of 95°C for 30', 58°C for 1 min, and 72°C for 1 min. This was followed by one cycle of 72°C for 5 min. An aliquot of the PCR reaction was run on 8% PAGE gel to verify size and quantity of PCR product, and most importantly to ensure the presence of a single PCR band. Pyrosequencing of the PCR product was done using sequencing with primer 1341.

b) **Microsatellite genotyping** was performed for SP-B-linked microsatellites, AAGG, D2S22 32, and D2S388, as described previously [31].

III. **SP-C analysis:**

Genotyping for SP-C markers CA138(A/C) and CA186 (A/G) was performed as previously described [51] using the PCR-cRFLP method.

IV. **SP-D analysis:**

Simplex or multiplex PCR for DA11 and DA160 of SP-D was performed using primer pair 1301/1305 for DA11 and primer pair 1306/1307 for DA160. PCR profiles for SP-D both simplex and multiplex reaction were as follows: 95°C for 2 min followed by 35 cycles of 95°C for 30', 59°C for 30', and 72°C for 30'. This was followed by one cycle of 72°C for 5 min. An aliquot of the PCR reaction was run on 8% PAGE gel to verify size and quantity of PCR product, and most importantly to ensure the presence of a single PCR band in simplex PCR or two bands in multiplex PCR. These SNPs were pyrosequenced either in simplex or multiplex forward assays using oligos 1301 and 1308, respectively.

2.5. *Statistical analysis*

PedCheck program was used to check the compatibility of genotypes at each marker locus within families, prior to analysis [44]. Marker loci with incompatible parental and offspring genotypes, were treated as missing in those families. Association between alleles of surfactant protein genes and BPD was tested using the Transmission Disequilibrium Test (TDT) [56] and Family Based Association Test FBAT [24,34,47].

TDT analysis was performed using GENEHUNTER [33] (Whitehead Institute for Biomedical Research, MIT) to determine, transmission of individual surfactant protein (SP-A, SP-B, SP-C, and SP-D) marker alleles and SP-B-linked microsatellite marker alleles from heterozygous parents to affected offspring, and also to test for transmission of SP-B haplotypes of two, three, and four marker loci. In this analysis it is sufficient for only one parent to be heterozygous for the marker in order for a family to be considered informative and thus used in the study. Extended TDT (ETDT) analysis was performed to assess linkage of a multi-allele locus to the disease locus [7,52].

FBAT analysis was performed using the online program, http://www.biostat.harvard.edu/f̃bat/fbat.htm. Bi-allelic FBAT analysis was performed for SP SNP markers and selected microsatellites [24,34,47]. Multi-allelic analysis, where association of the entire locus with disease is examined, was also performed for marker loci with multiple alleles [24]. All of the FBAT analyses were performed assuming an additive model. The minimum size [minsize] of FBAT analyses was set to 4 indicating that the test statistic was not computed when the number of informative families available was fewer than 4. Informative families refer to families with non-zero contribution to the FBAT statistic. For each FBAT marker, the Z-statistic and the corresponding p value are listed. A significant p value and a positive Z-statistic are indicative of a susceptibility marker allele for disease, while a significant p value and a negative Z-statistic is indicative of a protective marker allele for disease. TDT and FBAT analyses were performed

for the following groups: BPD 28D, BPD 36W, and BPD 28D/36W.

Stratified FBAT analyses were also performed within each group based on baby's steroid treatment, mother's steroid treatment, and baby's surfactant therapy. Haplotype analyses were performed using the FBAT program, which implements an EM based algorithm [25]. EM algorithm is a general method of finding the maximum-likelihood estimate of the parameters of an underlying distribution from a given data set when the data set is incomplete or has missing values. For single marker analysis, FBAT and TDT gave the same results, although some differences were observed in the haplotype analysis, perhaps due to subtle differences in the assumptions each method makes (see discussion). For both TDT and FBAT analysis, significant results were noted when p was smaller or approaching 0.01. Results were not corrected for multiple comparisons.

Inferred haplotype reconstruction was performed using FBAT, where the most likely haplotypes were assigned for each individual infant. FBAT compares the distribution of test statistics using the conditional offspring genotype distribution under the null hypothesis [25]. The offspring genotype probabilities were computed given the parental mating type.

3. Results

We examined transmission of individual SNPs for all five surfactant protein genes and SP-B-linked microsatellites markers (Table 2).

I. **Family-based allele association analysis:**

a) **SP SNPs marker alleles and SP-B-linked microsatellites:** Tables 4A and 4B, summarize the results of TDT and bi-allelic FBAT analysis, of SP-B and SP-B-linked markers, respectively. In both TDT and FBAT analysis, allele 6 of the AAGG marker was found to be transmitted more frequently to the affected offspring in BPD 28D/36W ($p = 0.011$, Table 4). The SP-B marker locus B-18(A/C) was found to be associated with BPD 36W group by both TDT and FBAT ($p = 0.018$), where the A allele was associated with protection and the C allele with susceptibility to disease. No significant associations were found for individual allelic variants of SP-A, SP-C, and SP-D by either TDT or FBAT.

b) **SP marker alleles stratified for pre-natal/neonatal treatments:** Stratified bi-allelic FBAT analysis was performed to investigate the possible effect of pre-

natal/neo-natal treatments on the outcome of BPD babies. Analyses were stratified according to the three co-variates: 1) lack or presence of pre-natal steroids, 2) lack or presence of post-natal steroids, and 3) lack or presence of surfactant therapy. Of significance is that in the combined BPD group and BPD 28D, allele 4 of the D2S388 marker locus was associated with protection from disease in babies that did not receive pre-natal steroid treatment ($p = 0.01$). When the effect of post-natal steroids was investigated, allele 7 of the AAGG marker was associated with a protective effect (in BPD 28D, $p = 0.014$) while allele AAGG 6 (in the combined BPD group, $p = 0.019$) and B-18 C (in the BPD 36W, $p = 0.018$) was associated with susceptibility to disease in babies treated with post-natal steroids. Allele AAGG 6 was also found to be related to susceptibility in the combined BPD group that received surfactant therapy ($p = 0.008$).

These results are in concordance with those of non-stratified analysis, as would be expected, since the majority of infants have received post-natal steroid treatment. On the other hand, in the smaller subgroup of BPD 28D (babies that did not receive post-natal surfactant therapy) allele 4 of the D2S388 marker was associated with a protective effect from disease ($p = 0.008$). The functional significance of these associations remains to be determined, particularly since the sample number of families under study is small for all, except for BPD at 28 days post-natal steroid treatment group.

II. **Family-based haplotype association analysis:**

a) **SP-A and SP-D SNP haplotypes:** Figure 3 depicts the findings of association between haplotypes of SP-A2 and SP-D genes and disease groups by TDT. In the BPD 28D group the DA160(G)SP-A2(1A^2) ($p = 0.005$) and the DA11(C)DA160(G)SP-A2(1A^2) ($p = 0.014$) haplotypes of SP-D/SP-A2 were not transmitted to the affected child. Therefore these haplotypes may act as protective factors in the BPD 28D subgroup. No significant associations were found for SP-A and SP-D haplotypes in the BPD 36W subgroup. Neither ETDT, nor bi and multi-allelic FBAT analysis showed any significant association between individual SP-A alleles and BPD disease subgroups.

b) **Haplotype analysis of SP-B SNP marker alleles and SP-B-linked microsatellite markers:** Haplotype analyses were performed by both TDT (generated by Genehunter) and FBAT for two, three, and four markers at the time, and by FBAT for five, six, and seven markers at the time. The results of TDT and FBAT analyses are listed in Tables 5A and 5B, respectively. Of note is that

Table 4A
TDT analysis of SP-B marker alleles

Disease group	Markers	allele	# of transmitted alleles	# of non-transmitted alleles	X^2	p-value	effect
BPD 36W[¥]	B-18(A/C)	C	14	4	5.56	0.0180	susceptibility
BPD 28D/36W[€]	AAGG 6	9	1	6.4	0.0114	susceptibility

Table 4B
Bi-allelic FBAT analysis of SP-B marker alleles

Disease group	marker	allele	Family #	Z	p value	minsize*	effect
BPD 36W[¥]	B-18(A/C)	C	11	2.357	0.0180	10	susceptibility
BPD 28D/36W[€]	AAGG 6	7	2.53	0.0114	10	susceptibility

Disease groups: 1) ¥ denotes BPD at 36 weeks post-conceptual age; 2) € denotes combined group of BPD at 28 days and BPD at 36 weeks post-conceptual age;
*Minsize denotes the minimum number of informative families necessary to compute the test statistics.
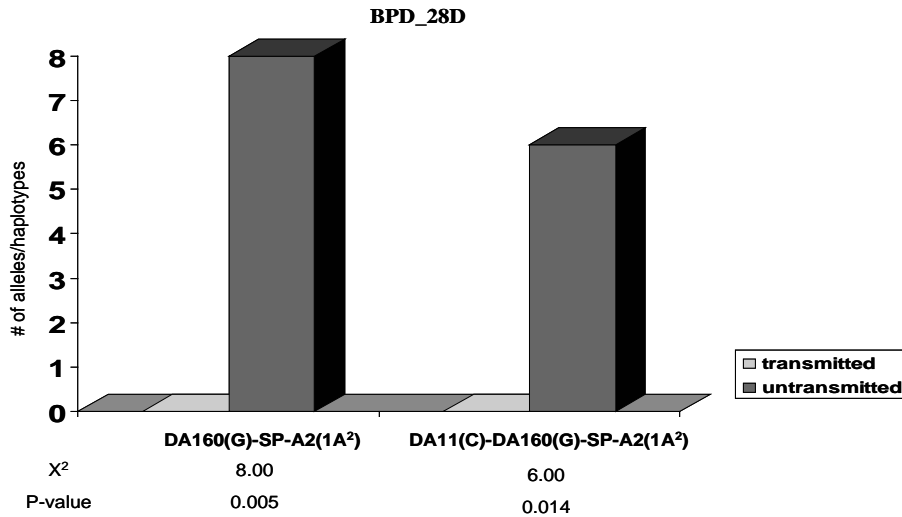

Fig. 3. TDT analysis of SP-A haplotypes: Significant haplotypes for BPD 28D/36W and BPD 36W are presented where p values approach 0.01. All 60 families were used in the analyses ($n = 52$ for BPD 28D; $n = 19$ for BPD 36W).

in most cases individual alleles found to be associated with either protection or susceptibility from disease in Table 4, were also found to form haplotypes that were associated with the same overall effect.

SP-B-linked microsatellite markers were found in several disease associated haplotypes. The frequency of transmission of haplotype AAGG 6/B-18 C to babies in the BPD 28D/36W group was increased by TDT ($p = 0.008$), as well as the AAGG 6/B-18 C/B1013 A haplotype ($p = 0.014$). Moreover, the frequency of transmission of the AAGG 7-containing haplotypes to babies in all three subgroups was decreased by TDT ($p = 0.02$), and the D2S2232 1/AAGG 7 haplotype was found less frequently in BPD 28D/36W group ($p = 0.014$).

Figure 4 provides a schematic presentation of haplotypes among different disease subgroups by the two different tests, TDT and FBAT. Presence of AAGG 6 allele in haplotypes was consistently associated with susceptibility to disease, while the presence of AAGG 7 allele was associated with a protective effect from disease. Although, AAGG 3 allele was found in three susceptibility haplotype combinations, this allele may not be informative as the AAGG 6 and AAGG 7 alleles may be, because it was also found in protection associated haplotypes (not shown: $p \leqslant 0.05$). The AAGG 3 containing haplotypes are, D2S2232 2/AAGG 3 found to associate with BPD 28D group by TDT, D2S388 5/D2S2232 2/AAGG 3 found to associate with both the BPD 28D and BPD 28D/36W by FBAT (in both $p = 0.01$), and

D2S388 5/D2S2232 2/AAGG 3/B-18 A found to associate with the BPD 28D/36W group by FBAT ($p = 0.01$).

SP-B SNP marker haplotype, B-18 C/B1013 A was associated with susceptibility by both TDT and FBAT analysis in BPD 36W ($p = 0.002$ and $p = 0.014$, respectively) whereas B-18 C/B1013 A/B1580 T and B-18 C/B1013 A/B1580 T/B9306 A haplotypes were associated with susceptibility in the same group only by TDT.

Overall, both individual allele and haplotype analysis of SP-B markers indicate an association of the microsatellite marker AAGG 6 allele with susceptibility to disease in infants affected with BPD 28D/36W. SNP marker B-18(A/C) alleles were associated with the more severely affected babies, where the C allele associated with susceptibility and the A allele with protection from BPD 36W.
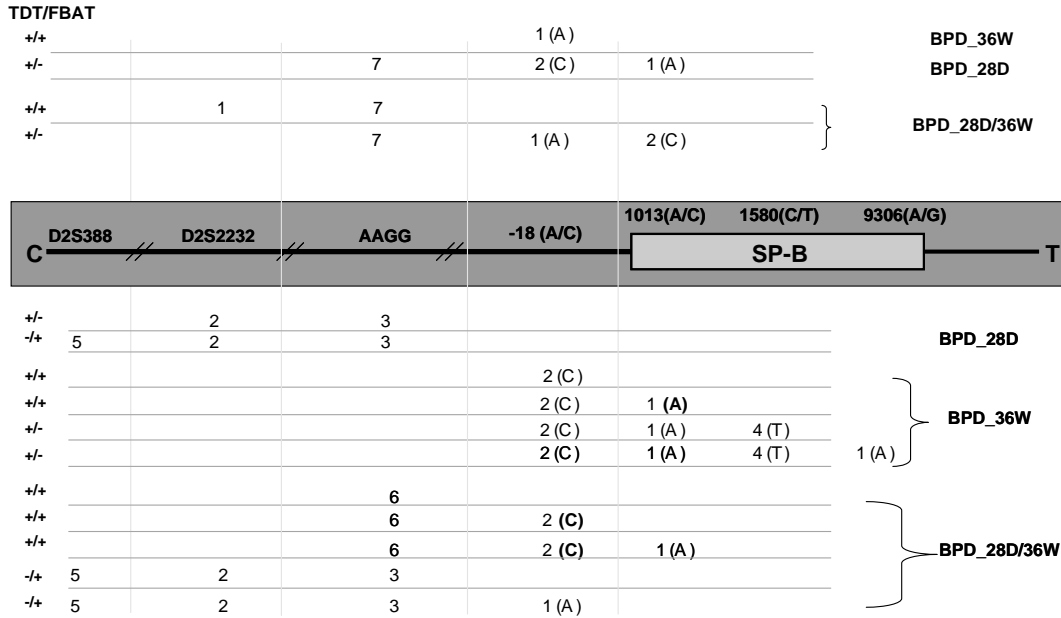
4. Discussion

BPD is a multifactorial disease of preterm infants where multiple risk factors contribute to permanent pathophysiologic changes. Numerous genes required for neonatal lung adaptation are likely to be involved in the etiology of this disease. Since surfactant proteins are essential for normal lung function and play a role in both gas exchange and innate immune mechanisms of the lung, they are likely to play a role in the pathogenesis of BPD. Susceptibility to RDS, the common precursor of BPD has been shown to be linked to surfactant proteins, and SP-A in particular [16–18, 20,21,48]. Recently, in a case-control study, an SP-B intron 4 deletion variant was identified as a risk factor in BPD [50]. In order to avoid spurious associations secondary to population stratification, characteristic of case-control studies, we performed family based association analyses. Such analyses are not confounded by population characteristics and provide a more powerful analysis even in small sample sizes [54]. The data from TDT and FBAT (using two powerful family based association tests) revealed, a) association of SP-B with susceptibility to BPD 36W; and b) association of the microsatellite marker AAGG 6 with susceptibility in the BPD 28D/36W group; c) Ten susceptibility and one protective haplotypes for SP-B and SP-linked microsatellite markers were detected for all three subgroups (BPD 28D, BPD 36W, and BPD 28D/36W); d) TDT analysis revealed two SP-A/SP-D susceptibility haplotypes for BPD 28D.

Of the 71 infants in the study, 52 had the less severe BPD 28D, while 19 infants were affected with the more severe BPD 36W. Although the sample size in BPD 36W may be small, the use of the family based association approach may mitigate the margin of error due to small sample size [54]. However, it should be noted that the majority of observations in the BPD 28D/36W group overlap with those for the BPD 28D sub-group, perhaps due to the predominance of the sample size of this group.

The SP-A2/SP-D haplotypes found to associate with protection from BPD 28D have not been previously studied. The lack of association with SP-A in the Finnish BPD study is most likely due to the fact that their study group consisted of more severely affected babies (86 infants had BPD at 36 weeks, and 21 had BPD at 28 days). The inflammatory components that appear to be at work in the early stages of the disease could be modulated by SP-A and SP-D [28], which may be consistent with their immune related functions [46]. Although it is unknown how the SP-A2(1A^2)/SP-D haplotypes may protect from BPD, we speculate that this protection is likely due to adequate levels of SP-A and/or SP-D at a critical developmental stage of the lung. *In-vitro* studies have shown genotype-dependent levels of SP-A mRNA [32] as well as differences among SP-A2 variants in 3'UTR mediated expression [59]. It is unlikely that the association reflects functional differences, because the SP-A2, 1A^2 variant is identical in the amino acid sequence of the mature protein with several other SP-A2 variants [11]. With regards to SP-D, the DA11(C/T) SNP was recently identified to affect oligomerization, function, and serum concentrations of the protein [35]. Individuals homozygous for the DA11(C/C) variant were found to have significantly lower SP-D serum levels than the DA11(T/T) homozygous individuals. Polymorphisms of SP-A were not investigated in this study [35].

SP-B SNP variant alleles and/or haplotypes on the other hand were only associated with the BPD 36W. Specifically, the C variant of the B-18(A/C) polymorphism within the 5'UTR of SP-B was found, either alone or in various haplotypes, to associate with susceptibility to BPD 36W by both TDT and bi- and multi-allelic FBAT. The finding of the SP-B intron 4 polymorphism association with BPD in a Finish study group [50], along with the present findings support a role of SP-B in BPD pathogenesis. Recently, a dual role for SP-B, as a protein essential for surfactant function and as an anti-inflammatory mediator, has been proposed that could explain involvement of this pro-

A. Protection from disease

TDT/FBAT

	D2S388	D2S2232	AAGG	-18 (A/C)	1013(A/C)	1580(C/T)	9306(A/G)	
+/+				1 (A)				BPD_36W
+/−			7	2 (C)	1 (A)			BPD_28D
+/+	1		7					BPD_28D/36W
+/−			7	1 (A)	2 (C)			

Schematic (middle): C — D2S388 // D2S2232 // AAGG // -18 (A/C) | 1013(A/C) 1580(C/T) 9306(A/G) | SP-B — T

	D2S388	D2S2232	AAGG	-18 (A/C)	1013(A/C)	1580(C/T)	9306(A/G)	
+/−		2	3					BPD_28D
−/+	5	2	3					
+/+				2 (C)				
+/+				2 (C)	1 (A)			BPD_36W
+/−				2 (C)	1 (A)	4 (T)		
+/−				2 (C)	1 (A)	4 (T)	1 (A)	
+/+			6					
+/+			6	2 (C)				
+/+			6	2 (C)	1 (A)			BPD_28D/36W
−/+	5	2	3					
−/+	5	2	3	1 (A)				

B. Susceptibility to disease

Fig. 4. Schematic presentation of susceptibility and protection haplotypes: Significant findings by both TDT and FBAT are presented in the Figure. A schematic representation of four SP-B SNPs B 18(A/C), B1013(A/C), B1580(C/T), B9306(A/G), and SP-B-linked microsatellite markers, is shown in the middle of the Figure. The centromeric (C) and telomeric (T) orientation of this DNA segment is also noted. The microsatellite marker AAGG is 27 kb from SP-B [31]. Markers D2S2232 and D2S388 are 120 kb and 195 kb from SP-B (according to the current version CH 2 map http://www.ensembl.org:80/Homo sapiens/geneview?gene=ENSG00000168878). Above the schematic presentation, protection haplotypes are shown (Table 5A) and below the schematic presentation susceptibility haplotypes are noted (Table 5B). The nucleotide for each SNP is also noted by a number: $A = 1$, $C = 2$, $G = 3$, and $T = 4$. For example in BPD 36W susceptibility haplotype B 18 C/B1013 A/B1580 T/B9306 A is also noted as haplotype 2/1/4/1. The alleles for each microsatellite marker locus have been defined previously [31] and are noted by numbers 1, 2, 3, etc. The haplotypes noted here are described in detail in Tables 4 and 5. Only significant findings where p value approaches 0.01 are presented. The plus (+) sign in the left hand column denotes a haplotype that was found by either TDT or FBAT, while the minus (−) sign denotes that a particular haplotype was not found to have a significant p value by either TDT or FBAT.

tein in BPD [13] at multiple levels. Regulatory roles in SP-B gene expression of SP-B that B-18(A/C) SNP and intron 4 variants [50] may have can only be speculated. Inadequate levels of SP-B may contribute to the progression of BPD especially in the later stages of the disease progression. Infants that require continued ventilatory support, as in BPD, experience transient episodes of surfactant dysfunction that are associated with deficiency in SP-B [43].

Mulltiallelic SP-B-linked microsatellite markers were also used to increase resolution in the definition of the genetic background of BPD subgroups. Of the three markers studied, AAGG and specifically allele AAGG 6 was found to associate with BPD 28D/36W. This microsatellite marker is located 26 kb upstream from SP-B, and allele 6 consists of 24 AAGG repeats [31]. No difference in transcription factor bind-

ing among the alleles of AAGG was found by current bioinformatics techniques (JASPAR transcription factor data base). It is currently unknown whether this marker imparts a direct functional effect on SP-B in relation to BPD, or whether it is linked to another gene and/or to a functional genetic variant(s) that is located in, or close to the SP-B gene.

To identify genes between the microsatellite marker AAGG locus and the SP-B SNPs an ensemble gene search (www.ensembl.org) was conducted. The only known gene located within this region is granulysin, a T cell activation gene. Although granulysin has not been implicated in the pathogenesis of BPD, this gene plays a role in immunity against intracellular pathogens, and may therefore be involved in the modulation of inflammatory responses in BPD [45]. However, the disease locus does not have to be limited to this region, and

Table 5A
TDT analysis of SP-B marker haplotypes

Disease group	Markers	haplotype	# transmitted	# non-transmitted	X^2	p-value	effect
BPD 28D/36W[€]	AAGG-B-18(A/C)	6 2	7	0	7	0.008	susceptibility
BPD 28D/36W	D2S2232-AAGG	1 7	0	6	6	0.014	protection
BPD 28D*	D2S2232-AAGG	2 3	13	3	6.25	0.013	susceptibility
BPD 36W[¥]	B-18(A/C)-B1013(A/C)	2 1	10	0	10	0.002	susceptibility
BPD 28D*	AAGG-B-18(A/C)-B1013(A/C)	7 2 1	8	21	5.83	0.016	protection
BPD 28D/36W	AAGG-B-18(A/C)-B1013(A/C)	6 2 1	6	0	6	0.014	susceptibility
BPD 28D/36W	AAGG-B-18(A/C)-B1013(A/C)	7 1 2	1	8	5.44	0.019	protection
BPD 36W	B-18(A/C)-B1013(A/C)-B1580(C/T)	2 1 4	6	0	6	0.014	susceptibility
BPD 36W	B-18(A/C)-B1013(A/C)-B1580(C/T)-B9306(A/G)	2 1 4 1	6	0	6	0.014	susceptibility

Disease groups: 1) *Denotes BPD at 28 days; 2) ¥ denotes BPD at 36 weeks post-conceptual age; 3)€ denotes combined group of BPD at 28 days and BPD at 36 weeks post-conceptual age;
For B-18(A/C), B1013(A/C), and B1580(C/T): 1 denotes A variant, 2 (C variant), and 4 (T variant).
transmitted and # non-transmitted denotes the # of haplotypes.

Table 5B
FBAT analysis of SP-B marker haplotypes

Disease group	marker	allele	Family #	Z	p value	minsize*	effect
BPD 36W[¥]	B-18(A/C)-B-1013(A/C)	2 1	7	2.554	0.014	4	susceptibility
BPD 28D*	D2S388-D2S2232-AAGG	5 2 3	5	2.558	0.010	4	susceptibility
BPD 28D/36W	D2S2232-AAGG	1 7	7	−2.376	0.017	4	protection
BPD 28D/36W[€]	AAGG-B-18(A/C)	6 2	6	2.333	0.019	4	susceptibility
BPD 28D/36W	AAGG-B-18(A/C)-B1013(A/C)	6 2 1	6	2.333	0.019	4	susceptibility
BPD 28D/36W[€]	D2S388-D2S2232-AAGG	5 2 3	5	2.558	0.010	4	susceptibility
BPD 28D/36W[€]	D2S388-D2S2232-AAGG-B-18(A/C)	5 2 3 1	5	2.558	0.010	4	susceptibility

Disease groups: 1) *Denotes BPD at 28 days; 2) ¥ denotes BPD at 36 weeks post-conceptual age; 3)€ denotes combined group of BPD at 28 days and BPD at 36 weeks post-conceptual age;
For B-18(A/C) 1 denotes A variant and 2 (C variant).
*Minsize denotes the minimum number of informative families necessary to compute the test statistics.

may in fact be much further upstream or downstream. Although association data may identify and implicate genes in the clinical course of a disease, functional analysis of variants of interest should be conducted to better understand their role in BPD pathogenesis.

Multilocus haplotype association analyses are advantageous because they provide a stronger power of association than the single nucleotide methods [66]. The results from these analyses confirmed and extended on the trends seen by individual SNP analysis. Combinations of B-18 C, B1013 A and B1580 T were found in susceptibility haplotypes by both TDT and FBAT. Moreover, significant three marker linkage disequilibrium for B-18(A/C), B1013(A/C), and B1580(C/T) has been reported in several populations [37]. In the Greek population the frequency of the B-18 C/B1013 A/B1580 T and B-18 A/B1013 C/B1580 C haplotypes of these three marker loci is the same (0.274) [37]. Therefore, association of the CAT haplotype of these marker loci with susceptibility to BPD in this study is not an artifact

of the overall population characteristic, but it may be rather disease related.

Although FBAT and TDT analyses identified similar alleles and/or haplotypes associated with BPD, different findings were also observed between these two analyses. For example, FBAT revealed haplotypes with additional upstream loci (D2S388 5/D2S2232 2/AAGG 3 and D2S388 5/D2S2232 2/AAGG 3/B-18 A) for BPD 28D and BPD28D/36W. The basis for the observed differences is currently unknown, but it may be due to differences in approach and/or the underlying assumptions of the two tests (TDT and FBAT). Haplotype TDT uses the maximum likelihood approach based on the Lander-Green hidden-Markov-model (HMM) [33], while the FBAT uses a weighted conditional approach [25]. Therefore, the differences in results are more likely to occur for low frequency haplotypes. Moreover, haplotype FBAT assumes no recombination among the marker loci [25] and haplotype TDT assumes no linkage disequilibrium [33]. Recent studies have used these two tests in complement for

either single nucleotide polymorphism analysis [5,40, 64] or haplotype analysis [8]. However, the partial discrepancies observed in the results by these two different family based association tests in these studies as well as the results presented in this study, point to the need for future studies to compare different analytical methods in parallel. Such efforts could identify differences in approach and assumption, understand the implications of such differences and clearly define advantages and disadvantages of each approach.

Due to the presumed overlap between BPD and RDS, it was hypothesized that similar genetic association findings would be seen in these two diseases [17,22]. It was, therefore, surprising that the B1580 T variant was found in the susceptibility haplotypes in the present study. This SNP codes for the non-glycosylated Ile (A TT) variant [58]. This variant has been associated with decreased risk of RDS in blacks in the presence of the SP-A1 allele (6A^3) [17], whereas the glycosylated Thr (ACT) variant has been shown to associate with susceptibility to RDS especially in the first born and male infant in a study of prematurely born twins [17, 41]. However, this is consistent with previous observations where factors that may protect from RDS, increase risk for BPD. For example, prenatal inflammation (chorioamnionitis) and increased post-natal lung inflammation associate with a decreased incidence of RDS among preterm babies, but these are associated with an increased risk for chronic lung injury following preterm delivery [60].

In summary, factors involved in predisposing infants to develop BPD following RDS are likely complex and presumably involve several loci with small effects, effectively making the traditional genetic studies difficult. In the present study, using two different family based association tests, we show that the SP-A/SP-D locus associates with BPD 28D, and the SP-B and SP-B-linked microsatellite marker loci with BPD 36W and BPD 28D/36W. However, larger sample size studies with various BPD subgroups are needed to extend these results and enable localization and characterization of genetic factors.

Acknowledgements

References

[1] E. Bancalari, G.E. Abdenour, R. Feller and J. Gannon, Bronchopulmonary dysplasia: clinical presentation, *J Pediatr* **95** (1979), 819–823.

[2] T.M. Berger, I.I. Bachmann, M. Adams and G. Schubiger, Impact of improved survival of very low-birth-weight infants on incidence and severity of bronchopulmonary dysplasia, *Biol Neonate* **86** (2004), 124–130.

[3] M. Blayney, E. Kerem, H. Whyte and H. O'Brodovich, Bronchopulmonary dysplasia: improvement in lung function between 7 and 10 years of age, *J Pediatr* **118** (1991), 201–206.

[4] V. Boggaram, Regulation of lung surfactant protein gene expression, *Front Biosci* **8** (2003), d751–764.

[5] N. Bougacha-Elleuch, A. Rebai, M. Mnif, H. Makni, M. Bellassouad, J. Jouida, M. Abid and A. Hammadi, Analysis of MHC genes in a Tunisian isolate with autoimmune thyroid diseases: implication of TNF-308 gene polymorphism, *J Autoimmun* **23** (2004), 75–80.

[6] E.C. Crouch, Surfactant protein-D and pulmonary host defense, *Respir Res* **1** (2000), 93–108.

[7] D. Curtis, Use of siblings as controls in case-control association studies, *Ann Hum Genet* **61**(Pt 4) (1997), 319–333.

[8] K.E. Deffenbacher, J.B. Kenyon, D.M. Hoover, R.K. Olson, B.F. Pennington, J.C. DeFries and S.D. Smith, Refinement of the 6p21.3 quantitative trait locus influencing dyslexia: linkage and association analyses, *Hum Genet* **115** (2004), 128–138.

[9] D.E. DeMello, S. Heyman, D.S. Phelps and J. Floros, Immunogold localization of SP-A in lungs of infants dying from respiratory distress syndrome, *Am J Pathol* **142** (1993), 1631–1640.

[10] D.E. DeMello, D.S. Phelps, G. Patel, J. Floros and D. Lagunoff, Expression of the 35 kDa and low molecular weight surfactant-associated proteins in the lungs of infants dying with respiratory distress syndrome, *Am J Pathol* **134** (1989), 1285–1293.

[11] S. DiAngelo, Z. Lin, G. Wang, S. Phillips, M. Ramet, J. Luo and J. Floros, Novel, non-radioactive, simple and multiplex PCR-cRFLP methods for genotyping human SP-A and SP-D marker alleles, *Dis Markers* **15** (1999), 269–281.

[12] E. Elahi and M. Ronaghi, Pyrosequencing: a tool for DNA sequencing analysis, *Methods Mol Biol* **255** (2004), 211–219.

[13] R. Epaud, M. Ikegami, J.A. Whitsett, A.H. Jobe, T.E. Weaver and H.T. Akinbi, Surfactant protein B inhibits endotoxin-induced lung inflammation, *Am J Respir Cell Mol Biol* **28** (2003), 373–378.

[14] M. Filippone, M. Sartor, F. Zacchello and E. Baraldi, Flow limitation in infants with bronchopulmonary dysplasia and respiratory function at school age, *Lancet* **361** (2003), 753–754.

[15] J. Floros, Sixty years of surfactant research, *Am J Physiol* **258** (1990), L238–240.

[16] J. Floros and R. Fan, Surfactant protein A and B genetic variants and respiratory distress syndrome: allele interactions, *Biol Neonate* **80**(Suppl 1) (2001), 22–25.

[17] J. Floros, R. Fan, S. Diangelo, X. Guo, J. Wert and J. Luo, Surfactant protein (SP) B associations and interactions with SP-A in white and black subjects with respiratory distress syndrome, *Pediatr Int* **43** (2001), 567–576.

[18] J. Floros, R. Fan, A. Matthews, S. DiAngelo, J. Luo, H. Nielsen, M. Dunn, I.H. Gewolb, J. Koppe, L. Van Sonderen, L. Farri-Kostopoulos, M. Tzaki, M. Ramet and J. Merrill, Family-based transmission disequilibrium test (TDT) and case-control

association studies reveal surfactant protein A (SP-A) susceptibility alleles for respiratory distress syndrome (RDS) and possible race differences, *Clin Genet* **60** (2001), 178–187.

[19] J. Floros and P. Kala, Surfactant proteins: molecular genetics of neonatal pulmonary diseases, *Annu Rev Physiol* **60** (1998), 365–384.

[20] R. Haataja and M. Hallman, Surfactant proteins as genetic determinants of multifactorial pulmonary diseases, *Ann Med* **34** (2002), 324–333.

[21] R. Haataja, R. Marttila, P. Uimari, J. Lofgren, M. Ramet and M. Hallman, Respiratory distress syndrome: evaluation of genetic susceptibility and protection by transmission disequilibrium test, *Hum Genet* **109** (2001), 351–355.

[22] R. Haataja, M. Ramet, R. Marttila and M. Hallman, Surfactant proteins A and B as interactive genetic determinants of neonatal respiratory distress syndrome, *Hum Mol Genet* **9** (2000), 2751–2760.

[23] R.R. Hoover and J. Floros, Organization of the human SP-A and SP-D loci at 10q22–q23. Physical and radiation hybrid mapping reveal gene order and orientation, *Am J Respir Cell Mol Biol* **18** (1998), 353–362.

[24] S. Horvath, X. Xu and N.M. Laird, The family based association test method: strategies for studying general genotype–phenotype associations, *Eur J Hum Genet* **9** (2001), 301–306.

[25] S. Horvath, X. Xu, S.L. Lake, E.K. Silverman, S.T. Weiss and N.M. Laird, Family-based tests for associating haplotypes with general phenotype data: application to asthma genetics, *Genet Epidemiol* **26** (2004), 61–69.

[26] A.H. Jobe, Antenatal factors and the development of bronchopulmonary dysplasia, *Semin Neonatol* **8** (2003), 9–17.

[27] A.H. Jobe and E. Bancalari, Bronchopulmonary dysplasia, *Am J Respir Crit Care Med* **163** (2001), 1723–1729.

[28] A.H. Jobe and M. Ikegami, Lung development and function in preterm infants in the surfactant treatment era, *Annu Rev Physiol* **62** (2000), 825–846.

[29] A.H. Jobe and M. Ikegami, Mechanisms initiating lung injury in the preterm, *Early Hum Dev* **53** (1998), 81–94.

[30] A.J. Jobe, The new BPD: an arrest of lung development, *Pediatr Res* **46** (1999), 641–643.

[31] P. Kala, M. Koptides, S. Diangelo, R.R. Hoover, Z. Lin, V. Veletza, D. Kouretas and J. Floros, Characterization of markers flanking the human SP-B locus, *Dis Markers* **13** (1997), 153–167.

[32] A.M. Karinch, D.E. DeMello and J. Floros, Effect of genotype on the levels of surfactant protein A mRNA and on the SP-A2 splice variants in adult humans, *Biochem J* **321**(Pt 1) (1997), 39–47.

[33] L. Kruglyak, M.J. Daly, M.P. Reeve-Daly and E.S. Lander, Parametric and nonparametric linkage analysis: a unified multipoint approach, *Am J Hum Genet* **58** (1996), 1347–1363.

[34] N.M. Laird, S. Horvath and X. Xu, Implementing a unified approach to family-based tests of association, *Genet Epidemiol* **19**(Suppl 1) (2000), S36–42.

[35] R. Leth-Larsen, P. Garred, H. Jensenius, J. Meschi, K. Hartshorn, J. Madsen, I. Tornoe, H.O. Madsen, G. Sorensen, E. Crouch and U. Holmskov, A common polymorphism in the SFTPD gene influences assembly, function, and concentration of surfactant protein D, *J Immunol* **174** (2005), 1532–1538.

[36] Z. Lin, C. Pearson, V. Chinchilli, S.M. Pietschmann, J. Luo, U. Pison and J. Floros, Polymorphisms of human SP-A, SP-B, and SP-D genes: association of SP-B Thr131Ile with ARDS, *Clin Genet* **58** (2000), 181–191.

[37] W. Liu, C.M. Bentley and J. Floros, Study of human SP-A, SP-B and SP-D loci: allele frequencies, linkage disequilibrium and heterozygosity in different races and ethnic groups, *BMC Genet* **4** (2003), 13.

[38] I.R. Makhoul, Neonatal lung injury and new BPD: need for a new terminology, *J Pediatr* **141** (2002), 150.

[39] V. Makri, B. Hospes, S. Stoll-Becker, A. Borkhardt and L. Gortner, Polymorphisms of surfactant protein B encoding gene: modifiers of the course of neonatal respiratory distress syndrome? *Eur J Pediatr* **161** (2002), 604–608.

[40] I. Manor, M. Corbex, J. Eisenberg, I. Gritsenkso, R. Bachner-Melman, S. Tyano and R.P. Ebstein, Association of the dopamine D5 receptor with attention deficit hyperactivity disorder (ADHD) and scores on a continuous performance test (TOVA), *Am J Med Genet B Neuropsychiatr Genet* **127** (2004), 73–77.

[41] R. Marttila, R. Haataja, M. Ramet, J. Lofgren and M. Hallman, Surfactant protein B polymorphism and respiratory distress syndrome in premature twins, *Hum Genet* **112** (2003), 18–23.

[42] R. Marttila, R. Haataja, M. Ramet, M.L. Pokela, O. Tammela and M. Hallman, Surfactant protein A gene locus and respiratory distress syndrome in Finnish premature twin pairs, *Ann Med* **35** (2003), 344–352.

[43] J.D. Merrill, R.A. Ballard, A. Cnaan, A.M. Hibbs, R.I. Godinez, M.H. Godinez, W.E. Truog and P.L. Ballard, Dysfunction of Pulmonary Surfactant in Chronically Ventilated Premature Infants, *Pediatr Res* (2004).

[44] J.R. O'Connell and D.E. Weeks, PedCheck: a program for identification of genotype incompatibilities in linkage analysis, *Am J Hum Genet* **63** (1998), 259–266.

[45] S.V. Pena, D.A. Hanson, B.A. Carr, T.J. Goralski and A.M. Krensky, Processing, subcellular localization, and function of 519 (granulysin), a human late T cell activation molecule with homology to small, lytic, granule proteins, *J Immunol* **158** (1997), 2680–2688.

[46] D.S. Phelps, Surfactant regulation of host defense function in the lung: a question of balance, *Pediatr Pathol Mol Med* **20** (2001), 269–292.

[47] D. Rabinowitz and N. Laird, A unified approach to adjusting association tests for population admixture with arbitrary pedigree structure and arbitrary missing marker information, *Hum Hered* **50** (2000), 211–223.

[48] M. Ramet, R. Haataja, R. Marttila, J. Floros and M. Hallman, Association between the surfactant protein A (SP-A) gene locus and respiratory-distress syndrome in the Finnish population, *Am J Hum Genet* **66** (2000), 1569–1579.

[49] M. Ronaghi, Pyrosequencing for SNP genotyping, *Methods Mol Biol* **212** (2003), 189–195.

[50] M. Rova, R. Haataja, R. Marttila, V. Ollikainen, O. Tammela and M. Hallman, Data mining and multiparameter analysis of lung surfactant protein genes in bronchopulmonary dysplasia, *Hum Mol Genet* **13** (2004), 1095–1104.

[51] M. Selman, H.M. Lin, M. Montano, A.L. Jenkins, A. Estrada, Z. Lin, G. Wang, S.L. DiAngelo, X. Guo, T.M. Umstead, C.M. Lang, A. Pardo, D.S. Phelps and J. Floros, Surfactant protein A and B genetic variants predispose to idiopathic pulmonary fibrosis, *Hum Genet* **113** (2003), 542–550.

[52] P.C. Sham and D. Curtis, An extended transmission/disequilibrium test (TDT) for multi-allele marker loci, *Ann Hum Genet* **59**(Pt 3) (1995), 323–336.

[53] A.T. Shennan, M.S. Dunn, A. Ohlsson, K. Lennox and E.M. Hoskins, Abnormal pulmonary outcomes in premature infants: prediction from oxygen requirement in the neonatal period, *Pediatrics* **82** (1988), 527–532.

[54] M.C. Shih and A.S. Whittemore, Tests for genetic association using family data, *Genet Epidemiol* **22** (2002), 128–145.

[55] R.S. Spielman and W.J. Ewens, The TDT and other family-based tests for linkage disequilibrium and association, *Am J Hum Genet* **59** (1996), 983–989.

[56] R.S. Spielman, R.E. McGinnis and W.J. Ewens, Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM), *Am J Hum Genet* **52** (1993), 506–516.

[57] G.T. Wagenaar, S.A. Ter Horst, M.A. Van Gastelen, L.M. Leijser, T. Mauad, P.A. Van der Velden, E. De Heer, P.S. Hiemstra, B.J. Poorthuis and F.J. Walther, Gene expression profile and histopathology of experimental bronchopulmonary dysplasia induced by prolonged oxidative stress, *Free Radic Biol Med* **36** (2004), 782–801.

[58] G. Wang, N.D. Christensen, B. Wigdahl, S.H. Guttentag and J. Floros, Differences in N-linked glycosylation between human surfactant protein-B variants of the C or T allele at the single-nucleotide polymorphism at position 1580: implications for disease, *Biochem J* **369** (2003), 179–184.

[59] G. Wang, X. Guo and J. Floros, Human SP-A 3'-UTR variants mediate differential gene expression in basal levels and in response to dexamethasone, *Am J Physiol Lung Cell Mol Physiol* **284** (2003), L738–748.

[60] K.L. Watterberg, L.M. Demers, S.M. Scott and S. Murphy, Chorioamnionitis and early lung inflammation in infants in whom bronchopulmonary dysplasia develops, *Pediatrics* **97** (1996), 210–215.

[61] T.E. Weaver and J.J. Conkright, Function of surfactant proteins B and C, *Annu Rev Physiol* **63** (2001), 555–578.

[62] B. Weber, A. Borkhardt, S. Stoll-Becker, I. Reiss and L. Gortner, Polymorphisms of surfactant protein A genes and the risk of bronchopulmonary dysplasia in preterm infants, *Turk J Pediatr* **42** (2000), 181–185.

[63] M.A. Wilder, Surfactant protein B deficiency in infants with respiratory failure, *J Perinat Neonatal Nurs* **18** (2004), 61–67.

[64] A.H. Wong, J. Trakalo, O. Likhodi, M. Yusuf, A. Macedo, M.H. Azevedo, T. Klempan, M.T. Pato, W.G. Honer, C.N. Pato, H.H. Van Tol and J.L. Kennedy, Association between schizophrenia and the syntaxin 1A gene, *Biol Psychiatry* **56** (2004), 24–29.

[65] J.R. Wright, P. Borron, K.G. Brinker and R.J. Folz, Surfactant Protein A: regulation of innate and adaptive immune responses in lung inflammation, *Am J Respir Cell Mol Biol* **24** (2001), 513–517.

[66] H. Zhao, R. Pfeiffer and M.H. Gail, Haplotype analysis in population genetics and association studies, *Pharmacogenomics* **4** (2003), 171–178.