Genetic association study for RSV bronchiolitis in infancy at the 5q31 cytokine cluster

J T Forton,[1,2] K Rowlands,[1] K Rockett,[1] N Hanchard,[1] M Herbert,[2] D P Kwiatkowski,[1,2,3] J Hull[1,2]

► Additional tables and figure are published online only at http://thorax.bmj.com/content/vol64/issue4

[1] The Wellcome Trust Centre for Human Genetics, University of Oxford, UK; [2] Department of Paediatrics, University of Oxford, UK; [3] The Wellcome Trust Sanger Institute, Hinxton, Cambridge, UK

Correspondence to: Dr J T Forton, University Department of Paediatrics, John Radcliffe Hospital, Headley Way, Headington, Oxford OX3 9DU, UK; julian.forton@paediatrics.ox.ac.uk

Received 22 May 2008
Accepted 3 December 2008
Published Online First 6 January 2009

ABSTRACT

Background: The pathophysiological basis of severe respiratory syncytial virus (RSV) bronchiolitis in infancy is poorly understood and has hindered vaccine development. Studies implicate the cell-mediated immune response in the pathogenesis of the disease. A recent twin study estimated a heritable contribution of 22% to RSV bronchiolitis. Genetic epidemiology provides a new approach to identifying important immune determinants of disease severity.

Methods: A comprehensive high-density gene-region association study for severe RSV bronchiolitis in infancy at 5q31 across 11 genes including the Th2-cytokine cluster was performed. A haplotype tagging approach was used to analyse genetic variation at 113 single nucleotide polymorphisms (SNPs) in 780 independent cases and 1045 controls. The study had sufficient power to detect small effects, perform extensive haplotype analysis and analyse both a principal phenotype and a refined age-limited phenotype enriched for first-exposure RSV infection.

Results: SNP associations were found at *IL4* and a highly significant risk haplotype was identified across *IL13 CNS-1* and *IL4* (odds ratio 1.69, $p<0.0001$), present in both case-control and family-based analyses. All associations were strongest for a phenotype limited to <6 months of age, implicating this locus in primary RSV disease. The same risk haplotype has previously been shown to be associated with increased *IL13* expression.

Conclusions: A haplotype at *IL13–1L4*, which is associated with increased *IL13* production, confers an increased risk of severe primary RSV bronchiolitis in early infancy. This study, together with previous studies implicating the same locus in atopic sensitisation, suggests that primary RSV bronchiolitis and atopy share a genetic contribution at the *IL13–IL4* locus.

Respiratory syncytial virus (RSV) bronchiolitis is responsible for between 18 000–75 000 hospital admissions and 200–500 deaths annually in the USA.[1] The pathophysiological basis of severe RSV bronchiolitis in infancy is poorly understood and this has hindered vaccine development. Studies consistently implicate the host response as an important determinant of disease severity, and many immunological studies have linked the Th1/Th2 spectrum of T cell differentiation in disease pathogenesis.[2–4]

The possibility of a genetic vulnerability is supported by a recent twin study comparing monozygotic with dizygotic twins, which showed RSV bronchiolitis in infancy to have a heritable contribution of 22%.[5]

Genetic epidemiology provides a new approach to dissecting the pathogenesis of complex disease traits. Study design is paramount, and many early studies have been difficult to replicate because of small sample size, consequent lack of power, phenotypic heterogeneity between studies and failure to accommodate for population substructure or multiple testing. Confidence in results can be gained through replication in independent studies.[6]

Animal studies of RSV bronchiolitis have been used extensively to investigate the immune response to RSV infection.[7 8] Less is known about the immune response of affected infants. Genetic epidemiology represents a powerful approach to identifying important immune determinants of severity in human disease.

Three recent genetic association studies of RSV bronchiolitis have identified associations for promoter polymorphisms in the Th2 cytokine *IL4*.[9–11] These studies were performed in small cohorts with limited power and sampled a small number of candidate polymorphisms only. Moderate associations were reported without accommodation for multiple testing.

In this study we performed a high-density comprehensive gene-region association study for severe RSV bronchiolitis in infancy at 5q31, a region containing the Th2 cytokine cluster and many other important immune genes. We applied a haplotype tagging approach to sample variation at 113 single nucleotide polymorphisms (SNPs) across the genes *IL3*, *GMCSF*, *P4HA2*, *RIL*, *SLC22A4*, *SLC22A5*, *IRF-1*, *IL5*, *RAD50*, *IL13* and *IL4*, and across all known intergenic regulatory elements. We used 780 independent cases and 1045 population controls of documented white European ancestry, with sufficient power to detect small effects, perform extensive haplotype analysis and accommodate for multiple testing. Our large sample size gives us sufficient power to concentrate specifically on first exposure RSV in the immunologically naïve infant by performing analysis on an age-limited phenotype (<6 months of age). This enabled us to dissect out the phenotype of primary RSV disease form RSV-induced wheeze or atopic wheeze precipitated by RSV infection that may present towards the end of the first year of life. We identified a highly significant risk haplotype across *IL13 CNS-1* and *IL4* which we have previously shown to be associated with increased *IL13* expression.[12]

METHODS

DNA samples and DNA extraction

The Oxford RSV DNA Archive has been described elsewhere.[13] In this study, 782 cases,

Thorax: first published as 10.1136/thx.2008.102111 on 8 January 2009. Downloaded from http://thorax.bmj.com/ on April 27, 2024 by guest. Protected by copyright.

1045 cord blood controls and 673 families were used in association analysis.

Identification and selection of SNPs across 5q31

The public databases CHIP Bioinformatics[14] and Project Ensembl were searched for all known SNPs across 656 kB of the 5q31 gene region. PubMed was searched for relevant literature relating to SNP discovery, disease association and functional molecular genetics at 5q31.

Strategy in deciding which SNPs to genotype was directed by two aims: to comprehensively describe the haplotype diversity of the 5q31 gene region and to enrich for SNPs with potential functionality.

Under the hypothesis of evolutionary constraint, cross species conserved non-coding sequences (CNS) may contain functional DNA important in gene regulation. SNPs in CNS regions were identified by human-mouse sequence homology studies using the web-based program VISTA.[15]

Transcription factor (TF) binding sites were identified using the program Matinspector. SNPs predicted to disrupt TF sequence motifs were again considered as important candidates for association analysis.[16]

In selecting SNPs for genotyping, priority was given to those SNPs previously validated in a white European population with a predicted frequency of >5%, to those positioned within a gene, located in particular within the promoter, exons or within 300 base pairs of an intron/exon boundary, to those SNPs in or within 300 base pairs of a CNS, and to those SNPs at a potential TF binding site. All SNPs with a previous positive disease association in the literature were also included.

Case definition, study phenotypes and control samples

Cases

The Oxford RSV archive contains in excess of 1200 cases. For the current study, the following phenotypes were defined and studied.

The principal phenotype is defined by the following criteria: a clinical diagnosis of bronchiolitis (breathlessness, chest wall recession and inspiratory crepitations on auscultation), evidence of RSV on nasopharyngeal aspirate, requirement for oxygen and/or nasogastric feeds during hospital admission, age <1 year at time of hospital admission. Only children with two white European parents were included in the study to avoid the effects of population substructure.[17] This phenotype is largely consistent with previous studies of RSV disease in infancy at the IL4 locus, although these studies enrolled children up to 2 years of age.

The second phenotype used in this study is a subgroup—here called the refined phenotype—which, in addition to the above criteria, also excludes those children with known risk factors for RSV disease (prematurity, chronic lung disease, congenital heart disease) and is restricted to infants under 6 months of age at the time of illness. This phenotype enables us to study more clearly the impact of polymorphism at 5q31 in the immunologically naïve infant during primary RSV exposure. Demographic data on both phenotypes are shown in table 1.

Table 1 Demographic data for cases used in the principal phenotype and the refined phenotype

	Principal phenotype Median (mean)	Refined phenotype (age-limited) Median (mean)
Duration of admission (days)	4 (5.3)	4 (4.9)
Weight on admission (kg)	5.15 (5.47)	5.03 (5.18)
Age on admission (weeks)	10 (15.6)	8 (9.1)
RSV positive (%)	100%	100%
Oxygen used (%)	88%	86%
Duration of oxygen (days)	3 (3.78)	3 (3.6)
Tube feeding required (%)	60.9%	61.7%
Intravenous fluids required (%)	32.1%	31.9%
Ventilation required (%)	23.7%	14.5%
Parental smoking (%)	36.6%	33.7%
Number of older siblings	1 (1.19)	1 (1.19)
Gestation (weeks)	39 (38)	40 (39.4)
Pre-existing heart condition (%)	5.5%	0.0%
Pre-existing lung condition (%)	5.9%	0.0%
Oxygen-dependent CLD (%)	2.0%	0.0%

CLD, chronic lung disease; RSV, respiratory syncytial virus.

Controls

The 1045 population controls used in the study were sequential cord blood samples taken from white European infants born at the John Radcliffe Hospital, Oxford. Population controls can be used in association studies where the disease phenotype is comparatively rare in the general population. This is the favoured approach taken by the Wellcome Trust Case Control Consortium GWAS initiative[18] and is appropriate for severe RSV bronchiolits which has a frequency of 1–3% in the general population.

Haplotype tagging approach to association study

A haplotype tagging approach to association was implemented. In this approach, many SNPs are selected and genotyped in a small representative group of population samples and the haplotypic relationships between these SNPs is inferred. A subset of haplotype tagging SNPs (htSNPs) which together comprehensively describe all the genetic diversity of the region can then be selected and taken forward to genotyping in the cases and controls. This reduces genotyping costs when using large cohorts.

In this study, all selected SNPs were first genotyped in 32 representative white European family trios and common population-specific haplotypes were inferred using the algorithms Phamily and PHASE.[19] The pattern of linkage disequilibrium across the region was analysed using the programs HaploXT[20] and MARKER,[21] and revealed well demarcated haplotype blocks. An efficient subset of htSNPs was selected within each block independently, using methods described elsewhere.[22] A total of 48 htSNPs were selected and taken forward for genotyping in the case-control association study.

Study design

In total, 782 cases and 1045 controls were genotyped in this association study.

As a cost-saving strategy, all 48 htSNPS were first genotyped in 420 cases and 576 controls (cohort 1). htSNPs with the strongest positive associations in cohort 1, together with SNPs required to generate haplotypes across *IL4_IL13*, were then genotyped in the remaining 362 cases and 469 controls (cohort 2).

The primary analysis for this study relates to the seven SNPs genotyped in the total cohort (amounting to a total of 782 cases and 1045 controls). These were used in single SNP and haplotype analysis. The data for the 48 SNPs genotyped in cohort 1 were also analysed independently in an extensive haplotype analysis spanning 5q31.

Thorax: first published as 10.1136/thx.2008.102111 on 8 January 2009. Downloaded from http://thorax.bmj.com/ on April 27, 2024 by guest. Protected by copyright.

The seven htSNPs genotyped in the total cohort were also tested in a family-based association analysis in 673 of the cases where DNA from both parents was available.

Genotyping and data curation

Methods of genotyping and curation have been described elsewhere.[23] Briefly, all genotyping was performed using Sequenom MassArray using whole genome preamplified DNA. Primers and multiplexes were designed using SpectroDESIGNER. Curation of genotyping calls was implemented using SpectroTYPER. SNP assays were accepted if SNP frequency exceeded 5%, the assay was in Hardy-Weinberg equilibrium in controls (HWE χ^2 $p>0.05$) and genotyping success exceeded 80%.

Statistical analysis

Single SNP analysis

Only SNPs with a population frequency of >5% were included in the analysis. For each SNP, allelic association statistics were generated using a 2×2 χ^2 test (1df) with the odds ratio calculated for minor allele versus major allele. Genotype association statistics were generated using a 2×3 χ^2 test (2df).

Haplotype analysis

Phase 2.1.1 implementing haplotype probability assignments was used to generate haplotypes.[19] Each individual may be allocated several different haplotype pairs with different probabilities and these are incorporated into the association analysis such that a single individual may contribute to the overall frequency of several different haplotype pairs.

Similar haplotypes that differ at one or a small number of sites are likely to be related by genealogy and, by virtue of their shared ancestry, these haplotypes are likely to share other unidentified recent alleles. Grouping these haplotypes together by genealogy will therefore theoretically increase the power to detect unobserved variants in association analysis. This forms the basis of cladistic haplotype association analysis.

Haplotype clades within each haplotype block were generated by hierarchical clustering using the program NEIGHBOUR. All within-block haplotype analysis was performed using 2×2 χ^2 tests generated for each clade versus all other clades.

Family-based association statistics were calculated using the transmission disequilibrium test (TDT).[24]

RESULTS

Population haplotypes and selection of htSNPs

Using the criteria outlined above, 152 SNPs of interest were selected for genotyping in population samples; 113 SNPs satisfied Hardy-Weinberg equilibrium, had a genotyping success rate exceeding 80% (median 96.9%, mean 94.8%) and a population frequency of >5%. Supplementary figure 1 published online only shows the distribution of SNPs identified and genotyped. Information including genotyping data for the 113 successful SNP assays can be found in supplementary table 1 in the online supplement. Genotyping data for the 113 SNPs were used to generate long-range population haplotypes across the gene region. Patterns of pairwise LD across the region were

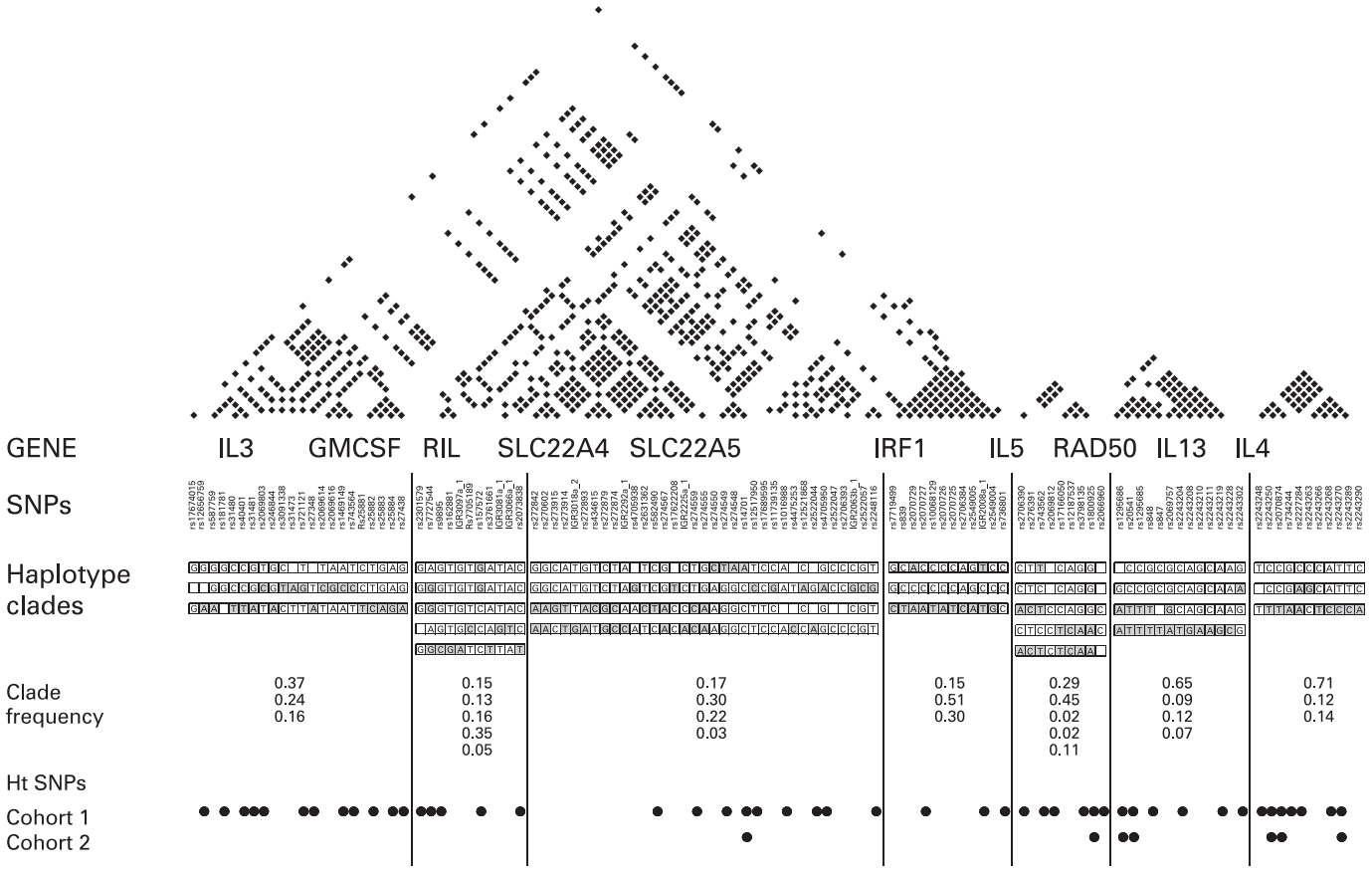


Figure 1 The 5q31 gene region used in association analysis. Annotations include pairwise r^2 statistics >0.3 for 113 single nucleotide polymorphisms (SNPs) across 11 genes, haplotype blocks, haplotype clades identified using hierarchical clustering and selected haplotype tagging SNPs (htSNPs) for case-controls 1 and 2.

Thorax: first published as 10.1136/thx.2008.102111 on 8 January 2009. Downloaded from http://thorax.bmj.com/ on April 27, 2024 by guest. Protected by copyright.

Table 2 Principal phenotype

			Cases			Controls			Genotype association		Allele association		
	Gene	SNP	Total	% failed	Freq	Total	% failed	Freq	χ^2 2df	χ^2 p Value	Odds ratio	χ^2 1df	χ^2 p Value
A	GMCSF	rs7727544	415	5.80	0.40	576	13.0	0.45	6.56	**0.038**	0.83 (0.69–1.01)	3.61	0.058
	SLC22A5	rs274549	415	2.70	0.18	576	2.6	0.13	8.20	**0.017**	1.39 (1.08–1.79)	6.43	**0.011**
	SLC22A5	rs14701	415	1.4	0.18	576	3.6	0.13	11.69	**0.003**	1.46 (1.13–1.87)	8.32	**0.004**
	IL13	rs1800925	415	1.9	0.18	576	4.9	0.21	2.14	0.343	0.85 (0.67–1.07)	1.79	0.181
	IL13	rs1295686	415	1.4	0.19	576	2.6	0.18	0.48	0.788	1.08 (0.86–1.36)	0.37	0.544
	IL13	rs20541	415	0.2	0.18	576	6.8	0.15	3.46	0.177	1.25 (0.98–1.59)	3.04	0.081
	CNS-1	rs2243302	415	1.45	0.11	576	2.78	0.13	2.39	0.303	0.8 (0.61–1.06)	2.19	0.139
	IL4	rs2243250	415	4.8	0.14	576	5.2	0.11	5.99	**0.050**	1.30 (0.99–1.71)	3.36	0.067
	IL4	rs2070874	415	3.6	0.15	576	6.8	0.12	8.87	**0.012**	1.32 (1.01–1.72)	3.87	**0.049**
	IL4	rs734244	415	5.1	0.15	576	5.2	0.12	6.34	**0.042**	1.31 (1.00–1.72)	3.62	**0.057**
	IL4	rs2227284	415	3.9	0.27	576	8.2	0.22	5.09	0.078	1.27 (1.03–1.58)	4.85	**0.028**
	IL4	rs2243268	413	1.0	0.15	576	2.1	0.12	3.27	0.195	1.26 (0.97–1.64)	2.82	0.093
	IL4	rs2243270	415	3.4	0.16	576	4.7	0.13	7.55	**0.023**	1.34 (1.03–1.73)	4.58	**0.032**
B	SLC22A5	rs14701	780	3.1	0.17	1045	3.6	0.14	7.52	**0.023**	1.21 (1.01–1.45)	3.91	**0.048**
	IL13	rs1800925	780	1.9	0.18	1045	3.4	0.19	0.45	0.800	0.96 (0.81–1.14)	0.19	0.665
	IL13	rs1295686	780	1.2	0.20	1045	1.9	0.18	3.00	0.223	1.15 (0.97–1.36)	2.59	0.108
	IL13	rs20541	782	3.6	0.19	1045	6.1	0.16	5.14	0.077	1.23 (1.03–1.46)	4.92	**0.027**
	IL4	rs2243250	780	3.3	0.15	1044	3.9	0.13	4.42	0.110	1.22 (1.00–1.48)	3.75	**0.053**
	IL4	rs2070874	779	3.0	0.15	1045	4.2	0.13	3.95	0.139	1.18 (0.98–1.43)	2.83	0.092
	IL4	rs2243270	780	5.1	0.16	1045	4.7	0.14	3.58	0.167	1.18 (0.97–1.42)	2.69	0.101

Single SNP association statistics for all SNPs with positive findings and those used in haplotype analysis: (A) Cohort 1. (B) Total cohort. Association statistics for all 48 htSNPs genotyped in cohort 1 are available in table 2A in the online supplement.
Freq, frequency.

analysed using the metric r^2 and demonstrate seven clearly demarcated juxtaposed haplotype blocks spanning the region (fig 1). Haplotype clades generated using hierarchical clustering revealed between three and five common haplotype motifs within each haplotype block. This small number of haplotype motifs describes between 75% and 94% of all observed haplotypes within each haplotype block. A total of 48 htSNPs were selected for genotyping in case-control 1. Pairwise r^2 statistics, haplotype blocks, haplotype clades and selected htSNPs are shown in fig 1.

Case-control study

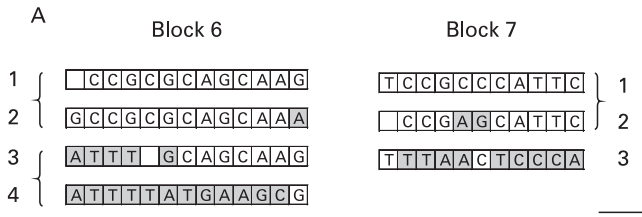
Forty-eight htSNPs were genotyped in 420 cases and 576 controls of documented white European ancestry in cohort 1.

Table 3 Refined phenotype

			Cases			Controls			Genotype association		Allele association		
	Gene	SNP	Total	% failed	Freq	Total	% failed	Freq	χ^2 2df	χ^2 p Value	Odds ratio	χ^2 1df	χ^2 p Value
A	GMCSF	rs7727544	218	6.9	0.41	576	13.0	0.45	1.68	0.432	0.85 (0.68–1.08)	1.61	0.205
	SLC22A5	rs274549	218	2.8	0.17	576	2.6	0.13	2.92	0.232	1.29 (0.95–1.76)	2.38	0.123
	SLC22A5	rs14701	218	1.4	0.16	576	3.6	0.13	4.24	0.120	1.31 (0.96–1.79)	2.72	0.099
	IL13	rs1800925	218	2.3	0.17	576	4.9	0.21	3.47	0.176	0.78 (0.59–1.05)	2.46	0.117
	IL13	rs1295686	218	1.4	0.19	576	2.6	0.18	0.17	0.920	1.06 (0.8–1.41)	0.09	0.759
	IL13	rs20541	218	0.5	0.18	576	6.8	0.15	1.65	0.439	1.21 (0.9–1.62)	1.37	0.242
	CNS-1	rs2243302	218	2.3	0.09	576	2.8	0.13	5.34	0.069	0.66 (0.46–0.96)	4.41	**0.036**
	IL4	rs2243250	218	3.7	0.17	576	5.2	0.11	11.77	**0.003**	1.64 (1.2–2.25)	9.19	**0.002**
	IL4	rs2070874	218	2.3	0.17	576	6.8	0.12	10.52	**0.005**	1.54 (1.13–2.1)	7.04	**0.008**
	IL4	rs734244	218	7.3	0.17	576	5.2	0.12	8.26	**0.016**	1.54 (1.12–2.12)	6.59	**0.010**
	IL4	rs2227284	218	5.5	0.28	576	8.2	0.22	7.49	**0.024**	1.37 (1.06–1.78)	5.51	**0.019**
	IL4	rs2243268	218	0.5	0.17	576	2.1	0.12	5.55	**0.062**	1.43 (1.05–1.94)	4.81	**0.028**
	IL4	rs2243270	218	4.6	0.19	576	4.7	0.13	12.19	**0.002**	1.6 (1.19–2.17)	9.00	**0.003**
B	SLC22A5	rs14701	408	2.9	0.17	1045	3.6	0.14	4.70	0.096	1.20 (0.96–1.51)	2.43	0.119
	IL13	rs1800925	408	2.2	0.19	1045	3.4	0.19	0.39	0.823	1.00 (0.81–1.23)	0.00	0.961
	IL13	rs1295686	408	0.7	0.21	1045	1.9	0.18	2.65	0.266	1.18 (0.96–1.45)	2.35	0.126
	IL13	rs20541	408	3.9	0.20	1045	6.1	0.16	4.78	**0.092**	1.27 (1.02–1.57)	4.55	**0.033**
	IL4	rs2243250	408	2.9	0.18	1044	3.9	0.13	12.05	**0.002**	1.48 (1.18–1.85)	11.39	**0.0007**
	IL4	rs2070874	408	3.2	0.17	1045	4.2	0.13	8.63	**0.013**	1.39 (1.11–1.74)	7.77	**0.005**
	IL4	rs2243270	408	6.6	0.18	1045	4.7	0.14	8.54	**0.014**	1.38 (1.10–1.72)	7.55	**0.006**

Single SNP association statistics for all SNPs with positive findings and those used in haplotype analysis: (A) Cohort 1. (B) Total cohort. Association statistics for all 48 htSNPs genotyped in cohort 1 are available in table 2B in the online supplement.
Freq, frequency.

Thorax: first published as 10.1136/thx.2008.102111 on 8 January 2009. Downloaded from http://thorax.bmj.com/ on April 27, 2024 by guest. Protected by copyright.



	Principal phenotype					Refined phenotype				
	Controls	Cases	OR	χ	p	Controls	Cases	OR	χ	p
B										
C C G C G C A G C A A G / G C C G C G C A G C A A A × T C C G C C C A T T C / C C G A G C A T T C	1584	1125	0.82 (0.71–0.95)	6.11	0.0134	1575	587	0.84 (0.7–1.01)	3.43	0.0640
A T T T G C A G C A A G / A T T T T A T G A A G C G × T C C G C C C A T T C / C C G A G C A T T C	233	183	1.06 (0.86–1.3)	0.25	0.6205	241	86	0.9 (0.7–1.17)	0.48	0.4870
C C G C G C A G C A A G / G C C G C G C A G C A A A × T T T A A C T C C C A	139	111	1.08 (0.83–1.39)	0.23	0.6287	140	61	1.13 (0.82–1.54)	0.44	0.5090
A T T T G C A G C A A G / A T T T T A T G A A G C G × T T T A A C T C C C A	126	123	1.33 (1.03–1.73)	4.55	0.0329	126	80	1.69 (1.26–2.27)	12.13	0.0005

Figure 2 Cross-block haplotypes were generated across the region spanning *IL13 CNS-1* and *IL4*. (A) Within-block haplotype clades for blocks 6 (*IL4*) and 7 (*IL13* and *CNS-1*). (B) Combination of within-block haplotype clades across blocks. The association seen for clade 3 in block 7 is only present when in combination with clades 3/4 in block 6. This risk haplotype has a stronger association than any single SNP or within-block haplotype. OR, odds ratio.

Genotyping success and association statistics for all 48 htSNPs for both principal and refined phenotypes are shown in supplementary table 2 available online. SNPs with positive associations in cohort 1 or those used later in haplotype analysis are shown in tables 2A and 3A for the principal and refined phenotypes, respectively. In the analysis of the principal phenotype, significant association results are apparent for one SNP in *GMCSF*, two SNPs in *SLC22A5* and five SNPs in *IL4*. In the analysis of the refined phenotype, the associations at *GMCSF* and *SLC22A5* are no longer present. However, the *IL4* association is stronger and apparent for seven SNPs spanning *CNS-1*, the *IL4* promoter and *IL4* coding region. The strongest effect is seen for the *IL4* promoter SNP rs2243250 (genotype: 2×3 χ^2, $p<0.003$; allele: OR 1.64; 2×2 χ^2, $p=0.0024$). Cladistic haplotype association analysis in all seven haplotype blocks for cohort 1 produced association results in keeping with single SNP statistics but revealed no stronger effects. These are shown in supplementary table 3 available online.

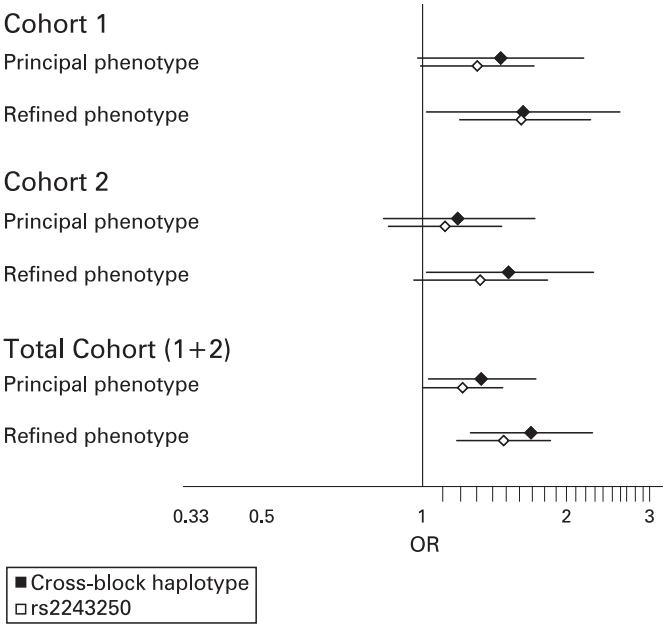


Figure 3 Summary of association statistics for cohort1 1, cohort 2 and total cohort data for the single SNP with the strongest association (rs2243250) and for the cross-block risk haplotype. In all datasets the effect is stronger for the cross-block haplotype than for the single SNP, and stronger for the refined age-limited phenotype than for the principal phenotype. The cross-block haplotype is independently significant in both cohort 1 and cohort 2. OR, odds ratio.

Seven SNPs from cohort 1 were genotyped in a further 362 cases and 469 controls in cohort 2. These seven SNPs were selected either because they had the strongest SNP associations at *IL4* and *SLC22A5* in cohort 1 or were necessary to generate haplotypes in the vicinity of *IL4–IL13*. These seven SNPs were therefore genotyped in a total cohort of 782 cases and 1045 controls. Results for the seven SNPs genotyped in the total cohort are shown in tables 2B and 3B for the principal and refined phenotypes, respectively. A highly significant result is seen for SNPs across *IL4* using the refined phenotype. Rs2243250 gives the strongest result with an odds ratio of 1.48 ($p<0.001$). Bonferroni correction for multiple testing is a conservative approach in the context of high LD as SNPs in this situation do not represent entirely independent tests. The results are, however, still significant when correction is made for 48 independent single locus tests (*IL4* rs2243250, $p=0.03$). Using methods which account for LD between tested SNPs,[25 26] the effective number of independent SNPs for this dataset is calculated at 25.1, giving an overall experiment-wide significance threshold required to keep the type I error rate at 5% of $p=0.002$. The results at the *IL4* locus ($p<0.001$) clearly remain significant after accommodation for multiple testing.

Cladistic haplotype analysis for the total cohort within block 7 produced association results in keeping with single SNP statistics but revealed no stronger effects, as the minor allele for

Thorax: first published as 10.1136/thx.2008.102111 on 8 January 2009. Downloaded from http://thorax.bmj.com/ on April 27, 2024 by guest. Protected by copyright.

Table 4 Transmission disequilibrium test (TDT) for principal phenotypes and refined phenotypes

Gene	SNP	Families	% failed	Complete families	Allele 2 transmitted	Allele 2 not transmitted	% transmission	p Value
Principal phenotypes								
SLC22A5	rs14701	658	5.6	554	139	129	51.9	0.54
IL13	rs1800925	659	4.1	582	162	177	47.8	0.42
IL13	rs1295686	660	3.3	598	197	177	52.7	0.30
IL13	rs20541	673	2.4	625	200	191	51.2	0.65
IL4	rs2243250	660	2.7	604	154	132	53.8	0.19
IL4	rs2070874	658	4.8	566	142	121	54.0	0.20
IL4	rs2243270	659	4.3	577	160	133	54.6	0.11
Refined phenotypes								
SLC22A5	rs14701	462	6.1	385	101	87	53.7	0.31
IL13	rs1800925	461	3.7	412	121	128	48.6	0.66
IL13	rs1295686	461	2.9	423	139	130	51.7	0.58
IL13	rs20541	473	2.2	442	139	138	50.2	0.95
IL4	rs2243250	461	2.9	419	117	95	55.2	0.13
IL4	rs2070874	461	5.3	393	107	83	56.3	0.08
IL4	rs2243270	462	4.5	402	120	94	56.1	0.08
Cross-block haplotype rs20541(T)-rs2070874(T)		362	0.0	362	64	44	59.3	0.05

all SNPs including rs2243250 are exclusively carried by clade 3 haplotypes.

It is recognised that haplotype block boundaries defined in block analysis are not absolute, with some haplotypes crossing block boundaries intact. It is therefore difficult to delineate the source of a positive association to the haplotype block which contains the association signal. In order to address whether the haplotype clade with association seen at *IL4* in block 7 extends across *CNS1* and *IL13*, cross-block haplotypes between blocks 6 and 7 were inferred with Phase 2.1.1 using five SNPs genotyped in the total cohort. Analysis of the five SNP cross-block haplotypes is shown in fig 2. For both the principal phenotype and the refined phenotype, the haplotype association signal detected across IL4 in block 7 segregates to a specific cross-block haplotype. The association in block 7 is stronger if it is in continuum with clades 3/4 in block 6 (principal phenotype OR 1.33 (1.03–1.73), p = 0.05; refined phenotype OR 1.69 (1.26–2.27), p<0.001) and absent if it is in continuum with clades 1/2. When the same cross-block haplotypes are generated and analysed in cohort 1 and cohort 2 independently, a significant association is found independently in both cohorts for the refined phenotype and a trend to significance with the principal phenotype. In all cases, the cross-block haplotype association is stronger for the refined phenotype than for the principal phenotype and stronger than for any single SNP, suggesting the source of the association signal is an unobserved SNP for which the cross-block haplotype is a sensitive marker. The results are summarised in fig 3.

Family analysis

The transmission disequilibrium test (TDT) was employed in a family-based study using 673 cases from the case-control study where both parents were available. TDT statistics for the seven SNPs genotyped in families are shown for the principal phenotype and refined phenotype in table 4. The family study is less well powered than the case-control study and, although no single SNP results are significant, a trend exists for both phenotypes at the *IL4* SNPs with a distortion in transmission of allele 2 at rs2243250 of 53.8% in the principal phenotype group and 55.2% (OR 1.23) in the refined phenotype group. The cross-block haplotype association seen in the case-control study was tested for the refined phenotype in the family study and showed a significant increase in transmission distortion to 59.3% (OR 1.46, p = 0.048). The magnitude of this effect is in keeping with that seen in the case-control study.

DISCUSSION

We have described a comprehensive large gene-region association study at 5q31 for a severe RSV bronchiolitis phenotype incorporating the genes *IL3*, *GMCSF, P4HA2, RIL, SLC22A4, SLC22A5, IRF-1, IL5, RAD50, IL13* and *IL4* and all known intergenic regulatory elements. We used a total of 780 independent cases and 1045 controls of documented white European ancestry with sufficient power to detect small effects, account for multiple testing, perform extensive haplotype analysis and analyse both a principal phenotype and a refined age-limited phenotype enriched for primary RSV exposure.

We defined a risk haplotype across *IL13*, *CNS-1* and *IL4*. The strongest single SNP effect (rs2243250) and the risk haplotype effect was seen for both the principal phenotype and for the refined age-limited phenotype (age <6 months). The findings were consistently stronger using the refined phenotype for which the risk haplotype carries an odds ratio of 1.69 (p<0.001). This remained significant even after conservative correction for multiple testing. The risk haplotype carried a stronger signal than any single SNP tested. This suggests that the source of the association signal is an unobserved SNP for which the risk haplotype is the most sensitive marker. A more complex epistatic effect resulting from the genetic interaction of more than one functional polymorphism may also explain this haplotype association.

We performed functional analysis using allele-specific transcript quantification at *IL13* in immortalised B cell lines, as described elsewhere.[12] This technique quantifies the relative expression of mRNA produced by the two copies of a gene within a single individual. Using many individuals, gene expression can be linked to haplotype background. The results showed that the RSV risk haplotype identified here was

Thorax: first published as 10.1136/thx.2008.102111 on 8 January 2009. Downloaded from http://thorax.bmj.com/ on April 27, 2024 by guest. Protected by copyright.

associated with an inducible upregulation of *IL13*. This effect is not explained by *IL13* promoter or exonic polymorphisms, suggesting that the risk haplotype, which spans other important regulatory regions including *CNS-1* (between *IL4* and *IL13*), may carry a more distal functional element.

Previous smaller genetic studies for RSV bronchiolitis at the *IL4–IL13* have identified positive associations for the single SNP rs2243250 and for haplotypes carrying rs2243250.[9–11] These studies were performed in small cohorts with limited power and moderate associations were reported without accommodation for multiple testing. Nevertheless, all positive associations for risk carried the minor allele T at rs2243250 and are consistent with our findings.

Previous studies have all defined the RSV phenotype using children under 2 years of age. RSV-positive children presenting with wheeze within the first 2 years of life may potentially have primary RSV bronchiolitis, RSV-induced wheeze or atopic wheeze precipitated by RSV infection. Refining the age phenotype is particularly important when studying the effects of the *IL13–IL4* locus on RSV disease, since variants at this locus have also been associated with atopy and asthma in later life.[27–29] Confounding due to over-representation of atopic infants may occur if an age-restricted phenotype is not applied. To capture a cleaner phenotype for *primary* RSV bronchiolitis, we have limited our main analysis to children aged <6 months (408 cases). We had the power to demonstrate that the effect at the *IL4–IL13* locus is even stronger in this immunologically naïve group of young infants. This strongly supports a specific association at *IL4–IL13* for *primary* severe RSV bronchiolitis.

How do the above findings integrate into the current understanding of the pathogenesis of severe primary RSV bronchiolitis? Many studies have implicated excess Th2/deficient Th1 immune responses in severe RSV disease.[2–4] Research into immunological responses to vaccines in early infancy reveals initial responses to be generically Th2-skewed with reciprocal immaturity in the Th1 cytotoxic response.[30] The rate at which the immune system matures in infancy is highly variable between individuals.[31] Those with delay in immune maturation are exposed to a period of susceptibility where Th1 cytotoxic responses to infections like RSV may be suboptimal. Recent epigenetic work looking at neonatal T cells has demonstrated incomplete silencing of the *IL13* locus in differentiating Th1 cells, suggesting that *IL13* may play an important role in the Th2 skew observed in early life.[32] The RSV risk haplotype identified here is associated with upregulation of *IL13* and may therefore have its effect on RSV susceptibility by contributing to the persistence of a Th2 environment in some infants.

Responses to primary RSV bronchiolitis and to allergens in early life may both be affected by modulation of the cytokine milieu. Delay in immune maturation has been implicated in the development of atopy with formation of Th2 memory in response to antigen exposure in infancy. Genetic studies for atopy and asthma have implicated the same locus at *IL13–IL4*, and specifically rs2243250.[27–29]

The epidemiological association between RSV bronchiolitis in infancy and atopic sensitisation and asthma has not been consistently demonstrated in prospective studies.[33 34] Both primary bronchiolitis and atopy are complex disease traits with numerous genetic and environmental contributions[35] and, in the context of multiple variables, it is conceivable that epidemiological studies may show conflicting results even if a concrete relationship exists. Our study, together with previous studies on atopy, suggests that susceptibility to *primary* severe RSV bronchiolitis and atopy share a genetic contribution at the *IL13–IL4* locus.

Acknowledgements: The authors thank Elham Sadighi Akha and Gaia Luoni for their help with genotyping.

Funding: This work was funded by a Wellcome Trust Clinical Research Training Fellowship (JTF) and the MRC(UK) (DK).

Competing interests: None.

Ethics approval: Ethics approval for all sample collection was obtained from the Oxford Regional ethics committee and the Multi Region ethics committee.

REFERENCES

1. **Shay DK,** Holman RC, Roosevelt GE, *et al*. Bronchiolitis-associated mortality and estimates of respiratory syncytial virus-associated deaths among US children, 1979–1997. *J Infect Dis* 2001;**183**:16–22.
2. **Legg JP,** Hussain IR, Warner JA, *et al*. Type 1 and type 2 cytokine imbalance in acute respiratory syncytial virus bronchiolitis. *Am J Respir Crit Care Med* 2003;**168**:633–9.
3. **Roman M,** Calhoun WJ, Hinton KL, *et al*. Respiratory syncytial virus infection in infants is associated with predominant Th-2-like response. *Am J Respir Crit Care Med* 1997;**156**:190–5.
4. **Bendelja K,** Gagro A, Bace A, *et al*. Predominant type-2 response in infants with respiratory syncytial virus (RSV) infection demonstrated by cytokine flow cytometry. *Clin Exp Immunol* 2000;**121**:332–8.
5. **Thomsen SF,** Stensballe LG, Skytthe A, *et al*. Increased concordance of severe respiratory syncytial virus infection in identical twins. *Pediatrics* 2008;**121**:493–6.
6. **Palmer LJ,** Cardon LR. Shaking the tree: mapping complex disease genes with linkage disequilibrium. *Lancet* 2005;**366**:1223–34.
7. **Culley FJ,** Pollott J, Openshaw PJ. Age at first viral infection determines the pattern of T cell-mediated disease during reinfection in adulthood. *J Exp Med* 2002;**196**:1381–6.
8. **Graham BS.** Immunological determinants of disease caused by respiratory syncytial virus. *Trends Microbiol* 1996;**4**:290–3.
9. **Choi EH,** Lee HJ, Yoo T, *et al*. A common haplotype of interleukin-4 gene IL4 is associated with severe respiratory syncytial virus disease in Korean children. *J Infect Dis* 2002;**186**:1207–11.
10. **Hoebee B,** Rietveld E, Bont L, *et al*. Association of severe respiratory syncytial virus bronchiolitis with interleukin-4 and interleukin-4 receptor alpha polymorphisms. *J Infect Dis* 2003;**187**:2–11.
11. **Puthothu B,** Krueger M, Forster J, *et al*. Association between severe respiratory syncytial virus infection and IL13/IL4 haplotypes. *J Infect Dis* 2006;**193**:438–41.
12. **Forton JT,** Udalova IA, Campino S, *et al*. Localization of a long-range cis-regulatory element of IL13 by allelic transcript ratio mapping. *Genome Res* 2007;**17**:82–7.
13. **Hull J,** Thomson A, Kwiatkowski D. Association of respiratory syncytial virus bronchiolitis with the interleukin 8 gene region in UK families. *Thorax* 2000;**55**:1023–7.
14. **Riva A,** Kohane IS. A web-based tool to retrieve human genome polymorphisms from public databases. *Proc AMIA Symp* 2001:558–62.
15. **Frazer KA,** Pachter L, Poliakov A, *et al*. VISTA: computational tools for comparative genomics. *Nucleic Acids Res* 2004;**32**(web server issue):W273–9.
16. **Cartharius K,** Frech K, Grote K, *et al*. MatInspector and beyond: promoter analysis based on transcription factor binding sites. *Bioinformatics* 2005;**21**:2933–42.
17. **Cardon LR,** Palmer LJ. Population stratification and spurious allelic association. *Lancet* 2003;**361**:598–604.
18. **Anon**. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* 2007;**447**:661–78.
19. **Stephens M,** Smith NJ, Donnelly P. A new statistical method for haplotype reconstruction from population data. *Am J Hum Genet* 2001;**68**:978–89.
20. **Abecasis GR,** Cookson WO, Cardon LR. Pedigree tests of transmission disequilibrium. *Eur J Hum Genet* 2000;**8**:545–51.
21. **Centre for Genomics and Global Health**. www.gmap.net.
22. **Forton J,** Kwiatkowski D, Rockett K, *et al*. Accuracy of haplotype reconstruction from haplotype-tagging single-nucleotide polymorphisms. *Am J Hum Genet* 2005;**76**:438–48.
23. **Luoni G,** Forton J, Jallow M, *et al*. Population-specific patterns of linkage disequilibrium in the human 5q31 region. *Genes Immun* 2005;**6**:723–7.
24. **Spielman RS,** McGinnis RE, Ewens WJ. Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). *Am J Hum Genet* 1993;**52**:506–16.
25. **Nyholt DR.** A simple correction for multiple testing for single-nucleotide polymorphisms in linkage disequilibrium with each other. *Am J Hum Genet* 2004;**74**:765–9.
26. **Li J,** Ji L. Adjusting multiple testing in multilocus analyses using the eigenvalues of a correlation matrix. *Heredity* 2005;**95**:221–7.
27. **Rosenwasser LJ.** Genetics of atopy and asthma: promoter-based candidate gene studies for IL-4. *Int Arch Allergy Immunol* 1997;**113**:61–4.
28. **van der Pouw Kraan TC,** van Veen A, Boeije LC, *et al*. An IL-13 promoter polymorphism associated with increased risk of allergic asthma. *Genes Immun* 1999;**1**:61–5.

Thorax: first published as 10.1136/thx.2008.102111 on 8 January 2009. Downloaded from http://thorax.bmj.com/ on April 27, 2024 by guest. Protected by copyright.

29. **Zhu S,** Chan-Yeung M, Becker AB, *et al*. Polymorphisms of the IL-4, TNF-alpha, and Fcε RIβ genes and the risk of allergic disorders in at-risk infants. *Am J Respir Crit Care Med* 2000;**161**:1655–9.
30. **Rowe J,** Macaubas C, Monger T, *et al*. Heterogeneity in diphtheria-tetanus-acellular pertussis vaccine-specific cellular immunity during infancy: relationship to variations in the kinetics of postnatal maturation of systemic th1 function. *J Infect Dis* 2001;**184**:80–8.
31. **Rowe J,** Macaubas C, Monger TM, *et al*. Antigen-specific responses to diphtheria-tetanus-acellular pertussis vaccine in human infants are initially Th2 polarized. *Infect Immun* 2000;**68**:3873–7.
32. **Webster RB,** Rodriguez Y, Klimecki WT, *et al*. The human IL-13 locus in neonatal CD4+T cells is refractory to the acquisition of a repressive chromatin architecture. *J Biol Chem* 2007;**282**:700–9.
33. **Sigurs N.** A cohort of children hospitalised with acute RSV bronchiolitis: impact on later respiratory disease. *Paediatr Respir Rev* 2002;**3**:177–83.
34. **Stein RT,** Sherrill D, Morgan WJ, *et al*. Respiratory syncytial virus in early life and risk of wheeze and allergy by age 13 years. *Lancet* 1999;**354**:541–5.
35. **Forton JT,** Kwiatkowski D. Searching for the regulators of gene expression. *Bioessays* 2006;**28**:1–5.

Lung alert

Gene expression: the key to treatment decisions in early-stage adenocarcinoma?

Recent evidence shows that adjuvant chemotherapy significantly improves the survival of patients with early-stage lung cancers. There are, however, groups of patients who do not fit neatly into prognostic categories according to stage, and tumour biology is likely to play an important role. This retrospective blinded validation study aimed to identify predictors of survival (classifiers) using gene expression profiling.

Data were collected about 442 patients with stage I, II or III lung adenocarcinoma, including gene expression data, clinical variables and outcome data (survival). To allow comparison between data sets from separate laboratories, standardised protocols were used for all aspects of data generation and collection. The data were split into two groups according to the laboratory which performed the initial analysis. One data set was used for ''training'' of classifiers and the other to validate these approaches. Each investigating team developed classifiers to predict the outcome for each patient, aiming to discover if these could be predicted based on gene expression alone or when combined with simple clinical information (sex, stage and age). Of the numerous classifiers generated, one showed the most consistent performance over the range of hypotheses. This method involved 100 gene clusters and ridge regression analysis to predict outcome. Prediction was poorer for stage I disease and better when clinical data were included in the analysis for all stages.

The study was not adequately powered to compare the classifiers with high significance, and used survival as the outcome measure rather than recurrence which may reflect tumour biology more accurately. However, the study has proved the principle of cooperation between institutions and, using standardised protocols, has generated a large data set which can be further analysed for better classifiers.

► Shedden K, Taylor JM, Enkemann SA, Tsao MS, *et al*. Director's Challenge Consortium for the Molecular Classification of Lung Adenocarcinoma. Gene expression-based survival prediction in lung adenocarcinoma: a multi-site, blinded validation study. *Nat Med* 2008;**14**:822–7.

T Hillman

Correspondence to: Dr T Hillman, ST3 Respiratory Medicine, Homerton University Hospital, London, UK; tobyh@doctors.org.uk

Thorax: first published as 10.1136/thx.2008.102111 on 8 January 2009. Downloaded from http://thorax.bmj.com/ on April 27, 2024 by guest. Protected by copyright.