SUPPLEMENTARY INFORMATION

Association between *ORMDL3*, *IL1RL1* and a deletion on chromosome 17q21 with asthma risk in Australia

Manuel A. R. Ferreira,[a] Allan F. McRae,[a] Sarah E. Medland,[a] Dale R. Nyholt,[a] Scott D. Gordon,[a] Margaret J. Wright,[a] Anjali K. Henders,[a] Pamela A. Madden,[b] Peter M. Visscher,[a] Naomi R. Wray,[a] Andrew C. Heath,[b] Grant W. Montgomery,[a] David L. Duffy,[a] and Nicholas G. Martin[a]

[a]Queensland Institute of Medical Research, Brisbane, Australia

[b]Washington University School of Medicine, St Louis, USA

Address for correspondence:

Manuel A R Ferreira

Queensland Institute of Medical Research,

P.O. Royal Brisbane Hospital, Brisbane 4029, Australia

E-mail: manuel.ferreira@qimr.edu.au

Phone: +61 7 3845 3552   Fax: +61 7 3362 0101

SUPPLEMENTARY FIGURE LEGENDS

Supplementary Figure 1. Population substructure analyses. (A) Results from the multidimensional scaling (MDS) analysis of identity-by-state (IBS) distances calculated between all pairs of individuals from the cohort analysed in this study (ASTHMA) and the eleven HapMap III populations. **(B)** Same analysis as in (A), but with the cluster of European descent magnified. **(C)** Plot restricted to the samples analysed in this study, with cases represented in red and controls in green. **(D)** Quantile-Quantile plot from the whole-genome SNP association analysis. λ: genomic inflation factor.

Supplementary Figure 2. Genome-wide association results for life-time asthma status. SNPs with a P-value $< 1\text{x}10^{-5}$ are highlighted in green.

Supplementary Figure 3. Visual inspection of logRRatio patterns for the 1q21 region identified in the whole-genome CNV analysis. Local logRRatio patterns were inspected for the 10 individuals identified by QuantiSNPv1.1 as carrying the 1q21 deletion. The software called a shorter deletion (~100-150 kb) in seven individuals (5 cases and 2 controls) and a larger overlapping deletion (~500-800 kb) in three asthma cases. The latter extend over a complex region of the genome with a number of gaps in the human assembly, for which there are no available SNP or CNV probes. Although five large deletions in this region have also been reported in the Database of Genomic Variants (http://projects.tcag.ca/variation/), the logRR patterns for the three larger deletions in our study are ambiguous and the CNV calls would need to be validated by molecular biology experiments. Patterns for the smaller seven deletions are also ambiguous for some individuals (eg. 8602902, 8628502 and 8898601) and so this region may also represent a technical artefact and was not considered further in our analysis.

Supplementary Figure 4. Visual inspection of logRRatio patterns for the 17q21 region identified in the whole-genome CNV analysis. Local logRRatio patterns were inspected for the 38 individuals identified by QuantiSNPv1.1 as carrying the 17q21 deletion. As with the 1q21 region, the software called a shorter deletion (~100-150 kb) in seven individuals (all cases) and a larger overlapping deletion (~300-400 kb) in 31 individuals (13 cases and 18 controls). However, for this region, logRRatio patterns are indeed consistent with the presence of a deletion in this region for all 38 individuals. It is nonetheless unclear whether the 38 segments identified represent one larger deletion or two smaller deletions separated by a short diploid segment, as data from a few individuals (eg. 8604701, 8665002) seem to suggest. A number of 100-200 kb deletions have been reported across this region in the Database of Genomic Variants (http://projects.tcag.ca/variation/), consistent with the presence of two distinct smaller CNVs in our data.

Supplementary Figure 5. Visual inspection of logRRatio patterns for the six 17q21 deletions identified in the replication panel of 604 individuals. Local logRRatio patterns were inspected for the 6 individuals identified by QuantiSNPv1.1 as carrying the 17q21 deletion in the replication panel. For these, the deletion spanned ~350 kb.

SUPPLEMENTARY TABLES

Supplementary Table 1. Characteristics of the study participants included in the whole-genome analysis.

QIMR study[1] [references]	**Cases**		**Controls**		**Assessment of lifetime asthma status**		
	N (female)	Age (range)	N (female)	Age (range)	Questionnaire item	Cases	Controls
Canberra [1]	18 (67%)	27 (19-47)	104 (69%)	26 (18-57)	How often have you had asthma?	"Only as a child", "Quite often"	"Never"
Alcohol [2]	356 (69%)	31 (18-78)	963 (66%)	29 (16-66)	How often have you had asthma?	"Sometimes", "Often"	"Never"
Alcohol Relatives [3]	117 (51%)	40 (19-67)	39 (59%)	38 (19-66)			
Aged [4]	35 (89%)	60 (50-80)	245 (83%)	61 (50-92)	Have you ever had asthma?	"Now", "Past", "Now and Past"	"No"
Asthma [5]	187 (67%)	38 (22-70)	-	-	Have you ever had asthma?	"Yes", "Yes, told to me by a doctor"	-
Adolescents [6]	273 (45%)	14 (12-23)	495 (55%)	14 (11-25)	Has a doctor ever diagnosed you as suffering from asthma?	"Yes"	"No"
TOTAL	986 (61%)	29 (12-80)	1846 (65%)	29 (11-92)	-	-	-

[1] Each original study ascertained multiple relatives per family. For this analysis, we selected one genotyped individual per family; in families having both affected and unaffected individuals, we selected the case only. A brief description of each study is provided below and details are available from the respective cited references. Some individuals provided asthma information in multiple studies. These were considered cases if asthma was reported at any time point, and controls if asthma was never reported at all time points.

Canberra study: Between 1980 and 1982, a questionnaire was mailed to all 5,967 pairs of twins registered with the Australian National Health and Medical Research Council Twin Registry who were older than 18 yr of age. These pairs were volunteers who had been recruited through schools, community groups and by media advertising throughout Australia. Of this group, 3,808 twin pairs returned completed questionnaires, which included items on age, sex, zygosity and a number of health symptoms, including asthma.

Alcohol study: The questionnaire used in the Canberra study did not include any questions about alcohol-related problems. Thus, this study was designed as a follow-up of the Canberra study, to explore the changes in drinking practices and history of alcohol problems in the original Canberra cohort, which was largely a random sample of the population. This follow-up questionnaire was mailed in 1988-1989 to all 3,808 complete twin pairs that participated in the 1980-1982 survey. A comparable asthma symptoms question was included in this new questionnaire.

Alcohol relatives study: Participants of the Alcohol study were asked to provide names of relatives who were prepared to provide information on their drinking practices. A slightly modified version of the questionnaire used in the Alcohol study was then mailed out to these relatives in 1989-1991, with 11,400 individuals returning a completed questonnaire, which included the same asthma item listed in the original Alcohol study.

Aged study: A study designed to cover a wide range of health issues affecting older people was undertaken as a multi-wave mailout between 1993 and 1995. A total of 2,281 twin pairs over 50 and listed in the Australian Twin Registry were invited to participate by completion of a 16-page questionnaire that included a range of health measures, namely asthma.

Asthma study: This study recruited 802 families registered on the Australian Twin Registry with at least on twin who had reported ever wheezing in previous studies conducted at QIMR (such as the Canberra and Alcohol studies) and by collaborators elsewhere in Australia. Participants completed a questionnaire that was designed to validate the diagnosis of asthma and to obtain data on respiratory symptoms, environmental exposures and family history of asthma. In addition, participants underwent clinical testing, including lung function ans skin prick tests.

Adolescents study: Twins were recruited in the context of an ongoing study of melanoma risk factors including benign melanocytic naevi , sun exposure time and pigmentation related variables. Twins were enlisted by contacting the principals of primary schools (first 7 years of education) in the greater Brisbane area, media appeals and by word of mouth. It is estimated that approximately 50 percent of the eligible birth cohort were recruited into the study. Twins were examined at age 12 years, and siblings at the same occasion if under 20 years of age. At the same time, twins and their parents completed questionnaires measuring melanoma risk factors, but that also included general health questions, such as asthma. Twins were again examined at age 14 and 16.

Supplementary Table 2. Quality control steps for the 7 waves of genotyping conducted at QIMR.

	Wave 1: ALCO CIDR	Wave 2: ALCO deCODE	Wave 3: MIG deCODE	Wave 4: EUTWIN	Wave 5: ADOL deCODE	Wave 6: GL CIDR	Wave 7: WH deCODE
Illumina Array	370K	370K	610K	317K	610K	610K	670K
No. genotype samples	4241	2611	999	462	4391	657	2360
No. samples used this study	633	396	541	417	758	20	5
No. genotyped SNPs	343955	344962	592385	318210	592392	589296	562193
SNP GenCall score <0.7	24494	27459	46931	NA	47418	36877	57589
SNP call rate <0.95	11584	7537	8038	5021	8447	12455	33459
SNP HWE P $<10^{-6}$	4318	1194	1221	67	2841	15474	1763
SNP MAF <0.01	7874	8976	33347	264	33347	28607	24509
# SNPs following QC	323093	321267	530922	312937	529379	531042	518948
% genotyped SNPs	93.93%	93.13%	89.62%	98.34%	89.36%	90.11%	92.31%

Following the QC of the individual projects the data from the seven waves of genotyping were integrated. After integration of the data sets the data were screened for missingness within individuals (>5%), pedigree and sex errors, and Mendelian errors (genotypes for all family members for a given SNP were removed on detection of errors), resulting in a sample of 16,507 individuals. After screening for non-European ancestry, the final sample consisted of 16,140 individuals.

Supplementary Table 3. Characteristics of the study participants included in the replication panel.

QIMR study[1] [references]	**Cases**		**Controls**	
	N (female)	Age (range)	N (female)	Age (range)
Canberra [E1]	2 (50%)	24 (19-28)	5 (60%)	26 (20-34)
Alcohol [E2]	33 (55%)	36 (19-71)	97 (47%)	30 (18-64)
Alcohol Relatives [E3]	28 (54%)	36 (21-66)	65 (55%)	37 (19-62)
Aged [E4]	3 (100%)	60 (53-70)	46 (61%)	59 (36-81)
Asthma [E5]	325 (51%)	25 (7-73)	-	-
TOTAL	391 (52%)	27 (7-73)	213 (53%)	38 (18-81)

[1] These studies, including the criteria for assessment of lifetime asthma status, are described **Supplementary Table** 1 above.

Supplementary Tables 4 and **5 are supplied in a separate file due to their large size.**

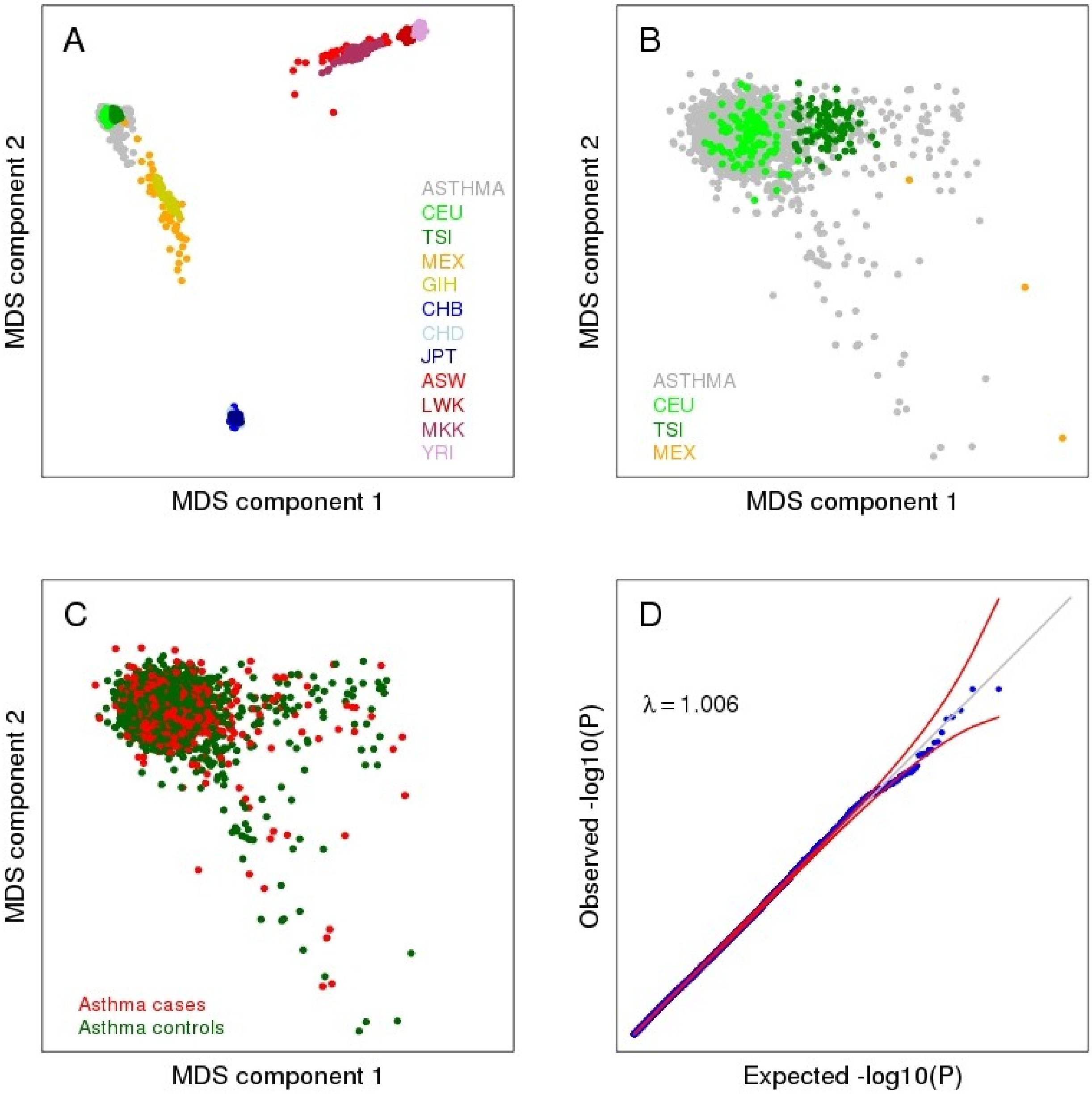
Supplementary Table 6. Follow-up of the top loci identified in the genome-wide SNP and CNV analyses.

Region	**Marker[1], reference allele**	**GWAS OR**	**Power**	**Allele Frequency**		**OR**	***P*-value**
				Cases	Controls		
Top hits from genome-wide SNP analysis							
6q16	rs12206223, C	1.54	55%	0.13	0.16	0.79	0.1505
7p14	rs17700752, G	0.71	60%	0.19	0.16	1.20	0.2584
22q12	rs4622835, C	1.29	55%	0.49	0.49	1.00	0.9828
10q26	rs7893667, A	1.31	60%	0.28	0.32	0.85	0.2254
5q31	rs31263, C	1.31	60%	0.28	0.30	0.90	0.4311
ILRL1 novel variant							
2q12	rs10197862, G	0.75	50%	0.14	0.22	0.57	2.4×10^{-4}
Top hit from genome-wide CNV analysis							
17q21	deletion	2.1	40%	5/391	1/213	2.7	0.3135

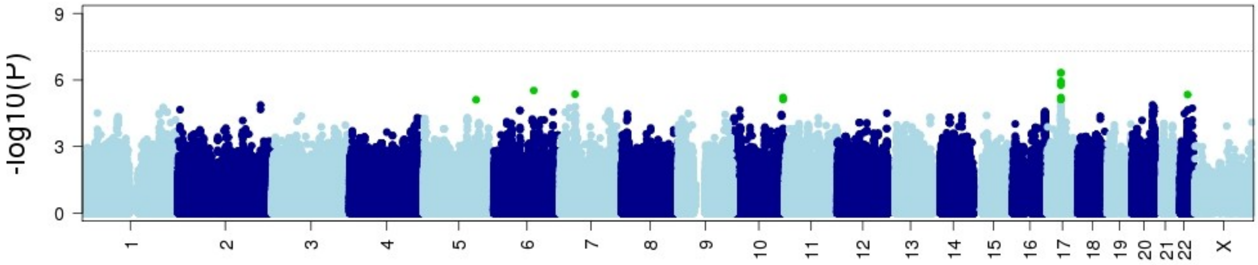
[1] All SNPs were imputed in the replication panel with high confidence (r^2>0.6) using MACH. The 17q21 deletion was identified with high confidence (Log Bayes Factor >10) using QuantiSNPv1.1 in 6 out of 604 samples and subsequently confirmed visually (cf. Figure E5).

SUPPLEMENTARY REFERENCES

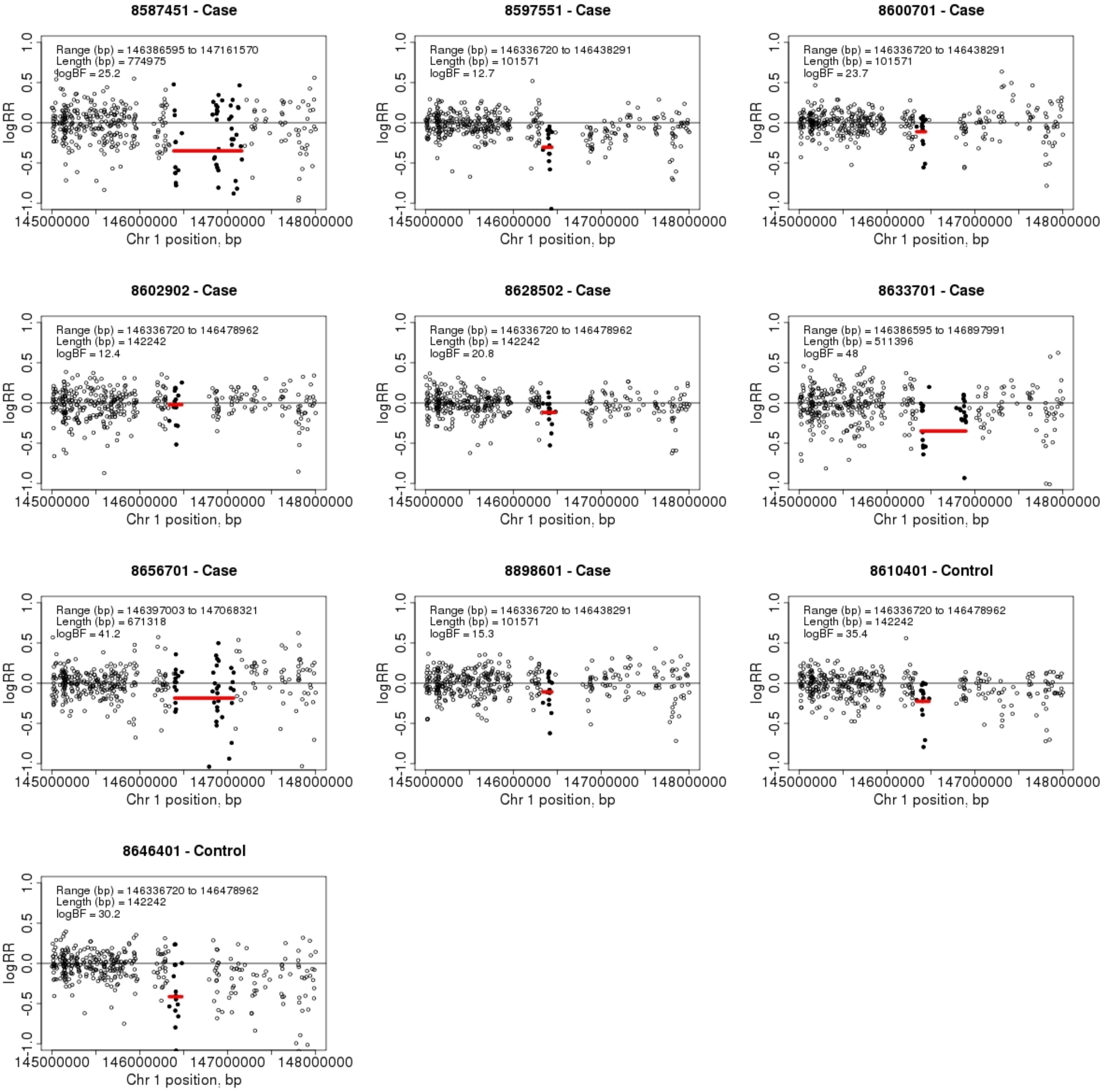
1. Duffy DL, Martin NG, Battistutta D, Hopper JL, Mathews JD. Genetics of asthma and hay fever in Australian twins. *Am Rev Respir Dis* 1990; **142**:1351-8.
2. Heath AC, Martin NG. Genetic influences on alcohol consumption patterns and problem drinking: results from the Australian NH&MRC twin panel follow-up survey. *Ann N Y Acad Sci* 1994; **708**: 72-85.
3. Lake RE, Eaves LJ, Maes HM, Heath AC, Martin NG. Further evidence against the environmental transmission of individual differences in Neuroticism from a collaborative study of 45,850 twins and relatives on two continents. *Behavior Genetics* 2000; **30**: 223-233.
4. Kirk KM, Martin NG. The short interpersonal reactions inventory, self-regulation and differentiation scales in an older Australian twin sample. *Pers Individ Dif* 1998; **25**: 591-604.
5. Ferreira MA, O'Gorman L, Le Souef P *et al:* Variance components analyses of multiple asthma traits in a large sample of Australian families ascertained through a twin proband. *Allergy* 2006; **61**: 245-53.
6. Zhu G, Montgomery GW, James MR *et al:* A genome-wide scan for naevus count: linkage to CDKN2A and to other chromosome regions. *Eur J Hum Genet* 2007; **15**: 94-102.
7. Dubois PC, Trynka G, Franke L et al: Multiple common variants for celiac disease influencing immune gene expression. Nat Genet 2010; 42: 295-302.
8. Barrett JC, Hansoul S, Nicolae DL et al: Genome-wide association defines more than 30 distinct susceptibility loci for Crohn's disease. Nat Genet 2008; 40: 955-62.



Supplementary Figure 1. Population substructure analyses. (A) Results from the multidimensional scaling (MDS) analysis of identity-by-state (IBS) distances calculated between all pairs of individuals from the cohort analysed in this study (ASTHMA) and the eleven HapMap III populations. **(B)** Same analysis as in (A), but with the cluster of European descent magnified. **(C)** Plot restricted to the samples analysed in this study, with cases represented in red and controls in green. (D) Quantile-Quantile plot from the whole-genome SNP association analysis. λ: genomic inflation factor.

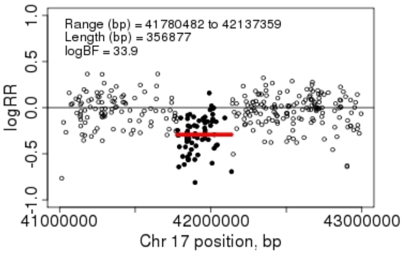


Supplementary Figure 2. Genome-wide association results for life-time asthma status. SNPs with a *P*-value < $1\text{x}10^{-5}$ are highlighted in green.

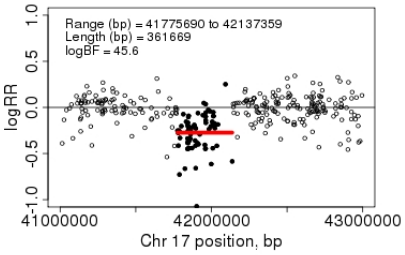


Supplementary Figure 3. Visual inspection of logRRatio patterns for the 1q21 region identified in the whole-genome CNV analysis. Local logRRatio patterns were inspected for the 10 individuals identified by QuantiSNPv1.1 as carrying the 1q21 deletion. The software called a shorter deletion (~100-150 kb) in seven individuals (5 cases and 2 controls) and a larger overlapping deletion (~500-800 kb) in three asthma cases. The latter extend over a complex region of the genome with a number of gaps in the human assembly, for which there are no available SNP or CNV probes. Although five large deletions in this region have also been reported in the Database of Genomic Variants (http://projects.tcag.ca/variation/), the logRR patterns for the three larger deletions in our study are ambiguous and the CNV calls would need to be validated by molecular biology experiments. Patterns for the smaller seven deletions are also ambiguous for some individuals (eg. 8602902, 8628502 and 8898601) and so this region may also represent a technical artefact and was not considered further in our analysis.

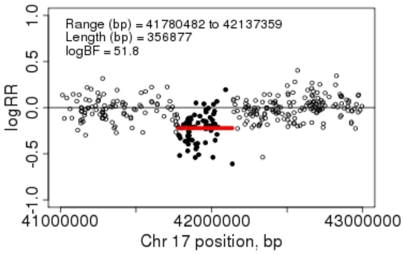
8132001 - Case



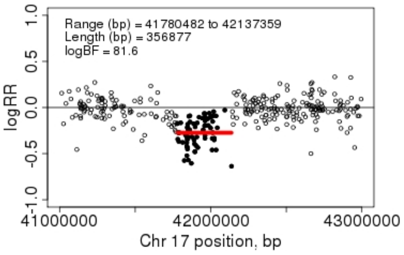
8586301 - Case



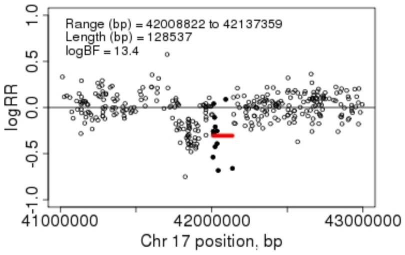
8588901 - Case



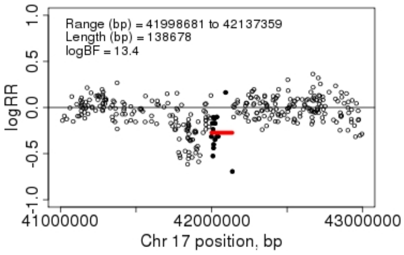
8598402 - Case



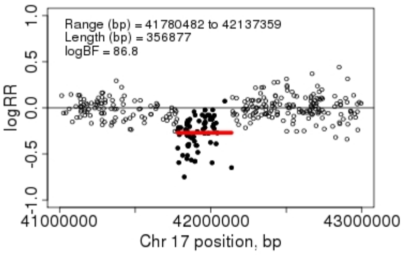
8604701 - Case



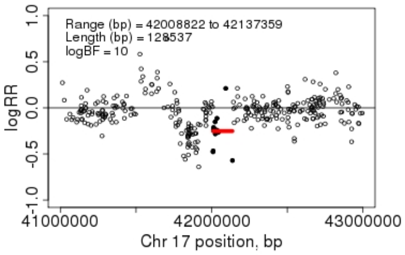
8613402 - Case



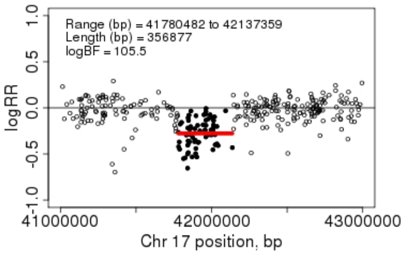
8614150 - Case



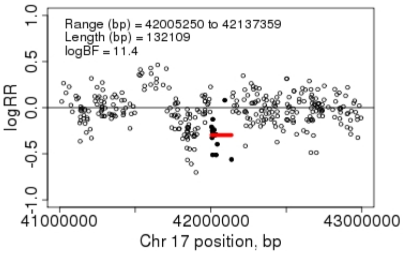
8618751 - Case



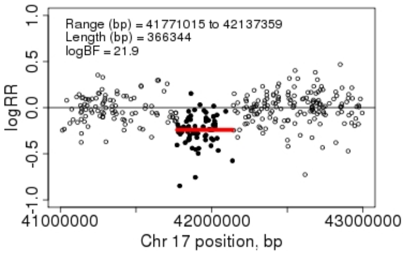
8619501 - Case



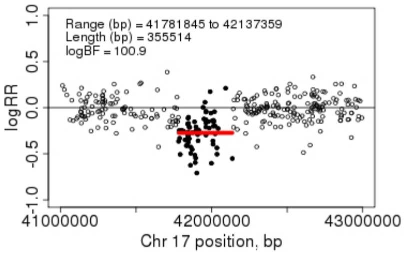
8626650 - Case



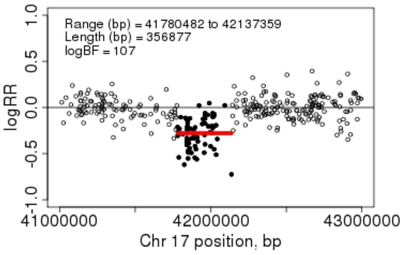
8633201 - Case



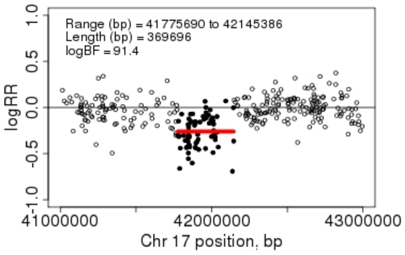
8644601 - Case



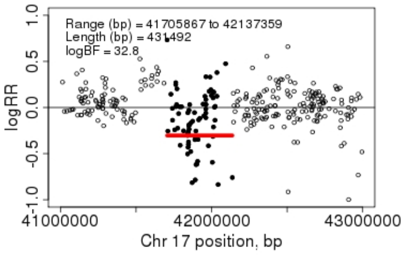
8652102 - Case



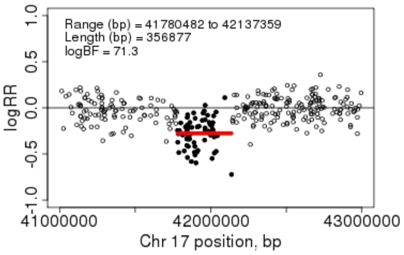
8656301 - Case



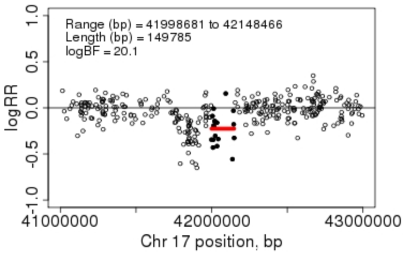
8656701 - Case



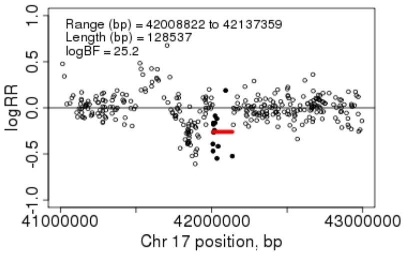
8656802 - Case



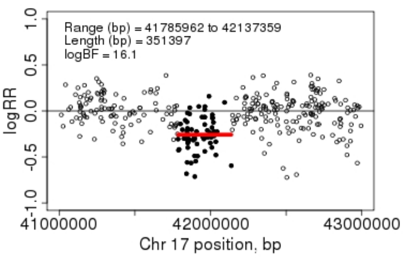
8665002 - Case



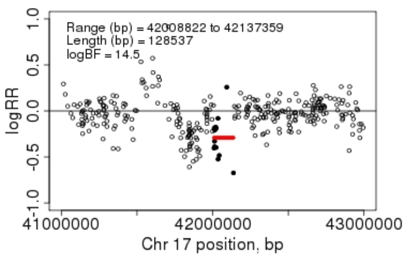
8681501 - Case



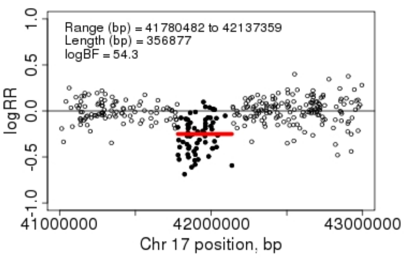
8896301 - Case



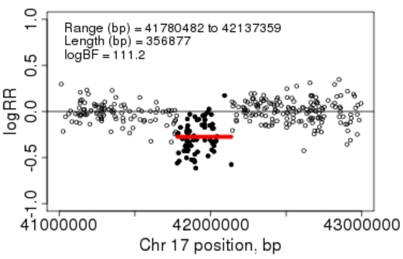
8915002 - Case



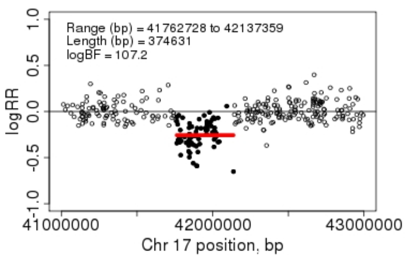
8102001 - Control



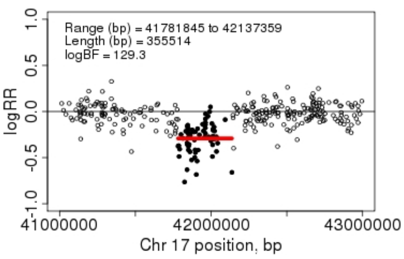
8160001 - Control



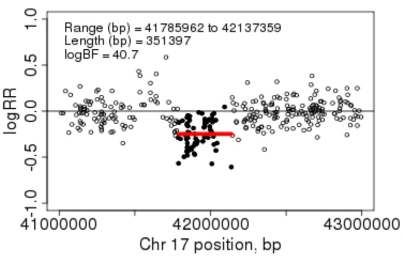
8233001 - Control



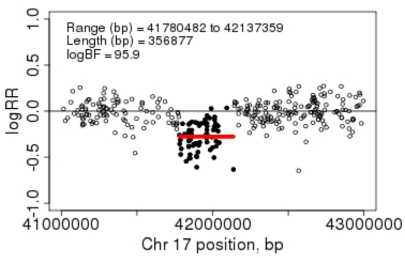
8263001 - Control



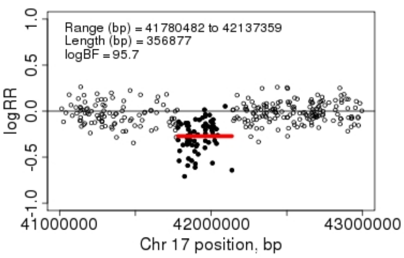
8343101 - Control



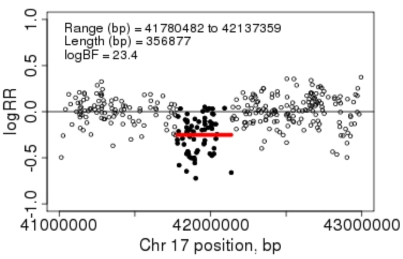
8447001 - Control



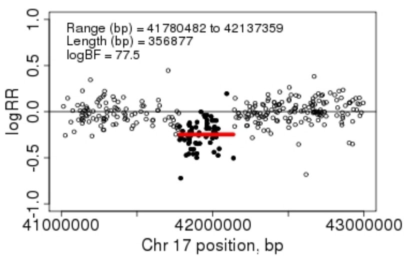
8469101 - Control



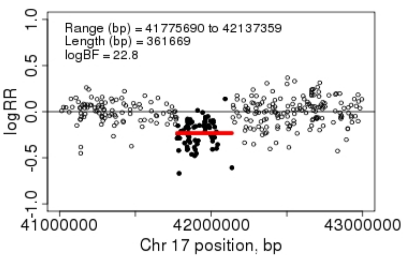
8474101 - Control



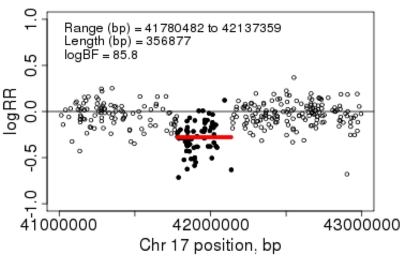
8586001 - Control



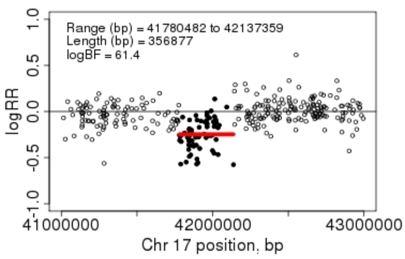
8586601 - Control



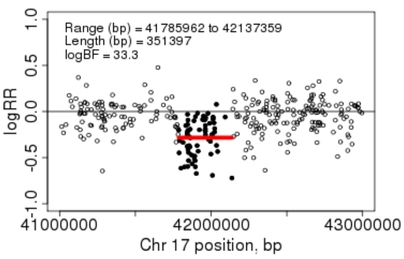
8587701 - Control



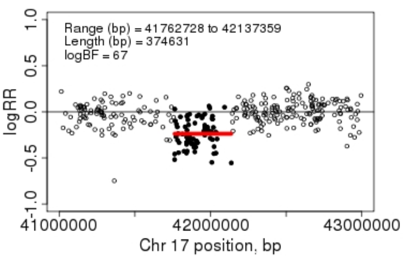
8591351 - Control



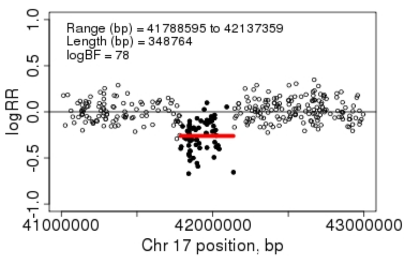
8601601 - Control



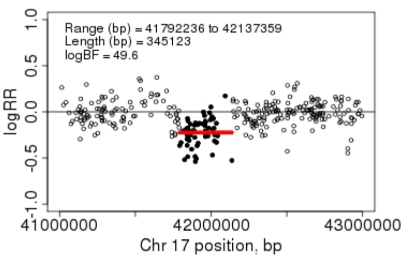
8608301 - Control

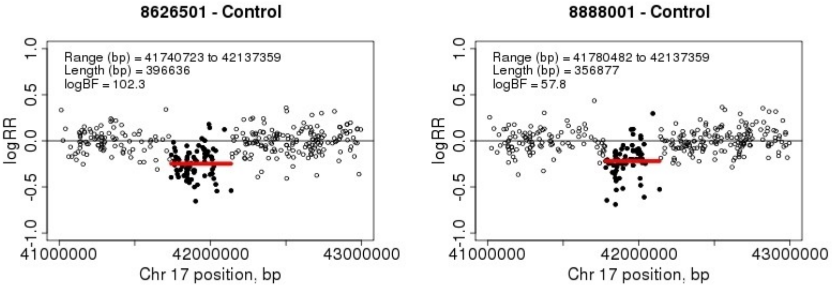


8616701 - Control

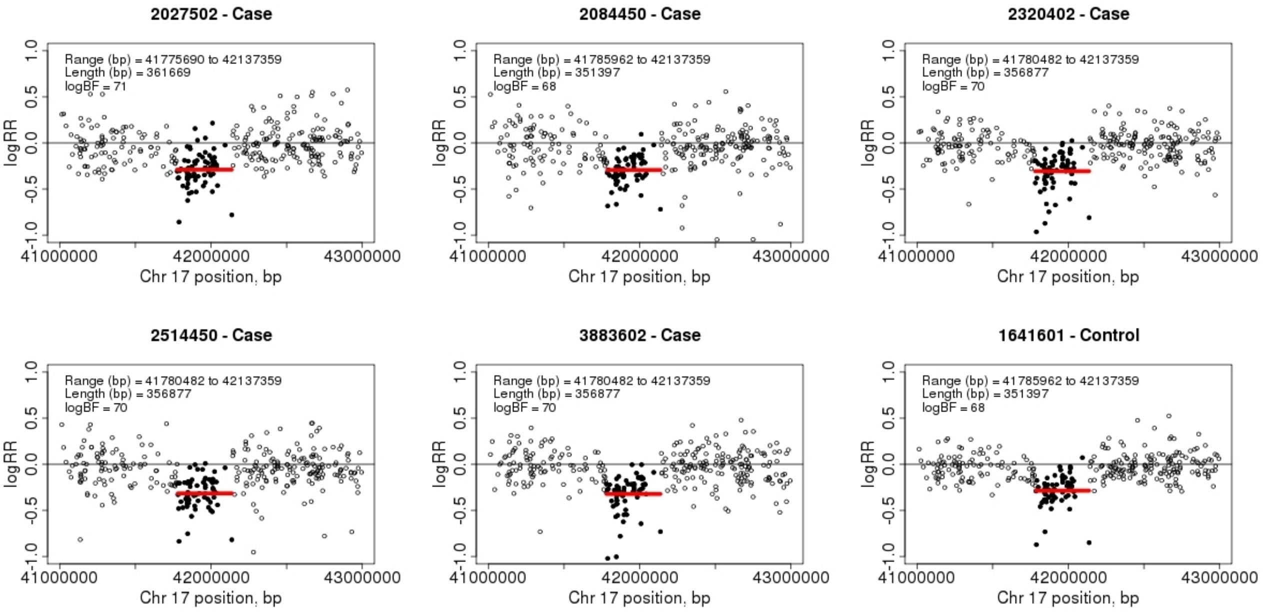


8617601 - Control





Supplementary Figure 4. Visual inspection of logRRatio patterns for the 17q21 region identified in the whole-genome CNV analysis. Local logRRatio patterns were inspected for the 38 individuals identified by QuantiSNPv1.1 as carrying the 17q21 deletion. As with the 1q21 region, the software called a shorter deletion (~100-150 kb) in seven individuals (all cases) and a larger overlapping deletion (~300-400 kb) in 31 individuals (13 cases and 18 controls). However, for this region, logRRatio patterns are indeed consistent with the presence of a deletion in this region for all 38 individuals. It is nonetheless unclear whether the 38 segments identified represent one larger deletion or two smaller deletions separated by a short diploid segment, as data from a few individuals (eg. 8604701, 8665002) seem to suggest. A number of 100-200 kb deletions have been reported across this region in the Database of Genomic Variants (http://projects.tcag.ca/variation/), consistent with the presence of two distinct smaller CNVs in our data.



Supplementary Figure 5. Visual inspection of logRRatio patterns for the six 17q21 deletions identified in the replication panel of 604 individuals. Local logRRatio patterns were inspected for the 6 individuals identified by QuantiSNPv1.1 as carrying the 17q21 deletion in the replication panel. For these, the deletion spanned ~350 kb.

Locus	**Genes**	**SNP**	**Previous reported association with Celiac disease** [7]			**Association with asthma in this study**					
			Reference Allele	OR	P-value	Minor Allele	Major Allele	MAF Cases	MAF Controls	OR	P-value
2p14	*PLEK*	rs17035378	?	1.14	8.0E-09	C	T	0.28	0.26	1.10	0.1362
3p22.3	*CCR4*	rs13314993	C	1.13	3.0E-09	G	T	0.44	0.44	1.01	0.8667
3q13.33	*CD80,KTELC1*	rs11712165	C	1.13	8.0E-09	G	T	0.38	0.39	0.97	0.5806
6q15	*BACH2,MAP3K7*	rs10806425	A	1.13	4.0E-10	A	C	0.42	0.42	1.02	0.7204
6q22.33	*PTPRK,THEMIS*	rs802734	G	1.17	3.0E-14	G	A	0.32	0.31	1.02	0.7495
10q22.3	*ZMIZ1*	rs1250552	?	1.12	9.0E-10	G	A	0.48	0.48	0.99	0.8918
11q24.3	*ETS1*	rs11221332	A	1.21	5.0E-16	T	C	0.2	0.21	0.96	0.5267
21q22.3	*ICOSLG*	rs4819388	?	1.14	2.0E-09	T	C	0.27	0.29	0.94	0.2808
2q12.1	*IL18RAP,IL18R1,IL1RL1,IL1RL2*	rs917997	A	1.19	1.0E-15	T	C	0.22	0.22	0.99	0.9145
2q33.2	*CTLA4,ICOS,CD28*	rs4675374	A	1.14	6.0E-09	T	C	0.19	0.2	0.95	0.4421
3p21.31	*CCR1,CCR2,CCRL2,CCR3,CCR5,CCR9*	rs13098911	A	1.30	3.0E-17	T	C	0.07	0.08	0.92	0.4132
3q25.33	*IL12A*	rs17810546	G	1.36	4.0E-28	G	A	0.12	0.12	0.97	0.7031
4q27	*IL2,IL21*	rs13151961	?	1.35	2.0E-27	G	A	0.17	0.17	1.01	0.8998
6p21.32	*HLA-DQA1,HLA-DQB1*	rs2187668	A	6.23	1.0E-50	T	C	0.15	0.14	1.05	0.5807
12q24.12	*SH2B3*	rs653178	G	1.20	7.0E-21	C	T	0.48	0.49	0.94	0.2734
1p36.32	*TNFRSF14,MMEL1*	rs3748816	?	1.12	3.0E-09	G	A	0.36	0.36	1.01	0.8428
1p36.11	*RUNX3*	rs10903122	?	1.12	2.0E-10	G	A	0.52	0.48	1.15	0.0142
1q32.1	Intergenic	rs296547	?	1.12	4.0E-09	T	C	0.36	0.36	0.98	0.7911
2p16.1	*REL,AHSA2*	rs13003464	G	1.15	4.0E-13	G	A	0.37	0.37	0.97	0.5446
6q23.3	*TNFAIP3*	rs2327832	G	1.23	4.0E-19	G	A	0.22	0.22	0.98	0.7377
18p11.21	*PTPN2*	rs1893217	G	1.17	3.0E-10	G	A	0.18	0.19	0.94	0.3937
8q24.21	Intergenic	rs9792269	?	1.14	3.0E-09	G	A	0.25	0.24	1.02	0.7315
16p13.13	*CIITA,SOCS1,CLEC16A*	rs12928822	?	1.16	3.0E-08	T	C	0.19	0.18	1.05	0.5273
1q31.2	*RGS1*	rs2816316	?	1.25	2.0E-17	C	A	0.18	0.18	0.98	0.7476
2q31.3	*ITGA4,UBE2E3*	rs13010713	G	1.13	5.0E-11	G	A	0.43	0.42	1.05	0.3836
3q28	*LPP*	rs1464510	A	1.29	3.0E-40	A	C	0.43	0.46	0.90	0.0609
6q25.3	*TAGAP*	rs1738074	A	1.16	3.0E-15	T	C	0.45	0.45	0.98	0.7516

Locus	**Genes**	**SNP**	**Previous reported association with Crohn's disease** [8]			**Association with asthma in this study**					
			Reference Allele	OR	P-value	Minor Allele	Major Allele	MAF Cases	MAF Controls	OR	P-value
9q32	*TNFSF15*	rs4263839	G	1.22	3.0E-10	A	G	0.32	0.32	1.03	0.6117
5p13.1	*PTGER4*	rs4613763	C	1.32	7.0E-27	C	T	0.12	0.13	0.91	0.2781
2q37.1	*ATG16L1*	rs3828309	G	1.25	2.0E-32	A	G	0.49	0.49	1.02	0.6918
1p31.3	*IL23R*	rs11465804	T	2.50	7.0E-63	G	T	0.07	0.07	1.01	0.9525
21q22.3	*ICOSLG*	rs762421	G	1.13	1.0E-09	G	A	0.38	0.39	0.99	0.8044
17q12	*ORMDL3*	rs2872507	A	1.12	5.0E-09	A	G	0.44	0.49	0.81	0.0002
11q13.5	*C11orf30*	rs7927894	T	1.16	1.0E-09	T	C	0.39	0.37	1.06	0.2805
21q21.1	Intergenic	rs1736135	T	1.18	7.0E-09	C	T	0.41	0.44	0.9	0.0512
12q12	*LRRK2,MUC19*	rs11175593	T	1.54	3.0E-10	T	C	0.01	0.02	0.92	0.7146
9p24.1	*JAK2*	rs10758669	C	1.12	3.0E-09	C	A	0.35	0.36	0.98	0.6932
3p21.31	*MST1*	rs3197999	A	1.20	1.0E-12	A	G	0.29	0.3	0.95	0.4022
5q33.1	*IRGM*	rs11747270	G	1.33	3.0E-16	G	A	0.06	0.07	0.85	0.1545
16q12.1	*NOD2*	rs2066847	C	3.99	3.0E-24	NA	NA	NA	NA	NA	NA
10q21.2	*ZNF365*	rs10995271	C	1.25	4.0E-20	C	G	0.37	0.39	0.91	0.0954
13q14.11	Unknown	rs3764147	G	1.25	2.0E-13	G	A	0.2	0.22	0.89	0.1053
6q21	Unknown	rs7746082	C	1.17	2.0E-10	C	G	0.29	0.28	1.06	0.3301
1q23.3	*ITLN1*	rs2274910	C	1.14	1.0E-09	T	C	0.33	0.32	1.01	0.8125
6p22.3	*CDKAL1*	rs6908425	C	1.21	9.0E-10	T	C	0.21	0.23	0.88	0.0627
8q24.13	Intergenic	rs1551398	A	1.08	5.0E-09	G	A	0.34	0.35	0.95	0.3446
18p11.21	*PTPN2*	rs2542151	G	1.35	5.0E-17	G	T	0.18	0.19	0.94	0.3949
10q24.2	*NKX2-3*	rs11190140	T	1.20	3.0E-16	T	C	0.49	0.49	0.98	0.7149
5q31.1	Intergenic	rs2188962	T	1.25	2.0E-18	T	C	0.44	0.43	1.04	0.4693
5q33.3	*IL12B*	rs10045431	C	1.11	4.0E-13	A	C	0.27	0.29	0.88	0.0483
6q27	*CCR6*	rs2301436	T	1.21	1.0E-12	T	C	0.45	0.47	0.94	0.2486
17q21.2	*STAT3*	rs744166	A	1.18	7.0E-12	G	A	0.43	0.42	1.04	0.4792

1p13.2	*PTPN22*	rs2476601	G	1.31	1.0E-08	A	G	0.1	0.09	1.07	0.5029
1q24.3	Intergenic	rs9286879	G	1.19	2.0E-09	G	A	0.23	0.25	0.89	0.0802
1q32.1	Intergenic	rs11584383	T	1.18	1.0E-11	C	T	0.3	0.31	0.96	0.5414
7p12.2	Intergenic	rs1456893	A	1.20	5.0E-09	G	A	0.32	0.31	1.04	0.5647
10p11.21	Intergenic	rs17582416	G	1.16	2.0E-09	G	T	0.33	0.34	0.99	0.8963