

ORIGINAL ARTICLE

# A new regulatory variant in the interleukin-6 receptor gene associates with asthma risk

JA Revez<sup>1</sup>, L Bain<sup>1</sup>, B Chapman<sup>1</sup>, JE Powell<sup>2</sup>, R Jansen<sup>3,4</sup>, DL Duffy<sup>1</sup>, JY Tung<sup>5</sup>, AAGC Collaborators<sup>6</sup>, BW Penninx<sup>3,4</sup>, PM Visscher<sup>2,7</sup>, EJC De Geus<sup>8,9</sup>, DI Boomsma<sup>4,8,9</sup>, DA Hinds<sup>5</sup>, NG Martin<sup>1</sup>, GW Montgomery<sup>1</sup> and MAR Ferreira<sup>1</sup>

The main genetic determinant of soluble interleukin 6 receptor (sIL-6R) levels is the missense variant rs2228145 that maps to the cleavage site of IL-6R. For each Ala allele, sIL-6R serum levels increase by  $\sim 20 \text{ ng ml}^{-1}$  and asthma risk by 1.09-fold. However, this variant does not explain the total heritability for sIL-6R levels. Additional independent variants in *IL6R* may therefore contribute to variation in sIL-6R levels and influence asthma risk. We imputed 471 variants in *IL6R* and tested these for association with sIL-6R serum levels in 360 individuals. An intronic variant (rs12083537) was associated with sIL-6R levels independently of rs4129267 ( $P = 0.0005$ ), a proxy single-nucleotide polymorphism for rs2228145. A significant and consistent association for rs12083537 was observed in a replication panel of 354 individuals ( $P = 0.033$ ). Each rs12083537:A allele increased sIL-6R serum levels by  $2.4 \text{ ng ml}^{-1}$ . Analysis of mRNA levels in two cohorts did not identify significant associations between rs12083537 and *IL6R* transcription levels. On the other hand, results from 16 705 asthmatics and 30 809 controls showed that the rs12083537:A allele increased asthma risk by 1.04-fold ( $P = 0.0419$ ). Genetic risk scores based on *IL6R* regulatory variants may prove useful in explaining variation in clinical response to tocilizumab, an anti-IL-6R monoclonal antibody.

*Genes and Immunity* advance online publication, 15 August 2013; doi:10.1038/gene.2013.38

**Keywords:** allergy; eQTL; expression; disease

## INTRODUCTION

Asthma is a chronic lung disease characterized by airway obstruction, inflammation and hyperresponsiveness.<sup>1</sup> It is caused by a combination of both genetic and environmental factors but their interaction is complex and still not fully understood.

Interleukin 6 (IL-6) is a cytokine known to be involved in allergic asthma<sup>2–5</sup> and it binds to its target cells through a complex of at least two distinct membrane-bound proteins, a 80-kD ligand-binding glycoprotein (IL-6R) and a 130-kD signal-transducing glycoprotein (gp130).<sup>6,7</sup> Although gp130 is expressed on most cells, expression of the membrane-bound form of IL-6R is limited to hepatocytes and some immune cells, such as neutrophils, monocytes, CD4+ T cells and B cells;<sup>8,9</sup> as such, only these cells are directly stimulated by IL-6. However, IL-6R also exists in a soluble form (sIL-6R) that is either produced by proteolytic cleavage of membrane-bound IL-6R or by alternative splicing.<sup>10–13</sup> sIL-6R can associate with free IL-6 and this complex is then recognized by the membrane-bound gp130, hence providing an alternative IL-6 signalling pathway available to cells that do not express membrane-bound IL-6R, a process known as trans-signalling.<sup>7</sup>

The main genetic determinant of sIL-6R levels is thought to be the nonsynonymous variant rs2228145 in exon 9,<sup>14</sup> which maps to amino acid position 358 on the main cleavage site of IL-6R.<sup>15</sup> Specifically, individuals who carry the Ala358 allele have increased sIL-6R levels<sup>14</sup>—which in turn are associated with increased levels of allergic inflammation in the lung<sup>16</sup>—and also increased risk of

asthma.<sup>17</sup> However, given that this variant explains  $\sim 19\%$  of the variation in sIL-6R levels<sup>18</sup> but the total heritability of this trait has been estimated at  $\sim 70\%$ ,<sup>19</sup> we hypothesized that additional variants in or near the *IL6R* gene contribute to sIL-6R variation independently of rs2228145 and may also influence asthma risk.

To test this hypothesis, in this study we (1) used data from the 1000 Genomes Project<sup>20</sup> to comprehensively impute common variants in or near the *IL6R* gene; (2) tested these variants for association with sIL-6R levels measured in 360 individuals; (3) followed-up the most associated variants in an independent sample of 354 individuals; and (4) tested the replicated variants for association with *IL6R* transcript levels in two independent cohorts ( $N = 851$  and  $5191$ ) and asthma risk ( $N = 47\,514$ ).

## RESULTS

Common variants affecting sIL-6R serum levels independently of rs4129267

To find new common variants associated with sIL-6R levels and possibly with asthma risk, we first measured sIL-6R serum levels in 360 asthmatics (Supplementary Table 1) who had been previously genotyped with Illumina 610K single-nucleotide polymorphism (SNP) arrays (Illumina Inc., San Diego, CA, USA).<sup>17</sup> sIL-6R measurements were performed in duplicate, with good agreement between assays ( $r = 0.95$ ). Assay and plate effects

<sup>1</sup>Queensland Institute of Medical Research, Brisbane, Queensland, Australia; <sup>2</sup>University of Queensland Diamantina Institute, Brisbane, Queensland, Australia; <sup>3</sup>Department of Psychiatry, VU University Medical Center, Amsterdam, The Netherlands; <sup>4</sup>Neuroscience Campus Amsterdam, Amsterdam, The Netherlands; <sup>5</sup>23andMe, Inc., Mountain View California, CA, USA; <sup>6</sup>The Australian Asthma Genetics Consortium (AAGC) authors are listed at the end of the article; <sup>7</sup>Queensland Brain Institute, Brisbane, Queensland, Australia; <sup>8</sup>Department of Biological Psychology, VU University Amsterdam, Amsterdam, The Netherlands and <sup>9</sup>EMGO institute for Health and Care Research, VU University Medical Center, Amsterdam, The Netherlands. Correspondence: Dr MAR Ferreira, Queensland Institute of Medical Research, Locked Bag 2000, Royal Brisbane Hospital, Herston, Queensland 4029, Australia.

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

Received 16 May 2013; revised 26 June 2013; accepted 3 July 2013

explained 0.7% ( $P=0.013$ ) and 13.6% ( $P=3.3 \times 10^{-20}$ ), respectively, of the sample variation in sIL-6R levels and hence these effects were removed before association analyses. Duplicate measurements were then averaged and the overall distribution normalized using a rank-based inverse-normal transformation. There were no significant effects of age ( $P=0.256$ ) or sex ( $P=0.932$ ) on sIL-6R levels.

The main genetic determinant of sIL-6R levels is thought to be the nonsynonymous variant rs2228145,<sup>14</sup> which is in high linkage disequilibrium ( $r^2=0.99$ ) with rs4129267 (Figure 1), a SNP that was directly genotyped in all individuals. We confirmed that the rs4129267 SNP has a significant impact on sIL-6R availability, increasing serum levels by  $19.7 \text{ ng ml}^{-1}$  for each copy of the T allele and explaining 30% of the interindividual variation ( $P=1.8 \times 10^{-29}$ , Supplementary Figure 1). However, given that the heritability of sIL-6R levels has been estimated at  $\sim 70\%$ ,<sup>19</sup> we hypothesized that additional variants in or near the *IL6R* gene contribute to variation in sIL-6R serum levels independently of rs4129267.

To test this hypothesis, we imputed 452 common variants in *IL6R* ( $\pm 50 \text{ kb}$ ) in the 360 asthmatics using data from the 1000 Genomes Project as the reference panel. These variants, as well as 19 additional *IL6R* SNPs that were present in the Illumina 610K array, were then tested for association with sIL-6R levels, after adjustment for the effect of rs4129267. Association results for all 471 SNPs are included in Supplementary Table 2.

The variant with strongest association with rs4129267-adjusted sIL-6R levels was rs12083537 (uncorrected  $P=0.0005$ , Supplementary Table 3 and Supplementary Figure 2); the association remained significant after correcting for the number of SNPs tested through 1 00 000 permutations ( $P=0.0496$ ). The rs12083537 variant is in low linkage disequilibrium with rs4129267 ( $r^2=0.03$ ) and has a minor allele (G) frequency of 0.23. An additional three independent ( $r^2 < 0.1$ ) SNPs had a significant (uncorrected  $P < 0.05$ ) association with sIL-6R levels after adjusting for both rs4129267 and rs12083537 (Supplementary Table 3). However, these associations did not remain significant after correction for multiple SNP testing ( $P > 0.05$ ).

#### Replication of the new association between rs12083537 and sIL-6R serum levels

To confirm that the observed association between rs12083537 and sIL-6R levels was real and not a technical or statistical artefact, we performed a replication study using the same experimental procedure in an additional set of 354 unrelated asthmatics (Supplementary Table 4), genotyped for both rs4129267 and rs12083537.

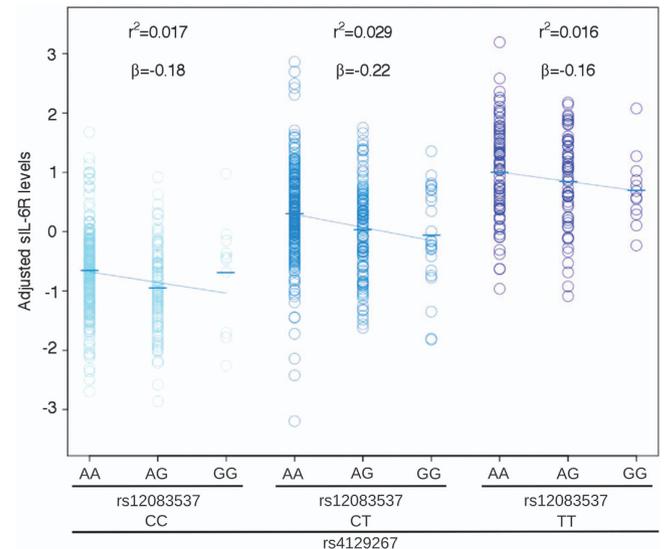
After adjusting for the effects of rs4129267, rs12083537 was significantly associated with sIL-6R levels ( $P=0.033$ ), with the same direction of effect observed in the discovery analysis (Supplementary Figure 3). These results thus confirm that

rs12083537 is a new quantitative trait locus for sIL-6R serum protein levels.

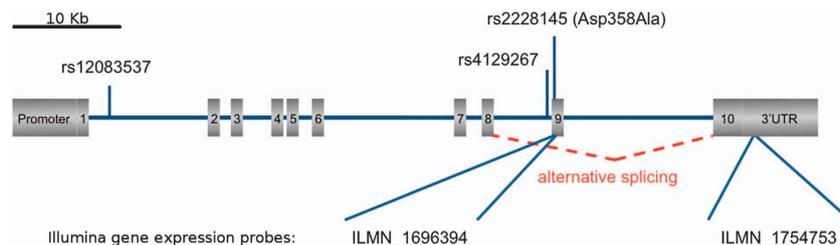
After combining data from the discovery and replication cohorts (combined  $N=714$ ), rs12083537 explained 2.2% of the variation in sIL-6R levels after adjustment for the effects of rs4129267 ( $P=6.2 \times 10^{-5}$ ). Before adjustment for rs4129267, rs12083537 explained 0.15% of the variation in sIL-6R levels. The rs12083537:G minor allele decreased sIL-6R levels similarly in the three rs4129267 genotype classes (Figure 2).

#### Association between rs12083537 with *IL6R* mRNA levels

The main genetic determinant of sIL-6R levels, rs2228145, is located on the main cleavage site of IL-6R and hence it is thought to affect protein levels through a post-transcriptional mechanism.<sup>15</sup> To provide some insight into the molecular mechanisms underlying the new rs12083537 association with sIL-6R protein levels, we investigated whether this variant influenced *IL6R* mRNA levels measured in whole blood in a population-based study of adolescent twins and their relatives ( $N=851$ ). There was no significant association between rs12083537 and *IL6R* mRNA levels in this cohort, either measured by a probe targeting the 3'-untranslated region



**Figure 2.** sIL-6R serum levels as a function of rs12083537 and rs12083537 genotype ( $N=714$ ). The association between sIL-6R levels (after adjusting for technical covariates) and rs12083537 genotype was tested using linear regression and is shown separately for each rs4129267 genotype class, including the proportion of variation in sIL-6R levels explained by rs12083537 ( $r^2$ ), the estimated regression coefficient ( $\beta$ ) and the corresponding regression line.



**Figure 1.** Configuration of the *IL6R* gene. The promoter, exons (1–10) and 3'-untranslated region (UTR) are shown in grey boxes, whereas introns are shown as solid horizontal lines. Also highlighted are the locations of the main genetic regulator of sIL-6R levels (rs2228145), a commonly tested variant in linkage disequilibrium (LD) with it (rs4129267), the novel regulatory variant identified in this study (rs12083537) and the two Illumina probes used to measure *IL6R* gene expression levels.

( $\beta = -0.050$  for the G allele, *s.e.* = 0.062,  $P = 0.4234$ ) or exon 9 ( $\beta = -0.062$  for the G allele, *s.e.* = 0.063,  $P = 0.3209$ ). On the other hand, we observed a significant association between rs4129267 (a proxy for rs2228145) and *IL6R* gene transcription (exon 9 probe,  $\beta = -0.120$  for T allele, *s.e.* = 0.051,  $P = 0.0187$ ). Counterintuitively, the rs4129267:T allele that increases sIL-6R protein levels was associated with decreased mRNA expression.

To confirm that the lack of association between rs12083537 and mRNA levels was not because of low power or incomplete coverage of *IL6R* expression patterns, we performed a similar analysis in a larger independent study ( $N = 5191$ ) with *IL6R* gene expression levels in whole blood measured by 32 individual Affymetrix probes, targeting exons 5, 7, 8 and the 3'-untranslated region. In this cohort, the rs12083537 variant was also not associated with *IL6R* transcription levels after a Bonferroni correction for the number of probes tested ( $P < 0.0016$ , Supplementary Table 5). The strongest association was with a probe mapping to the 3'-untranslated region ( $P = 0.0031$ ), with the rs12083537:A allele that increases sIL-6R protein levels being associated with decreased mRNA expression, as noted above for rs4129267.

#### Association between rs12083537 and asthma risk

The rs2228145:C (or rs4129267:T) allele that increases sIL-6R levels<sup>14</sup> also increases asthma risk.<sup>17</sup> We therefore tested the hypothesis that rs12083537 is also a genetic risk factor for asthma. Specifically, we hypothesized that, as for rs2228145, the rs12083537:A allele that is associated with higher sIL-6R levels would be predisposing for asthma. Association results were available for 47 514 individuals from three independent studies, including 16 705 doctor-diagnosed asthmatics and 30 809 controls. A consistent predisposing effect was observed for the rs12083537:A allele in all three studies (Table 1). Overall, the association between rs12083537 and asthma risk was weak but statistically significant (odds ratio = 1.039, 95% confidence interval = 1.002–1.078,  $P = 0.0419$ ). Results remained largely unchanged when the association was tested after adjusting for rs4129267 (odds ratio = 1.050, 95% confidence interval = 1.012–1.089,  $P = 0.0088$ ).

## DISCUSSION

We recently identified a variant (rs4129267) in the *IL6R* gene with a modest (odds ratio = 1.09) but genome-wide significant association with asthma risk.<sup>17</sup> This variant is in near complete linkage disequilibrium with the exon 9 missense variant rs2228145, and is associated with a 1.4-fold increase in sIL-6R levels for each copy of the T allele.<sup>22</sup> The goal of this study was to

test the hypotheses that (1) other variants in *IL6R* regulate gene expression and (2) that these also associate with asthma risk.

A single variant (rs12083537) located in intron 1 of *IL6R*, 2.9 kb away from exon 1, was found to associate with sIL-6R levels independently of rs4129267 and after correction for multiple SNP testing. A significant and consistent association for this variant was then observed in our replication panel. Furthermore, at the time of submission, we also noted that Ferreira *et al.* independently reported two new genetic associations for sIL-6R levels, one of which (rs1386821) is correlated with rs12083537 ( $r^2 = 0.94$ ). The alleles that increase sIL-6R levels for these two SNPs (rs12083537:A and rs1386821:T) are in phase and hence our results are consistent with those of Ferreira *et al.*<sup>23</sup> Together, these results establish rs12083537, or a variant in linkage disequilibrium with it, as a new regulatory variant for sIL-6R serum levels.

The rs12083537 variant is located in a region with high H3K4Me1 lysine methylation in four cell lines, including lung fibroblasts and keratinocytes.<sup>24</sup> This histone modification is associated with enhancers and promoters and hence it is plausible that rs12083537 may influence gene transcription. However, analysis of mRNA levels extracted from whole blood in two independent studies did not identify any significant associations between rs12083537 and *IL6R* transcription levels, after correcting for multiple testing. Thus, the molecular mechanism underlying the association between rs12083537 and sIL-6R levels remains to be elucidated. A potential caveat of our analysis was that we identified the association between rs12083537 and sIL-6R protein levels in asthma patients but then analysed mRNA levels in two studies that included mostly nonasthmatics. However, this difference is unlikely to explain the lack of association between rs12083537 and mRNA levels because *IL6R* regulatory variants have been shown to influence protein levels similarly in population-based samples,<sup>22</sup> healthy individuals<sup>14,25</sup> and patients with different diseases.<sup>23,25,26</sup> Gene expression studies in individual immune cell types or upon allergen stimulation may help characterize the function of this variant in greater detail.

Interestingly, we observed that the rs4129267:T allele that is associated with increased sIL-6R protein levels<sup>22</sup> was associated with decreased *IL6R* mRNA levels, which may appear contradicting. The Illumina probe for which we observed an association with rs4129267 maps to exon 9 (containing the IL-6R transmembrane domain), which is skipped in the main differentially spliced *IL6R* isoform that directly produces sIL-6R. Therefore, this probe measures the levels of the full-length isoform, which necessarily decrease if splicing of exon 9 increases. Given that the rs4129267:T allele, or an allele associated with it (for example, rs2228145:C), strongly increases the splicing of exon 9,<sup>23</sup> it is in fact not surprising that consequently this allele is associated with decreased mRNA levels of the full-length isoform. Unlike our

**Table 1.** Association between rs12083537 and asthma risk in three independent studies

Study	N cases	N controls	Before conditioning on rs4129267			After conditioning on rs4129267 <sup>a</sup>		
			OR <sup>b</sup>	s.e.	P-value	OR <sup>b</sup>	s.e.	P-value
AAGC	2110	3857	1.095	0.0528	0.0846	1.109	0.0527	0.0507
23andMe	4230	10 842	1.036	0.0333	0.2888	1.044	0.0334	0.2012
GABRIEL <sup>c</sup>	10 365	16 110	1.028	0.0249	0.2726	1.041	0.0247	0.1052
All samples	16 705	30 809	1.039	0.0186	0.0419	1.050	0.0186	0.0088

<sup>a</sup>For the AAGC and 23andMe studies, rs4129267 genotype (coded as allelic dosage: 0, 1 or 2) was included as a covariate in a logistic regression analysis modelling the association between asthma status and rs12083537 allelic dosage. Only genome-wide association study (GWAS) summary statistics were available for the GABRIEL; as such, for this study, the conditional analysis was performed using the approach of Yang *et al.*<sup>21</sup> implemented in Genome-wide Complex Trait Analysis (GCTA). <sup>b</sup>Odds ratio (OR) corresponds to the major allele, A. <sup>c</sup>For the GABRIEL study, results are shown for rs1386821 (T allele), a proxy SNP ( $r^2 = 0.94$ ) for rs12083537. The rs1386821:T allele is in phase with the rs12083537:A allele.

results, Ferreira *et al.*<sup>23</sup> did not observe a significant effect of rs2228145 on the levels of the full-length isoform. However, as the authors noted, splicing of exon 9 is a rare event. Therefore, the lack of association in that study likely reflected the low power provided by a small sample size ( $N=88$ ) to detect small differences in the levels of the full-length isoform between genotype classes.

Lastly, we found rs12083537 to be significantly associated with asthma risk. Based on our analysis, each rs12083537:A allele increased sIL-6R serum levels by  $2.4 \text{ ng ml}^{-1}$  and increased asthma risk by 1.04-fold. For comparison, each rs4129267:T allele increased sIL-6R serum levels by  $19.7 \text{ ng ml}^{-1}$  and increased asthma risk by 1.09-fold.<sup>17</sup> The association between rs12083537 and asthma risk was consistent across the three cohorts analysed but only marginally statistically significant in the overall analysis. Replication of our results by well-powered studies is therefore warranted; over 23 000 cases and 23 000 controls (or, for example, 18 000 cases and 36 000 controls) are required to achieve 80% power to detect ( $\alpha=0.05$ ) the effect we observed.

In summary, our study demonstrates that at least two independent genetic variants influence sIL-6R serum levels. Consequently, pharmacogenetic studies of clinical response to tocilizumab, an anti-IL-6R monoclonal antibody approved to treat rheumatoid arthritis<sup>27</sup> and suggested to treat asthma,<sup>16,17</sup> should consider the effects of not only the major genetic determinant of sIL-6R levels (rs2228145), but also of other confirmed regulatory variants such as rs12083537. For this purpose, multi-SNP scores that summarize an individual's combined genetic risk of having high sIL-6R levels may prove useful.

## MATERIALS AND METHODS

### Study participants

Genetic variants in or near the *IL6R* gene were tested for association with sIL-6R protein levels, mRNA levels and asthma risk in the cohorts described below. Informed consent was obtained from all participants and the study protocols were approved by the appropriate Ethics Committees.

**Association with IL-6R protein levels: discovery cohort.** To find new variants in the *IL6R* gene that affect sIL-6R levels, we first studied 360 unrelated asthmatic subjects from the QIMR 1995–1998 Asthma and Allergy study, which is described in detail elsewhere.<sup>28</sup> Briefly, 3073 subjects were recruited from 802 families that were registered on the Australian Twin Registry and had at least one twin who previously reported a history of wheezing in studies conducted at QIMR 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 and skin prick tests. For the present study, we selected 360 unrelated individuals with available DNA and a serum sample, and who answered 'Yes, told to me by a doctor' to the question 'Have you ever had asthma?'. Demographics and clinical characteristics for this discovery sample are summarized in Supplementary Table 1. These individuals were included in a recent genome-wide association study (GWAS) of asthma.<sup>17</sup>

**Association with IL-6R protein levels: replication cohort.** To confirm associations between SNPs and IL-6R protein levels observed in the discovery sample, we studied an additional sample of 354 unrelated asthmatics recruited through the QIMR Asthma Genetics Study between 2011 and 2013. In this study, individuals with a self-reported doctor diagnosis of asthma were recruited from the general community through media appeals (newspapers, radio, television and social media) to participate in a study of genetic risk factors for asthma. Participation included completing a brief online questionnaire about asthma symptoms, medication and severity, and the provision of a blood sample for genetic and immune function testing. Blood was collected at a local pathology laboratory and shipped to QIMR by overnight courier. Demographics and clinical characteristics for this replication sample are summarized in Supplementary Table 4.

**Association with IL6R mRNA levels.** To provide some insight into the molecular mechanisms underlying SNP–protein associations, SNPs confirmed to associate with sIL-6R protein levels were then tested for association with mRNA levels measured in 851 individuals from 262 families who participated in a genome-wide gene expression study reported recently.<sup>29</sup> Briefly, adolescent MZ and DZ twins, their siblings, and their parents have been recruited over a 21-year period into an ongoing study of the genetic and environmental factors influencing pigmented nevi and the associated risk of developing skin cancer and cognition (Twin Moles study).<sup>30</sup> Results obtained in this cohort were confirmed by analysing expression levels from an additional 5191 individuals who participated in two large-scale longitudinal studies: the Netherlands Twin Register (NTR;  $N=3170$ ) and the Netherlands Study of Depression and Anxiety (NESDA;  $N=2021$ ); henceforth we refer to this as the NTR–NESDA expression quantitative trait locus (eQTL) study. The individual NTR and NESDA cohorts are described in detail elsewhere.<sup>31,32</sup>

**Association with asthma risk.** SNPs associated with sIL-6R protein levels were also tested for association with asthma risk in 21 039 unrelated individuals of European ancestry who participated in two independent studies. First, 5967 individuals who participated in the Australian asthma GWAS,<sup>17</sup> including 2110 doctor-diagnosed asthmatics and 3857 controls. Sample ascertainment and patient characteristics for this study are described in detail in Ferreira *et al.*<sup>17</sup> Second, 15 072 customers of 23andMe, Inc., a personal genetics company, who had been genotyped as part of the 23andMe Personal Genome Service. We selected 4230 cases and 10 842 controls of European ancestry who had taken a survey about asthma. Cases (mean age 48, 54% female, 55% with onset of asthma at age  $\leq 16$  years) gave positive responses to the question 'Have you ever been diagnosed by a doctor with asthma or bronchial asthma?', and also checked a box indicating that they had ever had 'allergic rhinitis (stuffed or dripping nose caused by allergies)'. Controls (mean age 49, 39% females) gave negative responses to both questions. Association results from these 6340 cases and 14 699 controls were meta-analysed with publicly available results from 10 365 asthmatics and 16 110 controls included in the GABRIEL GWAS,<sup>33</sup> for a combined total sample size of 47 514 individuals, including 16 705 cases and 30 809 controls.

### sIL-6R serum protein measurements

A 40-ml venous blood sample was collected from all consenting subjects. After spinning 10 ml of blood in a serum tube at 805 *g* for 10 min, the serum layer was extracted and stored at  $-80^\circ\text{C}$  until analysis. ELISA kits (R&D Systems, Minneapolis, MN, USA) were used to measure sIL-6R levels according to the manufacturer's procedures. The optical density was determined using a BioTek PowerWave XS2 microplate spectrophotometer (BioTek Instruments, Inc., Winooski, VT, USA) at both 450 and 540 nm wavelengths. Results and standard curves were acquired with BioTek Gen5 2.0 Data Analysis Software. sIL-6R levels were measured in duplicate for each sample.

### IL6R mRNA measurements

Whole blood for expression profiling was collected from 851 individuals from the Twin Moles study and processed as described in detail by Powell *et al.*<sup>29</sup> Briefly, total RNA was extracted from PAXgene tubes using the QIAGEN whole blood gene RNA purification kit (QIAGEN, Valencia, CA, USA). RNA quality and concentration was estimated with Agilent Bioanalyzer (Agilent, Santa Clara, CA, USA) and subsequently converted to complementary DNA, amplified and purified using the Ambion Illumina TotalPrep RNA Amplification Kit (Ambion, Austin, TX, USA). Expression profiles were generated by hybridizing 750 ng of cRNA to Illumina HumanHT-12 v4.0 Beadchip according to Illumina whole-genome gene expression direct hybridization assay guide. Beadchips were then washed and stained and subsequently scanned to obtain fluorescence intensities. Samples were scanned using an Illumina Bead Array Reader. Samples were randomized across chips and chip positions, with check for balance across families, sex and generation. For this study, we restricted our analysis to two probes targeting the *IL6R* gene, specifically ILMN\_1754753 on the 3'-end and ILMN\_1696394 on exon 9 (Supplementary Table 6).

Gene expression association results were confirmed by analysing a further 5191 individuals from the NTR–NESDA eQTL study. In this study, venous blood samples were drawn in the morning after an overnight fast; heparinized whole blood was transferred within 20 min of sampling into PAXgene tubes and stored at  $-20^\circ\text{C}$ . RNA extraction and analysis was

performed at the Rutgers University Cell and DNA Repository (Piscataway, NJ, USA). RNA was extracted using QIAGEN Universal liquid handling system (PAXgene extraction kits following the manufacturer's protocol). Total RNA was measured by spectroscopy (Trinean DropSense, Gentbrugge, Belgium) to determine purity and concentration, whereas RNA fidelity was measured by the Agilent Bioanalyzer analysis. RNA samples were hybridized to Affymetrix (Santa Clara, CA, USA) U219 array plates (GeneTitan) that contain 530 467 probes, each 25 bases in length. Array hybridization, washing, staining and scanning were carried out in an Affymetrix GeneTitan System following the manufacturer's protocol. Nonuniquely mapping probes (hg19) and probes containing a polymorphic SNP based on snp137 (UCSC) were removed. Expression values were obtained using RMA normalization implemented in Affymetrix Power Tools (APT, v 1.12.0). A total of 32 probes targeting *IL6R* were selected for analysis (Supplementary Table 5).

### *IL6R* SNP genotyping

*For association with sIL-6R protein levels: discovery cohort.* Genome-wide SNP data for the 360 individuals were obtained using Illumina 610-Quad Beadchip as part of the Australian Asthma Genetics Consortium GWAS, which is described in detail elsewhere.<sup>17</sup> For this study, we selected 254 post quality control (minor allele frequency >1%, call rate >95%, Hardy-Weinberg equilibrium  $P > 10^{-6}$ ) SNPs in or within 1000 kb of *IL6R*, including rs4129267. To expand the genetic coverage of *IL6R*, these 254 directly genotyped SNPs were then used to impute 530 common variants (minor allele frequency >= 1%) from the 1000 Genomes Project (March 2012 release, 1092 samples of all ancestries) using Impute2.<sup>34</sup> After quality control (minor allele frequency >1%, imputation information score >0.5), there were 471 SNPs located in or within 50 kb of *IL6R* available for analysis, of which 19 had been directly genotyped.

*For association with sIL-6R protein levels: replication cohort.* Genotypes for two SNPs (rs4129267 and rs12083537) were obtained for the 354 individuals with Sequenom MassARRAY iPLEX platform; 12.5 ng of DNA isolated from buffy coats with a salt precipitation method were used per assay. SNP sequences were downloaded from the National Centre for Biotechnology Information (<http://www.ncbi.nlm.nih.gov/>) and cross-checked. Design of the PCR and iPLEX Extension assays was done using the Sequenom Design Suite (<https://mysequenom.com/>). The assays were then manually adjusted to increase the level of multiplexing. SNPs were typed using iPLEXTM Gold chemistry and analysed using a Sequenom MassARRAY Compact Mass Spectrometer (Sequenom Inc., San Diego, CA, USA). The PCR, SAP and iPLEX reactions were performed using half the volume of reagents recommended in the manufacturer's specifications. The post-PCR products were spotted on a Sequenom SpectroChip 2, and the data were processed and analysed using Sequenom MassARRAY TYPER 4.0 software. The primers used to amplify both SNPs are included in Supplementary Table 7.

*For association with IL6R mRNA levels.* DNA samples from the Twin Moles study ( $N = 851$ ) were genotyped by the Scientific Services Division at deCODE Genetics (Reykjavik, Iceland), using the Illumina 610-Quad Beadchip. Genotypes were called with the Illumina BeadStudio software. Full details of genotyping procedures are given in Medland *et al.*<sup>27</sup> DNA extraction for the NTR-NESDA eQTL study ( $N = 5191$ ) has been described before.<sup>31</sup> Genotyping was performed on multiple SNP array platforms, including Perlegen 5.0, Illumina 370, Illumina 660, Illumina Omni Express 1M and Affymetrix 6.0. Standard preimputation quality control filters were applied first within and then between chip platforms; subsequently, data were merged into a single data set. Genome-wide SNP imputation was done with Impute2,<sup>34</sup> using the 1000 Genomes phase I Interim Jun 2011 release. This included the rs12083537 variant analysed in this study, which was imputed with high confidence (information score >0.9).

*For association with asthma risk.* We accessed individual-level genotype data for a single SNP (rs12083537) from the Australian GWAS ( $N = 5967$ ) and the 23andMe cohort ( $N = 15072$ ). This SNP was imputed in the Australian GWAS with high confidence (information score 0.97), using Impute2 and genotype data from the combined 1000 Genomes (60 individuals with northern and western European ancestry from the Centre d'Etude du Polymorphisme Humain collection (CEU), March 2010 release) and HapMap 3 (955 individuals from 11 populations, February 2009 release) projects as the reference panel. In the 23andMe cohort,

rs12083537 was imputed with high confidence ( $r^2 = 0.95$ ) using the August 2010 release of the 1000 Genomes reference haplotypes.

### Statistical analyses

*Quality control of sIL-6R ELISA data.* Each sample was measured in duplicate (two assays), with up to 80 serum samples included per ELISA plate. No outlier observations (located six s.d. from the mean) were observed for either assay. The same procedure was used for the discovery and replication cohorts to estimate and remove the effects of both technical and biological effects on sIL-6R levels. Briefly, we (1) used linear regression to adjust for assay and plate effects; (2) normalized the distribution using a rank-based inverse-normal transformation; and (3) tested and, if significant ( $P < 0.05$ ), adjusted for the effects of age and sex. Subsequent association analyses were carried out using the residuals obtained from (3).

*Association analyses.* Individual SNPs in *IL6R*, directly genotyped or imputed, were tested for association with sIL-6R protein levels (discovery and replication cohorts), mRNA levels and asthma risk. In all cases, we used additive allelic tests of association; for imputed SNPs, we used the inferred allelic dosage.

The association between the 471 SNPs in *IL6R* and protein levels was tested using linear regression in R. Association analyses were performed before and after adjusting sIL-6R levels for the effects of rs4129267, a SNP highly correlated with the missense variant rs2228145 ( $r^2 = 0.99$ ), the main genetic determinant of sIL-6R protein serum levels.<sup>22</sup> To adjust for the effects of rs4129267, we (1) performed a linear regression of sIL-6R levels (dependent variable) on rs4129267 allelic dosage (independent variable); (2) extracted the residuals from this analysis, which represent rs4129267-adjusted sIL-6R levels; and (3) performed a linear regression of rs4129267-adjusted sIL-6R levels (dependent variable) on the allelic dosage of individual SNPs (independent variable). To correct the observed asymptotic association  $P$ -value ( $P_{\text{obs}}$ ) of a given SNP for the number of (and correlation between) SNPs tested, we (1) permuted rs4129267-adjusted sIL-6R levels between the 360 asthmatics; (2) tested the permuted sIL-6R levels for association individually with each of the 471 SNPs and recorded the asymptotic  $P$ -value for the most significantly associated SNP ( $P_{\text{min}}$ ); (3) repeated steps (1) and (2) 1 000 000 times; and (4) calculated for each SNP the empirical association  $P$ -value corrected for the number of SNPs tested as the proportion of 1 000 000 replicates for which  $P_{\text{min}} \leq P_{\text{obs}}$ .

Association with mRNA levels in the Twin Moles study was tested using MACH2QTL,<sup>35,36</sup> which takes into account the family structure of the data. In the NTR-NESDA eQTL study, association with mRNA levels was tested using a linear mixed model in R to correct for family structure (as random effects) as well as sex, age, body mass index, smoking status, technical covariates and ancestry (as fixed effects).

Lastly, association with asthma risk was tested using logistic regression in R, including covariates to adjust for sample origin in the AAGC study and, in the 23andMe study, for age, sex and ancestry. Association analyses conditional on rs4129267 were performed in both studies by including this SNP (coded as allelic dosage: 0, 1 or 2) as a covariate in the logistic regression model. Only GWAS summary statistics were available for the GABRIEL study; as such, for this study, the conditional analysis was performed using the approach of Yang *et al.*<sup>21</sup> implemented in Genome-wide Complex Trait Analysis (GCTA). Fixed-effects meta-analysis of results from the Australian GWAS, 23andMe and the publicly available GABRIEL GWAS was performed with METAL.<sup>37</sup>

### CONTRIBUTORS

Melanie C Matheson, Guy B Marks, Jennie Hui, Peter Le Souëf, Patrick Danoy, Svetlana Baltic, Dale R Nyholt, Mark Jenkins, Catherine Hayden, John Beilby, Faang Cheah, Pamela A Madden, Andrew C Heath, John L Hopper, Bill Musk, Stephen R Leeder, Eugene H Walters, Alan James, Graham Jones, Michael J Abramson, Colin F Robertson, Shyamali C Dharmage, Matthew A Brown and Philip J Thompson.

Affiliation details for this group can be found in the supplementary information.

### CONFLICT OF INTEREST

The authors declare no conflict of interest.

## ACKNOWLEDGEMENTS

We thank all the participants of the Asthma and Twin moles studies: Ann Eldridge, Marlene Grace, Kerrie McAloney (sample collection); Melinda Richter, Lisa Bowdler, Steven Crooks (DNA processing); David Smyth, Harry Beeby, Daniel Park (IT support). Funding was provided by the Australian National Health and Medical Research Council (NHMRC, 613627). The NTR/NESDA data described in this paper were funded by the US National Institute of Mental Health (RC2 MH089951, PI Sullivan) as part of the American Recovery and Reinvestment Act of 2009. MARF is supported by a Career Development Fellowship from the NHMRC.

## REFERENCES

- Murphy DM, O'Byrne PM. Recent advances in the pathophysiology of asthma. *Chest* 2010; **137**: 1417–1426.
- Doganci A, Sauer K, Karwot R, Finotto S. Pathological role of IL-6 in the experimental allergic bronchial asthma in mice. *Clin Rev Allergy Immunol* 2005; **28**: 257–270.
- Neveu WA, Allard JL, Raymond DM, Bourassa LM, Burns SM, Bunn JY et al. Elevation of IL-6 in the allergic asthmatic airway is independent of inflammation but associates with loss of central airway function. *Respir Res* 2010; **11**: 28.
- Rincon M, Irvin CG. Role of IL-6 in asthma and other inflammatory pulmonary diseases. *Int J Biol Sci* 2012; **8**: 1281–1290.
- Wang J, Homer RJ, Chen Q, Elias JA. Endogenous and exogenous IL-6 inhibit aeroallergen-induced Th2 inflammation. *J Immunol* 2000; **165**: 4051–4061.
- Hibi M, Murakami M, Saito M, Hirano T, Taga T, Kishimoto T. Molecular cloning and expression of an IL-6 signal transducer, gp130. *Cell* 1990; **63**: 1149–1157.
- Taga T, Hibi M, Hirata Y, Yamasaki K, Yasukawa K, Matsuda T et al. Interleukin-6 triggers the association of its receptor with a possible signal transducer, gp130. *Cell* 1989; **58**: 573–581.
- Chalaris A, Rabe B, Paliga K, Lange H, Laskay T, Fielding CA et al. Apoptosis is a natural stimulus of IL6R shedding and contributes to the proinflammatory trans-signaling function of neutrophils. *Blood* 2007; **110**: 1748–1755.
- Oberg HH, Wesch D, Grussel S, Rose-John S, Kabelitz D. Differential expression of CD126 and CD130 mediates different STAT-3 phosphorylation in CD4+ CD25- and CD25high regulatory T cells. *Int Immunol* 2006; **18**: 555–563.
- Horiuchi S, Koyanagi Y, Zhou Y, Miyamoto H, Tanaka Y, Waki M et al. Soluble interleukin-6 receptors released from T cell or granulocyte/macrophage cell lines and human peripheral blood mononuclear cells are generated through an alternative splicing mechanism. *Eur J Immunol* 1994; **24**: 1945–1948.
- Lust JA, Donovan KA, Kline MP, Greipp PR, Kyle RA, Maihle NJ. Isolation of an mRNA encoding a soluble form of the human interleukin-6 receptor. *Cytokine* 1992; **4**: 96–100.
- Mullberg J, Schooltink H, Stoyan T, Gunther M, Graeve L, Buse G et al. The soluble interleukin-6 receptor is generated by shedding. *Eur J Immunol* 1993; **23**: 473–480.
- Rose-John S, Heinrich PC. Soluble receptors for cytokines and growth factors: generation and biological function. *Biochem J* 1994; **300**: Pt 2 281–290.
- Galicía JC, Tai H, Komatsu Y, Shimada Y, Akazawa K, Yoshie H. Polymorphisms in the IL-6 receptor (IL-6R) gene: strong evidence that serum levels of soluble IL-6R are genetically influenced. *Genes Immun* 2004; **5**: 513–516.
- Mullberg J, Oberthur W, Lottspeich F, Mehl E, Dittrich E, Graeve L et al. The soluble human IL-6 receptor. Mutational characterization of the proteolytic cleavage site. *J Immunol* 1994; **152**: 4958–4968.
- Doganci A, Eigenbrod T, Krug N, De Sanctis GT, Hausding M, Erpenbeck VJ et al. The IL-6R alpha chain controls lung CD4+ CD25+ Treg development and function during allergic airway inflammation in vivo. *J Clin Invest* 2005; **115**: 313–325.
- Ferreira MA, Matheson MC, Duffy DL, Marks GB, Hui J, Le Souef P et al. Identification of IL6R and chromosome 11q13.5 as risk loci for asthma. *Lancet* 2011; **378**: 1006–1014.
- Rafiq S, Frayling TM, Murray A, Hurst A, Stevens K, Weedon MN et al. A common variant of the interleukin 6 receptor (IL-6r) gene increases IL-6r and IL-6 levels, without other inflammatory effects. *Genes Immun* 2007; **8**: 552–559.
- Raggi P, Su S, Karohl C, Veledar E, Rojas-Campos E, Vaccarino V. Heritability of renal function and inflammatory markers in adult male twins. *Am J Nephrol* 2010; **32**: 317–323.
- Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, Handsaker RE et al. An integrated map of genetic variation from 1,092 human genomes. *Nature* 2012; **491**: 56–65.
- Yang J, Ferreira T, Morris AP, Medland SE, Madden PA, Heath AC et al. Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat Genet* 2012; **44**: 369–375, S1–3.
- Melzer D, Perry JR, Hernandez D, Corsi AM, Stevens K, Rafferty I et al. A genome-wide association study identifies protein quantitative trait loci (pQTLs). *PLoS Genet* 2008; **4**: e1000072.
- Ferreira RC, Freitag DF, Cutler AJ, Howson JM, Rainbow DB, Smyth DJ et al. Functional IL6R 358Ala allele impairs classical IL-6 receptor signaling and influences risk of diverse inflammatory diseases. *PLoS Genet* 2013; **9**: e1003444.
- Dunham I, Kundaje A, Aldred SF, Collins PJ, Davis CA, Doyle F et al. An integrated encyclopedia of DNA elements in the human genome. *Nature* 2012; **489**: 57–74.
- Cortes A, Hadler J, Pointon JP, Robinson PC, Karaderi T, Leo P et al. Identification of multiple risk variants for ankylosing spondylitis through high-density genotyping of immune-related loci. *Nat Genet* 2013; **45**: 730–738.
- Hingorani AD, Casas JP. The interleukin-6 receptor as a target for prevention of coronary heart disease: a mendelian randomisation analysis. *Lancet* 2012; **379**: 1214–1224.
- Medland SE, Nyholt DR, Painter JN, McEvoy BP, McRae AF, Zhu G et al. Common variants in the trichohyalin gene are associated with straight hair in Europeans. *Am J Hum Genet* 2009; **85**: 750–755.
- Ferreira MA, O'Gorman L, Le Souef P, Burton PR, Toelle BG, Robertson CF et al. Variance components analyses of multiple asthma traits in a large sample of Australian families ascertained through a twin proband. *Allergy* 2006; **61**: 245–253.
- Powell JE, Henders AK, McRae AF, Caracella A, Smith S, Wright MJ et al. The Brisbane Systems Genetics Study: genetical genomics meets complex trait genetics. *PLoS One* 2012; **7**: e35430.
- Zhu G, Montgomery GW, James MR, Trent JM, Hayward NK, Martin NG 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.
- Boomsma DI, Willemsen G, Sullivan PF, Heutink P, Meijer P, Sondervan D et al. Genome-wide association of major depression: description of samples for the GAIN Major Depressive Disorder Study: NTR and NESDA biobank projects. *Eur J Hum Genet* 2008; **16**: 335–342.
- Willemsen G, Vink JM, Abdellaoui A, den Braber A, van Beek JH, Draisma HH et al. The Adult Netherlands Twin Register: twenty-five years of survey and biological data collection. *Twin Res Hum Genet* 2013; **16**: 271–281.
- Moffatt MF, Gut IG, Demenais F, Strachan DP, Bouzigon E, Heath S et al. A large-scale, consortium-based genomewide association study of asthma. *N Engl J Med* 2010; **363**: 1211–1221.
- Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet* 2009; **5**: e1000529.
- Li Y, Willer C, Sanna S, Abecasis G. Genotype imputation. *Annu Rev Genomics Hum Genet* 2009; **10**: 387–406.
- Li Y, Willer CJ, Ding J, Scheet P, Abecasis GR. MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. *Genet Epidemiol* 2010; **34**: 816–834.
- Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genome-wide association scans. *Bioinformatics* 2010; **26**: 2190–2191.

Supplementary Information accompanies this paper on Genes and Immunity website (<http://www.nature.com/gene>)