npg

ARTICLE

Determination of the real effect of genes identified in GWAS: the example of *IL2RA* in multiple sclerosis

Marie-Claude Babron[1,2], Hervé Perdry[3,4], Adam E Handel[5], Sreeram V Ramagopalan[5], Vincent Damotte[6,7,8], Bertrand Fontaine[6,7,8], Bertram Müller-Myhsok[9], George C Ebers[5] and Françoise Clerget-Darpoux*[,3,4]

Genome-wide association studies (GWAS), although efficient to detect genes involved in complex diseases, are not designed to measure the real effect of the genes. This is illustrated here by the example of IL2RA in multiple sclerosis (MS). Association between IL2RA and MS is clearly established, although the functional variation is still unknown: the effect of IL2RA might be better described by several SNPs than by a single one. This study investigates whether a pair of SNPs better explains the observed linkage and association data than a single SNP. In total, 522 trio families and 244 affected sib-pairs were typed for 26 IL2RA SNPs. For each SNP and pairs of SNPs, the phased genotypes of patients and controls were compared to determine the SNP set offering the best risk discrimination. Consistency between the genotype risks provided by the retained set and the identical by descent allele sharing in affected sib-pairs was assessed. After controlling for multiple testing, the set of SNPs rs2256774 and rs3118470, provides the best discrimination between the case and control genotype distributions (P-corrected=0.009). The relative risk between the least and most at-risk genotypes is 3.54 with a 95% confidence interval of [2.14–5.94]. Furthermore, the linkage information provided by the allele sharing between affected sibs is consistent with the retained set (P=0.80) but rejects the SNP reported in the literature (P=0.006). Establishing a valid modeling of a disease gene is essential to test its potential interaction with other genes and to reconstruct the pathophysiological pathways.
European Journal of Human Genetics (2012) **20**, 321–325; doi:10.1038/ejhg.2011.197; published online 16 November 2011

INTRODUCTION

The *IL2RA* gene is important in immune regulation and was considered a very good candidate for autoimmunity since Sharfe *et al*[1] described a severe immune disorder in a patient with a mutation in this gene. Analysis of this patient suggested that expression of the IL-2R α chain is necessary not only for the functional responses of mature T cells but also for autoreactive T lymphocytes, in preventing their generation and/or controlling their activity. Evidence for an association between *IL2RA* and type 1 diabetes (T1D) was then shown by Vella *et al*[2] through the analysis of 20 *IL2RA* tag SNPs in a large case–control and family collection. It is now well established that the *IL2RA* gene is involved in the pathophysiology of several autoimmune diseases: multiple sclerosis, T1D, systemic lupus erythematosus, Graves' disease and ANCA-associated vasculitis.[3]

For multiple sclerosis (MS), the association with *IL2RA* was first suggested by Matesanz *et al*[4] who compared the genotypes of 346 MS patients *versus* 413 controls for four SNPs in this gene. The association was clearly established in the genome-wide study carried out by the International Multiple Sclerosis Genetics Consortium[5] on large UK and US collections. It was then replicated in several other populations of Caucasian origin.[6–9] *IL2RA* was, before confirmation of its involvement on large samples, a very good candidate gene and is a perfect illustration of Peltonen's 'old suspect found guilty'.[10]

Association between *IL2RA* and MS was consistently reflected by the SNPs rs2104286 and rs12722489, located within the first intron of the gene. Analysis of these two SNPs in 15 Caucasian populations (totaling 11 019 unrelated cases, 13 616 controls and 2811 trios) did not show any heterogeneous effect across populations.[11] The authors concluded that the association observed with rs12722489 was a result of its linkage disequilibrium with rs2104286, although rs2104286 is probably not directly involved in the disease process.

Several other studies focusing on *IL2RA* and testing several SNPs in this gene also concluded to an association of MS with rs2104286.[12,13] Moreover, functional studies showed that rs2104286 was associated with differences in surface expression of level proteins,[13] and in levels of a soluble form of *IL-2* receptors.[14,15]

However, when an association signal has been obtained on a SNP, there is still a long way before identifying the causal variants, and deciphering their mode of action and their connection with the other pieces of the pathway.[16,17] This is well illustrated by the HLA story. Although association between the MHC region and many diseases was identified almost 40 years ago,[18] for most of those diseases, there is still no clear understanding of the functional variation involved in the pathological process. For MS although HLA-DRB1*1501 is the most strongly associated susceptibility allele in many populations, the differential of risk cannot be summarized by a single allele, or even a single haplotype.[19] This is also clearly demonstrated in celiac disease

[1]Inserm, U946, Variabilité Génétique et Maladies Humaines, Paris, France; [2]Univ Paris-Diderot, UMR-S946, Paris, France; [3]Univ Paris-Sud, UMR-S669, Villejuif, France; [4]Inserm, U669, Génomique Épidémiologique Intégrative, Villejuif, France; [5]Wellcome Trust Centre for Human Genetics and Department of Clinical Neurology, University of Oxford, Oxford, UK; [6]Inserm, U975, CRICM, Paris, France; [7]Univ Pierre et Marie Curie, UMR-S975, Paris, France; [8]Assistance Publique-Hôpitaux de Paris, Département de Neurologie, Groupe Hospitalier Pitié-Salpêtrière, Paris, France; [9]Department of Neurology, Max-Planck Institute of Psychiatry, Munich, Germany
*Correspondence: Dr F Clerget-Darpoux, INSERM UMRS 669, Génomique épidémiologique intégrative, Hôpital Paul Brousse – Bat 15/16, 16 avenue Paul-Vaillant-Couturier, 94807 Villejuif, France. Tel: +33 1 45 59 53 63; Fax: +33 1 45 59 53 31; E-mail: Francoise.Clerget@inserm.fr
Received 10 February 2011; revised 6 September 2011; accepted 16 September 2011; published online 16 November 2011

with the major role of a DQA-DQB heterodimer encoded in *cis* or *trans*[20] and in rheumatoid arthritis with the involvement of several amino acids.[21] This complexity is certainly not restricted to the MHC but also to the other susceptibility genes. Some studies already suggested that the role of *IL2RA* is better captured by several SNPs rather than a single one.[22,23]

The contrast between cases and controls is a good tool for the estimation of the genotype relative risks (GRRs). However, when the observed marker is not the actual causal variant, but is in disequilibrium with it, the marker genotype relative risks give a distorted image of the 'true' GRRs, that is, those associated with the genotypes for causal variation. Other sources of information must be used and can be crucial in assessing the real effect of the unobserved causal variant.[24]

In particular, in a candidate gene study, the familial data provide complementary information to the one of cases–controls.[25–29] In rheumatoid arthritis, the parental allele sharing of affected sibs clearly demonstrated that individuals were misclassified in terms of risks when using only the information provided by association studies on *HLA* and *PTPN22*.[30,31]

Our aim, here, is to show how combining the two sources of information, association and linkage, makes it possible to better measure the genetic effect of *IL2RA* in MS. This is accomplished through the use of trios and multiplex families from the collaboration of the Réseau Français de la Génétique de la Sclérose En Plaques (REFGENSEP) and the Canadian Collaborative Project on Genetic Susceptibility to MS (CCPGSMS) groups.

SUBJECTS AND METHODS
Subjects
Patients, all of European descent, either belonged to 523 trio families (one affected child with two living parents) or to 245 multiplex families (at least two affected sibs, with or without living parents, and with or without unaffected siblings). Trio family samples were collected through REFGENSEP, whereas samples for multiplex families were obtained from both REFGENSEP and CCPGSMS. All patients were reviewed by a board-certified neurologist and diagnosed according to the criteria of Poser *et al*.[32] Only patients diagnosed with definite MS were included. A total of 3226 individuals were available for typing, including 768 index cases, defined as the affected child in trio families, and as one random affected child in multiplex families. All individuals signed informed consent in accordance with the European Union and Country Laws and the Helsinki Convention.

TagSNP selection and genotyping
Genotypes for a set of 161 SNPs covering the *IL2RA* gene and its promoter region were extracted from the CEU population data from HapMap Release 22 (April 2007). TagSNPs were selected using Haploview v4.1[33] with the following constraints: pairwise tagging, MAF > 0.10 and $r^2 < 0.8$. Twenty-six SNPs were retained following this process for typing in both sample sets (see Supplementary Table 1).

Genotyping was performed in Munich and Oxford following the same procedure. Genotyping of SNPs was performed using the Sequenom Mass EXTEND protocol (www.sequenom.com). Only conservative and moderate genotyping calls were accepted in this study. Samples having aggressive or low probability quality genotypes were reanalysed. All genotypes were generated blind to pedigree structure and disease status of the individual.

Quality control, Mendelian checking and testing for homogeneity of the two sets of index cases, REFGENSEP and CCPGSMS, were performed using PLINK software.[34]

Statistical analysis
Our objective is to better model the role of *IL2RA* in MS by finding the set of SNPs that offers the best discrimination in terms of genetic risk between patients and controls and that is compatible with the allele sharing observed in pairs of affected sibs. This is carried out in three steps:

(1) In the first step, the SNP set that most significantly distinguish the phased genotypes of index cases and controls is searched.
(2) The second step consists in calculating the GRR of each phased genotype at the SNP set. The values of the GRR and the allele/haplotype frequencies provide a genetic model, which must be consistent with the other information available from the data.
(3) In the third step, the allele sharing distributions in sib-pairs expected under the genetic model determined in step 2 is computed, conditional to the index patient genotype. The observed distributions are compared with the expected ones, in order to test whether the genetic model correctly predicts the sib-pair allele sharing.

These three steps are detailed below. The full procedure was also applied to the SNP of the literature.

SNP-set selection. In this first step, the phased genotypes (diplotypes) of index cases and controls were compared for different SNP sets. Control diplotypes were constructed from the untransmitted parental haplotypes of trio families, assuming the Hardy–Weinberg equilibrium, as the untransmitted parental haplotypes of trio families have been shown to represent those of the general population.[35] All possible diplotypes of each index case and control were determined using our EMphase software. Similar to Clayton's program SNPHAP,[36] EMphase estimates haplotype frequencies using an EM algorithm to calculate maximum likelihood estimates. It also allows for missing genotypes. For each individual, EMphase provides the list of possible haplotype configurations together with their respective probabilities. In addition to the functionalities of SNPHAP, EMphase can also treat data from both unrelated individuals and trio families.

The combination test[37] was used to determine the SNP sets significantly distinguishing the phased genotypes of index cases and controls. Briefly, the combination test consists in testing the contrast between case and controls diplotypes observed in all possible subsets of *n* SNPs, from one SNP (*n* possible subsets) to the *n* SNPs taken together. For a given subset, an association χ^2 is calculated and its nominal *P*-value estimated by permuting case and control diplotypes. The global significance of the complete procedure, based on the minimum of all the *P*-values, is then assessed by permutation to alleviate the problems relating to multiple, non-independent tests. A full exploration of the 26 SNPs in *IL2RA* would require more than 63 million tests ($2^{26} - 1$ tests). Here, only sets of one or two SNPs were considered (351 subsets) in order to limit the number of tests. Note that finding that a pair of SNPs significantly discriminate cases and controls does not imply the absence of more complex relationships, which could have been evidenced if sets of more than two SNPs had been considered.

Genetic modeling. The genotype relative risks (GRR_{ij}) of phased genotypes G_{ij} were obtained using Bayes' rule knowing the haplotype frequencies *i* and *j* for the set of SNPs given by the combination test, or the allele frequencies for the SNP of the literature. The genetic model corresponding to one SNP is described by two allele frequencies, and two GRRs for the three genotypes, with an arbitrary genotype taken as reference (GRR = 1). For a set of two SNPs, the genetic model is described by 4 haplotype frequencies, and 9 GRRs for the 10 phased genotypes, again with a GRR of 1 for the reference genotype.

Test of genetic model on affected sib pairs. For a genetic model described in terms of GRR_{ij} and allele/haplotype frequency p_i, it is possible to determine the expected number of alleles shared identical by descent (IBD) by the affected sib of a patient, conditional on the genotype of this patient (index patient). Details on the calculation are given in the Supplementary Appendix. Different models give different expectations,[25] which can then be compared with what is observed in the data to test whether a particular genetic model fits the sib-pair data or not.

In each affected sib-pair, one affected was considered as the index patient. The IBD status of each sib-pair (ie, the number *n* = 0, 1 or 2 of IBD shared parental alleles), was determined using Merlin.[38] Merlin computes the

probability of each possible IBD state; if the probability of a state exceeds a threshold of 0.9, this state is assigned to the sib-pair.

Observed and expected IBD distribution conditional on each index genotype G_{ij} was then compared using a χ^2-test. Index genotypes with similar risks were pooled in order to avoid small-expected cell sizes.

RESULTS

All SNPs had a genotyping rate greater than 98%. One REFGENSEP trio and one CCPGSMS multiplex family were excluded after detection of more than one Mendelian incompatibility. When only one incompatibility was detected, the entire family was considered unknown for that SNP. After quality control, 522 REFGENSEP trio families and 244 sib-pairs (143 CCPGSMS and 101 REFGENSEP) were retained for the analysis.

As there was no evidence for heterogeneity between the REFGEN-SEP and CCPGSMS index cases, the two sets were pooled. The sample consisted thus of 766 index case diplotypes and 522 control diplotypes derived from the untransmitted haplotypes identified in the trio families. The last two columns of Supplementary Table 1 report the minor allele frequencies observed in the 1044 untransmitted parental alleles (controls) and 1532 index alleles (cases).

All possible 351 combinations of up to 2 SNPs among the 26 genotyped SNPs were tested for association by the combination test.

The SNP reported in the literature, rs2104286 (G/A), shows nominal significance ($P=0.03$), but is not significant once corrected for multiple testing. The relative risk between the lowest (GG) and highest-risk genotype (AA) is 1.56 (Table 1). This value is close to that reported in the IMSGC meta-analysis[11] of 1.48 with a 95% confidence interval of (1.32–1.64).

A single set of two SNPs, rs2256774 (G/A) and rs3118470 (C/T) was found to be significantly associated after correction for multiple testing (P-corrected$=0.009$; nominal $P=8.10^{-5}$). Taken individually, rs3118470 is found to be nominally associated ($P=0.002$), whereas rs2256774 is not ($P=0.57$). However, the combined presence of rs2256774 and rs3118470 strongly reinforces the signal for association of rs3118470 with MS. The four possible haplotypes for these two SNPs and their frequencies in cases and controls are presented in Supplementary Table 2. Haplotype frequencies in controls were used to calculate the GRR and the 95% confidence interval of all possible diplotypes taking as reference the most frequent diplotype AT-GT (Table 2). There is a tendency, although not significant, given the small sample size, for carriers of rs2256774-A and rs3118470-T in *trans* (diplotype AC-GT) to be more at risk than *cis* carriers (diplotype AT-GC; $P=0.30$). The AT-AT genotype has the smallest GRR (0.70), whereas two genotypes GC-GT and AC-AC have a high GRR of 2.80 and 2.37, respectively. The relative risk between the two most (GC-GT and AC-AC) and the least (AT-AT) at-risk genotypes is estimated to 3.54. The 95% confidence interval of this relative risk, obtained with a bootstrap procedure (10^6 replicates) is (2.14–5.94).

If a SNP or a set of SNPs correctly represents the *IL2RA* effect, it should be consistent with the linkage information, measured by the allele sharing between affected sibs. The expected distributions of the IBD allele sharing conditional on the index patient genotype are computed under the two genetic models, for the literature SNP, rs2102486, and the retained pair of SNPs, rs2256774 and rs3118470, respectively.

For SNP rs2102486, the rare genotype GG was pooled with the heterozygote AG. Thus, two IBD distributions are computed, one for the sibs of index patient AA, the other for the sibs of index patient GG or AG (Table 3). The IBD distribution computed on the GRR for rs2102486 is strongly rejected with $P=0.006$ ($\chi^2=14.53$ with 4 d.f.),

Table 1 Genotype relative risk for rs2104286, with genotype GG as reference

Genotype	Internal controls	Index cases	GRR	95% IC
GG	0.061	0.044	1	—
GA	0.355	0.302	1.19	0.95–1.49
AA	0.584	0.654	1.56	1.30–1.87

Table 2 Genotype relative risk for phased genotypes at SNPs rs2256774 and rs3118470

Diplotype	Internal controls	Index cases	GRR	95% IC
AT-AT	0.146	0.089	0.70	0.47–1.04
AC-GC	0.031	0.021	0.78	0.37–1.61
AT-GC	0.025	0.021	0.94	0.43–2.03
AT-GT	0.253	0.221	1.00	—
AC-AT	0.180	0.166	1.05	0.74–1.50
GT-GT	0.134	0.135	1.15	0.79–1.68
AC-GT	0.174	0.219	1.43	1.02–2.02
AC-AC	0.048	0.099	2.37	1.43–3.93
GC-GT	0.010	0.023	2.80	1.01–7.74
GC-GC	0.000	0.005	n.c.	n.c.

Abbreviation: n.c., not calculated.
Phased genotype AT-GT is taken as reference.

Table 3 Observed and expected identity by descent distributions for rs2104286

Genotype	IBD2	IBD1	IBD0
GG, AG			
Observed	26	29	18
Expected	16.6	36.6	19.8
AA			
Observed	22	79	31
Expected	34.9	65.98	31.2

showing that rs2102486 cannot explain the data observed at the linkage level.

For the pair of SNPs rs2256774 and rs3118470, the 10 genotypes were grouped in three classes, corresponding to GRR less than, close to, and greater than that of the reference genotype, leading to three IBD distributions according to the genotype of the index case (Table 4). In contrast to the result obtained for rs2102486, the IBD distribution computed on the GRR for the combination of two SNPs fits the observations ($\chi^2=3.14$, 6 d.f., $P=0.80$).

DISCUSSION

The affected sib-pair data show that the SNP rs2104286, presented in the literature as the SNP the most strongly associated to MS, incorrectly represents the IL2RA effect. Contrarily, the 2-SNP model built on rs3118470–rs2256774 is consistent with our data. These two SNPs are in low LD both in the HapMap CEU data ($r^2=0.09$) and in our data ($r^2=0.15$). Interestingly, the SNP rs3118470 was found associated to T1D,[39,40] and with different levels of expression of IL2RA in a functional study.[41] This SNP is also in strong LD with

Table 4 Observed and expected identity by descent distributions for the SNPs rs2256774 and rs3118470

Genotype	IBD2	IBD1	IBD0
AT-AT, AC-GC, AT-GC			
Observed	8	18	9
Expected	6.5	18.5	10
AT-GT, AC-AT, GT-GT			
Observed	24	50	25
Expected	23.1	50.1	25.8
AC-GT, AC-AC, GC-GT			
Observed	16	40	15
Expected	22.2	34.4	14.4

rs10795791 and rs4147359 (HapMap r^2=0.93 and 0.72, respectively), which were shown to be associated both with MS and T1D.[12] Regarding rs2256774, it was found associated to higher levels of rubella antibody[42] and is in strong LD (HapMap r^2=0.63) with rs11594656, which was shown to be correlated with IL2RA expression.[13,15] These studies only considered single-SNP effect and did not take into account the genetic complexity of IL2RA effect. This complexity was underlined by Perera et al[23] reporting that SNP rs791589 remains associated after adjustment on rs210486. However, the SNP pair rs2104286–rs791589, as all SNP pairs containing rs2104286 (25 pairs in total), were tested with the combination test. None were retained at the 5% significance level, after adjustment for multiple testing. Note also that rs3118470 and rs2256774 are in low LD with rs210486 (r^2=0.11 and 0.19, respectively) and not in LD with rs791589.

Our study illustrates that the strength of the association signal obtained for a SNP is not a direct measure of the gene effect. The more complex is the variability of the gene expression the less representative is the SNP signal. Indeed, a gene with a high capacity to differentiate genotypic risk may lead to these weak signals. This is well illustrated in the evaluation of the effect of *PTPN22* in rheumatoid arthritis.[31] The differential risk was 2.7 for the most strongly associated SNP of *PTPN22* but reached 4.7 for a combination of three SNPs in this gene. The IBD allele sharing between sibs was compatible with this classification of genotypic risks but not with that built on the most strongly associated SNP of *PTPN22*. Another illustration may be given on celiac disease for which the OR of the most strongly associated SNP in the HLA region is 7,[43] while it is equal to 25 when the two genes *HLA-DQA1* and *HLA-DQB1*, which are likely to be functional, are considered simultaneously.[44] For *IL2RA* in MS, the relative risk between the least and most at-risk genotypes reported in the literature for the SNP rs2104286 is 1.48 (1.32–1.64), whereas it reaches 3.54 (2.14–5.94) for the joint information provided by SNPs rs2256774 and rs3118470.

Our study population may appear small compared with the populations used in genome-wide association studies (GWAS). However, our sample is large enough to show that rs2104286 poorly discriminates index cases and controls when compared with the pair of SNPs rs2256774–rs3118470. In addition, our 205 sib-pair sample is large enough to demonstrate that rs2104286 incorrectly represents the true effect of IL2RA. Presently, these two SNPs are the best markers of the true functional IL2RA variation, which probably still remains unobserved. They can prove to be of interest in the future, in

particular to test for interaction with other genes or in functional studies. Increasing our sample size might reveal that a combination involving a greater number of SNPs is even better.

Genome-wide association studies are widely used tools for the analysis of multifactorial diseases. In MS, although many genes have been detected through this approach, our knowledge of MS genetics remains incomplete.[45] However, the missing information, often termed 'missing heritability', is measured by using information on associated SNPs, which most often underrepresent the effect of the corresponding genes. Consequently, even if we are convinced that other susceptibility factors for MS remain to be detected, we also believe that a part of the missing information is due to a poor measure of the gene effect. Gene modeling requires adopting a true candidate gene strategy. Its success depends not only on the sample size but also on the type of information available. A good candidate strategy should also take advantage of linkage information. Indeed, when affected sibpair data are available, the observed IBD allele sharing distributions, conditional on the genotype of index patients, must be consistent with the expectations.

Several biological pathways linking the genes detected through GWAS are now proposed for MS susceptibility.[45] Establishing a good modeling and thus a good genotypic risk classification for each gene is an essential step before pathway reconstruction. In particular, the power of testing the potential interaction between two genes depends on the way the information is extracted for these genes. The apparent contradiction in the study of multifactorial diseases between the evidence for biological interaction between genes and the absence of statistical evidence is, in our view, a demonstration that the associated SNP poorly represents the gene effects. This also highlights the interest of linkage information in genetic model selection. Despite the spectacular success of large-scale association studies, the collection of familial data should not be neglected.

CONFLICT OF INTEREST
The authors declare no conflict of interest.

1 Sharfe N, Dadi HK, Shahar M, Roifman CM: Human immune disorder arising from mutation of the alpha chain of the interleukin-2 receptor. *Proc Natl Acad Sci USA* 1997; **94**: 3168–3171.
2 Vella A, Cooper JD, Lowe CE et al: Localization of a type 1 diabetes locus in the IL2RA/CD25 region by use of tag single-nucleotide polymorphisms. *Am J Hum Genet* 2005; **76**: 773–779.
3 Carr EJ, Clatworthy MR, Lowe CE et al: Contrasting genetic association of IL2RA with SLE and ANCA-associated vasculitis. *BMC Med Genet* 2009; **10**: 22.

4 Matesanz F, Caro-Maldonado A, Fedetz M et al: IL2RA/CD25 polymorphisms contribute to multiple sclerosis susceptibility. *J Neurol* 2007; **254**: 682–684.

5 Hafler DA, Compston A, Sawcer S et al: Risk alleles for multiple sclerosis identified by a genome wide study. *N Engl J Med* 2007; **357**: 851–862.

6 Matesanz F, Fernandez O, Alcina A: Genome wide study of multiple sclerosis. *N Engl J Med* 2007; **357**: 2200; author reply 2200–2201.

7 Ramagopalan SV, Anderson C, Sadovnick AD, Ebers GC: Genome wide study of multiple sclerosis. *N Engl J Med* 2007; **357**: 2199–2200; author reply 2200–2191.

8 Rubio JP, Stankovich J, Field J et al: Replication of KIAA0350, IL2RA, RPL5 and CD58 as multiple sclerosis susceptibility genes in Australians. *Genes Immun* 2008; **9**: 624–630.

9 Weber F, Fontaine B, Cournu-Rebeix I et al: IL2RA and IL7RA genes confer suscept-ibility for multiple sclerosis in two independent European populations. *Genes Immun* 2008; **9**: 259–263.

10 Peltonen L: Old suspects found guilty—the first genome profile of multiple sclerosis. *N Engl J Med* 2007; **357**: 927–929.

11 IMSGC: Refining genetic associations in multiple sclerosis. *Lancet Neurol* 2008; **7**: 567–569.

12 Alcina A, Fedetz M, Ndagire D et al: IL2RA/CD25 gene polymorphisms: uneven associa-tion with multiple sclerosis (MS) and type 1 diabetes (T1D). *PLoS One* 2009; **4**: e4137.

13 Dendrou CA, Plagnol V, Fung E et al: Cell-specific protein phenotypes for the autoimmune locus IL2RA using a genotype-selectable human bioresource. *Nat Genet* 2009; **41**: 1011–1015.

14 Maier LM, Anderson DE, Severson CA et al: Soluble IL-2RA levels in multiple sclerosis subjects and the effect of soluble IL-2RA on immune responses. *J Immunol* 2009; **182**: 1541–1547.

15 Maier LM, Lowe CE, Cooper J et al: IL2RA genetic heterogeneity in multiple sclerosis and type 1 diabetes susceptibility and soluble interleukin-2 receptor production. *PLoS Genet* 2009; **5**: e1000322.

16 Clerget-Darpoux F, Elston RC: Are linkage analysis and the collection of family data dead? Prospects for family studies in the age of genome-wide association. *Hum Hered* 2007; **64**: 91–96.

17 Lander ES, Schork NJ: Genetic dissection of complex traits. *Science* 1994; **265**: 2037–2048.

18 Jersild C, Svejgaard A, Fog T: HL-A antigens and multiple sclerosis. *Lancet* 1972; **1**: 1240–1241.

19 Lincoln MR, Ramagopalan SV, Chao MJ et al: Epistasis among HLA-DRB1, HLA-DQA1, and HLA-DQB1 loci determines multiple sclerosis susceptibility. *Proc Natl Acad Sci USA* 2009; **106**: 7542–7547.

20 Sollid LM, Markussen G, Ek J, Gjerde H, Vartdal F, Thorsby E: Evidence for a primary association of celiac disease to a particular HLA-DQ alpha/beta heterodimer. *J Exp Med* 1989; **169**: 345–350.

21 Gregersen PK, Silver J, Winchester RJ: The shared epitope hypothesis. An approach to understanding the molecular genetics of susceptibility to rheumatoid arthritis. *Arthritis Rheum* 1987; **30**: 1205–1213.

22 Cavanillas ML, Alcina A, Nunez C et al: Polymorphisms in the IL2, IL2RA and IL2RB genes in multiple sclerosis risk. *Eur J Hum Genet* 2010; **18**: 794–799.

23 Perera D, Stankovich J, Butzkueven H et al: Fine mapping of multiple sclerosis susceptibility genes provides evidence of allelic heterogeneity at the IL2RA locus. *J Neuroimmunol* 2009; **211**: 105–109.

24 Bourgain C, Genin E, Cox N, Clerget-Darpoux F: Are genome-wide association studies all that we need to dissect the genetic component of complex human diseases? *Eur J Hum Genet* 2007; **15**: 260–263.

25 Clerget-Darpoux F, Babron MC, Prum B, Lathrop GM, Deschamps I, Hors J: A new method to test genetic models in HLA associated diseases: the MASC method. *Ann Hum Genet* 1988; **52** (Pt 3): 247–258.

26 Owen MJ, McGuffin P: Association and linkage: complementary strategies for complex disorders. *J Med Genet* 1993; **30**: 638–639.

27 Bickeboller H: Investigation of linkage and association - issues on study design. *Int Stat Rev* 2000; **68**: 75–81.

28 Houwing-Duistermaat JJ, Uh HW, van Houwelingen HC: A new score statistic to test for association given linkage in affected sibling pair-control designs. *BMC Proc* 2007; **1** (Suppl 1): S39.

29 Thomson G, Barcellos LF, Valdes AM: Searching for additional disease loci in a genomic region. *Adv Genet* 2008; **60**: 253–292.

30 du Montcel ST, Michou L, Petit-Teixeira E et al: New classification of HLA-DRB1 alleles supports the shared epitope hypothesis of rheumatoid arthritis susceptibility. *Arthritis Rheum* 2005; **52**: 1063–1068.

31 Bourgey M, Perdry H, Clerget-Darpoux F: Modeling the effect of PTPN22 in rheumatoid arthritis. *BMC Proc* 2007; **1**(Suppl 1): S37.

32 Poser CM, Paty DW, Scheinberg L et al: New diagnostic criteria for multiple sclerosis: guidelines for research protocols. *Ann Neurol* 1983; **13**: 227–231.

33 Barrett JC, Fry B, Maller J, Daly MJ: Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 2005; **21**: 263–265.

34 Purcell S, Neale B, Todd-Brown K et al: PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 2007; **81**: 559–575.

35 Thomson G: Mapping disease genes: family-based association studies. *Am J Hum Genet* 1995; **57**: 487–498.

36 Clayton DG: SNPHAP: a program for estimating frequencies of large haplotypes of SNPs http://www-genecimrcamacuk/clayton/software/snphaptxt, 2002.

37 Jannot AS, Essioux L, Reese MG, Clerget-Darpoux F: Improved use of SNP information to detect the role of genes. *Genet Epidemiol* 2003; **25**: 158–167.

38 Abecasis GR, Cherny SS, Cookson WO, Cardon LR: Merlin—rapid analysis of dense genetic maps using sparse gene flow trees. *Nat Genet* 2002; **30**: 97–101.

39 Qu HQ, Montpetit A, Ge B, Hudson TJ, Polychronakos C: Toward further mapping of the association between the IL2RA locus and type 1 diabetes. *Diabetes* 2007; **56**: 1174–1176.

40 Kawasaki E, Awata T, Ikegami H et al: Genetic association between the interleukin-2 receptor-alpha gene and mode of onset of type 1 diabetes in the Japanese population. *J Clin Endocrinol Metab* 2009; **94**: 947–952.

41 Qu HQ, Verlaan DJ, Ge B et al: A cis-acting regulatory variant in the IL2RA locus. *J Immunol* 2009; **183**: 5158–5162.

42 Pankratz VS, Vierkant RA, O'Byrne MM, Ovsyannikova IG, Poland GA: Associations between SNPs in candidate immune-relevant genes and rubella antibody levels: a multigenic assessment. *BMC Immunol* 2010; **11**: 48.

43 van Heel DA, Franke L, Hunt KA et al: A genome-wide association study for celiac disease identifies risk variants in the region harboring IL2 and IL21. *Nat Genet* 2007; **39**: 827–829.

44 Margaritte-Jeannin P, Babron MC, Bourgey M et al: HLA-DQ relative risks for coeliac disease in European populations: a study of the European Genetics Cluster on Coeliac Disease. *Tissue Antigens* 2004; **63**: 562–567.

45 Oksenberg JR, Baranzini SE: Multiple sclerosis genetics—is the glass half full, or half empty? *Nat Rev Neurol* 2011; **6**: 429–437.

Supplementary Information accompanies the paper on European Journal of Human Genetics website (http://www.nature.com/ejhg)