Filter-free exhaustive odds ratio-based genome-wide interaction approach pinpoints evidence for interaction in the HLA region in psoriasis.

Laura Grange, Jean-François Bureau, Iryna Nikolayeva, Richard Paul, Kristel Van Steen, Benno Schwikowski, Anavaj Sakuntabhai

To cite this version:
Laura Grange, Jean-François Bureau, Iryna Nikolayeva, Richard Paul, Kristel Van Steen, et al.. Filter-free exhaustive odds ratio-based genome-wide interaction approach pinpoints evidence for interaction in the HLA region in psoriasis.. BMC Genetics, BioMed Central, 2015, 16, pp.11. 10.1186/s12863-015-0174-3. pasteur-01370988

HAL Id: pasteur-01370988
https://hal-pasteur.archives-ouvertes.fr/pasteur-01370988
Submitted on 23 Sep 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Filter-free exhaustive odds ratio-based genome-wide interaction approach pinpoints evidence for interaction in the HLA region in psoriasis

Laura Grange1,2,3, Jean-François Bureau1,2, Iryna Nikolayeva1,4,5, Richard Paul1,2, Kristel Van Steen6,7, Benno Schwikowski4* and Anavaj Sakuntabhai1,2*

Abstract

Background: Deciphering the genetic architecture of complex traits is still a major challenge for human genetics. In most cases, genome-wide association studies have only partially explained the heritability of traits and diseases. Epistasis, one potentially important cause of this missing heritability, is difficult to explore at the genome-wide level. Here, we develop and assess a tool based on interactive odds ratios (I\textsubscript{OR}), Fast Odds Ratio-based sCan for Epistasis (FORCE), as a novel approach for exhaustive genome-wide epistasis search. I\textsubscript{OR} is the ratio between the multiplicative term of the odds ratio (OR) of having each variant over the OR of having both of them. By definition, an I\textsubscript{OR} that significantly deviates from 1 suggests the occurrence of an interaction (epistasis). As the I\textsubscript{OR} is fast to calculate, we used the I\textsubscript{OR} to rank and select pairs of interacting polymorphisms for P value estimation, which is more time consuming.

Results: FORCE displayed power and accuracy similar to existing parametric and non-parametric methods, and is fast enough to complete a filter-free genome-wide epistasis search in a few days on a standard computer. Analysis of psoriasis data uncovered novel epistatic interactions in the HLA region, corroborating the known major and complex role of the HLA region in psoriasis susceptibility.

Conclusions: Our systematic study revealed the ability of FORCE to uncover novel interactions, highlighted the importance of exhaustiveness, as well as its specificity for certain types of interactions that were not detected by existing approaches. We therefore believe that FORCE is a valuable new tool for decoding the genetic basis of complex diseases.

Keywords: Genome-wide interaction studies, Epistasis, Plink, MBMDR, IOR

Background

During the past decade, many genome-wide association studies (GWAS) have aimed to identify new genetic factors determining susceptibility to a variety of diseases [1,2]. Although promising and sometimes successful, these large-scale studies have only led to modest advances [3]. One explanation is that the underlying model that single SNPs contribute independently to the complex trait may frequently be too simple. Rather, complex traits are likely to result from a complex interplay between genes, notably epistatic gene-environment and gene-gene interactions [4].

The principal obstacles in a genome-wide search for epistasis are statistical power to overcome the limitations of multiple testing and the computational time of the search itself. Over the past decades, many tools have been developed for epistasis detection using various statistical methods [5,6], including those based on regression [7-11], linkage disequilibrium and haplotypes [12,13], and Bayesian approaches [14,15]. Alternative approaches are based on data-filtering, machine-learning and data mining [16-19]. Here, we present an approach that detects pairwise epistasis on a genome-wide scale based on the
classical interaction odds ratio (IOR). Introduced by Piegorsch et al. in 1994 [20], this approach has mainly been used for the detection of gene-environment interactions in case-only designs [21]. VanderWeele et al. [22] showed how the use of IOR can help reveal mechanistic actions in case-only designs. Firstly, we report on the first efficient implementation of an approach for genome-wide epistasis detection, which we call FORCE (Fast Odds Ratio sCan for Epistasis). Due to its mathematical simplicity, the approach is suitable for exhaustive unfiltered epistasis analysis; i.e., the exact value of the IOR statistic can be evaluated for all pairs of genotyped SNPs in reasonable time on a standard computer. We introduce the mathematics to compute exact P-values for the most extreme values of IOR.

Secondly, we describe the application of FORCE to the Welcome Trust Case Control Consortium (WTCCC) data on psoriasis, and analyze the previously unknown statistical interactions we found in the light of already-known results.

Lastly we ask whether the statistical interactions detected by FORCE were found due to its exhaustiveness and/or its underlying genetic model, and we present evidence for both. We show that the restriction of FORCE to analyzing only certain SNPs selected according to their marginal effect on psoriasis (as previously described by Knight et al. [23]) strongly limits the statistical significance of the results. We then benchmark the performance of FORCE and other popular methods to detect simulated epistatic interactions, always using exhaustive search. Under different common models for interaction and noise, FORCE consistently detects certain types of interactions better than other approaches.

**Methods**

**Definition of interaction odds ratio (IOR)**

For any given pair of SNPs, the interaction odds-ratio statistic IOR is calculated from a pair of 2×2 contingency tables. These tables are derived from 3×3 tables of all allele combinations, by collapsing them according to a dominant or recessive model (see Table 1). Following preliminary evidence that the dominant model allowed more efficient detection of epistasis (Table 2), all analyses were performed using this dominant genetic model.

We define the following odds ratios:

\[
OR_1 = \frac{\beta \epsilon}{\alpha \gamma}, \quad OR_2 = \frac{\beta \eta}{\alpha \delta}, \quad OR_{1,2} = \frac{\beta \eta}{\alpha \delta}, \quad \text{and } IOR = \frac{OR_{1,2}}{OR_1 \cdot OR_2} = \frac{\alpha \delta \eta}{\beta \gamma \theta}
\]

Note that IOR is undefined when the denominator of this expression becomes zero. For formal consistency, we therefore added a pseudocount of 1 to each cell of the two contingency tables.

**Statistical significance: Empirical and exact P-values**

Note that an IOR of x equals an IOR of 1/x after exchanging counts between cases and controls. We define universal IOR, u(IOR):

\[
u(IOR) = \frac{1}{IOR} \quad \text{if } IOR \leq 1 \quad \text{and } u(IOR) = IOR \quad \text{if } IOR > 1.
\]

This definition allows us to express significant deviations of u(IOR) from the expectation of 1 using a one-tailed P-value.

Pairs with high u(IOR) were identified by the straightforward algorithm that computes u(IOR) for each pair of given SNPs. Our C implementation encodes, in a pre-processing step, all data related to any given SNP into a bit string, and then uses fast logical and bit-counting function to compute u(IOR) for all pairs.

Marginal empirical P-values for any given pair of SNPs were calculated as the proportion of u(IOR) values from randomly generated permutations of case–control labels that were larger than or equal to the value of u(IOR) obtained for the same pair in real data. The number of permutations performed (1000 for simulated data, 100,000 for real data) was adapted to the number of tests performed in these two scenarios.

Exact P-values were calculated using

\[
p = Pr(u(IOR) \geq x) = 2 \cdot \sum_{\text{all } (\alpha',\gamma',\epsilon',\eta') \text{ such that } IOR_{1,2} \geq x} \left( \frac{n_{\alpha'\gamma'\epsilon'\eta'}}{n_{\alpha\gamma\epsilon\eta}} \right) \frac{\binom{\alpha + \gamma}{\alpha'} \binom{\gamma + \epsilon + \eta}{\gamma'} \binom{\gamma + \epsilon + \eta}{\epsilon'} \binom{\gamma + \epsilon + \eta}{\eta'}}{\binom{\alpha + \gamma + \epsilon + \eta + \gamma}{\alpha'}}
\]

and computed by the straightforward algorithm with four nested loops to cover all required parameter tuples \((\alpha',\gamma',\epsilon',\eta')\). Each inner loop only visits those parameter values that correspond to possible tuples with \(\alpha' + \gamma' + \epsilon' + \eta' = \alpha + \gamma + \epsilon + \eta\), given the parameter values in the outer loop. Summed are those terms with \(u(IOR) \geq x\).

**Table 1 Contingency table under a dominant model**

| SNP1 | SNP2     | Cases | Controls |
|------|----------|-------|----------|
| AA   | BB       | \(\alpha\) | \(\beta\) |
| AA   | Bb or bb | \(\gamma\) | \(\delta\) |
| Aa or aa | BB     | \(\epsilon\) | \(\zeta\) |
| Aa or aa | Bb or bb | \(\eta\) | \(\theta\) |

Major alleles are respectively A and B for each SNP and minor alleles a and b. The risk allele is assumed to be the minor allele.

Application of FORCE to psoriasis data

To evaluate FORCE, we assessed its performance on the WTCCC psoriasis dataset. Initial GWAS and further analyses performed on these data are described in [24]. Following general practice for pre-processing, we excluded potentially low-quality SNP data from further analysis.
Specifically, we discarded i) any individual whose total missing rate was above 0.05, ii) any SNP with a frequency of missing data above 0.05, and iii) any SNP with minor allele frequency below 0.05. After pre-processing, our dataset consisted of 2,618 cases, 2,737 controls and 491,191 SNPs, corresponding to approximately $1.2 \times 10^{11}$ SNP pairs. We excluded pairs with a genomic distance of less than 100 kb to avoid pairs in linkage disequilibrium.

In addition, we found that low row and cell counts in the contingency table (Table 1) can lead to extreme but frequently not significant values of $u(I_{OR})$. For the purposes of this study, we excluded 3,521,114 SNP pairs with a total count of less than 50 in any row, or less than 5 in any cell of the contingency table. In addition to FORCE, we performed PLINK (FastEpistasis mode) on the top-ranked 500 pairs to compare the results obtained with both methods.

### Comparison of exhaustive FORCE with semi-exhaustive and conditional search

To assess the utility of exhaustive search, we constructed a reference dataset of SNPs previously implicated in psoriasis. We started with a set of 34 SNPs from two previous reviews on psoriasis genetics [25,26] that were part of our psoriasis dataset. After applying quality control thresholds (described above), 18 SNPs remained.

Following general practice for genome-wide approaches, for exhaustive and semi-exhaustive searches, we used a genome-wide significance threshold of $p = \frac{2 \times 10^{-6}}{10^{6}} = 10^{-13}$, which is based on a model of the human genome with $10^{6}$ independent SNPs [27].

### Comparison of FORCE with other approaches on simulated datasets

We simulated datasets of 10 biallelic SNPs over 200 cases and 200 controls following the Hardy-Weinberg equilibrium model. Interactions were simulated according to six different previously described models without main effect [28] (Table 3). These models represent pure epistasis effects, and not confounding main effects. Model 1 is an interaction effect in which high risk of disease occurs when inheriting heterozygous genotypes at either locus (Aa or Bb) but not both. Model 2 represents high risk of disease when inheriting two high-risk alleles that could be A and/or B. Models 3–6 correspond to the epistasis model discovery method described by Moore et al. [29]. Each of these models represents an interaction effect without any main effects. Allele frequencies are $p = 0.25$ and $q = 0.75$ for model 3 and 4, $p = 0.1$ and $q = 0.9$ for model 5 and 6.

For each of the six models, we generated 100 datasets in each of the 16 conditions of the presence or absence of four of the most commonly encountered sources of noise: missing data (MS), genotyping errors (GE), genetic heterogeneity (GH), and phenocopy (PC).

For GH, two independent interactions were simulated instead of one, each interaction being risk-associated in half of the affected cases. When PC was simulated, interaction affected the trait for half of the cases, emulating an unknown environmental effect. GE and MS were simulated at 5%, as previously described [28].

An epistatic pair of SNPs was considered as detected if the empirical $P$-value was below 0.001, i.e., below 0.05 after Bonferroni correction. Power was estimated as $n/100$, where $n$ is the number of datasets with detection (s). When two pairs ($P_1$, $P_2$) of SNPs were simulated, detection was counted under one of three different conditions: D1) when $P_1$ and $P_2$ were detected, D2) when $P_1$ was detected, or D3) when $P_1$ or $P_2$ was detected. Family-wise error rate (FWER) was calculated as $m/100$, where $m$

| Genetic model | Test | Method 1 | Method 2 | Method 3 | Method 4 | Method 5 | Method 6 |
|---------------|------|----------|----------|----------|----------|----------|----------|
| Dominant      | Power | 1        | 0.97     | 0.96     | 1        | 0.93     | 0.99     |
|               | FWER  | 0.05     | 0.02     | 0.02     | 0.06     | 0.05     | 0.04     |
| Recessive     | Power | 0.93     | 0.96     | 0.01     | 0.01     | 0        | 0        |
|               | FWER  | 0.04     | 0.07     | 0.02     | 0.02     | 0.03     | 0.01     |

| Genetic model | Test | Method 1 | Method 2 | Method 3 | Method 4 | Method 5 | Method 6 |
|---------------|------|----------|----------|----------|----------|----------|----------|
| Dominant      | Power | 1        | 0.97     | 0.96     | 1        | 0.93     | 0.99     |
|               | FWER  | 0.05     | 0.02     | 0.02     | 0.06     | 0.05     | 0.04     |
| Recessive     | Power | 0.93     | 0.96     | 0.01     | 0.01     | 0        | 0        |
|               | FWER  | 0.04     | 0.07     | 0.02     | 0.02     | 0.03     | 0.01     |

Marginal penetrances for each genotype are identical as we simulate pure epistasis effects.

---

**Table 2**: Power and Family-wise error rate (FWER) for detection of the functional pair using a dominant or recessive transmission assumption in 6 different epistasis models

| Genetic model | Test | Model 1 | Model 2 | Model 3 | Model 4 | Model 5 | Model 6 |
|---------------|------|---------|---------|---------|---------|---------|---------|
| Dominant      | Power | 1       | 0.97    | 0.96    | 1       | 0.93    | 0.99    |
|               | FWER  | 0.05    | 0.02    | 0.02    | 0.06    | 0.05    | 0.04    |
| Recessive     | Power | 0.93    | 0.96    | 0.01    | 0.01    | 0       | 0       |
|               | FWER  | 0.04    | 0.07    | 0.02    | 0.02    | 0.03    | 0.01    |

**Table 3**: Penetrances and allele frequencies ($p,q$) used to simulate the interaction models – from Ritchie [28]

| Model 1 | Model 2 | Model 3 |
|---------|---------|---------|
|          | BB  | Bb  | bb  | BB  | Bb  | bb  | BB  | Bb  | bb  | BB  | Bb  | bb  |
| AA 0.01 | 0   | 0.10| 0   | AA 0.01| 0   | 0.10| AA 0.01| 0   | 0.10| AA 0.01| 0   | 0.10|
| Aa 0.10 | 0.10| 0   | 0.10| Aa 0.10| 0   | 0.10| Aa 0.10| 0   | 0.10| Aa 0.10| 0   | 0.10|
| aa 0.10 | 0.10| 0   | 0.10| aa 0.10| 0   | 0.10| aa 0.10| 0   | 0.10| aa 0.10| 0   | 0.10|

| Model 4 | Model 5 | Model 6 |
|---------|---------|---------|
|          | BB  | Bb  | bb  | BB  | Bb  | bb  | BB  | Bb  | bb  | BB  | Bb  | bb  |
| AA 0.01 | 0   | 0.09| 0   | AA 0.01| 0   | 0.09| AA 0.01| 0   | 0.09| AA 0.01| 0   | 0.09|
| Aa 0.01 | 0.08| 0.09| 0.01| Aa 0.01| 0.08| 0.09| Aa 0.01| 0.08| 0.09| Aa 0.01| 0.08| 0.09|
| aa 0.01 | 0.03| 0.09| 0.03| aa 0.01| 0.03| 0.09| aa 0.01| 0.03| 0.09| aa 0.01| 0.03| 0.09|

$p = 0.25$, $q = 0.75$ $p = 0.1$, $q = 0.9$ $p = 0.1$, $q = 0.9$
is the number of datasets for which at least one pair other than the simulated pair was detected.

Results
FORCE enables exhaustive unfiltered epistasis analysis
The FORCE method for epistasis detection is based on the choice of a dominant or recessive model that collapses combinations of allele counts into two 2x2 incidence tables (see Methods). Interactions are then detected as extreme values of the I\textsubscript{OR} statistic. We implemented the FORCE method for epistasis in C language [30]. Due to its mathematical simplicity and efficient implementation, the computation of I\textsubscript{OR} could be performed rapidly, compared to other approaches (4.3 days on a single core of a standard computer). Table 4 shows running times of different methods selected for this study.

Identification of statistically strong interactions requires exhaustive search
To assess the value of exhaustive search, we first evaluated the performance of a conventional, non-exhaustive approach of constraining the analysis to pairs of SNPs that were previously shown to have main effects associated with the phenotype. We therefore performed a constrained analysis on all pairs of 18 high-quality SNPs that had main effects on psoriasis in previous GWA studies (see Methods). Table 5 gives the best 25 hits obtained through this approach when evaluated on the WTCCC dataset on psoriasis [24] (the results of all pairs are shown in Additional file 1: Table S1). None of the 153 pairs reached a significant interaction P-value below a genome-wide significance threshold of 10\textsuperscript{-13}.

A more comprehensive approach, to which we will here refer to as semi-exhaustive, constrains only one of the SNPs in a pair to a set of previously identified SNPs [8]. Table 6 shows, for each of the 18 previously identified “main effects” SNPs, the highest-scoring interactors, according to the FORCE and PLINK FastEpistasis statistics. Note that FORCE and PLINK identified a few genome-wide significant interactions with P-values as low as 10\textsuperscript{-20}.

Table 4 Average time needed to exhaustively test one/all 1.25x10\textsuperscript{11} pairs among 500,000 SNPs using a single-core CPU computer

| Software               | Time for one/all SNPs (single core) |
|------------------------|-------------------------------------|
| MB-MDR [16]            | 5x10\textsuperscript{-3} s/20 years [31] |
| PLINK Epistasis [7]    | 2x10\textsuperscript{-4} s/289 days [5] |
| PLINK FastEpistasis [8] | 2x10\textsuperscript{-5} s/29 days [32] |
| FORCE                  | 3x10\textsuperscript{-6} s/43 days |
| GWIS - 3 filters [33]  | 1.6x10\textsuperscript{-6} s/2.2 days [33] |
| GWIS - 1 filter [33]   | 3.8x10\textsuperscript{-7} s/0.5 days [33] |

We included the recent GWIS approach that is described as ‘exhaustive’, but uses filtering to avoid computing test statistics for all pairs of SNPs.

Finally, the relatively low computational complexity required for the FORCE statistic allowed us to perform exhaustive analysis of all SNP pairs in the psoriasis dataset. The results are shown in Table 7 (100 best hits shown in Additional file 1: Table S2). Strikingly, the best resulting P-values are another 20 orders of magnitude lower than the P-values identified by semi-exhaustive search. This shows that a large number of the most significant interactions are missed by the semi-exhaustive approach, and hence that the possibility of discovering the statistically best-supported interactions requires an exhaustive approach. Interestingly, FORCE and PLINK identify distinct interactions.

FORCE pinpoints interactions beyond main effects in the HLA region
We also analyzed the exhaustive FORCE results with regard to previous studies, which have detected numerous main effects [24-26], but only few weak statistical interactions [24,34,35]. We assessed the performance of FORCE using the WTCCC psoriasis dataset, which contains 2,618 cases, 2,737 controls and 491,191 SNPs. Table 7 shows the 25 best FORCE hits. Twenty-one out of 25 SNP pairs involve SNPs located in the HLA region on chromosome 6, which is consistent with the known strong involvement of the HLA region in psoriasis. Interestingly, certain SNP pairs found to be statistically significant by FORCE did not reach genome-wide significance when using PLINK FastEpistasis.

It is well known that SNPs with main effects may falsely appear to be interacting [36]. To avoid such artifacts in our analysis, we removed those SNPs that displayed a univariate statistical association P-value of 10\textsuperscript{-5} or less [24]. The results show three highly significant interactions involving SNPs from the HLA region that display no main effect (Table 8). In the absence of correlation between the SNPs we claim that these findings provide evidence of interactive effects involved in psoriasis susceptibility. This confirms that FORCE is able to uncover novel statistical interactions in the HLA region that have not been detected before using conventional approaches.

FORCE systematically detects interactions missed by other approaches
Besides its exhaustiveness, the other characteristic feature of the FORCE approach is the use of the I\textsubscript{OR} statistic for genome-wide epistasis analysis. To study the extent to which the choice of this statistic contributed to the identification of novel statistical interactions, we used datasets that contained different simulated epistatic interactions between SNPs without main effects, according one of six models of Ritchie [28], and none or one of the four sources of noise: Genotyping Error (GE), Missing Data (MS), Genetic Heterogeneity (GH), Phenocopy (PC) (see
Methods for details). We then evaluated the power of FORCE and three other popular epistasis detection methods (PLINK Epistasis [7] and PLINK FastEpistasis [8] using default parameters, and MB-MDR [16], using recommended parameters [37]) to detect the simulated interactions. We used a significance threshold of 0.001. Figure 1 shows the results for all epistatic models for the case of no noise.

Under all six models, FORCE and MB-MDR consistently showed power close to 1. The situation became more interesting in the presence of noise. Figure 2 shows the power of the tested methods for all six models in the presence of one type of noise (numerical values for are given in Tables 9, 10, 11 and 12). While the results for Genotyping Errors (GE) and Missing Data (MS) were very similar to the no-noise scenario, the presence of Genetic Heterogeneity (GH, independent of the definition of “detection”) or Phenocopy (PC) revealed larger differences among the different approaches. Firstly, we noted that, with GH and PC, all approaches lose power. Secondly, we observed that different approaches worked consistently better than others, depending on the interaction model. For interaction models 1 and 2, MB-MDR dominated all other approaches; FORCE dominated the other approaches for interaction models 3–6.

**Discussion**

This study introduces the FORCE approach for genome-wide epistasis analysis. On the basis of the Interaction Odds Ratio ($I_{OR}$) statistic, it performs a genome-wide search for epistatic interactions between pairs of SNPs in a reasonable time on a standard laptop computer. The search is exhaustive and filter-free; i.e., the result is guaranteed to reflect the most extreme $I_{OR}$ values over all possible interactions. Exhaustive search using FORCE is possible because of the computational simplicity of the $I_{OR}$ statistic.

| First GWAS SNP | Second GWAS SNP | FORCE $I_{OR}$ | Empirical p-value | PLINK FastEpistasis p-value |
|----------------|----------------|----------------|-------------------|-----------------------------|
| rs10484554 6p21.33 | rs27524 5q15 | 6.846 | 0.008882 | 0.003095 |
| rs10484554 6p21.33 | rs3134792 6p21.33 | 1.068 | 0.3014 | 0.007746 |
| rs2201841 1p31.3 | rs3213094 5q33.3 | 4.737 | 0.02952 | 0.012373 |
| rs3134792 6p21.33 | rs4795067 17q11 | 3.188 | 0.07419 | 0.012783 |
| rs20541 5q31 | rs17716942 2q24 | 6.987 | 0.008212 | 0.013389 |
| rs702873 2p16 | rs4795067 17q11 | 3.414 | 0.06466 | 0.014129 |
| rs10484554 6p21.33 | rs4795067 17q11 | 2.597 | 0.1071 | 0.018096 |
| rs610604 6q23 | rs17716942 2q24 | 6.591 | 0.01025 | 0.023261 |
| rs3213094 5q33.3 | rs12580100 12q13.2 | 5.132 | 0.02349 | 0.028993 |
| rs4649203 1p31.3 | rs240993 6q21 | 2.270 | 0.1319 | 0.037791 |
| rs4649203 1p31.3 | rs702873 2p16 | 1.237 | 0.266 | 0.041136 |
| rs3134792 6p21.33 | rs27524 5q15 | 11.840 | 0.000581 | 0.014183 |
| rs702873 2p16 | rs2546890 5q33.3 | 0.804 | 0.37 | 0.041729 |
| rs27524 5q15 | rs17716942 2q24 | 5.280 | 0.02158 | 0.045812 |
| rs610604 6q23 | rs6701216 1q21 | 4.289 | 0.03837 | 0.057701 |
| rs2201841 1p31.3 | rs2546890 5q33.3 | 2.587 | 0.1077 | 0.059206 |
| rs27524 5q15 | rs7993214 13q14.11 | 3.596 | 0.05793 | 0.059609 |
| rs3134792 6p21.33 | rs213094 5q33.3 | 2.610 | 0.1062 | 0.072999 |
| rs702873 2p16 | rs2201841 1p31.3 | 1.669 | 0.1964 | 0.083717 |
| rs10484554 6p21.33 | rs12580100 12q13.2 | 2.535 | 0.1113 | 0.086631 |
| rs4649203 1p31.3 | rs6701216 1q21 | 3.518 | 0.06072 | 0.086671 |
| rs2201841 1p31.3 | rs27524 5q15 | 1.666 | 0.1968 | 0.088785 |
| rs4112788 1q21.3 | rs7993214 13q14.11 | 1.546 | 0.2137 | 0.090419 |
| rs240993 6q21 | rs7993214 13q14.11 | 1.896 | 0.1685 | 0.096038 |
| rs6701216 1q21 | rs8016947 14q13 | 1.087 | 0.2971 | 0.100508 |
| GWAS-identified SNP | Highest-scoring interactor with GWAS-identified SNP | PLINK FastEpistasis Z-score |
|---------------------|---------------------------------------------------|-----------------------------|
| rs number | Chromosomal location | Risk allele OR | Single association p-value | rs number | Chromosomal location | u(I\_OR) | u(I\_OR) | Empirical p-value | Exact p-value | rs number | Chromosomal location | Exact p-value |
|rs10484554 | 10q21.33 | 4.66 | 4.0E-214 | rs4151664 | 6p21.33 | 2.97 | <10E-06 | 7.86E-10 |
|rs2546890 | 5q33.3 | 1.54 | 1.0E-20 | rs7525345 | 1p31.1 | 2.53 | 7.10E-04 | 2.17E-06 |
|rs6701216 | 1q21 | 1.45 | 6.2E-05 | rs2156892 | 22q11.22 | 2.5 | <10E-06 | 1.30E-13 |
|rs4112788 | 1q21.3 | 1.41 | 6.5E-09 | No pair meeting all inclusion criteria |
|rs7993214 | 13q14.11 | 1.41 | 2.0E-06 | No pair meeting all inclusion criteria |
|rs3213094 | 5q33.3 | 1.39 | 5.0E-11 | No pair meeting all inclusion criteria |
|rs17716942 | 2q24 | 1.29 | 1.1E-13 | rs16928722 | 10q22.1 | 2.69 | 4.10E-06 | 2.78E-06 |
|rs20541 | 5q31.1 | 1.27 | 5.0E-15 | No pair meeting all inclusion criteria |
|rs240993 | 6q21 | 1.25 | 5.3E-20 | rs4727157 | 7q21.12 | 2.78 | <10E-06 | 1.88E-20 |
|rs4795057 | 17q11 | 1.19 | 4.0E-11 | No pair meeting all inclusion criteria |
|rs8016947 | 1q13 | 1.19 | 1.5E-11 | No pair meeting all inclusion criteria |
|rs610604 | 6q23 | 1.19 | 7.0E-07 | rs17585537 | 3p26.2 | 2.69 | <10E-06 | 6.47E-19 |
|rs12580100 | 1q13.2 | 1.17 | 1.0E-06 | rs7565742 | 2q31.2 | 3.39 | <10E-06 | 6.80E-20 |
|rs4649203 | 1p31 | 1.13 | 6.8E-08 | No pair meeting all inclusion criteria |
|rs2201841 | 1p31.3 | 1.13 | 3.0E-08 | No pair meeting all inclusion criteria |
|rs27524 | 5q15 | 1.13 | 2.6E-11 | No pair meeting all inclusion criteria |
|rs702873 | 2p16 | 1.12 | 3.6E-09 | No pair meeting all inclusion criteria |
|rs3134792 | 6p21.33 | NR | 1.0E-09 | rs1062070 | 6p21.32 | 2.88 | <10E-06 | 2.85E-10 |

*Bold data are genome-wide significant interactions.*
Wu et al. [38] introduced a haplotype-based measure based on the following term:

\[ I_{GH} = \frac{OR_{G_1H_1}}{OR_{G_1H_2} OR_{G_2H_1}} \]

where \( OR_{G_iH_i} \) is the odds ratio for both risk haplotypes when carried together, compared to the baseline haplotypes; \( OR_{G_1H_1} \) and \( OR_{G_2H_i} \) are the odds ratios for each risk haplotype, respectively, compared to the baseline haplotype.

Although both methods are based on odds ratios, the methods differ in several respects. First, and most significantly, Wu’s method uses haplotypes, which typically require the statistical inference of haplotypes. Even though this design was shown to be better powered than classical

Table 7 FORCEx Exhaustive search top hits, and PLINK FastEpistasis results in WTCCC psoriasis data

| SNP pair description | Epistasis search results |
|----------------------|-------------------------|
| **SNP1** | **SNP2** | **FORCE** | **PLINK FastEpistasis** |
| rs4151664 | 6p21.33 | rs4151664 | 6p21.33 | 0.008 | 0.076 | 1.22E-22 | 5.23 | 0.002 |
| rs9267532 | rs2254556 | 6p21.33 | rs2254556 | 6p21.33 | 0.008 | 0.076 | 1.22E-22 | 5.23 | 0.002 |
| rs9267532 | rs2254556 | 6p21.33 | rs2254556 | 6p21.33 | 0.008 | 0.076 | 1.22E-22 | 5.23 | 0.002 |

*Bold data are genome-wide significant interactions.*
Figure 1: Power of different approaches to detect simulated epistatic interactions across the six epistasis models by Ritchie [28]. Purple: FORCE – Green: MB-MDR – Blue: PLINK Epistasis – Red: PLINK FastEpistasis. Refer to Table 3 for the definitions of the 6 interaction models.

Figure 2: Power of different approaches to detect simulated epistatic interactions across the six epistasis models by Ritchie [28], in the presence of noise. Comparison of the power of four methods to detect interaction in the presence of one source of noise. GH: Genetic heterogeneity – GE: Genotyping errors – MS: Missing data – PC: Phenocopy. When GH is simulated, three different ways of calculating power are employed: the power of detecting both pairs in the same dataset, the power of detecting the first (fixed) pair and the power to detect either of the two epistatic pairs. Purple: FORCE – Green: MB-MDR – Blue: PLINK Epistasis – Red: PLINK FastEpistasis.
genotype-based statistics, the additional calculations are computationally costly. As a result, FORCE can perform an exhaustive genome-wide epistasis search in a few days on a single compute core while, in practice, Wu et al.'s method only allows a limited number of SNP pairs to be tested.

In addition to the different statistics themselves, the approaches to calculating significance differ. FORCE relies on an exact P-value that requires too much time to be calculated exhaustively for all SNP pairs. Instead, P-values are calculated only for pairs with the highest I\textsubscript{OR}. Conversely, Wu et al. used an approximate, chi-square distribution-based, P-value which can be applied to each investigated pair of the search.

Our study on WTCCC psoriasis data suggests that the computational effort for exhaustive testing is currently not just a luxury. The popular class of conditional analyses focuses only on possible interactions of previously implicated SNPs – often the only option to perform large-scale analysis in reasonable time. When comparing conditional and exhaustive FORCE analyses, we found that the conditional approach only detects interactions of vastly weaker statistical significance.

### Table 9: Power and family-wise error rate (FWER) of FORCE, MBMDR, Plink Epistasis and Plink FastEpistasis on 6 epistasis models with or without noise

| Model | FORCE | MBMDR | Plink Epistasis | FastEpistasis |
|-------|-------|-------|-----------------|---------------|
| No noise |   |       |                 |               |
| Power\textsuperscript{a} | 1 | 0.97 | 0.96 | 1 | 0.93 | 0.99 |
| FWER\textsuperscript{b} | 0.05 | 0.02 | 0.02 | 0.06 | 0.05 | 0.04 |
| GE | FORCE |       |                 |               |
| Power | 0.99 | 0.99 | 0.97 | 1 | 0.95 | 0.99 |
| FWER | 0.03 | 0.03 | 0.08 | 0.03 | 0.03 | 0.04 |
| MBMDR |       |       |                 |               |
| Power | 1 | 1 | 1 | 0.99 | 0.85 | 1 |
| FWER | 0 | 0 | 0.01 | 0.02 | 0 | 0 |
| Plink Epistasis |       |       |                 |               |
| Power | 0 | 1 | 0.28 | 0.99 | 0.87 | 0.99 |
| FWER | 0 | 0 | 0.07 | 0.02 | 0 | 0 |
| FastEpistasis |       |       |                 |               |
| Power | 0.01 | 1 | 0.31 | 1 | 0.74 | 0.92 |
| FWER | 0.04 | 0 | 0.09 | 0.05 | 0 | 0 |
| MS | FORCE |       |                 |               |
| Power | 0.99 | 0.96 | 0.95 | 1 | 0.93 | 0.99 |
| FWER | 0.07 | 0.02 | 0.02 | 0.03 | 0.03 | 0.06 |
| MBMDR |       |       |                 |               |
| Power | 1 | 1 | 1 | 0.99 | 0.8 | 0.94 |
| FWER | 0 | 0 | 0 | 0 | 0 | 0 |
| Plink Epistasis |       |       |                 |               |
| Power | 0 | 1 | 0.26 | 1 | 0.8 | 0.97 |
| FWER | 0.03 | 0.02 | 0.08 | 0.02 | 0 | 0 |
| FastEpistasis |       |       |                 |               |
| Power | 0 | 1 | 0.29 | 1 | 0.67 | 0.91 |
| FWER | 0.06 | 0.01 | 0.1 | 0.04 | 0 | 0 |
| PC | FORCE |       |                 |               |
| Power | 0.08 | 0.55 | 0.13 | 0.33 | 0.23 | 0.37 |
| FWER | 0.03 | 0.05 | 0.05 | 0.06 | 0.04 | 0.07 |
| MBMDR |       |       |                 |               |
| Power | 0.72 | 0.98 | 0.12 | 0.09 | 0.11 | 0.16 |
| FWER | 0 | 0 | 0 | 0 | 0 | 0 |
| Plink Epistasis |       |       |                 |               |
| Power | 0 | 0.95 | 0.01 | 0.21 | 0.19 | 0.28 |
| FWER | 0.04 | 0.03 | 0.05 | 0.06 | 0.01 | 0 |
| FastEpistasis |       |       |                 |               |
| Power | 0 | 0.99 | 0.01 | 0.23 | 0.07 | 0.21 |
| FWER | 0.07 | 0.03 | 0.04 | 0.06 | 0.01 | 0.01 |

\textsuperscript{a}In bold, power higher than 50%. \textsuperscript{b}In bold, FWER lower than 5%.

Grange et al. BMC Genetics (2015) 16:11 Page 9 of 13
|                | Model 1 |         | Model 2 |         | Model 3 |         | Model 4 |         | Model 5 |         | Model 6 |         |
|----------------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|
|                | Both    | First   | Either  | Both    | First   | Either  | Both    | First   | Either  | Both    | First   | Either  |
| FORCE          | Power<sup>a</sup> | 0.01  | 0.08  | 0.14  | 0.38  | 0.6  | 0.82  | 0.03  | 0.19  | 0.34  | 0.16  | 0.39  | 0.62  |
|                | FWER<sup>b</sup> | 0.02  | 0.07  | 0.07  | 0.04  | 0.02  | 0.02  | 0.02  | 0.02  | 0.02  | 0.02  | 0.02  |
| MBMDR          | Power   | 0.75  | 0.86  | 0.97  | 0.96  | 0.98 | 1    | 0    | 0.09  | 0.17  | 0.01  | 0.07  | 0.13  |
|                | FWER    | 0     | 0     | 0     | 0     | 0    | 0    | 0    | 0     | 0     | 0     | 0     |
| Plink Epistasis| Power   | 0     | 0     | 0     | 0.91  | 0.96 | 1    | 0    | 0.02  | 0.03  | 0.05  | 0.29  | 0.52  |
|                | FWER    | 0.01  | 0.02  | 0.01  | 0.01  | 0.05  | 0    | 0    | 0     | 0     | 0.07  | 0.14  | 0.25  |
| FastEpistasis  | Power   | 0     | 0     | 0     | 0.96  | 0.98 | 1    | 0    | 0.02  | 0.04  | 0.08  | 0.31  | 0.54  |
|                | FWER    | 0.02  | 0.03  | 0.01  | 0.07  | 0.01  | 0    | 0    | 0     | 0     | 0.07  | 0.14  | 0.34  |

<sup>a</sup>In bold, power higher than 50%.<sup>b</sup>In bold, FWER lower than 5%.
Table 11 Power of FORCE detection method, impact of various sources of noise and combinations of them for the 6 epistatic models

| Type of noise         | Model 1 | Model 2 | Model 3 | Model 4 | Model 5 | Model 6 |
|-----------------------|---------|---------|---------|---------|---------|---------|
| No noise              | 1.000   | 0.977   | 0.966   | 1.000   | 0.933   | 0.998   |
| Genotype errors (GE)  | 0.990   | 0.990   | 0.977   | 1.000   | 0.950   | 0.998   |
| Phenocopy (PC)        | 0.081   | 0.549   | 0.128   | 0.323   | 0.228   | 0.367   |
| Missing data (MS)     | 0.990   | 0.966   | 0.952   | 1.000   | 0.933   | 0.998   |
| GE + PC               | 0.050   | 0.618   | 0.183   | 0.300   | 0.213   | 0.352   |
| GE + MS               | 0.950   | 0.982   | 0.966   | 1.000   | 0.913   | 0.998   |
| PC + MS               | 0.060   | 0.522   | 0.201   | 0.301   | 0.201   | 0.26   |
| GE + PC + MS          | 0.090   | 0.549   | 0.201   | 0.460   | 0.120   | 0.352   |

GE: Genotyping errors – GH: Genetic heterogeneity – MS: Missing data – PC: Phenocopy. In case of GH, power is calculated in 3 different ways as the proportion of datasets in which both, the first or either of the interacting pairs are detected. *In bold, power higher than 50%.

Table 12 Family-wise error rate (FWER) of FORCE for the 6 epistatic models and 16 noise conditions tested

| Family-wise error rate | Model 1 | Model 2 | Model 3 | Model 4 | Model 5 | Model 6 |
|------------------------|---------|---------|---------|---------|---------|---------|
| No noise               | 0.050   | 0.020   | 0.020   | 0.060   | 0.050   | 0.040   |
| Genotype errors (GE)   | 0.030   | 0.030   | 0.080   | 0.030   | 0.030   | 0.040   |
| Genetic heterogeneity (GH) | 0.020   | 0.070   | 0.070   | 0.040   | 0.020   | 0.020   |
| Phenocopy (PC)         | 0.030   | 0.050   | 0.050   | 0.060   | 0.040   | 0.070   |
| Missing data (MS)      | 0.070   | 0.020   | 0.020   | 0.030   | 0.030   | 0.060   |
| GE + GH                | 0.050   | 0.070   | 0.040   | 0.030   | 0.010   | 0.070   |
| GE + PC                | 0.050   | 0.050   | 0.010   | 0.020   | 0.030   | 0.020   |
| GE + MS                | 0.020   | 0.010   | 0.060   | 0.040   | 0.040   | 0.070   |
| GH + PC                | 0.050   | 0.050   | 0.030   | 0.030   | 0.030   | 0.030   |
| GH + MS                | 0.040   | 0.070   | 0.050   | 0.030   | 0.030   | 0.010   |
| PC + MS                | 0.030   | 0.030   | 0.060   | 0.060   | 0.040   | 0.030   |
| GE + GH + PC           | 0.070   | 0.030   | 0.060   | 0.030   | 0.020   | 0.020   |
| GE + GH + MS           | 0.070   | 0.050   | 0.040   | 0.010   | 0.050   | 0.050   |
| GH + PC + MS           | 0.040   | 0.020   | 0.060   | 0.060   | 0.050   | 0.050   |
| GE + GH + PC + MS      | 0.020   | 0.070   | 0.060   | 0.040   | 0.050   | 0.060   |

GE: Genotyping errors – GH: Genetic heterogeneity – MS: Missing data – PC: Phenocopy. *In bold, FWER > 0.05.
Our systematic study on small simulated datasets indicates that FORCE not only "goes farther" than existing approaches because of its exhaustive search, but also detects fundamentally different types of interactions, in particular in the biologically more relevant models 3–6. In two out of six models of epistatic interaction described by Ritchie [28], and across the different sources of noise in the data, FORCE consistently displayed a good power of detection compared to other approaches. Interestingly, each of the four approaches is always less efficient than another for at least one model associated with one type of noise.

Finally, by applying FORCE to WTCCC psoriasis data, we were able to detect statistical interactions between SNPs in the HLA region, even after the exclusion of all SNPs with main effects. To our knowledge this constitutes the first demonstration that the genetic structure of the HLA region cannot be understood by the analysis of main effects alone and that more than one interacting locus exists in that region.

Conclusions
Together, the different elements of our study suggest that FORCE represents a valuable new addition to the arsenal of genome-wide epistasis detection approaches for case–control studies. As with other approaches, the additionally detected interactions are a priori of a statistical nature, and require detailed analysis and follow-up.

Beyond this, our study has provided an example for the need for exhaustive epistasis analysis. In the future, exhaustive analysis will be facilitated by the ever-increasing computational power available to biological research. On one hand, this may enable the exhaustive calculation of FORCE P-values, which can be expected to lead to a potentially much enlarged set of statistically significant interactions. On the other hand, more computational power, as well as algorithmic improvements, may also render exhaustive analysis under those models of interactions feasible for which running times are prohibitive today. Finally, we believe that these improvements are necessary for the integration of different types of interactions and other types of large-scale data, which may ultimately be key to understanding the genetic basis of complex diseases.

Additional file

Additional file 1: Table S1. Epistasis analysis among GWAS hits: all 153 pairs of the conditional search, Table S2. FORCE Exhaustive search top 100 hits on psoriasis data.

Abbreviations
OR: Odds ratio; IOR: Interaction odds ratio; HLA: Human leukocyte antigen; GWAS: Genome wide association study; SNP: Single nucleotide polymorphism; WTCCC: The welcome trust case control consortium; GE: Genotyping error; MS: Missing data; GH: Genetic heterogeneity; PC: Phenocopy; FWER: Family-wise error rate.

Competing interests
The authors declare that they have no competing interests.

Authors’ contributions
LG participated in the study design, carried out the analyses, interpreted results and drafted the manuscript. JFB participated in the study design, interpreted result and drafted the manuscript. IN contributed to the program, RP participated in interpreting results and was involved in drafting and revising the manuscript, KVS conceived and implemented the FORCE method, analyzed and interpreted data, and was involved in drafting and revising the manuscript, AS conceived of the study, participated in its design and coordination and helped to draft the manuscript. All authors read and approved the final manuscript.

Acknowledgements
We would like to thank the WTCCC for permission to use the psoriasis genome-wide data set. This article is linked to a project funded by “ANR-11-BSV1-027-01”. LG is supported by Ministère de la Recherche, IN is supported by the French Government’s Investissement d’Avenir program, Laboratoire d’Excellence “Integrative Biology of Emerging Infectious Diseases” (grant n°ANR-10-LABX-62-IBED).

Author details
1Department of Genomes and Genetics, Institut Pasteur, Functional Genetics of Infectious Diseases Unit, Paris 75015, France. 2CNRS URA3012, Paris 75015, France. 3Université Paris Diderot, Paris 75013, France. 4Department of Genomes and Genetics, Institut Pasteur, Systems Biology Lab, Paris 75015, France. 5Université Paris-Descartes, Sorbonne Paris Cité, Paris, France. 6Systems and Modeling Unit, Montefiore institute, University of Liège, Liège, Belgium. 7Bioinformatics and Modeling, GiGA-R, University of Liège, Liège, Belgium.

Received: 29 September 2014 Accepted: 23 January 2015

Published online: 06 February 2015

References
1. Visscher PM, Brown MA, McCarthy MI, Yang J. Five years of GWAS discovery. Ann J Hum Genet. 2012;90(1):7–24.
2. Welter D, MacArthur J, Morales J, Burdett T, Hall P, Junkins H, et al. The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. Nucleic Acids Res. 2014;42(Database issue):D1001–1006.
3. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, et al. Finding the missing heritability of complex diseases. Nature. 2009;461(7265):747–53.
4. Mackay TF. Epistasis and quantitative traits: using model organisms to study gene-gene interactions. Nat Rev Genet. 2014;15(1):22–33.
5. Steen KV. Travelling the world of gene-gene interactions. Brief Bioinform. 2012;13(1):11–19.
6. Wei WH, Hemani G, Haley CS. Detecting epistasis in human complex traits. Nat Rev Genet. 2014;15(1):72–22.
7. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007;81(3):559–75.
8. Schupbach T, Xenarios I, Bergmann S, Kapur K. FastEpistasis: a high performance computing solution for quantitative trait epistasis. Bioinformatics. 2010;26(11):1468–9.
9. Hemani G, Theodoridis A, Wei W, Haley C. Epistasis scans parallelized on consumer level graphics cards. Bioinformatics. 2011;27(11):1462–5.
10. Lange K, Papp JC, Sinheimer JS, Srip racha R, Zhou H, Sobel EM. Mendel: the Swiss army knife of genetic analysis programs. Bioinformatics. 2013;29(12):1568–70.
11. Han X, Yang C, Yang Q, Xue H, Fan X, Tang NL, et al. BOOST: a fast approach to detecting gene-gene interactions in genome-wide case–control studies. Am J Hum Genet. 2010;87(3):325–40.
12. Kam-Thong T, Czamara D, Tsuda K, Borgwardt K, Lewis CM, Erhardt-Lehmann A, et al. EPiBLASTER-fast exhaustive two-locus epistasis detection strategy using graphical processing units. Eur J Hum Genet. 2011;19(4):465–71.

13. Prabh U, Pe'er I. Ultrafast genome-wide scan for SNP-SNP interactions in common complex disease. Genome Res. 2012;22(11):2230–40.

14. Yi N, Hakamim V, Pasche B. Bayesian analysis of genetic interactions in case–control studies, with application to adiponectin genes and colorectal cancer risk. Ann Hum Genet. 2011;75(1):90–104.

15. Zhang Y, Liu JS. Bayesian inference of epistatic interactions in case–control studies. Nat Genet. 2007;39(9):1167–73.

16. Luz Calle ML, Urea V, Van Steen K. MB-MDR: Model-Based Multifactor Dimensionality Reduction for detecting interactions in high-dimensional genomic data. Vic; Spain: Universitat de Vic; 2008.

17. Schwarz DF, Koniig IR, Ziegler A. On safari to Random Jungle: a fast implementation of Random Forests for high-dimensional data. Bioinformatics. 2010;26(14):1752–8.

18. Ueki M, Tamya G. Ultrahigh-dimensional variable selection method for whole-genome gene–gene interaction analysis. BMC Bioinformatics. 2012;13:72.

19. Xie M, Li J, Jiang T. Detecting genome-wide epistases based on the clustering of relatively frequent items. Bioinformatics. 2012;28(1):5–12.

20. Piegorsch WW, Weinberg CR, Taylor JA. Non-hierarchical logistic models and case-only designs for assessing susceptibility in population-based case–control studies. Stat Med. 1994;13(2):153–62.

21. Thomas D. Gene–environment-wide association studies: emerging approaches. Nat Rev Genet. 2010;11(4):259–72.

22. VanderWeele TJ, Hernandez-Diaz S, Hernan MA. Case-only gene–environment interaction studies: when does association imply mechanistic interaction? Genet Epidemiol. 2010;34(4):327–34.

23. Knight J, Spain SL, Capon F, Hayday A, Nestle FO, Clop A, et al. Conditional analysis identifies three novel major histocompatibility complex loci associated with psoriasis. Hum Mol Genet. 2012;21(23):5185–92.

24. Strange A, Capon F, Spencer CC, Knight J, Wiisle ME, Allen KH, et al. A genome-wide association study identifies new psoriasis susceptibility loci and an interaction between HLA-C and ERAP1. Nat Genet. 2010;42(11):985–90.

25. Chandran V. The genetics of psoriasis and psoriatic arthritis. Clin Rev Allergy Immun. 2013;44(2):149–56.

26. Oka A, Mabuchi T, Ozawa A, Inoko H. Current understanding of human genetics and genetic analysis of psoriasis. J Dermatol. 2012;39(3):231–41.

27. Ziegler A, Konig IR, Thompson JR. Biostatistical aspects of genome-wide association studies. Born J. 2008;50(1):18–28.

28. Ritchie MD, Hahn LW, Moore JH. Power of multifactor dimensionality reduction for detecting gene–gene interactions in the presence of genotyping error, missing data, phenocopy, and genetic heterogeneity. Genet Epidemiol. 2003;24(2):150–7.

29. Moore JH, Hahn LW, Ritchie MD. Application of Genetic Algorithms to the Discovery of Complex Models for Simulation Studies in Human Genetics. Proc Genet Evol Comput Conf. 2002;2002:1150–5.

30. Kernighan BW, Ritchie DM. The C programming language. Second edition., ed. Prentice Hall; 1988.

31. Cattaert T, Calle ML, Dudek SM, Mahachie John JM, Van Lishout F, Urea V, et al. Model-based multifactor dimensionality reduction for detecting epistasis in case–control data in the presence of noise. Ann Hum Genet. 2011;75(1):78–89.

32. Cordell HJ. Detecting gene–gene interactions that underlie human diseases. Nat Rev Genet. 2009;10(6):392–404.

33. Goudsby B, Ravlinson D, Wang Q, Shi F, Ferra H, Campbell RM, et al. GWIS-model-free, fast and exhaustive search for epistatic interactions in case–control GWAS. BMC Genomics. 2013;14 Suppl 3:S10.

34. Riveira-Munoz E, He SM, Escaramis G, Stuart PE, Huffmeier U, Lee C, et al. Meta-analysis confirms the LCE3C-LCE3B deletion as a risk factor for psoriasis in several ethnic groups and finds interaction with HLA-Cw6. J Invest Dermatol. 2011;131(5):1105–9.

35. Veal CD, Clough RL, Barber RC, Mason S, Tillman D, Ferry B, et al. Identification of a novel psoriasis susceptibility locus at 1p and evidence of epistasis between PSORS1 and candidate loci. J Med Genet. 2001;38(1):17–13.

36. Ueki M, Cordell HJ. Improved statistics for genome-wide interaction analysis. PLoS Genet. 2012;8(4):e1002625.

37. Mahachie John JM, Van Lishout F, Van Steen K. Model-Based Multifactor Dimensionality Reduction to detect epistasis for quantitative traits in the presence of error-free and noisy data. Eur J Hum Genet. 2011;19(6):696–703.

38. Wu X, Dong H, Luo L, Zhu Y, Peng G, Reveille JD, et al. A novel statistic for genome-wide interaction analysis. PLoS Genet. 2010;6(9):e1001131.