

SCIENTIFIC REPORTS



OPEN

Discovering Genetic Factors for psoriasis through exhaustively searching for significant second order SNP-SNP interactions

Kwan-Yeung Lee¹, Kwong-Sak Leung¹, Nelson L. S. Tang^{1,2} & Man-Hon Wong¹

In this paper, we aim at discovering genetic factors of psoriasis through searching for statistically significant SNP-SNP interactions exhaustively from two real psoriasis genome-wide association study datasets (phs000019.v1.p1 and phs000982.v1.p1) downloaded from the database of Genotypes and Phenotypes. To deal with the enormous search space, our search algorithm is accelerated with eight biological plausible interaction patterns and a pre-computed look-up table. After our search, we have discovered several SNPs having a stronger association to psoriasis when they are in combination with another SNP and these combinations may be non-linear interactions. Among the top 20 SNP-SNP interactions being found in terms of pairwise p-value and improvement metric value, we have discovered 27 novel potential psoriasis-associated SNPs where most of them are reported to be eQTLs of a number of known psoriasis-associated genes. On the other hand, we have inferred a gene network after selecting the top 10000 SNP-SNP interactions in terms of improvement metric value and we have discovered a novel long distance interaction between *XXbac-BPG154L12.4* and *RNU6-283P* which is not a long distance haplotype and may be a new discovery. Finally, our experiments with the synthetic datasets have shown that our pre-computed look-up table technique can significantly speed up the search process.

Psoriasis is a common polygenic chronic inflammatory skin disease affecting up to 3% of population¹. Recently, genome-wide association study (GWAS) provided the first opportunity to have a comprehensive screen for susceptibility genes and up to 50 loci had been reported^{1,2}. The discovery of novel susceptibility genes by GWAS included genes coding for key cytokines involved in Th17 activation (like *IL-12B*, *IL23A* and *IL23R*) and *NF-κB* pathway also contributed to the aetiology (*TNFAIP3* and *TNIP1*)³. In addition, macrophage and dendritic cells are also involved. *HLA-C* locus had been a known susceptibility gene before the GWAS era⁴. Exaggerated expression of keratinocyte antigens (like *LCE3D*) was also identified as susceptibility genes⁵.

Although many predisposition genes for psoriasis have been identified, individually they accounted for very small effect size, for example their odds ratios were typically less than 1.2⁶. While heritability of psoriasis had been estimated to be as high as 60–90%, the genetic risk of all variants added up could only account for one-fourth of susceptibility due to genetics. This phenomenon is called missing heritability⁷. One possible source of unaccounted risk is interactions (including gene-gene and gene-environment interaction), which is the risk of certain genotype may be altered to a large extent in the presence of another risk factor, which is also known as non-additive effects^{6,8,9}. Various algorithms have been proposed to detect the interactions between SNPs or genes in GWAS data^{10,11}. They all faced the difficulties of large search space, exponential increase in SNP combination with increasing level of interaction, and limited statistical basis of the proposed methods. On the other hand, examples of epistasis were found in model organisms and human diseases^{9,12,13}. We had proposed a biological framework of gene-gene interaction and suggested that typical samples size (thousands of cases and controls) should have sufficient power to detect such gene-gene interaction with simulation data¹⁴. There were also some suggestions of epistasis in psoriasis but replication in subsequent studies were lacking^{3,8}.

¹Department of Computer Science and Engineering, the Chinese University of Hong Kong, Hong Kong, China.

²Department of Chemical Pathology, the Chinese University of Hong Kong, Hong Kong, China. Correspondence and requests for materials should be addressed to K.-Y.L. (email: kylee@cse.cuhk.edu.hk) or N.L.S.T. (email: nelsontang@cuhk.edu.hk)

	SNP ₁	SNP ₂	...	SNP _m	Status
Sample ₁	1	1	...	1	T
Sample ₂	2	1	...	2	F
⋮	⋮	⋮	⋮	⋮	⋮
Sample _{n-1}	3	2	...	2	F
Sample _n	3	2	...	2	F

Table 1. This table shows an example of a GWAS dataset.

Original Genotype	Encode Value
Missing Data	0
Major Allele, Major Allele	1
Major Allele, Minor Allele	2
Minor Allele, Minor Allele	3

Table 2. This table shows the encoding scheme for SNP genotype.

Genotype		Count
Major Major	Case	400
	Ctrl	373
Major Minor	Case	415
	Ctrl	260
Minor Minor	Case	97
	Ctrl	38

Figure 1. In this figure, the table shows the distribution of case and control under different genotypes of SNP rs3130048 in dataset phs000019.v1.p1. The 1 d.f. chi-square p -value and odds ratio of rs3130048 are 3.96×10^{-6} and 1.6021 respectively.

Biological pathways are regulated by the interactions among bio-molecules constructed according to the genetic instructions stored in deoxyribonucleic acid (DNA). A single nucleotide polymorphism (SNP) is a variation at a specific DNA position among a population of organisms which may affect the structure of these bio-molecules. In a genome-wide association study (GWAS), DNA sequences of a large population of patient samples (cases) of a particular genetic disease and healthy samples (controls) are collected and researchers can discover disease-associated SNPs through comparing the DNA sequences between cases and controls¹⁵. These DNA sequences can be arranged into a matrix A where each column (except the last column) corresponds to a SNP and each row corresponds to a sample as shown in Table 1. Each entry $A_{i,j}$ corresponds to the genotype of i^{th} sample at j^{th} SNP under the encoding scheme shown in Table 2. Each sample is either labelled as ‘case’ or ‘control’ through the value of the last column of the matrix. A traditional approach for finding statistically significant SNPs which has been widely adapted in many GWAS researches is to perform statistical tests after building a contingency table for each column of matrix A .

As previously discussed, the cause of many genetic diseases can be better explained through certain combinations of SNPs (i.e. SNP-SNP interactions) rather than a number of independent SNPs alone^{16,17}. Although many SNPs are weakly associated to the genetic diseases when they are analysed independently, some of them may show a stronger association only when they are analysed in combination with other SNPs. A typical example of this phenomenon is shown in Figs 1–3 and Supplementary Fig. S1. In Figs 1 and 2, rs3132486 and rs3130048 from dataset phs000019.v1.p1 are both weakly associated to psoriasis with a 1 degree of freedom (d.f.) chi-square p -value less significant than 1×10^{-6} and an odds ratio smaller than 1.7 if they are analysed independently. Meanwhile in Fig. 3, the combination of rs3132486 and rs3130048 are significantly associated to psoriasis with

Genotype		Count
Major	Case	300
Major	Ctrl	163
Major	Case	452
Minor	Ctrl	331
Minor	Case	163
Minor	Ctrl	180

Figure 2. In this figure, the table shows the distribution of case and control under different genotypes of SNP rs3132486 in dataset phs000019.v1.p1. The 1 d.f. chi-square p -value and odds ratio of rs3132486 are 2.06×10^{-5} and 1.6810 respectively.

			rs3130048		
			Major Major	Major Minor	Minor Minor
rs3132486	Major	Case	130	144	25
	Major	Ctrl	112	46	3
	Major	Case	193	206	53
	Minor	Ctrl	189	122	17
	Minor	Case	76	64	19
	Minor	Ctrl	70	92	18

Figure 3. In this figure, the table shows the distribution of case and control under genotype of the combination of SNP rs3132486 and rs3130048 in dataset phs000019.v1.p1. The 1 d.f. chi-square p -value and odds ratio of the combination of rs3132486 and rs3130048 are 2.52×10^{-14} and 2.2719 respectively.

a 1 d.f. chi-square p -value of 2.52×10^{-14} and an odds ratio of 2.2719. Therefore, rs3132486 and rs3130048 may have a potential non-linear interaction associated to psoriasis. Similar but much weaker phenomena have also been observed between SNPs among two known psoriasis associated gene-gene interactions (*HLA-C*, *IL12B*)⁸ and (*HLA-C*, *TNFAIP3*)³ and are shown in Supplementary Figs S2 and S3 respectively.

In previous GWAS researches, filtering out a large proportion of statistically insignificant SNPs or performing search with greedy and stochastic search algorithms are two popular techniques for shrinking the search space of SNP-SNP interactions^{18,19}. Any researcher who performs his/her analysis with a search space confined to statistically significant single SNPs only will prematurely filtered out some disease-associated SNPs which are statistically insignificant on their own similar to the ones shown in Figs 1 and 2. Meanwhile, greedy and stochastic search selectively explore the search space under the guidance of heuristic functions thus they may also fail to detect some disease-associated but statistically insignificant SNPs as well. However, performing exhaustive search for interactions among k different SNPs (i.e. k^{th} order SNP-SNP interaction) would take a time complexity of $O(m^kn)$ where m is the number of SNPs and n is the number of samples. Given the fact that there are around 38 million SNPs with one million tag SNPs across the whole human genome²⁰ and a GWAS dataset often contains thousands of samples and hundreds of thousands of SNPs, this problem is difficult to be solved without any advanced computational technique. Driven by the continuous growth of computational power, it is now becoming possible to perform exhaustive search for 2^{nd} order SNP-SNP interactions in a GWAS dataset within a reasonable amount of time^{21–28}. However, existing exhaustive search algorithms are not driven by any biological knowledge and they evaluate the disease association of each SNP combination solely based on heuristic or statistical parameters. Therefore, they may found statistical significant SNP-SNP interactions which may not be biologically interpretable.

We have developed an exhaustive search algorithm driven by eight biological plausible SNP-SNP interaction and applied it on two psoriasis GWAS datasets (phs000019.v1.p1 and phs000982.v1.p1). We have first discovered

a number of statistically significant SNP-SNP interactions which may have a stronger association to psoriasis than their component SNPs measured independently similar to the ones shown in Fig. 3 where the SNPs involved in these interactions are far less significant when they are considered individually. After that, we have discovered 27 novel potential psoriasis-associated SNPs among the top 20 statistically significant SNP-SNP interactions in terms of p -value and improvement metric value. Most of these novel potential psoriasis-associated SNPs are reported to be Expression quantitative trait loci (eQTLs) of known psoriasis-associated genes like *HLA-B* and *HLA-C* in GTEx Portal²⁹. After mapping the nearest gene to each SNP involved in top 10000 SNP-SNP interactions in terms of improvement metric value from both dataset, we have constructed a disease-associated gene network. In our network, almost half of the gene-gene interactions inferred are consistent with existing literature. Meanwhile, some of the remaining gene-gene interactions are potentially due to the long-distance haplotype interactions presence in the HLA region of Chromosome 6. Furthermore, we have discovered an interaction between 2 SNPs located in gene *XXbac-BPG154L12.4* and *RNU6-283P* of the HLA loci, which doesn't correspond to any short-distance or even long-distance haplotype interactions and hasn't been reported in existing literature as well. Therefore, the interaction between *XXbac-BPG154L12.4* and *RNU6-283P* is a potential new discovery. In addition, we have shown that counting contingency table through a pre-computed look-up table is effective in speeding up the process of exhaustive search.

Results

GWAS datasets. We have downloaded two psoriasis GWAS datasets namely phs000019.v1.p1 and phs000982.v1.p1 from the database of Genotypes and Phenotypes (dbGaP). Data pre-processing has been performed on these two datasets to remove low quality SNPs and samples with Plink³⁰. The parameters of data cleansing can be found as Supplementary Tables S1 and S2 which followed the common recommendations from NCBI³¹.

After data cleansing, there are 352945 SNPs and 1593 samples (cases: 917, controls: 676) in dataset phs000019.v1.p1 and there are 790527 SNPs and 2689 samples (cases: 1363, controls: 1326) in dataset phs000982.v1.p1. The genotypes of every SNP in these two datasets are encoded as 0, 1, 2, 3 according to the encoding scheme shown in Table 2.

Measurement metric for ranking and filtering SNP-SNP interaction. In this paper, SNP-SNP interactions are prioritised with the following two measurement metrics. The pairwise p -value of a 2nd order SNP combination under a particular genotype interaction pattern is referring to the 1 d.f. chi-square p -value of its 2×2 contingency table founded from a real psoriasis dataset with the procedure illustrated in the Methods section. Additionally, we have defined another measurement metric called improvement metric value to prioritise statistically significant SNP-SNP interactions where its two components SNPs are far less associated to psoriasis when they are analysed independently. It compares the pairwise p -value of a 2nd order SNP combination under a particular genotype interaction pattern against the standalone 1 d.f. chi-square p -value of both of its component SNPs and can be calculated by equation 1. If a 2nd order SNP combination has a higher improvement metric value, its component SNPs are far more statistically significant when they are considered as a SNP combination rather than separately considered as two independent SNPs.

$$\text{Improvement_Metric_Value} = \text{MIN} \left(\frac{\text{SNP1 } p\text{-value}}{2^{\text{nd}} \text{ Order SNP } p\text{-value}}, \frac{\text{SNP2 } p\text{-value}}{2^{\text{nd}} \text{ Order SNP } p\text{-value}} \right) \quad (1)$$

Exhaustive search on psoriasis datasets phs000019.v1.p1 and phs000982.v1.p1. We have performed exhaustive search on the two cleansed psoriasis GWAS datasets to discover biologically plausible and statistically significant 2nd order SNP-SNP interactions. There are 62284910040 and 312466073601 unique 2nd order SNP combinations in datasets phs000019.v1.p1 and phs000982.v1.p1 respectively. Each SNP combination is subjected to eight 1 d.f. chi-square statistic tests corresponding to eight genotype interaction patterns and eight pairwise p -value is thus calculated. Therefore, there are 498279280320 and 2499728588808 statistical tests performed on phs000019.v1.p1 and phs000982.v1.p1 datasets respectively. Among these tests, there are 3058119 and 59810682 statistically significant pairwise SNP-SNP interactions found in datasets phs000019.v1.p1 and phs000982.v1.p1 which have a pairwise p -value smaller than 1×10^{-13} and 1×10^{-14} respectively. We have sorted these statistically significant SNP-SNP interactions by their pairwise p -value and improvement metric value separately for further analysis.

Analysis on top 20 most statistical significant SNP-SNP interaction. After sorting the interactions found in datasets phs000019.v1.p1 and phs000982.v1.p1 by their pairwise p -value, we have selected the top 20 statistical significant SNP-SNP interactions in terms of pairwise p -value which are listed in Supplementary Tables S3 and S4 respectively.

Among the interactions shown in Supplementary Table S3, there are six SNPs which are already reported to be associated to psoriasis in existing literature: rs12191877^{3,32-35}, rs1265078^{32,33,36}, rs2894207³², rs3130467³², rs3130517³² and rs3130573³³. According to GTEx Portal²⁹, there are nine SNPs acting as eQTLs of the following four known psoriasis associated genes, *HLA-C*^{37,38}, *HCP5*^{39,40}, *PSORS1C1*^{41,42} and *MICB*⁴³. First, rs2244027 and rs2894176 are found to be eQTLs of *HLA-C*. After that, rs2516417, rs2516510, rs2523708 and rs2844502 are found to be eQTLs of *HCP5*. Then, rs9262492 and rs9262498 are found to be eQTLs of *PSORS1C1*. Finally, rs2534666 is found to be eQTLs of *MICB*. Therefore, these nine SNPs have a high potential to be associated to psoriasis and may be new discoveries.

On the other hand among the interactions we have selected in dataset phs000982.v1.p1, there are 3 psoriasis associated SNPs which are already reported in existing literature: rs13203895⁴⁴, rs10484554^{39,45–47} and rs17728338^{47–50}. Meanwhile, rs4349859 and rs4418214 have already been found to be strongly associated to *HLA-b27* allele⁵¹ and *HIV* infection⁵² respectively in other studies. According to GTEx Portal²⁹, rs45533135 is an eQTLs of *MICA*. Since *HIV*⁵³, *HLA-B*³⁷ and *MICA*^{54–56} are strongly associated to psoriasis, these three SNPs have a high potential to be associated to psoriasis and may be new discoveries.

Analysis on SNP-SNP interactions with the top 20 improvement metric value. After sorting the statistically significant SNP-SNP interactions found in both datasets by their improvement metric value, we have selected top 20 statistically significant SNP-SNP interactions in terms of improvement metric value and are listed in Supplementary Tables S5 and S6. We can observe that these SNP-SNP interactions have a pairwise *p*-value much smaller than the standalone *p*-value of their component SNPs. Therefore, these interactions may be non-linear and further verification through wet-lab experiments should be performed in the future.

In Supplementary Table S5, there are nine SNPs which are present in Supplementary Table S3. Meanwhile, there are six SNPs rs9380237, rs7756521, rs2853950, rs2844645, rs7773175 and rs8365 which are not found in Supplementary Table S3. Among these six SNPs, rs7773175³², rs9380237³² and rs2853950⁵⁷ are literature reported psoriasis associated SNPs. According to GTEx Portal²⁹, there are three SNPs which are eQTLs of the following two known psoriasis associated genes, *PSORS1C1*^{41,42} and *HLA-DQB1*^{58,59}. First, rs7756521 and rs2844645 are found to be eQTLs of *PSORS1C1*. After that, rs8365 is found to be an eQTL of *HLA-DQB1*. The association between these three SNPs and psoriasis may be new discoveries.

Meanwhile there are no common SNPs between Supplementary Tables S6 and S4. Among the SNPs found in Supplementary Table S6, we observed that there is a SNP rs1576 which is reported to be associated to psoriasis in existing literature^{60,61}. Meanwhile, SNPs rs1265112 and rs746647 are reported to be in complete linkage disequilibrium ($r^2 = 1.00$) with the SNP rs1576 in an existing literature⁶². Therefore, SNPs rs1265112 and rs746647 can both be considered as a proxy SNP of a literature-reported psoriasis associated SNP and they are not new discoveries. According to GTEx Portal²⁹, there are ten SNPs which are eQTLs of the following two known psoriasis associated genes, *HLA-C*^{37,38} and *MICA*^{54–56}. Four of them are found to be reported as eQTLs of *HLA-C* namely rs2517985, rs1265079, rs1265114 and rs1265067. Meanwhile, six of them are found to be reported as eQTLs of *MICA* namely rs4358666, rs2395491, rs4624908, rs7754026, rs13194571 and rs7775117. The association between these ten eQTL SNPs and psoriasis may be new discoveries.

Further analysis on the component SNPs of the SNP-SNP interactions being discovered with CADD SNP annotation. By referring to the genome assembly GRCh37 published by Genome Reference Consortium, the genomic position of every SNP can be retrieved. After knowing the genomic position of every SNP, we have annotated every component SNP of the top 20 SNP-SNP interactions in terms of improvement metric value through CADD version 1.3⁶³. The genomic position of these SNPs and their nearest genes (if available) are shown in Supplementary Table S7. Among these SNPs, rs13191519 is located at an intron region of a RNA gene *XXbac-BPG248L24.13* (also known as *LOC105375015*). Since *LOC105375015* is reported to be associated to HIV and AIDS progression⁶⁴ and HIV is associated to psoriasis, rs13191519 may be associated to psoriasis as well. Meanwhile, rs3094205 is located at the upstream region of *CDSN* and *CDSN* is reported to be associated to psoriasis in an existing literature^{65,66}. Therefore these two SNPs are likely to be associated to psoriasis and may be a new discovery.

Predicting gene-gene interactions with CADD SNP annotation. By making an assumption that if SNP_i and SNP_j have an SNP-SNP interaction, $Gene_i$ and $Gene_j$ will have a gene-gene interaction where $Gene_i$ and $Gene_j$ are the closest genes to SNP_i and SNP_j respectively, we can predict gene-gene interactions based on the SNP-SNP interactions we have found after annotated with CADD.

Analysis on common gene-gene interactions predicted by statistically significant SNP-SNP interactions with top 10000th ranking in improvement metric value. We have selected top 10000 statistically significant SNP-SNP interactions in terms of improvement metric value from both datasets. Then we have predicted a number of gene-gene interactions based on these SNP-SNP interactions. As shown in Supplementary Fig. S4, we have predicted 3501 unique gene-gene interactions from dataset phs000019.v1.p1 and 430 unique gene-gene interactions from dataset phs000982.v1.p1. There are 62 common gene-gene interactions between these two datasets. After excluding 2 self-looping interactions found on gene *XXbac-BPG248L24.13* and *CCHCR1*, there are 60 interactions left and are listed in Supplementary Table S8.

Among these 60 common gene-gene interactions, there are 29 unique genes. Through referencing Ensembl release 75, the positions of these 29 genes under GRCh37 can be found (see Supplementary Table S9). Among these 29 genes, there are 14 genes which are reported by existing literature to be associated to psoriasis: *HLA-B*³⁷, *HLA-C*^{37,38}, *PSORS1C1* (which was previous named as *SEEK1*)^{41,42}, *CCHCR1*^{67,68}, *HCP5*^{39,40}, *CDSN*^{65,66}, *USP8P1*⁴⁴, *MICA*^{54,69}, *PSORS1C3*^{65,70}, *HCG27*⁷¹, *POU5F1*⁷², *WASF5*^{D44}, *MICB*⁴³, *C2*⁷³. On the other hand, these 60 common gene-gene interactions can be arranged into a network with a circular layout as shown in Fig. 4. Each gene is represented by a node and each common gene-gene interaction is represented by a black edge. If a gene is reported in an existing literature, its corresponding node will be highlighted in grey colour. If a predicted gene-gene interaction is reported as a direct interaction or an indirect interaction through an intermediate gene in one of the following seven biomolecule interaction databases: String⁷⁴, RAID⁷⁵, lncRNA2target⁷⁶, lncReg⁷⁷, InBio_Map⁷⁸, HPRD⁷⁹ and BioGRID⁸⁰, a new edge with a colour corresponding to the reporting database will be added to the network. Furthermore, the gene network can also be arranged with a linear layout (see Supplementary Fig. S5), where the genes are laid down according to their genomic position on Chromosome 6. The recombination

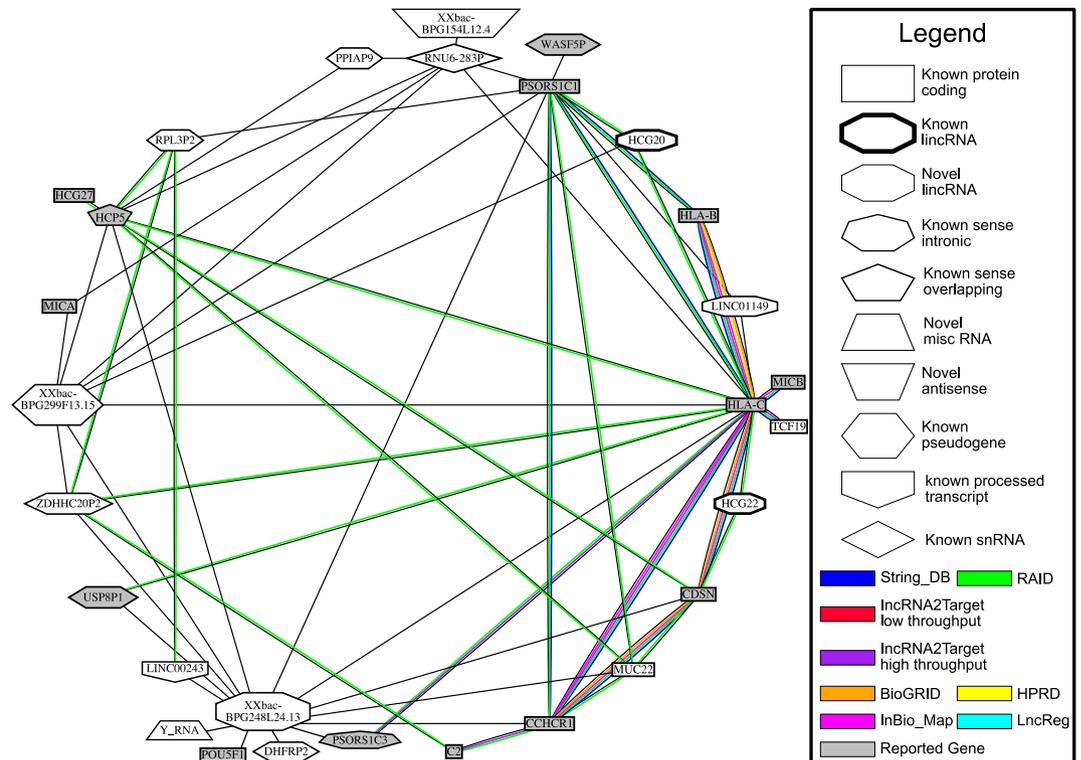


Figure 4. This figure shows a gene network constructed from the 60 common gene-gene interactions predicted from the top 10000 SNP-SNP interactions in terms of improvement metric value found from datasets phs000019.v1.p1 and phs000982.v1.p1 with a circular layout. Genes which are already reported to be associated to psoriasis in existing literatures are coloured in grey colour. Meanwhile, if a gene-gene interaction is supported by a database, a thickened edge with a specific colour will be added to the network.

rate and SNP pairs with significant linkage disequilibrium score ($r \geq 0.9$) along Chromosome 6 from position 30734602 to 32233615 are shown in Supplementary Fig. S5a,b respectively. The linkage disequilibrium score is calculated with the samples from phs000982.p1.v1 using Plink³⁰.

As seen in Fig. 4, 28 out of 60 common gene-gene interactions predicted can be verified by one or more than one existing databases as direct or indirect gene-gene interactions. Since almost half of the gene-gene interactions are supported by existing databases, our gene network is generally supported. Furthermore, there is a long non-coding RNA (lincRNA) gene *XXbac-BPG248L24.13* acting as a hub and is interacting with eight literature reported psoriasis-associated genes namely *PSORS1C3*, *CDSN*, *POU5F1*, *PSORS1C1*, *CCHCR1*, *HLA-C*, *USP8P1* and *HCP5*. Therefore, *XXbac-BPG248L24.13* may be a new discovery. However, since *XXbac-BPG248L24.13* and two other literature reported psoriasis-associated genes *HLA-C* and *USP8P1* are located in the same LD region as shown in Supplementary Fig. S5b. Therefore, *XXbac-BPG248L24.13* may be a false discovery and the SNPs being mapped to *XXbac-BPG248L24.13* by CADD may be proxy SNPs of the disease-associated SNPs located in *HLA-C* or *USP8P1*. Further verification through wet-lab experiments should be performed. Meanwhile, there is a pseudogene *XXbac-BPG299F13.15* which is located in the same LD region as *HLA-C* and is interacting with 4 other literature reported psoriasis-associated genes (*PSORS1C1*, *MICA*, *HCP5*, *HLA-C*). Similar to *XXbac-BPG248L24.13*, *XXbac-BPG299F13.15* may be a discovery and its effect can only be verified through wet-lab experiments. In Supplementary Fig. S5b, there is long-distance gene-gene interactions between *HCP5* and *MUC22*. Since there is a long-distance haplotype between *HCP5* and *MUC22*, the gene-gene interaction predicted has a high potential to be representing the effect of a long-distance haplotype. Similarly, *LINC00243* and *PSORS1C1* have a predicted gene-gene interaction which corresponds to a long-distance haplotype. Clearly, long-distance disease-associated haplotype can be found through our exhaustive search algorithm as well. Finally, the gene-gene interaction predicted between *XXbac-BPG154L12.4* and *RNU6-283P* may be a novel discovery. Although *RNU6-283P* and a known psoriasis-associated gene *HLA-B* are both located at the same LD region, *XXbac-BPG154L12.4* is not in any LD region which contains any known psoriasis associated gene. Therefore, *XXbac-BPG154L12.4* may be interacting with a proxy gene of a known psoriasis-associated gene *HLA-B*. Furthermore, there is a strong recombination site in between *XXbac-BPG154L12.4* and any other known psoriasis associated gene. Therefore, the interaction between *XXbac-BPG154L12.4* and *RNU6-283P* cannot be simply explained as a long-distance haplotype and it may be a new discovery.

Simulations on speeding up counting of contingency table with a pre-computed look-up table. We have performed simulations to compare the time of counting contingency table of each SNP-SNP interaction under every pattern with pre-computed look-up table instead of naively counting the number of

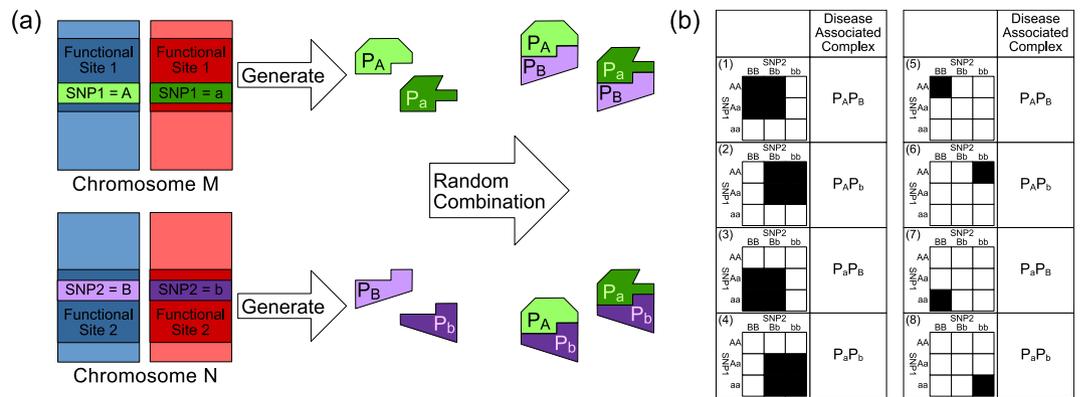


Figure 5. In this figure, part (a) shows the bio-molecule interaction mechanism behind a 2^{nd} order SNP-SNP interaction¹⁴ where SNP_1 and SNP_2 are both having genotype (*major, minor*). Meanwhile, part (b) shows the eight biologically plausible 2^{nd} order genotype interaction patterns and their corresponding disease-associated complexes. Major alleles are represented by upper-case letters (i.e. A, B) and minor alleles are represented by lower-case letters (i.e. a,b).

cases and controls under black and white genotype under every pattern. In this simulation, we have executed our program under the synthesis datasets generated by us. The average run-time of our program with or without pre-computed look-up table under datasets with different numbers of SNPs and heritabilities are shown in Supplementary Table S10. By observing Supplementary Table S10, we can observe that our program executed with the pre-computed look-up table is at least 8 times faster under datasets with 1000 SNPs and at least 10 times faster under datasets with 5000 and 10000 SNPs. This shows that the pre-computed look-up table is an effective mean to accelerate the counting of contingency table. Since the loading time of the pre-computed look-up table is constant under any dataset inputted, the impact of loading pre-computed look-up table on run-time is far more significant under a smaller dataset. Therefore, the speed-up under datasets with 1000 SNPs is less significant than the speed-up under dataset with 1000 and 10000 SNPs.

Discussion

This project studies the genetic risk factors of psoriasis. Psoriasis is a chronic inflammatory dermatitis characterised by hyperproliferation of the epidermis. It is a common dermatitis affecting up to 3% of the general population¹. A strong role of genetic predisposition in its etiology has been confirmed by recent family-based linkage studies and twin studies. Recent GWAS also confirmed a number of predisposition SNPs, particularly in Chromosome 6 *HLA* loci region^{1,2}. In this paper, we aim at demonstrating SNP-SNP interactions also play an important role in causing psoriasis. Our results have shown that interactions between SNPs are present in psoriasis patients. Although these interacting SNPs are mostly found in the extended region of *HLA* in Chromosome 6, there are SNP-SNP interactions spanned across recombination hotspots which excluded the possibility that the epistasis is due to long-distance haplotype effect. Existing gene/protein interaction database also confirmed that the bio-molecule products of these genes in the extended region of *HLA* in Chromosome 6 are indeed interacting with each other.

Performing GWAS through analysing the independent effect of every SNP did not provide an adequate explanation for the heritability of psoriasis. Previous studies have shown that monozygotic twins were more likely to be affected together than dizygotic twins⁸¹. Its heritability is as high as 90% which is one of the highest among common diseases and it can equally affect both male and female⁸². On the other hand, the best predisposition SNPs found in *HLA-C* could only increase disease risk by around 4 folds^{4,39,83}. This level of odds ratio is not sufficient to account for the high heritability of psoriasis. Other mechanisms must be involved. Gene-gene interactions or gene-environmental interactions are among the most likely explanations for this phenomenon of missing heritability.

Unlike other existing GWAS researches, we mainly focus on searching SNP-SNP interactions which can be explained as interactions between two bio-molecules. Although there are many potential mechanisms which can explain the interaction between two loci, interaction between 2 bio-molecules is the most valid and feasible hypothesis. With this idea in mind, we can introduce new constraints to the genotype interaction patterns to shrink the search space. Given any 3×3 genotype table of any 2 SNPs, there are total 2^9 possible genotype interaction patterns available. Such exhaustive enumeration of interaction patterns is unnecessary as most of these patterns are not biologically interpretable under our assumption. Based on our previous analysis on the distribution of bio-molecule complexes between two different bio-molecules, we have shown that the search space can be restricted to only 8 genotype interaction patterns as shown in Fig. 5b. After we have applied our algorithm to two psoriasis GWAS datasets, we have identified interaction between bio-molecules generated from the extended *HLA* region in Chromosome 6.

Among the gene-gene interactions we have found, some are obviously associated to psoriasis. For example, *HLA-C* and *HLA-B*, these 2 genes are probably expressed together and their protein products can interact with antigens on cell surface. It is possible some alleles of the *HLA* lead to a more intense inflammatory response.

For example, *HLA-b27* has been known to be a genetic factor of psoriasis. Similar *HLA* alleles have already been found in other diseases like HIV infection and Diabetes.

Knowledge of interaction between other bio-molecules in the extended *HLA* loci is still less understood. However, genome-wide protein-protein interactions were evident in numerous wet experiment. In fact, it is not uncommon to have interacting gene partners (like ligands and receptors) located in nearby genomic region, with the benefit that they could be regulated simultaneously during organism development. This genomic arrangement leads to various example of gene clusters, like those of cytokines, chemokines and their receptor etc.

Methods

Restricting search space of SNP-SNP interaction with biologically plausible genotype interaction pattern. The number of cases and controls under each genotype of any 2^{nd} order SNP combination (SNP_i, SNP_j) can be arranged into a 3×3 genotype table similar to the table shown in Fig. 3, where each cell in the 3×3 table corresponds to a genotype of (SNP_i, SNP_j). A genotype in a 3×3 table can be labelled as a high-risk or low-risk genotype through applying statistical or heuristic algorithms like multi-factor dimensionality reduction (MDR) algorithm and its derivatives⁸⁴. However, the patterns of the genotype label generated by these algorithms may not be biologically interpretable. In this paper, we have applied eight 2^{nd} order biological plausible SNP-SNP interaction patterns¹⁴ for labelling genotypes into high-risk and low risk genotypes. The principles and assumptions in deriving these eight SNP-SNP interaction patterns are shown in Fig. 5a and are explained below.

1. SNP_1 and SNP_2 are located within two different functional sites $Site_1$ and $Site_2$ respectively. Major alleles are represented by upper-case letters (i.e. A, B) and minor alleles are represented by lower-case letters (i.e. a, b).
2. SNP_1 and SNP_2 can affect their respective functional sites and cause each site to produce at most two different subtypes of bio-molecules. For example, bio-molecules p_A is generated from $Site_1$ with SNP_1 having an major allele.
3. The bio-molecules generated from $Site_1$ and $Site_2$ can randomly dock with each other to form at most four different bio-molecule complexes. For example, complex $p_A p_B$ is composed by bio-molecules p_A and p_B generated from $Site_1$ and $Site_2$ respectively.
4. A bio-molecule complex is associated to a genetic disease if (1) Only its solo presence (i.e. no other bio-molecules are present) or (2) Its presence can either promote or inhibit a disease.

The eight biologically plausible SNP-SNP interaction patterns are shown in Fig. 5b. Considering pattern 1 in Fig. 5b, if $p_A p_B$ is the only disease-associated bio-molecule complex and its presence can either promote or inhibit a disease (i.e. condition 4b), samples carrying genotype {"AA", "BB"}, {"AA", "Bb"}, {"Aa", "BB"} and {"Aa", "Bb"} obviously would have a different level of disease risk comparing to samples carrying other genotypes. After labelling these two groups of genotypes with two different colours, pattern 1 can hence be defined. On the other hand considering pattern 5 in Fig. 5b, if $p_A p_B$ is the only disease-associated bio-molecule complex and only its solo presence can either promote or inhibit a disease (i.e. condition 4a), samples carrying genotype AA, BB would have a different level of disease risk comparing to samples with other genotypes. Similarly, other patterns shown in Fig. 5b can be also defined through a similar argument shown above.

Since other genotype interaction patterns are not biologically plausible, we can reduce our search space on genotype interaction patterns from 2^9 to the eight patterns shown in Fig. 5b and thus significantly reduce the size of the search space.

Finding statistically significant SNP-SNP interactions with exhaustive search. In Supplementary Fig. S6, the process of converting the 3×3 table of a 2^{nd} order SNP combination (SNP_i, SNP_j) into a 2×2 contingency table is being demonstrated. Considering a 2^{nd} order SNP combination (SNP_i, SNP_j) after its genotypes being labelled according to genotype interaction pattern 1 in Fig. 5b, the number of cases and controls having black genotypes are summed up as $N_{D,B}$ and $N_{H,B}$ respectively. Meanwhile the number of cases and controls of white genotypes are summed up as $N_{D,W}$ and $N_{H,W}$ respectively. This summation process is shown in Supplementary Fig. S6a. After that, the counts of cases and controls $N_{D,B}$, $N_{H,B}$ and $N_{D,W}$ and $N_{H,W}$ can then be arranged into a 2×2 contingency table as shown in Supplementary Fig. S6b. Finally, the 2×2 contingency table of every SNP-SNP interaction is subjected to statistical tests and the SNP-SNP interaction found to be statistically significant are analysed. Imputation of SNPs on selected chromosomes were carried out on Michigan Imputation Server (<https://imputationserver.sph.umich.edu>) which is based on Minimac3 imputation algorithm⁸⁵.

Accelerating counting of contingency table with a pre-computed look-up table. We propose to accelerate the exhaustive search process through a pre-computed look-up table. For each SNP_i , it corresponds to a vector $\overrightarrow{SNP_i}$ (i.e. the i th column of matrix A in Table 1). Each vector $\overrightarrow{SNP_i}$ can be spliced into two different vectors $\overrightarrow{SNP_{i,case}}$ and $\overrightarrow{SNP_{i,ctrl}}$, where case vector $\overrightarrow{SNP_{i,case}}$ only has genotypes of SNP_i from cases and control vector $\overrightarrow{SNP_{i,ctrl}}$ has genotypes of SNP_i from controls. Each genotype can be considered as a 2 bit integer and thus every p genotypes in a vector can be combined into a $2p$ bit integer g . Given any possible pair of integer g , the distribution of black and white genotypes in cases and controls under every genotype interaction pattern can be pre-computed and stored in a look-up table located at the main memory. Therefore, the 2×2 contingency table of any pairs of SNP can be found without direct counting. Instead, the distribution of black and white genotypes in cases under any pair of SNP SNP_i and SNP_j can be found by retrieving and summing the distribution of black and white genotypes of every corresponding pair of integer g between case vectors $\overrightarrow{SNP_{i,case}}$ and $\overrightarrow{SNP_{j,case}}$ from the look-up table in the main memory. Meanwhile, the distribution of black and white genotypes in controls can be obtained in a similar fashion. This significantly accelerates the time needed to build the contingency table of each SNP-SNP interaction.

References

- Harden, J. L., Krueger, J. G. & Bowcock, A. M. The immunogenetics of Psoriasis: A comprehensive review, <https://doi.org/10.1016/j.jaut.2015.07.008> (2015).
- Anbunathan, H. & Bowcock, A. M. The Molecular Revolution in Cutaneous Biology: The Era of Genome-Wide Association Studies and Statistical, Bigg Data, and Computational Topics, <https://doi.org/10.1016/j.jid.2016.03.047> (2017).
- Nair, R. P. *et al.* Genome-wide scan reveals association of psoriasis with IL-23 and NF-kappaB pathways. *Nat. genetics*, <https://doi.org/10.1038/ng.311> (2009).
- Strange, A. *et al.* A genome-wide association study identifies new psoriasis susceptibility loci and an interaction between HLA-C and ERAP1. *Nat. Genet.*, <https://doi.org/10.1038/ng.694> (2010).
- De Cid, R. *et al.* Deletion of the late cornified envelope LCE3B and LCE3C genes as a susceptibility factor for psoriasis. *Nat. Genet.*, <https://doi.org/10.1038/ng.313> (2009).
- Ray-Jones, H., Eyre, S., Barton, A. & Warren, R. B. One SNP at a Time: Moving beyond GWAS in Psoriasis, <https://doi.org/10.1016/j.jid.2015.11.025> (2016).
- Golan, D., Lander, E. S. & Rosset, S. Measuring missing heritability: Inferring the contribution of common variants. *Proc. Nat. Acad. Sci.*, <https://doi.org/10.1073/pnas.1419064111> (2014).
- Wei, W. H., Massey, J., Worthington, J., Barton, A. & Warren, R. B. Genotypic variability-based genome-wide association study identifies non-additive loci HLA-C and IL12B for psoriasis. *J. Hum. Genet.*, <https://doi.org/10.1038/s10038-017-0350-6> (2018).
- Kuzmin, E. *et al.* Systematic analysis of complex genetic interactions. *Sci.*, <https://doi.org/10.1126/science.aao1729> (2018).
- Niel, C., Sinoquet, C., Dina, C. & Rocheleau, G. A survey about methods dedicated to epistasis detection, <https://doi.org/10.3389/fgene.2015.00285> (2015).
- Ritchie, M. D. Finding the epistasis needles in the genome-wide haystack. In *Epistasis: Methods and Protocols*, https://doi.org/10.1007/978-1-4939-2155-3_2 (2014).
- Chen, A., Liu, Y., Williams, S. M., Morris, N. & Buchner, D. A. Widespread epistasis regulates glucose homeostasis and gene expression. *PLoS Genet.*, <https://doi.org/10.1371/journal.pgen.1007025> (2017).
- Yip, D. K.-S. *et al.* A network approach to exploring the functional basis of gene-gene epistatic interactions in disease susceptibility. *Bioinforma.*, <https://doi.org/10.1093/bioinformatics/bty005> (2017).
- Chu, S. K., Xu, S. G., Xu, F. & Tang, N. L. S. Gene-gene Interaction Analysis by IAC (Interaction Analysis by Chi-Square) - A Novel Biological Constraint-based Interaction Analysis Framework. In *Proceedings of the 9th International Joint Conference on Biomedical Engineering Systems and Technologies*, <https://doi.org/10.5220/0005654601420150> (2016).
- Hirschhorn, J. N. & Daly, M. J. Genome-wide association studies for common diseases and complex traits, <https://doi.org/10.1038/nrg1521> (2005).
- Gerke, J., Lorenz, K. & Cohen, B. Genetic interactions between transcription factors cause natural variation in yeast. *Sci.*, <https://doi.org/10.1126/science.1166426> (2009).
- Moore, J. H., Asselbergs, F. W. & Williams, S. M. Bioinformatics challenges for genome-wide association studies, <https://doi.org/10.1093/bioinformatics/btp713> (2010).
- Wei, W. H., Hemani, G. & Haley, C. S. Detecting epistasis in human complex traits, <https://doi.org/10.1038/nrg3747> (2014).
- Li, P., Guo, M., Wang, C., Liu, X. & Zou, Q. An overview of SNP interactions in genome-wide association studies. *Briefings Funct. Genomics*, <https://doi.org/10.1093/bfpg/elu036> (2015).
- Altshuler, D. M. *et al.* An integrated map of genetic variation from 1,092 human genomes. *Nat.*, <https://doi.org/10.1038/nature11632> (2012).
- Wienbrandt, L. *et al.* FPGA-based acceleration of detecting statistical epistasis in GWAS. In *Procedia Computer Science*, <https://doi.org/10.1016/j.procs.2014.05.020> (2014).
- Zhu, Z. *et al.* Development of GMDR-GPU for Gene-Gene Interaction Analysis and Its Application to WTCCC GWAS Data for Type 2 Diabetes. *PLoS ONE*, <https://doi.org/10.1371/journal.pone.0061943> (2013).
- Goudey, B. *et al.* GWIS - model-free, fast and exhaustive search for epistatic interactions in case-control GWAS. *BMC Genomics*, <https://doi.org/10.1186/1471-2164-14-S3-S10> (2013).
- Yung, L. S., Yang, C., Wan, X. & Yu, W. GBOOST: A GPU-based tool for detecting gene-gene interactions in genome-wide case control studies. *Bioinforma.*, <https://doi.org/10.1093/bioinformatics/btr114> (2011).
- González-Domínguez, J., Schmidt, B., Kässens, J. C. & Wienbrandt, L. Hybrid CPU/GPU acceleration of detection of 2-SNP epistatic interactions in GWAS. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, https://doi.org/10.1007/978-3-319-09873-9_57 (2014).
- Hu, X. *et al.* SHEsisEpi, a GPU-enhanced genome-wide SNP-SNP interaction scanning algorithm, efficiently reveals the risk genetic epistasis in bipolar disorder. *Cell Res.*, <https://doi.org/10.1038/cr.2010.68> (2010).
- González-Domínguez, J. & Schmidt, B. GPU-accelerated exhaustive search for third-order epistatic interactions in case-control studies. *J. Comput. Sci.*, <https://doi.org/10.1016/j.jocs.2015.04.001> (2015).
- Hemani, G., Theodoridis, A., Wei, W. & Haley, C. EpiGPU: Exhaustive pairwise epistasis scans parallelized on consumer level graphics cards. *Bioinforma.*, <https://doi.org/10.1093/bioinformatics/btr172> (2011).
- Lonsdale, J. *et al.* The Genotype-Tissue Expression (GTEx) project, <https://doi.org/10.1038/ng.2653> (2013).
- Purcell, S. *et al.* PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *The Am. J. Hum. Genet.*, <https://doi.org/10.1086/519795> (2007).
- Anderson, C. A. *et al.* Data quality control in genetic case-control association studies. *Nat. Protoc.*, <https://doi.org/10.1038/nprot.2010.116> (2010).
- Fang, S., Fang, X. & Xiong, M. Psoriasis prediction from genome-wide SNP profiles. *BMC Dermatol.*, <https://doi.org/10.1186/1471-5945-11-1> (2011).
- Climer, S., Templeton, A. R. & Zhang, W. Allele-Specific Network Reveals Combinatorial Interaction That Transcends Small Effects in Psoriasis GWAS. *PLoS Comput. Biol.*, <https://doi.org/10.1371/journal.pcbi.1003766> (2014).
- Feng, B. J. *et al.* Multiple loci within the major histocompatibility complex confer risk of psoriasis. *PLoS Genet.*, <https://doi.org/10.1371/journal.pgen.1000606> (2009).
- Elder, J. T. Genome-wide association scan yields new insights into the immunopathogenesis of psoriasis, <https://doi.org/10.1038/gene.2009.11> (2009).
- Li, M. *et al.* Deletion of the late cornified envelope genes LCE3C and LCE3B is associated with psoriasis in a Chinese population. *J. Investig. Dermatol.*, <https://doi.org/10.1038/jid.2011.86> (2011).
- Gladman, D. D., Anhorn, K. A., Schachter, R. K. & Mervart, H. HLA antigens in psoriatic arthritis. *The J. rheumatology* (1986).
- Nair, R. P. *et al.* Sequence and haplotype analysis supports HLA-C as the psoriasis susceptibility 1 gene. *Am. j. human genetics*, <https://doi.org/10.1086/503821> (2006).
- Liu, Y. *et al.* A genome-wide association study of psoriasis and psoriatic arthritis identifies new disease loci. *PLoS Genetics*, <https://doi.org/10.1371/journal.pgen.1000041> (2008).
- Li, X. L., Yu, H. & Wu, G. S. Investigating the genetic association of HCP5, SPATA2, TNIP1, TNFAIP3 and COG6 with psoriasis in Chinese population. *Int. J. Immunogenetics*, <https://doi.org/10.1111/iji.12150> (2014).
- Holm, S. J., Carlén, L. M., Mallbris, L., Ståhle-Bäckdahl, M. & O'Brien, K. P. Polymorphisms in the SEEK1 and SPR1 genes on 6p21.3 associate with psoriasis in the Swedish population. *Exp. Dermatol.*, <https://doi.org/10.1034/j.1600-0625.2003.00048.x> (2003).

42. Rahman, P. *et al.* Association of SEEK1 and psoriatic arthritis in two distinct Canadian populations. *Annals Rheum. Dis.*, <https://doi.org/10.1136/ard.2004.031765> (2005).
43. Knight, J. *et al.* Conditional analysis identifies three novel major histocompatibility complex loci associated with psoriasis. *Hum. Mol. Genet.*, <https://doi.org/10.1093/hmg/dds344> (2012).
44. Zhou, F. *et al.* Epigenome-wide association data implicates DNA methylation-mediated genetic risk in psoriasis. *Clin. Epigenetics*, <https://doi.org/10.1186/s13148-016-0297-z> (2016).
45. Denny, J. C. *et al.* Systematic comparison of phenome-wide association study of electronic medical record data and genome-wide association study data. *Nat. Biotechnol.*, <https://doi.org/10.1038/nbt.2749> (2013).
46. Lysell, J., Padyukov, L., Kockum, I., Nikamo, P. & Ståhle, M. Genetic association with ERAP1 in psoriasis is confined to disease onset after puberty and not dependent on HLA-C*06. *J. Invest. Dermatol.*, <https://doi.org/10.1038/jid.2012.280> (2013).
47. Villarreal-Martínez, A. *et al.* Candidate gene polymorphisms and risk of psoriasis: A pilot study. *Exp. Ther. Medicine*, <https://doi.org/10.3892/etm.2016.3066> (2016).
48. Bowes, J. *et al.* Confirmation of TNIP1 and IL23A as susceptibility loci for psoriatic arthritis. *Annals Rheum. Dis.*, <https://doi.org/10.1136/ard.2011.150102> (2011).
49. Yang, Q. *et al.* Investigation of 20 non-HLA (human leucocyte antigen) psoriasis susceptibility loci in Chinese patients with psoriatic arthritis and psoriasis vulgaris. *Br. J. Dermatol.*, <https://doi.org/10.1111/bjd.12142> (2013).
50. Das, S. *et al.* Fine mapping of eight psoriasis susceptibility loci. *Eur. J. Hum. Genet.*, <https://doi.org/10.1038/ejhg.2014.172> (2015).
51. Evans, D. M. *et al.* Interaction between ERAP1 and HLA-B27 in ankylosing spondylitis implicates peptide handling in the mechanism for HLA-B27 in disease susceptibility. *Nat. Genet.*, <https://doi.org/10.1038/ng.873> (2011).
52. International HIV Controllers Study, T. I. H. C. *et al.* The major genetic determinants of HIV-1 control affect HLA class I peptide presentation. *Science (New York, N.Y.)*, <https://doi.org/10.1126/science.1195271> (2010).
53. Arnett, F. C., Reveille, J. D. & Duvic, M. Psoriasis and psoriatic arthritis associated with human immunodeficiency virus infection. *Rheum. diseases clinics North Am* (1991).
54. Choi, H. B., Han, H., Youn, J. I., Kim, T. Y. & Kim, T. G. MICA 5.1 allele is a susceptibility marker for psoriasis in the Korean population. *Tissue Antigens*, <https://doi.org/10.1034/j.1399-0039.2000.560609.x> (2000).
55. González, S. *et al.* Polymorphism in MICA rather than HLA-B/C genes is associated with psoriatic arthritis in the Jewish population. *Hum. Immunol.*, [https://doi.org/10.1016/S0198-8859\(01\)00242-7](https://doi.org/10.1016/S0198-8859(01)00242-7) (2001).
56. Chang, Y. T. *et al.* A study of candidate genes for psoriasis near HLA-C in Chinese patients with psoriasis. *Br. J. Dermatol.*, <https://doi.org/10.1046/j.1365-2133.2003.05166.x> (2003).
57. Wu, X. *et al.* A Novel Statistic for Genome-wide interaction analysis. *PLoS Genet.*, <https://doi.org/10.1371/journal.pgen.1001131> (2010).
58. Zhang, X. *et al.* HLA-DQA1 and DQB1 alleles are associated with genetic susceptibility to psoriasis vulgaris in Chinese Han. *Int J Dermatol* (2004).
59. Ozawa, A. *et al.* HLA class I and II alleles and susceptibility to generalized pustular psoriasis: Significant associations with HLA-Cw1 and HLA-DQB1*0303. *J. Dermatol.*, <https://doi.org/10.1111/j.1346-8138.1998.tb02461.x> (1998).
60. Lesueur, F. *et al.* ADAM33, a new candidate for psoriasis susceptibility. *PLoS ONE*, <https://doi.org/10.1371/journal.pone.0000906> (2007).
61. Asumalahti, K. *et al.* Coding haplotype analysis supports HCR as the putative susceptibility gene for psoriasis at the MHC PSORS1 locus. *Hum. molecular genetics*, <https://doi.org/10.1093/hmg/11.5.589> (2002).
62. Chantarangsu, S. *et al.* Genome-wide association study identifies variations in 6p21.3 associated with nevirapine-induced rash. <https://doi.org/10.1093/cid/cir403> (2011).
63. Kircher, M. *et al.* A general framework for estimating the relative pathogenicity of human genetic variants. *Nat. Genet.* <https://doi.org/10.1038/ng.2892> (2014).
64. Matzaraki, V., Kumar, V., Wijmenga, C. & Zhernakova, A. The MHC locus and genetic susceptibility to autoimmune and infectious diseases. <https://doi.org/10.1186/s13059-017-1207-1> (2017).
65. Chang, Y. T. *et al.* Psoriasis vulgaris in Chinese individuals is associated with PSORS1C3 and CDSN genes. *Br. J. Dermatol.*, <https://doi.org/10.1111/j.1365-2133.2006.07420.x> (2006).
66. Orrù, S. *et al.* Psoriasis is associated with a SNP haplotype of the corneodesmosin gene (CDSN). *Tissue Antigens*, <https://doi.org/10.1034/j.1399-0039.2002.600403.x> (2002).
67. Suomela, S. *et al.* Clinical associations of the risk alleles of HLA-Cw6 and CCHCR1*WWCC in psoriasis. *Acta Dermato-Venerologica*, <https://doi.org/10.2340/00015555-0184> (2007).
68. Gandhi, G., Buttar, B. S., Albert, L., Hasan, Q. & Aggarwal, R. K. Psoriasis-associated genetic polymorphism in North Indian population in the CCHCR1 gene and in a genomic segment flanking the HLA-C region. *Dis. Markers*, <https://doi.org/10.3233/DMA-2011-0851> (2011).
69. Romphruk, A. V. *et al.* Major histocompatibility complex class I chain-related gene A in Thai psoriasis patients: MICA association as a part of human leukocyte antigen-B-Cw haplotypes. *Tissue Antigens*, <https://doi.org/10.1111/j.0001-2815.2004.00238.x> (2004).
70. Holm, S. J. *et al.* HLA-Cw*0602 associates more strongly to psoriasis in the Swedish population than variants of the novel 6p21.3 gene PSORS1C3. *Acta Dermato-Venerologica*, <https://doi.org/10.1080/00015550410023527> (2005).
71. Zhu, K. J. *et al.* Psoriasis regression analysis of MHC loci identifies shared genetic variants with vitiligo. *PLoS ONE*, <https://doi.org/10.1371/journal.pone.0023089> (2011).
72. Chang, Y. T. *et al.* The genetic polymorphisms of POU5F1 gene are associated with psoriasis vulgaris in Chinese. <https://doi.org/10.1016/j.jdermsci.2007.01.003> (2007).
73. Dewald, G., Lange, C. E., Schmeel, E. & Kreysel, H. W. HLA-linked complement polymorphisms (C2, BF) in psoriasis. *Arch. Dermatol. Res.*, <https://doi.org/10.1007/BF00417201> (1983).
74. Jensen, L. J. *et al.* STRING 8—a global view on proteins and their functional interactions in 630 organisms. *Nucleic Acids Res*, <https://doi.org/10.1093/nar/gkn760> (2009).
75. Zhang, X. *et al.* RAID: A comprehensive resource for human RNA-associated (RNA-RNA/RNA-protein) interaction. *RNA*. <https://doi.org/10.1261/rna.044776.114> (2014).
76. Jiang, Q. *et al.* LncRNA2Target: a database for differentially expressed genes after lncRNA knockdown or overexpression. *Nucleic Acids Res.*, <https://doi.org/10.1093/nar/gku1173> (2015).
77. Zhou, Z., Shen, Y., Khan, M. R. & Li, A. LncReg: A reference resource for lncRNA-associated regulatory networks. *Database*, <https://doi.org/10.1093/database/bav083> (2015).
78. Intomics A/S. InBio_Map, <https://www.intomics.com/inbio/map> (2016).
79. Peri, S. *et al.* Human protein reference database as a discovery resource for proteomics. *Nucleic Acids Res.*, <https://doi.org/10.1093/nar/gkh070> (2004).
80. Stark, C. *et al.* BioGRID: a general repository for interaction datasets. *Nucleic Acids Res.*, <https://doi.org/10.1093/nar/gkj109> (2006).
81. Lönnerberg, A. S. *et al.* Heritability of psoriasis in a large twin sample. *Br. J. Dermatol.*, <https://doi.org/10.1111/bjd.12375> (2013).
82. Wuepper, K. D., Coulter, S. N. & Haberman, A. Psoriasis vulgaris: a genetic approach. *The J. investigative dermatology*, <https://doi.org/10.1111/1523-1747.ep12505638> (1990).
83. Stuart, P. E. *et al.* Genome-wide Association Analysis of Psoriatic Arthritis and Cutaneous Psoriasis Reveals Differences in Their Genetic Architecture. *Am. J. Hum. Genet.*, <https://doi.org/10.1016/j.ajhg.2015.10.019> (2015).

84. Gola, D., Mahachie John, J. M., Van Steen, K. & König, I. R. A roadmap to multifactor dimensionality reduction methods. *Briefings in Bioinforma.*, <https://doi.org/10.1093/bib/bbv038> (2016).
85. Das, S. *et al.* Next-generation genotype imputation service and methods. *Nat. Genet.*, <https://doi.org/10.1038/ng.3656> (2016).

Acknowledgements

Datasets phs000019.v1.p1. and phs000982.v1.p1. used for the analyses described in this paper were obtained from the database of Genotypes and Phenotypes (dbGaP). Dataset phs000019.v1.p1 was collected by Dr. James T. Elder (University of Michigan, Ann Arbor, MI), Gerald G. Krueger (University of Utah, Salt Lake City, UT), Anne Bowcock (Washington University, St. Louis, MO) and Gonçalo R. Abecasis (University of Michigan, Ann Arbor, MI). Data collection was funded by the National Institutes of Health, the Foundation for the National Institutes of Health, and the National Psoriasis Foundation. Support for genotyping of samples was provided through the Genetic Association Information Network (GAIN). For a description of the dataset, phenotypes, genotype data and quality control procedures see Nair *et al.* (2009) *Nature Genetics* 41:200-204. Dataset phs000982.v1.p1 was collected by James T. Elder, University of Michigan, with collaborators Dr. Dafna Gladman, University of Toronto and Dr. Proton Rahman, Memorial University of Newfoundland, providing samples. Data collection was supported by grants from the National Institutes of Health, the Canadian Institute for Health Research, and the Krembil Foundation. Additional support was provided by the Babcock Memorial Trust and by the Barbara and Neal Henschel Charitable Foundation. JTE is supported by the Ann Arbor Veterans Affairs Hospital.

Author Contributions

Kwan-Yeung Lee conducted the project and developed the the algorithms used in this study under the supervision of Man-Hon Wong and Kwong-Sak Leung and interpreted the biological meaning of the experimental results under the guidance of Nelson L.S. Tang who pioneered the interaction models. Kwan-Yeung Lee also ran the experiments and wrote the main manuscript text. All authors discussed the results and reviewed the manuscript thoroughly.

Additional Information

Supplementary information accompanies this paper at <https://doi.org/10.1038/s41598-018-33493-w>.

Competing Interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018