

RESEARCH ARTICLE

Open Access

# Genetic variation in the odorant receptors family 13 and the mhc loci influence mate selection in a multiple sclerosis dataset

Pouya Khankhanian, Pierre-Antoine Gourraud, Stacy J Caillier, Adam Santaniello, Stephen L Hauser, Sergio E Baranzini, Jorge R Oksenberg\*

## Abstract

**Background:** When selecting mates, many vertebrate species seek partners with major histocompatibility complex (MHC) genes different from their own, presumably in response to selective pressure against inbreeding and towards MHC diversity. Attempts at replication of these genetic results in human studies, however, have reached conflicting conclusions.

**Results:** Using a multi-analytical strategy, we report validated genome-wide relationships between genetic identity and human mate choice in 930 couples of European ancestry. We found significant similarity between spouses in the MHC at class I region in chromosome 6p21, and at the odorant receptor family 13 locus in chromosome 9. Conversely, there was significant dissimilarity in the MHC class II region, near the *HLA-DQA1* and *-DQB1* genes. We also found that genomic regions with significant similarity between spouses show excessive homozygosity in the general population (assessed in the HapMap CEU dataset). Conversely, loci that were significantly dissimilar among spouses were more likely to show excessive heterozygosity in the general population.

**Conclusions:** This study highlights complex patterns of genomic identity among partners in unrelated couples, consistent with a multi-faceted role for genetic factors in mate choice behavior in human populations.

## Background

When selecting mates, individuals of many vertebrate species favor partners with heterologous major histocompatibility complex (MHC) genes [1-6]. These genes have immune-recognition and response functions, thus this behavior can be interpreted as a mechanism developed through evolution to prevent inbreeding and increase MHC diversity, imparting more robust immune systems to offspring [7-10]. However, translation of these observations to human mating is not straightforward and the dependence of human mate selection on genetic factors, including the MHC, remains controversial [11-18]. Individuals in some populations, European American couples from Utah [13,14] and Hutterites [19], have been found to favor MHC-dissimilar mates. Other populations show no strong evidence of MHC-

selective mating, including Yorubans in Nigeria [13,14], South Amerindians [20], Dutch [21], Japanese [22], Swedish [23], Uruguayans [24], and Caucasians [25]. On the other hand, evidence for preference of MHC-similar mates was seen in Tohoku Japanese [22] but only when considering extended haplotypes composed of alleles of the HLA genes A, B, C, DR, and DQ in linkage disequilibrium. MHC similarity among mates was also seen in a multi-population study [26]. "Facial preference" studies generally show no preference for MHC similarity or dissimilarity, although a preference for heterozygosity could not be ruled out [27-31]. Early "sweaty t-shirt" studies suggested that MHC-dissimilarity mediates odor preference for a potential mate, but follow-up studies highlighted a variety of confounding factors in this phenomenon, such as genetic background, sex, and contraceptive use [31-36], providing a partial explanation for the conflicting results.

The recent availability of highly efficient genome-wide genotyping platforms affords deeper marker saturation

\* Correspondence: jorge.oksenberg@ucsf.edu  
Department of Neurology, University of California, San Francisco, CA 94143-0435, USA

in regions of interest as well as the performance of hypothesis-neutral screens. In this study, a screen with 309,100 single nucleotide polymorphisms (SNPs) in 930 unrelated couples of European ancestry was used to assess genetic similarity between spouses. Quality-controlled genotypic data was obtained from the International Multiple Sclerosis Genetics Consortium (IMSGC) from a study performed to identify multiple sclerosis (MS) susceptibility genes [37]. In a hypothesis-neutral approach, we looked for similarity at the genome-wide level, at the regional level, and at the individual SNP level. A Benjamini-Hochberg [38] tail-area-based correction for multiple comparisons was applied to strictly control for false positives. An excess of SNP-level similarity was observed in class I of the MHC, and in a locus on chromosome 9 near eight consecutive functional odorant receptor genes. The results are consistent with a significant but multifaceted role for genetic factors influencing mate selection in humans.

## Results

Genetic similarity/dissimilarity between spouses was assessed using three parallel approaches. Relatedness coefficient  $R_1$  (defined in methods) measures similarity across genetic regions, while  $R_2$  measures similarity at individual SNPs. A third approach seeks the preponderance extreme  $R_2$  values across regions. In this way, we allowed for a region to exhibit similarity and dissimilarity independently. This is important for gene-rich regions such as the MHC, which could potentially have an intricate role in mate preference. Relatedness coefficients are positive when spouses are more similar than random pairs of individuals, and negative when spouses are more dissimilar.

### Genome-Wide Similarity

Genome wide, the partners in the 930 couples of European descent included in this study are more genetically similar than expected by chance ( $R_1 = 0.00152$ , positive values of  $R_1$  indicate genetic similarity;  $p < 10^{-6}$ , using  $10^6$  permutations). In apparent contrast, Chaix et al. [13] reported no genetic similarity in the 28 spouses from the HapMap CEU population ( $R_1 = -0.00016$ ,  $p = 0.739$ ). Is their population different from ours? To answer this question, similarity ( $R_1$ ) was measured in 100,000 randomly selected sets of 28 couples from our dataset. The random set of 28 couples had  $R_1$  lower (less similarity) than that observed in the Chaix et al. study in 7.06% of the permutations. Although the power to assess differences in genetic identity patterns between the two populations is limited, it appears that the two populations are different. Further, the Chaix et al. results may be influenced by a few outlying couples [14].

### Hypothesis-Neutral Results

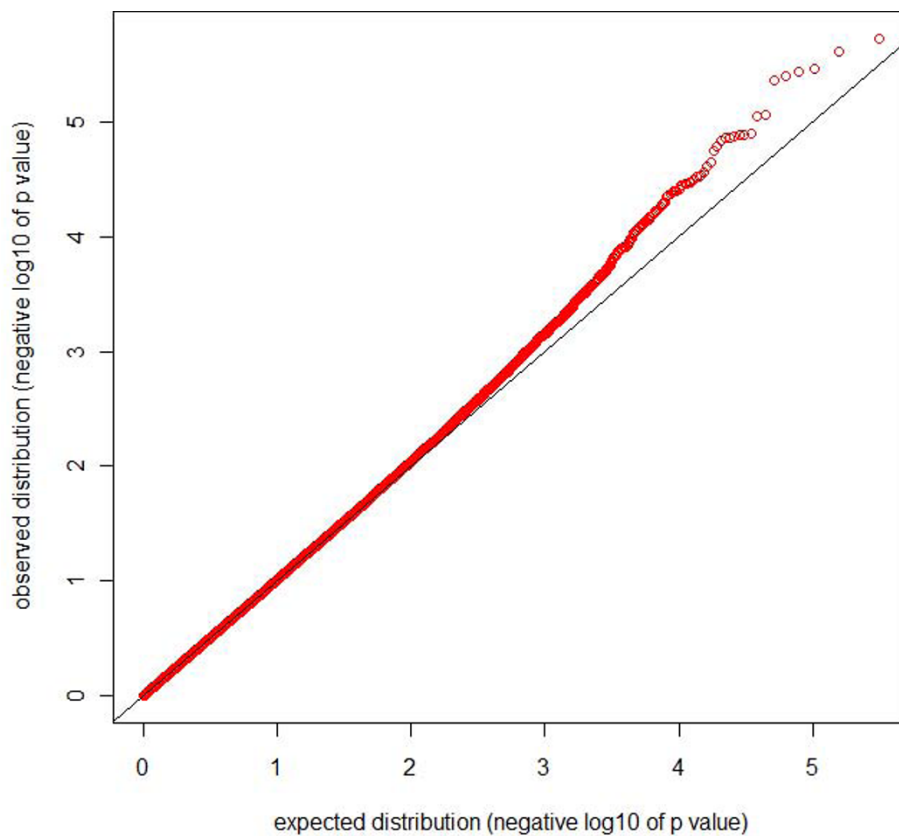
Using sliding windows of 3.6 Mb (in 100 Kb increments), there were 21,665 regions with at least 300 SNPs. The top 100 regions exhibiting excessive similarity or dissimilarity (extreme values of  $R_1$ ) are shown in Additional file 1: Table S1. After correction for multiple-hypothesis testing, none of these regions remain statistically significant ( $fdr \approx 1$ ). Additional file 2: figure S1 shows a relationship between  $R_1$  and recombination rate similar to that found in figure 2 of Chaix et al. [13], with more extreme values of  $R_1$  seen in regions of lower recombination rates. Hence, an examination of  $R_1$  at the locus level yielded no significant results after correction for multiple comparisons.

On the other hand, using Pearson Correlation ( $R_2$ ) for SNPs that showed significant similarity or dissimilarity among couples in the screen, 38 individual SNPs passed the genome-wide significance threshold ( $fdr < 0.1$ ) (Figure 1, Additional file 3: Table S2) after applying a Benjamini-Hochberg correction for multiple comparisons ( $n = 309,100$  SNPs). A large proportion (33 of 38) of these SNPs exhibit spousal identity, in line with the genome-wide observations. Of these, 10 SNPs came from a region upstream from the 8 consecutive odorant receptor family 13 genes on chromosome 9.

An examination of the Fisher meta value at the regional level is shown in Additional file 4: Table S3. The top 100 regions exhibiting an abundance of significant SNPs with spousal identity are concentrated in two areas. The first is the odorant receptor (OR) 13 region on chromosome 9 (13-17.4 Mbp), and the other is on chromosome 11 (61.6-63.6 Mbp). The top 100 regions showing abundance of dissimilar SNPs (Additional file 5: Table S4) are also concentrated in two areas, chromosome 2 (13-17.4 Mbp) and chromosome 9 (36.9-38 Mbp). The abundance of similarity or dissimilarity found in these regions was not significant after correction for multiple comparisons.

### The MHC

Using  $R_1$ , we observed an overall trend of similarity across the entire MHC that is not significant ( $R_1 = 0.0051$ , uncorrected  $p = 0.076$ ). When considering the results emerging from the MHC region, however, it is important to bear in mind the origin of our dataset: couples with an offspring affected with MS, an autoimmune disease with a genetic component, and a known strong association with the MHC, specifically with the relatively common HLA-DRB1\*15:01 allele. Other genes within the HLA class I region have been also proposed to be independently linked to MS by conferring protection. In apparent contrast, Chaix et al reported significant dissimilarity in this region ( $R = -0.043$ ,  $p = 0.015$ ) for the 28 HapMap couples. Once again we ask whether their population was different from ours. When  $R_1$  was measured in



**Figure 1 P-P plot of SNP level results.** *p* values of correlation among couples are plotted for all SNPs as a function of the normal distribution. The black line is equal to the expectation on H0. Overall, the data follows the normal distribution with an excess in the tail at  $p < 10^{-4}$ .

100,000 random sets of 28 couples chosen from our study, a value of  $R_1$  lower than -0.043 was only observed 0.853% of the time, suggesting significant differences between the two populations and reflecting the greater diversity of the dataset used in this study as well as the sampling variability of the HapMap samples [14].

Using  $R_2$ , no individual SNP from the MHC passed the genome-wide threshold of significance. The top MHC SNP *rs2844731* ( $R_2 = 0.113$ ,  $p = 0.00063$ ,  $fdr = 0.42$ ) was 291st on the list (sorted by significance) (Additional File 3: Table S2). The closest non-pseudogene is HLA-E.

Using the Fisher meta value, the MHC region as a whole showed a greater excess of significantly similar SNPs than the rest of the genome (Fisher meta value =  $7.7 \times 10^{-10}$ ,  $p = 0.013$ ). Upon breaking down the MHC into three classes, significant genetic identity among couples was found in class I (Fisher meta value =  $1.0 \times 10^{-8}$ ,  $p = 0.029$ ) (Table 1). Much of the similarity was due to a series of markers near *HLA-E* and *RPP21* (a gene involved in maturation of rRNA), and to a lesser extent to another series of markers near *MDC1* (a mediator of DNA damage checkpoint), including the non-

synonymous SNP *rs9262152* (Figure 2). We chose to perform PCR-based genotyping on this single SNP because of its relatively high correlation within the MHC ( $R_2 = 0.106$ ,  $p = 0.0012$ ) and its non-synonymous nature. Observed similarity among couples at *rs9262152* was confirmed by PCR ( $R_2 = 0.091$ ,  $p = 0.037$ ,  $n = 387$  couples) in a subset of the same couples but results were not replicated in independent couples ( $R_2 = -0.0047$ ,  $p = 0.5$ ,  $n = 393$ ), indicating that regional class I similarity is not a consequence of this SNP.

Genetic similarity among couples was not ubiquitous throughout class I; a locus of dissimilarity was found at 31.0 Mbp, within a gene-rich region including *GTF2H4* (general transcription factor involved in nucleotide excision repair), *SFTPG* (surfactant associated protein), *DPCR1* (diffuse panbronchiolitis critical region) and *VARS* (valyltRNA synthetase 2). The identity pattern detected using Affymetrix arrays was confirmed by a dense Illumina custom array of 1536 MHC SNPs performed on 920 of the 930 couples (Additional file 6: Figure S2, Additional file 7: Table S5) [39].

Next, imputation of 6 classical HLA alleles (A, B, C, DRB1, DQA1, and DQB1) was performed using this

**Table 1 Regional spousal identity**

Similarity only	SNPs	Fisher meta Value	Background Mean Fisher Value	P value
HLA region	458	$7.7 * 10^{-10}$	0.15	0.013
HLA class I region	280	$1.0 * 10^{-8}$	0.043	0.029
HLA class II region	90	0.82	0.35	0.83
HLA class III region	88	$6.4 * 10^{-6}$	0.36	0.09
OR13 region chr. 9	25	$3.8 * 10^{-43}$	0.48	$8.8 * 10^{-5}$
Dissimilarity only	SNPs	Fisher meta Value	Background Mean Fisher Value	P value
HLA region	246	0.11	0.61	0.20
HLA class I region	140	0.43	0.56	0.39
HLA class II region	63	0.063	0.59	0.12
HLA class III region	43	0.24	0.60	0.27
OR13 region chr. 9	60	0.24	0.41	0.48
Combined Similarity and Dissimilarity	SNPs	Fisher meta Value	Background Mean Fisher Value	P value
HLA region	704	$7.5 * 10^{-9}$	0.51	0.0025
HLA class I region	420	$7.3 * 10^{-7}$	0.033	0.23
HLA class II region	153	0.38	0.49	0.44
HLA class III region	131	$2.6 * 10^{-5}$	0.53	0.047
OR13 region chr. 9	85	$7.7 * 10^{-29}$	0.52	0.00056
Position	Recombination rate cM/Mb	Start	End	
HLA region	1.57	29,700,000	33,300,000	
HLA class I region	1.38	29,700,000	31,500,000	
HLA class II region	2.43	32,500,000	33,300,000	
HLA class III region	1.12	31,500,000	32,500,000	
OR13 region chr. 9	0.69	104,000,000	104,500,000	

The MHC region as a whole shows similarity between couples. When broken down into classes, class I shows significant similarity between couples. Of the three MHC classes, class II shows the most dissimilarity between couples, albeit not statistically significant. HLA-B and HLA-DRA denote the boundaries between the three MHC classes. UCSC Build 35 coordinates are shown.

dense coverage (Additional file 8: Table S6). There was significant dissimilarity in two of the three class II genes, *DQA1* ( $p = 0.001$ ) and *DQB1* ( $p = 0.044$ ).

### Functional Odorant Receptor Family 13

Functional odorant receptor (OR) families (13 and 18) are located in clusters on various chromosomes. Only one such cluster (from family 13) was large enough for analysis. A 500 Kbp region covering 8 consecutive *OR13* genes (*F1*, *C4*, *C3*, *C8*, *C5*, *C2*, *C9*, and *D1*) is bounded by *SMC2* and *NIPSNAP3A*. This region exhibits the greatest abundance of SNP-level allelic similarity between couples in the entire genome (Fisher meta value =  $3.85 * 10^{-43}$ ,  $p = 8.8 * 10^{-5}$ ). The bulk of genetic similarity is found in a non-coding region toward the centromere, 200-300 Kbp upstream from the OR13F gene cluster (Figure 3), even though there was adequate SNP coverage near the coding regions. The similarity in

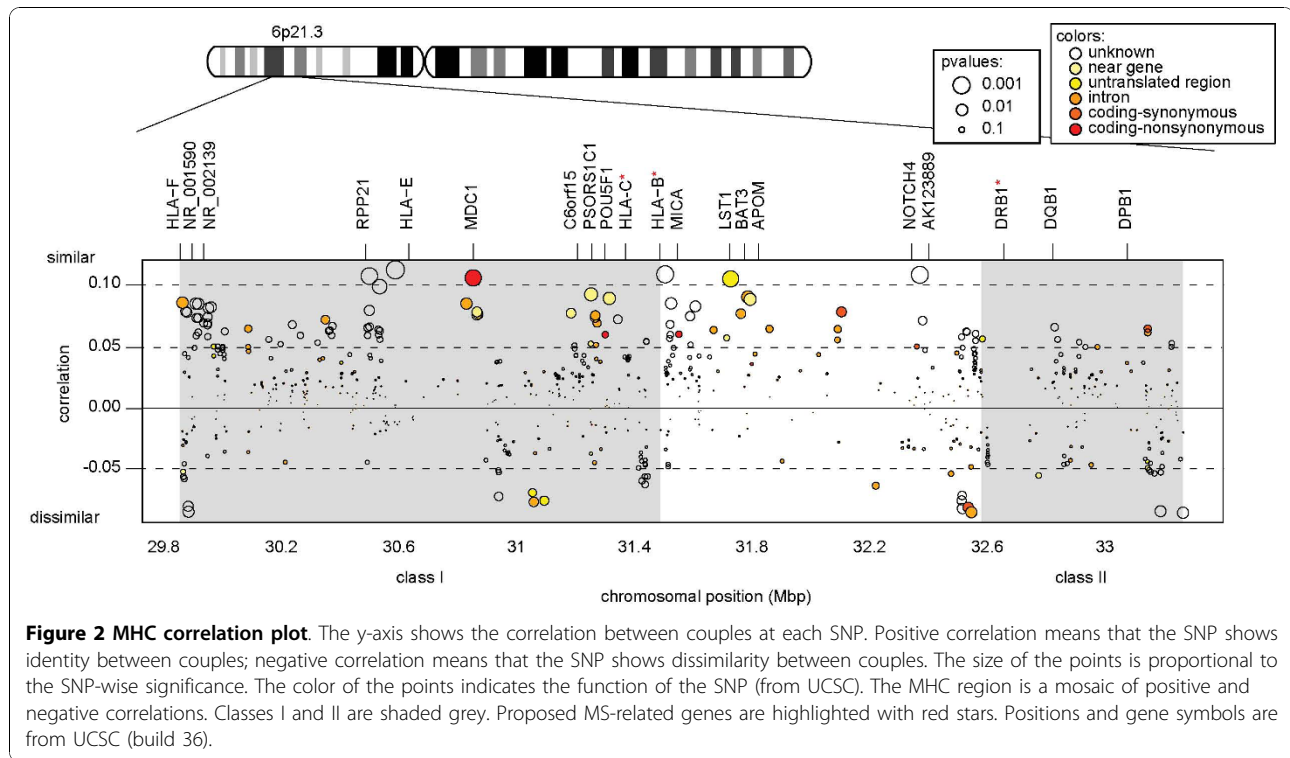
one SNP from this region (*rs1450686*) was confirmed by PCR-based genotyping in 387 couples from the original analysis ( $R^2 = 0.17$ ,  $p = 0.00056$ ) and 393 new couples ( $R^2 = 0.11$ ,  $p = 0.016$ ).

### Multiple-Sclerosis-Associated SNPs

Eleven non-MHC SNPs covered in this study are reportedly associated with MS risk. Similarity between couples ( $R^2$ ) is not significant at these SNPs, after a correction for 11 multiple hypotheses (Additional file 9: Table S7). Interestingly, 9 of 11 SNPs (including 2 that reach an uncorrected level of significance) show a trend for dissimilarity between spouses.

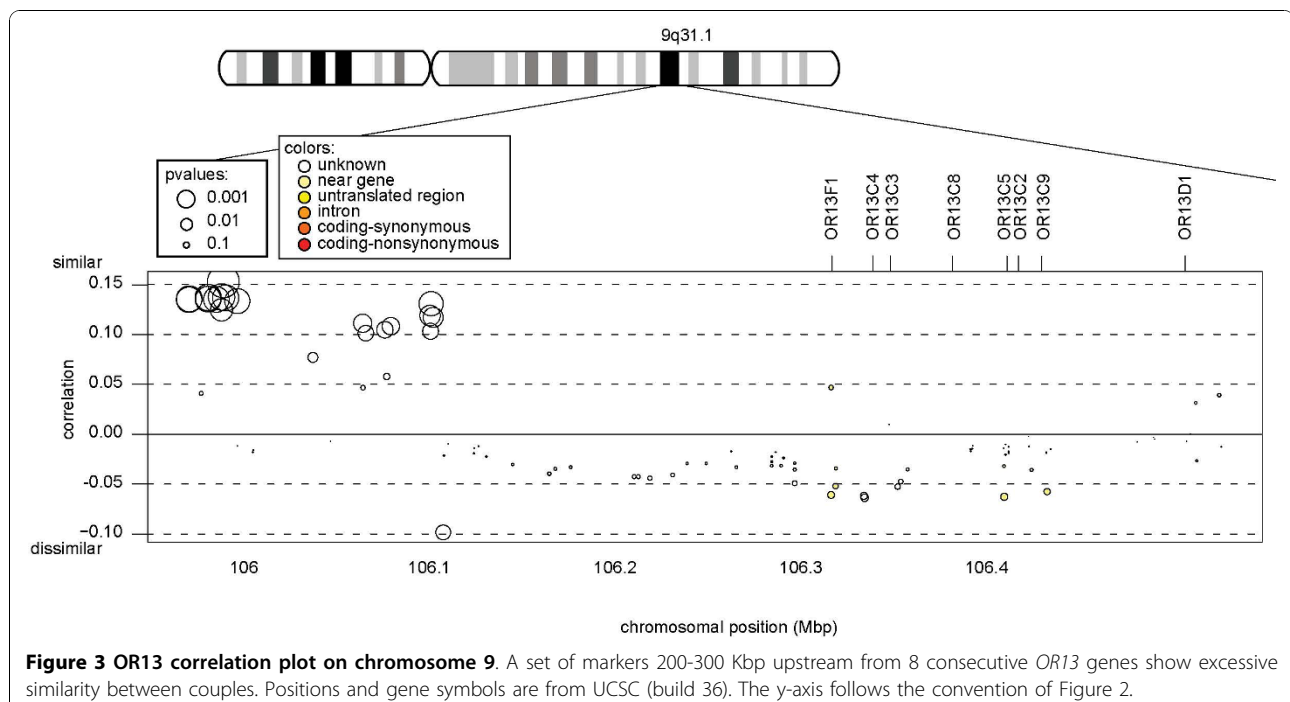
### Association with Observed Homozygosity in the General Population

If humans tend to select mates that are similar to self at certain genetic loci, then we would expect those loci



would show excessive homozygosity in the population over time (Additional file 10: Figure S3). To test this hypothesis, we searched for excessive homozygosity (versus Hardy-Weinberg equilibrium) in unrelated individuals from the HapMap CEU dataset. For this analysis,

we filtered out SNPs with minor allele frequency (MAF) less than 1%. We considered all markers showing significant similarity or dissimilarity between couples ( $p < 0.024$ ) and significant deviation from Hardy-Weinberg equilibrium ( $p < 0.024$ ) (Hardy-Weinberg disequilibrium



$p$  values calculated with the software Plink [40]). The two  $p$  value cutoffs were chosen because they combine by Fisher's method to yield  $p < 0.005$ . Markers showing similarity among couples in the screening dataset were 5 times more likely to show excessive homozygosity in the HapMap population than markers showing dissimilarity between couples ( $X^2 = 12.6021$ ,  $df = 1$ ,  $p < 0.004$ , chi-square test), further validating our observations (Table 2).

## Discussion

We performed a genome-wide analysis of genomic identity in 930 White couples of European ancestry, and report (i) a significant similarity in couples' genotypes at MHC class I (ii) novel significant similarity among couples in SNPs linked to the *Odorant receptor family 13* region on chromosome 9; (iii) 38 SNPs whose alleles showed significant correlation between couples ( $q$  value  $< 0.1$ ), 10 of which are upstream from eight consecutive *OR13* genes on chromosome 9. This, to our knowledge, is the first genome-wide study of its size with regard to human mate selection. We report a complex but statistically significant role for genetic similarity in mate choice, in particular the genes of the MHC and odorant receptors. In the MHC the interaction is different from class to class and from gene to gene, indicating that the disparity in the literature regarding the role of the MHC in mate choice may be resolved with inspection of smaller genetic windows.

### MHC, Mate Selection, and Multiple Sclerosis

In mice, MHC genotypes have been shown to be a determining factor in mate selection in some strains but not in others [6]. Although more strains of mice prefer the scent of MHC-dissimilar individuals when selecting mates [41], or rather the scent of mice with MHC dissimilar to parental MHC [5], it has been shown that they prefer the scent of MHC-similar mice when selecting a nesting partner [42]. Humans overwhelmingly tend to pick one person as both mate and nesting partner; it is difficult then to resolve the two. Difficulty in

extrapolation across species is exacerbated by the differences in reproductive strategies of mice (who tend to have more pregnancies per lifetime, with several pups per pregnancy) and humans (few pregnancies per lifetime, usually one child per pregnancy). Another difference is exposed when the proposed mechanism for murine MHC detection in mate selection is examined. Rodent mate selection is grounded in odorant sensations [43]. Specifically, mice respond to the scent of urinary proteins, which are naturally abundant in mice but mainly indicative of disease in humans [44-46].

Our results contribute to this ongoing exploration by finding an abundance of genetic similarity among couples (parents of children with MS) across class I of the MHC. Within class I, much of the similarity occurs near the HLA-E gene. HLA-E is a member of the non-classical class I genes (Ib), with characteristically low polymorphism across primates. Assortative mating may be driving the relatively low polymorphism in this gene. One of the drivers of high polymorphism in HLA genes is "crossing over" events during meiosis. Crossing over fails to increase polymorphism in homozygous individuals. Homozygous individuals are more frequent in the population when mating is associative.

This study sampled much older couples than the majority of previous studies. For 73% of the couples in this study, the mating event (successful breeding) occurred in the 1950's and 1960's. Just as we recognize that our findings of MHC class I similarity may be specific to the ethnicity of our population, we must allow that the results may be specific to this post-war generation.

When considering these results in the context of human mate selection, it is important to note that MS is a complex genetic disease strongly associated with the MHC class II gene *HLA-DRB1*. Our dataset is thus enriched for the set of common risk alleles of this gene (*DRB1\*15:01*, and to a lesser extent *DRB1\*03:01*) potentially leading to biased observations, but we report significant dissimilarity at this locus. While *HLA-DRB1* confers the greatest susceptibility to MS, it has been proposed that other HLA genes, in this case conferring resistance (*HLA-C* [47] and *HLA-B* [39]), may exist in the MHC class I region. Interestingly, we observed primarily similarity in this locus. Altogether, the observed patterns of identity in the parents of the affected individuals appear to conflict with what is expected in an MS dataset; we assume that risk genes would show similarity among parents. This assumption is subject to debate, as the mechanisms of MHC-mediated genetic risk to MS are not well understood. Furthermore, it is important to note that MHC class I similarity in parents could conceivably lead to viral susceptibility in offspring, and that the Epstein-Barr virus has been linked to MS [48].

**Table 2 Excessive homozygosity in the general population**

Observed (expected)	Excessive heterozygosity	Excessive homozygosity
Dissimilar among couples	16 (7.97)	20 (28.03)
Similar among couples	13 (21.03)	82 (73.97)

SNPs that are dissimilar among couples in the IMSGC dataset tend to show excessive heterozygosity in the HapMap CEU samples, while markers that are similar among IMSGC couples show excessive homozygosity in the HapMap samples (chi-square  $p < 0.004$ , expected values shown in parentheses). In all, 131 SNPs showed significant similarity or dissimilarity among IMSGC couples ( $p < 0.024$ ) and significant deviation from Hardy-Weinberg equilibrium in the independent HapMap CEU dataset ( $p < 0.024$ ).

Similar studies in other trio format datasets from other diseases with MHC etiology will be of great value in testing the hypothesis that parents of disease individuals display a different pattern of MHC identity than parents of healthy children. Overall, the mosaic-like statistically significant pattern of association between MHC and mate choice at the class level (and also at the gene level) is remarkably complex, linked perhaps to the extreme functional diversity across the MHC and abundance of hot-spots and warm-spots of genetic recombination distributed differently among individuals [49]. This complexity may underlie the conflicting reports of genetic identity in the literature focused on a single or limited number of variants.

The consequences of any deviation from random mating for human disease, particularly autoimmunity, are unknown, but regions of extreme similarity or dissimilarity among parents of affected individuals may be related to the presence of susceptibility and/or resistance loci. Furthermore, random mating is a commonly accepted assumption for most statistical genetic models usually performed to assess genetic association. If non-random mating exists, association studies should correct for the expected departure from Hardy Weinberg equilibrium assumptions.

### Olfaction

Olfaction, the proposed mechanism for MHC recognition, is involved in a variety of mating-related behaviors. While there is no genetic evidence to directly link odorant receptors to mate selection, the connection between olfaction and sexual behavior is well established. Olfaction is a necessary step in the mating mechanism for rodents and the kin-recognition mechanism in both humans and rodents. Specifically, mice respond to the scent of naturally abundant urinary proteins [44-46]. Odorant bulb removal eliminates mating behavior in male mice and hamsters, while eliminating maternal behavior in female mice [43,50,51]. For humans, body odor detection is a mechanism for kin recognition and mate preference, especially for females [32,33,52,53]. Human body odor influences general mood, attention state, and females' proclivity towards males [54-57]. Although most are non-functional, odorant receptors (OR) represent the largest family of genes in the human genome. There are two human OR families, 13 and 18, with known ligands (having a variety of perceived odors including sweat) [58]. It remains unclear whether the genetic similarity observed between mates leads to sexual attraction through olfaction or simply implies similarities in odor recognition and food and other smell preferences that would be practical concerns for human couples who live and eat together.

The genetic identity in the region was found 200-300 Kb upstream from the cluster of 8 consecutive OR13 genes. This would indicate that it is not protein identity, but rather some form of long-range gene regulation that is associated with mate selection. Long-range enhancers found in gene deserts are known to act at distances of hundreds of kilo base pairs [59-64]. The bulk of genetic similarity is 10-100 Kbp downstream from *SMC2*.

### Conclusions

We have seen a complex role for genetic similarity in mate choice, in particular genes of the MHC and odorant receptor family 13. Regarding the MHC, the observed interaction is not ubiquitous throughout the 3.6 Mbp region. Rather, the interaction is different from class to class, and from gene to gene. As higher resolution scans and sample populations of other ancestries, environments, and phenotypes (in particular non-MHC diseases) become available, deeper analysis of the roles of individual genes and functional pathways in mate selection and implications for health and disease will become possible.

### Methods

#### Datasets

a) 334,923 Single Nucleotide Polymorphism (SNPs) from the International Multiple Sclerosis Genetics Consortium (IMSGC) dataset, typed by Affymetrix 500K GeneChip in 931 European and European American couples (1862 individuals). Samples came from collections the University of California, San Francisco (417 couples), the Cambridge University Hospital Multiple Sclerosis Center (453 couples), and the Brigham and Women's Hospital (61 couples). The mean age was 71.2 (sd = 9.5) for males and 68.6 (sd = 9.2) for females. For 73% of couples, the mating event (successful breeding) occurred in the 1950's and 1960s. Divorce status and length of marriage was not recorded. One couple was removed because one parent was affected with multiple sclerosis (MS). 930 couples, each with a child afflicted with MS were used for analysis presented here. We included only autosomal SNPs with minor allele frequency above 5%, missing no more than 10% of genotypes in males or females (309,100 SNPs). This dataset is available on dbGAP (phs000139.v1.p1) [37].

b) 1,536 SNPs in the extended MHC region from the International MHC and Autoimmunity Genetics Network (IMAGEN) [39] dataset, typed by Illumina chip in 920 of 930 IMSGC couples. 1,078 SNPs passed quality control, 94 of which were in common between the two datasets.

c) A set of HLA types for 6 of the main HLA genes imputed from the IMAGEN SNP genotypes by the original authors.

d) Candidate SNPs were genotyped by PCR in 387 couples from the IMSGC dataset and 393 new European American couples. Both datasets were collected using the same inclusion criteria and have a similar distribution of age and ethnicity. Notably, these 393 new couples are also parents of children with MS.

#### Relatedness Analysis

The following approaches for measuring genetic similarity were used in parallel.

a) In the IMSGC dataset, a relatedness coefficient  $R1$  was defined for each couple at each variant as a ratio of probabilities of identity in state  $R1 = (Qc - Qm)/(1 - Qm)$ , where  $Qc$  is the proportion of identical variants between the two spouses and  $Qm$  is the mean proportion of identical variants in the sample (an average over all possible pairs). For SNPs, the proportion of identical variants between two individuals is 1 if both individuals are homozygous for the same allele, 0.5 if either individual is heterozygous, and 0 otherwise. For larger genetic regions of at least 300 SNPs,  $Qc$  and  $Qm$  were averaged across all SNPs in the region. The significance of  $R1$  was assessed using a permutation approach: the two-sided  $p$  value is the proportion of permutations (where spouses are shuffled) in which the mean  $R1$  of permuted couples is more extreme than the mean  $R1$  of real couples. 100,000 permutations were performed. This approach was used for HLA types and regions of SNPs by Chaix et al. [13] and the relatedness coefficient is discussed by Rousset [65].

b) In the IMSGC dataset, another relatedness measure  $R2$  was defined at individual SNPs across all couples as the Pearson correlation between fathers' and mothers' genotypes (recoded as 0, 1, and 2). A comparison of  $R2$  and  $R1$  at individual SNPs is discussed in Additional file 11: Text S1. Significance was assessed using two approaches, one based on permutation and another based on the genome as a background (Additional file 12: Text S2). Both methods yield similar results; we present the latter method.

c) In the IMSGC dataset, we next tested the hypothesis that an abundance of significant positive similarity exists among SNPs in a given region, say the MHC. For that purpose we combined the  $p$  values of all SNPs in the region which were similar between couples ( $R2 > 0$  from part b) using Fisher's method [66]. A low Fisher meta value indicates a more significant finding. The Fisher meta value of the given region was contrasted against all other equally sized non-centromeric regions (looking only at SNPs exhibiting similarity,  $R2 > 0$ ) from the entire genome (excluding the chromosome containing the candidate region) with a lower recombination rate than the candidate region if the candidate region has lower than average recombination rate (or a higher

recombination rate than the candidate region if the candidate region has a higher than average recombination rate). The  $p$  value is assigned to each candidate region as the percent of regions across the genome that had Fisher meta values smaller than the Fisher meta value of the candidate region. In the hypothesis-neutral approach (where all regions genome-wide are considered), we apply a Benjamini Hochberg correction for multiple comparisons.

We also tested the hypothesis that an abundance of significant negative similarity exists among SNPs in a given region. For this, we repeated the above steps using only SNPs showing dissimilarity ( $R2 < 0$ ). In this way, we allowed for a region to exhibit similarity and dissimilarity independently. This is important for gene-rich regions such as the MHC, which could potentially have a multifaceted role in mate preference.

d) Validation of observations in the MHC region was performed using the IMAGEN study. For each SNP, significance of the similarity score was assessed by 50,000 permutations. Regional scoring of the 3 MHC classes (Fisher meta value) was done in the same manner as the IMSGC. However, the procedure for assigning significance to the Fisher meta value was necessarily different from that used for the IMSGC dataset; with the IMAGEN dataset, we did not have the entire genome to use as a background. Instead, we created a regional background by shuffling couples 50,000 times, each time calculating the Fisher meta value on each of the 3 MHC classes. The percent of random Fisher meta values from the background that are lower than the observed Fisher meta value of each region is reported as the  $p$  value of that region.

e) For each of the six HLA genes from the IMAGEN study, a similarity score was calculated as follows. Each gene was given one point for each couple that shares one common allele at that gene, and two points for each couple that has both alleles in common. The similarity score for each gene was the total number of points across 920 couples. Couples were reassigned 20,000 times and similarity scores recalculated, creating a background distribution.  $P$  values were assigned using a cumulative normal distribution, with mean and standard deviation assessed from the background. Normality of the background distributions was assessed by visual inspection and the Anscombe-Glynn test of kurtosis.

f) For SNPs genotyped by PCR, significance of the similarity measure (Pearson correlation) was assessed by using a normal distribution with mean and standard deviation estimated from 200,000 random measurements (where the spouses were randomly re-assigned). Normality of the background distributions was assessed by visual inspection and the Anscombe-Glynn test of kurtosis.



### Linkage Disequilibrium

We checked that results are not affected by varying SNP density and linkage pattern across the genome by re-doing our analyses on reduced sets of approximately independent SNPs. Haplotype block tagging SNPs were selected genome-wide at  $r^2$  thresholds of 0.25, 0.5, and 0.75 using software Plink.

### Control for Ethnic Diversity

An earlier study with a large sample size ( $n = 1,017$  couples) that also found HLA similarity between couples suggested a possible confounding issue. The existence of ancestral or ethnic stratification with characteristic HLA types may influence the degree of genetic identity between couples [26]. Ancestry-related mate selection would appear as HLA-related selection because HLA is an excellent ancestry marker [67]. In our study, this issue is addressed foremost by comparing each candidate region or SNP to the entire genome. As a second layer of control, genome-wide pair-wise IBD distances (calculated with software Plink [40]) were used to cluster patients (using Ward agglomeration via the `hclust` function in R [68]) (Additional file 13: Figure S4). Outlying clusters of Mediterranean, Hispanic, Ashkenazi, and Eastern European couples were removed. All analysis was repeated on this smaller dataset of 803 couples. Results were largely unchanged.

In this paper, we chose to present the results from all 930 couples. When filtering for a more homogeneous western European population, we are removing a number of inter-group spouses (i.e. where one spouse is western European and the other is not). Inter-group mating events are a real phenomenon that we want to capture in the analysis.

### PCR

PCR validations and replications were done with a made-to-order Applied Biosystems TaqMan SNP genotyping assay, and carried out in 384-well plates using Applied Biosystems TaqMan genotyping Master Mix on an ABI Prism 7900HT Sequence Detection System using SDS 2.1 software.

### Imputation of HLA Alleles

Imputation of HLA Alleles from dense SNP coverage was performed by the authors of the IMAGEN paper [39]. HLA genotypes (at 2 digit resolution) were imputed from SNPs in the MHC using a recently developed approach [69]. The training database was from a previously created map of 7,500 SNPs, deletion insertion polymorphisms, and HLA alleles for 182 Utah residents (29 extended families containing 45 unrelated parent-offspring trios) of European ancestry in the Centre d'Etude du Polymorphisme Humain collection [70]. Up

to 40 SNPs were used to impute each HLA allele. Note the significant overlap between the training dataset used here to impute HLA types and the dataset used by Chaix et al to assess HLA similarity between spouses.

### Additional material

**Additional file 1: Table S1. Top similarities at the regional level using R1.** Each of the top regions was compared against all regions from the genome with lower recombination rate if the region is lower than average, or higher recombination rate if the region is higher than average. On the far right, we see what results would look like using the more homogeneous subset (803 couples of Western European descent) of the population (see Methods).

**Additional file 2: Figure S1. Average relatedness coefficient R1 between spouses at 3.6 Mb regions throughout the genome versus recombination rate.** More extreme values of R1 are seen in regions of lower recombination rates. The HLA region is shown in red. Compare to Figure 2 of Chaix et al.

**Additional file 3: Table S2. Genome-wide SNP-level results.** Approximately 4,000 most highly correlated SNPs among the 930 IMSGC couples. This includes positive and negative correlation. All SNPs with a one-tailed  $p$  value of 0.01 or better are highlighted by filter.

**Additional file 4: Table S3. Top similarities at the regional level using Fisher values.** Regions exhibiting an abundance of significantly similar SNPs ( $R2 > 0$ ).

**Additional file 5: Table S4. Top dissimilarities at the regional level using Fisher values.** Regions exhibiting an abundance of significantly dissimilar SNPs ( $R2 < 0$ ).

**Additional file 6: Figure S2. Validation on MHC results with the IMAGEN dataset.** In the screening IMSGC dataset, the MHC region (663 SNPs) was identified in the candidate-region approach as a mosaic of similarity and dissimilarity. 920 of the 930 couples were re-genotyped by a dense custom Illumina Platform (IMAGEN dataset: 1,078 SNPs passed quality control). (A) The pattern of similarity found in the IMAGEN dataset is comparable to that found in the screening (Figure 2). (B) 150 MHC SNPs were in common between IMAGEN and IMSGC. For each SNP passing quality control (94 SNPs), similarity between couples was calculated separately in both datasets. The correspondence of similarity scores between the two datasets was high ( $r^2 = 0.94$ ).

**Additional file 7: Table S5. IMAGEN regional p values.** Regional scoring of the 3 MHC classes (Fisher meta value) was done in the same manner as the IMSGC. P values are obtained by shuffling couples 50,000 times.

**Additional file 8: Table S6. Imputed classical HLA alleles.** Two of the class II genes, *DQA1* and *DQB1* showed significant dissimilarity between couples. Two-digit allele designations were used.

**Additional file 9: Table S7. Multiple-Sclerosis-associated SNPs.** Spousal identity ( $R2$ ) and uncorrected  $p$ -value for 11 SNPs associated with multiple sclerosis. After correction for 11 multiple comparisons, the spousal identity is not statistically significant.

**Additional file 10: Figure S3. Parental similarity versus offspring heterozygosity.** When parents choose mates that are similar to self at a given SNP, the result is excessive homozygosity in the children (an excess of homozygous genotypes at that SNP). Conversely, when parents choose mates that are dissimilar to self, the result is excessive heterozygosity in the children. In a simulation, random genotypes for 22,500 SNPs (2,500 with each MAF  $\epsilon(0.05, 0.1, 0.15, 0.2, 0.25, 0.3, 0.35, 0.4, 0.45)$ ) were generated for 1,000 sets of parents. At each SNP, the similarity measure (Pearson correlation) was calculated between the vectors of parental genotypes (shown on the y-axis). For each SNP, the genotypic frequencies of the offspring of the 1,000 sets of parents were calculated based on Mendelian inheritance. The observed frequency of heterozygotes in the offspring was divided by the expected frequency of heterozygotes, assuming Hardy Weinberg equilibrium (x-axis). A value higher than 1 on the x-axis means that offspring have a greater than

expected frequency of heterozygotes, while a value smaller than 1 on the x-axis means that offspring display excessive homozygosity. These plots show that SNPs which show similarity between parents (high values on the y-axis) are more likely to show excessive homozygosity in the offspring (low values on the x-axis). To extend the concept: if parents select mates that are similar to self at a given SNP, over many generations we expect excessive homozygosity in the general population compared to Hardy Weinberg equilibrium.

**Additional file 11: Text S1. Comparison of two measures of similarity.**

**Additional file 12: Text S2. Comparison of two methods of assessing significance of Pearson Correlation as a measure of similarity.**

**Additional file 13: Figure S4. Hierarchical clustering.** Using IBD distances calculated in software Plink, Ward agglomerative clustering (done in R) reveals a large cluster (A) of Scandinavian and western Europeans on the left. Smaller clusters on the right include (B) Eastern European (Russian and Polish) Ashkenazi Jews, (C) Mediterranean/Western European, (D) Hispanic with some Mediterranean, (E) Mediterranean, (F) non-Ashkenazi Eastern European. Self-reported ethnicity data was available for about 1/3 of the samples. This data is shown below the clusters. A red dot on the "Polish" row means that the person reports being Polish. A black dot means that the person did not report being Polish. The grey background means that no self-reported data was available for that person. Just above the self-reported ethnicity rows (black and red) is a single row showing cohort. Each sample belonged to one of three cohorts (UCSF = green, BWH = black, CMS = red). Note that nearly all samples from the non-western European group (B-F) came from the UCSF cohort.

#### Acknowledgements

We thank the International Multiple Sclerosis Genetics (IMSGC) and International MHC Autoimmunity Genetics Network (IMAGEN) consortia, which were responsible for the original data collection as well as the first level of quality control analysis. Sarah Hill provided editorial assistance. This work was supported by the National Multiple Sclerosis Society (NMSS) Collaborative Research Award CA 1035-A-7 (JRO and SEB), NMSS RG2901 (JRO), and RO1 NS26799 (SLH and JRO), which had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. SEB is a Harry Weaver Neuroscience Scholar of the NMSS. The International MS Genetics Consortium (Supported by RO1 NS049477) and the IMAGEN Consortium (supported by U19 AI067152) provided the data.

#### Authors' contributions

JRO, SLH, SEB, PAG, and PK conceived and designed the study. SJC carried out the rt-PCR assays. AS recoded HLA types and provided data management. PK performed the statistical analysis. JRO, SEB, and PK wrote the manuscript. All authors read and approved the final manuscript.

Received: 4 May 2010 Accepted: 10 November 2010

Published: 10 November 2010

#### References

- Freeman-Gallant CR, Meguerdichian M, Wheelwright NT, Sollecito SV: **Social pairing and female mating fidelity predicted by restriction fragment length polymorphism similarity at the major histocompatibility complex in a songbird.** *Mol Ecol* 2003, **12**:3077-3083.
- Knapp LA, Robson J, Waterhouse JS: **Olfactory signals and the MHC: a review and a case study in Lemur catta.** *Am J Primatol* 2006, **68**:568-584.
- Olsson KH, Grahm M, Lohm J, Langefors A: **MHC and kin discrimination in juvenile Arctic charr, *Salvelinus alpinus* (L.).** *Anim Behav* 1998, **56**:319-327.
- Olsson M, Madsen T, Nordby J, Wapstra E, Ujvari B, Wittsell H: **Major histocompatibility complex and mate choice in sand lizards.** *Proc Biol Sci* 2003, **270**(Suppl 2):S254-256.
- Yamazaki K, Beauchamp GK, Kupniewski D, Bard J, Thomas L, Boyse EA: **Familial imprinting determines H-2 selective mating preferences.** *Science* 1988, **240**:1331-1332.
- Yamazaki K, Boyse EA, Mike V, Thaler HT, Mathieson BJ, Abbott J, Boyse J, Zayas ZA, Thomas L: **Control of mating preferences in mice by genes in the major histocompatibility complex.** *J Exp Med* 1976, **144**:1324-1335.
- Carrington M, Nelson GW, Martin MP, Kissner T, Vlahov D, Goedert JJ, Kaslow R, Buchbinder S, Hoots K, O'Brien SJ: **HLA and HIV-1: heterozygote advantage and B\*35-Cw\*04 disadvantage.** *Science* 1999, **283**:1748-1752.
- McMichael AJ, Phillips RE: **Escape of human immunodeficiency virus from immune control.** *Annu Rev Immunol* 1997, **15**:271-296.
- Penn DJ, Damjanovich K, Potts WK: **MHC heterozygosity confers a selective advantage against multiple-strain infections.** *Proc Natl Acad Sci USA* 2002, **99**:11260-11264.
- Penn DJ, Potts WK: **The Evolution of Mating Preferences and Major Histocompatibility Complex Genes.** *The American Naturalist*, 1999 - UChicago Press 1999.
- Havlicek J, Roberts SC: **MHC-correlated mate choice in humans: a review.** *Psychoneuroendocrinology* 2009, **34**:497-512.
- Beauchamp GK, Yamazaki K: **HLA and mate selection in humans: commentary.** *Am J Hum Genet* 1997, **61**:494-496.
- Chaix R, Cao C, Donnelly P: **Is mate choice in humans MHC-dependent?** *PLoS Genet* 2008, **4**:e1000184.
- Derti A, Cenik C, Kraft P, Roth F: **Absence of Evidence for MHC-Dependent Mate Selection within HapMap Populations.** *PLoS Genet* 2010, **6**.
- Grob B, Knapp LA, Martin RD, Anzenberger G: **The major histocompatibility complex and mate choice: inbreeding avoidance and selection of good genes.** *Exp Clin Immunogenet* 1998, **15**:119-129.
- Jordan WC, Bruford MW: **New perspectives on mate choice and the MHC.** *Heredity* 1998, **81**(Pt 3):239-245.
- Roberts SC, Little AC: **Good genes, complementary genes and human mate preferences.** *Genetica* 2008, **134**:31-43.
- Yamazaki K, Beauchamp GK: **Genetic basis for MHC-dependent mate choice.** *Adv Genet* 2007, **59**:129-145.
- Ober C, Weitkamp LR, Cox N, Dytch H, Kostyu D, Elias S: **HLA and mate choice in humans.** *Am J Hum Genet* 1997, **61**:497-504.
- Hedrick PW, Black FL: **HLA and mate selection: no evidence in South Amerindians.** *Am J Hum Genet* 1997, **61**:505-511.
- Giphart MJ, D'Amaro J: **HLA and reproduction?** *J Immunogenet* 1983, **10**:25-29.
- Ihara Y, Aoki K, Tokunaga K, Takahashi K, Juji T: **HLA and Human Mate Choice: Tests on Japanese Couples.** *Anthropological Science* 2000, **108**:199-214.
- Nordlander C, Hammarstrom L, Lindblom B, Smith CI: **No role of HLA in mate selection.** *Immunogenetics* 1983, **18**:429-431.
- Sans M, Alvarez I, Callegari-Jacques SM, Salzano FM: **Genetic similarity and mate selection in Uruguay.** *J Biosoc Sci* 1994, **26**:285-289.
- Jin K, Speed TP, Thomson G: **Tests of random mating for a highly polymorphic locus: application to HLA data.** *Biometrics* 1995, **51**:1064-1076.
- Rosenberg LT, Cooperman D, Payne R: **HLA and mate selection.** *Immunogenetics* 1983, **17**:89-93.
- Coetzee V, Barrett L, Greeff JM, Henzi SP, Perrett DI, Wade AA: **Common HLA alleles associated with health, but not with facial attractiveness.** *PLoS One* 2007, **2**:e640.
- Lie HC, Rhodes G, Simmons LW: **Genetic diversity revealed in human faces.** *Evolution* 2008, **62**:2473-2486.
- Roberts SC, Little AC, Gosling LM, Jones BC, Perrett DI, Carter V, Petrie M: **MHC-assortative facial preferences in humans.** *Biol Lett* 2005, **1**:400-403.
- Roberts SC, Little AC, Gosling LM, Perrett DI, Carter V, Jones BC, Penton-Voak IS, Petrie M: **MHC-heterozygosity and human facial attractiveness.** *Evol Hum Behav* 2005, **26**:213-226.
- Thornhill R, Gangestad S, Miller R, Scheyd G, McCollough J, Franklin M: **Major histocompatibility complex genes, symmetry, and body scent attractiveness in men and women.** *Behav Ecol* 2003, **14**:668-678.
- Jacob S, McClintock MK, Zelano B, Ober C: **Paternally inherited HLA alleles are associated with women's choice of male odor.** *Nat Genet* 2002, **30**:175-179.
- Roberts SC, Gosling LM, Carter V, Petrie M: **MHC-correlated odour preferences in humans and the use of oral contraceptives.** *Proc Biol Sci* 2008, **275**:2715-2722.
- Santos PS, Schinemann JA, Gabardo J, Bicalho Mda G: **New evidence that the MHC influences odor perception in humans: a study with 58 Southern Brazilian students.** *Horm Behav* 2005, **47**:384-388.

35. Wedekind C, Furi S: **Body odour preferences in men and women: do they aim for specific MHC combinations or simply heterozygosity?** *Proc Biol Sci* 1997, **264**:1471-1479.
36. Wedekind C, Seebeck T, Bettens F, Paepke AJ: **MHC-dependent mate preferences in humans.** *Proc Biol Sci* 1995, **260**:245-249.
37. IMSGC IMSGC: **Risk alleles for multiple sclerosis identified by a genome-wide study.** *N Engl J Med* 2007, **357**:851-862.
38. Benjamini YHY: **Controlling the false discovery rate: a practical and powerful approach to multiple testing.** *J R Statist Soc B* 1995, **57**:289-300.
39. International MHC and Autoimmunity Genetics Network (IMAGEN), Rioux JD, Goyette P, Vyse TJ, Hammarstrom L, Fernando MMA, Green T, De Jager PL, Foisy S, Wang J, de Bakker PIW, Leslie S, McVean G, Padyukov L, Alfredsson L, Annesse V, Hafler DA, Pan-Hammarstrom Q, Pirskanen R, Sawcer SJ, Compston AD, Cree BAC, Mirel DB, Daly MJ, Behrens TW, Klareskog L, Gregersen PK, Oksenberg JR, Hauser SL: **Mapping of Multiple Susceptibility Variants Within the MHC Region for Seven Immune-Mediated Diseases.** *Proc Natl Acad Sci USA* 2009.
40. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, Sham PC: **PLINK: a tool set for whole-genome association and population-based linkage analyses.** *Am J Hum Genet* 2007, **81**:559-575.
41. Potts WK, Wakeland EK: **Evolution of MHC genetic diversity: a tale of incest, pestilence and sexual preference.** *Trends Genet* 1993, **9**:408-412.
42. Potts WK: **Wisdom through immunogenetics.** *Nat Genet* 2002, **30**:130-131.
43. Murphy MR, Schneider GE: **Olfactory bulb removal eliminates mating behavior in the male golden hamster.** *Science* 1970, **167**:302-304.
44. Gosling LM, Roberts SC: **Scent-Marking by Male Mammals: Cheat-Proof Signals to Competitors and Mates.** *Advances in the Study of Behavior* 2001 - Academic Press Limited 2001.
45. Sherborne AL, Thom MD, Paterson S, Jury F, Ollier WE, Stockley P, Beynon RJ, Hurst JL: **The genetic basis of inbreeding avoidance in house mice.** *Curr Biol* 2007, **17**:2061-2066.
46. Yamaguchi M, Yamazaki K, Beauchamp GK, Bard J, Thomas L, Boyse EA: **Distinctive urinary odors governed by the major histocompatibility locus of the mouse.** *Proc Natl Acad Sci USA* 1981, **78**:5817-5820.
47. Yeo TW, De Jager PL, Gregory SG, Barcellos LF, Walton A, Goris A, Fenoglio C, Ban M, Taylor CJ, Goodman RS, Walsh E, Wolfish CS, Horton R, Traherne J, Beck S, Trowsdale J, Caillier SJ, Ivinson AJ, Green T, Pobywajlo S, Lander ES, Pericak-Vance MA, Haines JL, Daly MJ, Oksenberg JR, Hauser SL, Compston A, Hafler DA, Rioux JD, Sawcer S: **A second major histocompatibility complex susceptibility locus for multiple sclerosis.** *Ann Neurol* 2007, **61**:228-236.
48. Bagert BA: **Epstein-Barr virus in multiple sclerosis.** *Curr Neurol Neurosci Rep* 2009, **9**:405-410.
49. Cullen M, Peretto SP, Klitz W, Nelson G, Carrington M: **High-resolution patterns of meiotic recombination across the human major histocompatibility complex.** *Am J Hum Genet* 2002, **71**:759-776.
50. Gandelman R, Zarrow MX, Denenberg VH, Myers M: **Olfactory bulb removal eliminates maternal behavior in the mouse.** *Science* 1971, **171**:210-211.
51. Rowe FA, Edwards DA: **Olfactory bulb removal: influences on the mating behavior of male mice.** *Physiol Behav* 1972, **8**:37-41.
52. Herz RSEDC: **Differential use of sensory information in sexual behavior as a function of gender.** *Hum Nat* 1997, **8**:275-289.
53. Wedekind C: **The MHC and body odors: arbitrary effects caused by shifts of mean pleasantness.** *Nat Genet* 2002, **31**:237, author reply 237.
54. Chen D, Haviland-Jones J: **Rapid mood change and human odors.** *Physiol Behav* 1999, **68**:241-250.
55. Havlicek J, Saxton TK, Roberts SC, Jozifkova E, Lhota S, Valentova J, Flegr J: **He sees, she smells? Male and female reports of sensory reliance in mate choice and non-mate-choice contexts.** *Pers Individ Diff* 2008, **45**:565-570.
56. Jacob S, Kinnunen LH, Metz J, Cooper M, McClintock MK: **Sustained human chemosignal unconsciously alters brain function.** *Neuroreport* 2001, **12**:2391-2394.
57. Saxton TK, Lyndon A, Little AC, Roberts SC: **Evidence that androstadienone, a putative human chemosignal, modulates women's attributions of men's attractiveness.** *Horm Behav* 2008, **54**:597-601.
58. Malnic B, Godfrey PA, Buck LB: **The human olfactory receptor gene family.** *Proc Natl Acad Sci USA* 2004, **101**:2584-2589.
59. Kleinjan DA, Lettice LA: **Long-range gene control and genetic disease.** *Adv Genet* 2008, **61**:339-388.
60. Kleinjan DA, van Heyningen V: **Long-range control of gene expression: emerging mechanisms and disruption in disease.** *Am J Hum Genet* 2005, **76**:8-32.
61. Lettice LA, Horikoshi T, Heaney SJ, van Baren MJ, van der Linde HC, Breedveld GJ, Joosse M, Akarsu N, Oostra BA, Endo N, Shibata M, Suzuki M, Takahashi E, Shinka T, Nakahori Y, Ayusawa D, Nakabayashi K, Scherer SW, Heutink P, Hill RE, Noji S: **Disruption of a long-range cis-acting regulator for Shh causes preaxial polydactyly.** *Proc Natl Acad Sci USA* 2002, **99**:7548-7553.
62. Nobrega MA, Ovcharenko I, Afzal V, Rubin EM: **Scanning human gene deserts for long-range enhancers.** *Science* 2003, **302**:413.
63. Pfeifer D, Kist R, Dewar K, Devon K, Lander ES, Birren B, Korniszewski L, Back E, Scherer G: **Campomelic dysplasia translocation breakpoints are scattered over 1 Mb proximal to SOX9: evidence for an extended control region.** *Am J Hum Genet* 1999, **65**:111-124.
64. Wunderle VM, Critcher R, Hastie N, Goodfellow PN, Schedl A: **Deletion of long-range regulatory elements upstream of SOX9 causes campomelic dysplasia.** *Proc Natl Acad Sci USA* 1998, **95**:10649-10654.
65. Rousset F: **Inbreeding and relatedness coefficients: what do they measure?** *Heredity* 2002, **88**:371-380.
66. Fisher R: **Combining independent tests of significance.** *American Statistician* 1948, **2**:30.
67. Sebro R, Hoffman TJ, Lange C, Rogus JJ, Risch NJ: **Testing for non-random mating: evidence for ancestry-related assortative mating in the Framingham heart study.** *Genet Epidemiol* 2010.
68. Murtagh F: **Multidimensional Clustering Algorithms.** *COMPSTAT Lectures 4 Wuerzburg: Physica-Verlag* 1985.
69. Leslie S, Donnelly P, McVean G: **A statistical method for predicting classical HLA alleles from SNP data.** *Am J Hum Genet* 2008, **82**:48-56.
70. de Bakker PI, McVean G, Sabeti PC, Miretti MM, Green T, Marchini J, Ke X, Monsuur AJ, Whittaker P, Delgado M, Morrison J, Richardson A, Walsh EC, Gao X, Galver L, Hart J, Hafler DA, Pericak-Vance M, Todd JA, Daly MJ, Trowsdale J, Wijmenga C, Vyse TJ, Beck S, Murray SS, Carrington M, Gregory S, Deloukas P, Rioux JD: **A high-resolution HLA and SNP haplotype map for disease association studies in the extended human MHC.** *Nat Genet* 2006, **38**:1166-1172.

doi:10.1186/1471-2164-11-626

Cite this article as: Khankhanian et al.: Genetic variation in the odorant receptors family 13 and the mhc loci influence mate selection in a multiple sclerosis dataset. *BMC Genomics* 2010 **11**:626.

**Submit your next manuscript to BioMed Central and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at  
www.biomedcentral.com/submit

