nature genetics

# Global diversity and evidence for coevolution of *KIR* and *HLA*

Richard M Single[1,7], Maureen P Martin[2,7], Xiaojiang Gao[2], Diogo Meyer[3], Meredith Yeager[4,5], Judith R Kidd[6], Kenneth K Kidd[6] & Mary Carrington[2]

© 2007 Nature Publishing Group http://www.nature.com/naturegenetics

**The killer immunoglobulin-like receptor (*KIR*) gene cluster shows extensive genetic diversity, as do the *HLA* class I loci, which encode ligands for KIR molecules. We genotyped 1,642 individuals from 30 geographically distinct populations to examine population-level evidence for coevolution of these two functionally related but unlinked gene clusters. We observed strong negative correlations between the presence of activating *KIR* genes and their corresponding *HLA* ligand groups across populations, especially KIR3DS1 and its putative *HLA-B Bw4-80I* ligands ($r = -0.66$, $P = 0.038$). In contrast, we observed weak positive relationships between the various inhibitory *KIR* genes and their ligands. We observed a negative correlation between distance from East Africa and frequency of activating *KIR* genes and their corresponding ligands, suggesting a balance between selection on *HLA* and *KIR* loci. Most *KIR-HLA* genetic association studies indicate a primary influence of activating *KIR-HLA* genotypes in disease risk[1,2]; concomitantly, activating receptor-ligand pairs in this study show the strongest signature of coevolution of these two complex genetic systems as compared with inhibitory receptor-ligand pairs.**

The *KIR* genes located on chromosome 19q13.4 encode activating and inhibitory receptors that are expressed on natural killer (NK) cells and a subset of T cells. The *KIR* gene cluster shows extensive variability in terms of gene content across haplotypes, probably because of non-allelic homologous recombination occurring between pairs of homologous *KIR* genes[3]. Over 30 distinct *KIR* haplotypes have been identified based on gene content alone[4], among which the most salient distinction is the presence or absence of various activating receptor genes. *HLA* class I genes are characterized by extreme allelic polymorphism and encode molecules that bind T cell receptors on cytotoxic T lymphocytes, thereby initiating acquired immune responses. More recently, HLA allotypes have been shown to serve as ligands for KIR, establishing their importance in both the innate and acquired immune response. The extreme variability at *KIR* and *HLA* loci is thought to provide protection against a wide variety of pathogens, with different *KIR-HLA* combinations conferring protection against distinct diseases. Several models for the maintenance of this degree of diversity have been posited, including frequency-dependent selection, heterozygote advantage and selection varying in time and space[5–7]. Overall, the concept of different KIR-HLA combinations protecting against distinct diseases is supported by disease association data[1,2]. For example, activating *KIR* genotypes tend to be protective against infectious diseases, but they are associated with susceptibility to autoimmune disease, whereas inhibitory genotypes are associated with protection against inflammatory diseases. The receptor-ligand specificity between KIR and HLA class I represents a classic example of genetic epistasis, where the presence of genes or alleles encoding corresponding receptor-ligand pairs is necessary for functional activity, but the presence of one without the other has no influence on effector cell activity.

Comparisons of *KIR* genes across primate species have demonstrated the rapid evolution of this receptor family—possibly more rapid than the evolution of the genes encoding their major histocompatibility complex (MHC) class I ligands[4,8,9]. For example, the inhibitory KIR specificities for particular amino acids (lysine or asparagine) at position 80 of the MHC-C molecule are preserved between humans and chimpanzees[8], but the MHC-C$^{lys80}$ specificity is mediated by KIRs with differing numbers of extracellular domains in chimpanzees and humans, suggesting particularly rapid evolution of the *KIR* genes. The extensive diversity of the *HLA* and *KIR* gene loci and the central role of their interactions in modulating immune responses are expected to favor the coevolution of genotypic combinations of these two loci in order to maintain appropriate functional interaction. Evidence of *HLA-KIR* coevolution has also been suggested in disease studies[1,2,10] as well as in the comparative genetic studies across primate species. However, direct evidence from human population studies, which could pinpoint receptor-ligand combinations that are major factors in their coevolution, has been lacking. Significant correlations between frequencies of specific *KIR* genes and *HLA* alleles

[1]The Department of Mathematics and Statistics, University of Vermont, Burlington, Vermont 05405, USA. [2]Laboratory of Genomic Diversity, SAIC-Frederick, National Cancer Institute-Frederick, Frederick, Maryland 21702, USA. [3]Departamento de Genética e Biologia Evolutiva, Universidade de São Paulo, Rua do Matão 277, São Paulo, SP 05608-900 Brazil. [4]Core Genotyping Facility, Advanced Technology Program, SAIC-Frederick, Inc., National Cancer Institute-Frederick, Maryland 21702, USA. [5]Division of Cancer Epidemiology and Genetics, National Cancer Institute, US National Institutes of Health, Department of Health and Human Services, Rockville, Maryland 20852, USA. [6]Department of Genetics, Yale University School of Medicine, 333 Cedar Street, New Haven, Connecticut 06510, USA. [7]These authors contributed equally to this work. Correspondence should be addressed to M.C. (carringt@ncifcrf.gov).

Received 5 March; accepted 13 June; published online 12 August 2007; doi:10.1038/ng2077

**Table 1 Populations, geographic regions, sample sizes and carrier frequencies**

| No. | Population[a] | Region[b] | *n* | Carrier frequencies | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | C-group1 | C-group2 | *Bw4* | *Bw4-80I* | *KIR2DL1* | *KIR2DL2* | *KIR2DL3* | *KIR3DL1* | *KIR2DS1* | *KIR2DS2* | *KIR3DS1* |
| 1 | Biaka | AFR | 69 | 69.4 | 72.6 | 78.7 | 49.2 | 97.1 | 58.0 | 80.3 | 100.0 | 1.4 | 56.5 | 2.9 |
| 2 | Ethiopian | AFR | 31 | 72.4 | 75.9 | 80.0 | 76.7 | 90.3 | 80.6 | 77.4 | 96.8 | 25.8 | 77.4 | 22.6 |
| 3 | Hausa | AFR | 37 | 63.6 | 84.8 | 78.8 | 66.7 | 100.0 | 36.7 | 96.7 | 100.0 | 13.8 | 32.1 | 3.4 |
| 4 | Ibo | AFR | 48 | 63.4 | 82.9 | 77.5 | 75.0 | 95.7 | 60.9 | 76.1 | 100.0 | 8.3 | 58.3 | 6.3 |
| 5 | Mbuti | AFR | 38 | 67.6 | 70.3 | 73.0 | 45.9 | 94.4 | 65.7 | 61.8 | 97.2 | 27.8 | 65.7 | 11.1 |
| 6 | Yoruba | AFR | 75 | 66.7 | 84.1 | 65.1 | 55.6 | 100.0 | 38.4 | 92.0 | 98.7 | 14.7 | 34.7 | 12.0 |
| 7 | Adygei | EUR | 54 | 71.7 | 79.2 | 77.8 | 57.4 | 100.0 | 50.0 | 92.5 | 98.1 | 35.2 | 50.0 | 35.2 |
| 8 | CEPH_UT | EUR | 90 | 84.4 | 54.4 | 62.2 | 25.6 | 96.7 | 56.7 | 88.9 | 94.4 | 40.0 | 58.9 | 38.9 |
| 9 | Danish | EUR | 50 | 82.9 | 61.0 | 50.0 | 20.0 | 98.0 | 52.0 | 96.0 | 98.0 | 36.7 | 52.0 | 38.0 |
| 10 | European | EUR | 91 | 81.6 | 65.5 | 56.0 | 26.2 | 97.8 | 46.2 | 92.3 | 93.3 | 37.8 | 46.7 | 40.0 |
| 11 | Finns | EUR | 35 | 77.1 | 62.9 | 51.4 | 25.7 | 100.0 | 39.4 | 97.0 | 90.9 | 54.5 | 39.4 | 53.1 |
| 12 | Irish | EUR | 94 | 88.2 | 58.8 | 65.5 | 19.5 | 98.9 | 52.1 | 91.5 | 96.8 | 28.0 | 52.1 | 31.9 |
| 13 | Russian | EUR | 47 | 84.8 | 60.9 | 67.4 | 26.1 | 93.5 | 43.5 | 86.4 | 97.9 | 23.4 | 43.5 | 21.3 |
| 14 | Druze | SWA | 116 | 73.9 | 80.2 | 68.5 | 57.4 | 96.5 | 71.9 | 78.4 | 96.5 | 23.7 | 71.7 | 30.7 |
| 15 | Yemenites | SWA | 43 | 83.9 | 80.6 | 69.0 | 58.6 | 97.7 | 46.5 | 95.2 | 97.7 | 31.0 | 50.0 | 35.7 |
| 16 | Ami | EAS | 40 | 97.5 | 32.5 | 5.0 | 0.0 | 92.3 | 62.5 | 84.6 | 100.0 | 37.5 | 60.0 | 32.5 |
| 17 | Atayal | EAS | 42 | 100.0 | 9.8 | 0.0 | 0.0 | 100.0 | 0.0 | 100.0 | 100.0 | 33.3 | 0.0 | 34.1 |
| 18 | Cambodian | EAS | 22 | 88.9 | 61.1 | 68.8 | 12.5 | 90.9 | 45.0 | 85.7 | 95.2 | 42.9 | 42.9 | 47.6 |
| 19 | Hakka | EAS | 40 | 100.0 | 30.0 | 55.3 | 31.6 | 100.0 | 33.3 | 97.4 | 97.5 | 50.0 | 33.3 | 52.5 |
| 20 | Han_SF | EAS | 59 | 100.0 | 37.5 | 68.1 | 40.4 | 100.0 | 13.6 | 100.0 | 98.3 | 37.3 | 13.6 | 35.6 |
| 21 | Han_Taiwan | EAS | 48 | 100.0 | 38.3 | 79.1 | 51.2 | 97.9 | 19.1 | 93.6 | 97.9 | 34.8 | 19.1 | 37.0 |
| 22 | Japan | EAS | 49 | 95.5 | 18.2 | 59.5 | 38.1 | 100.0 | 8.5 | 100.0 | 95.7 | 47.8 | 8.5 | 48.9 |
| 23 | Micronesia | OCE | 36 | 52.8 | 88.9 | 8.6 | 2.9 | 97.1 | 42.9 | 97.1 | 97.1 | 50.0 | 42.9 | 51.4 |
| 24 | Nasioi | OCE | 22 | 95.2 | 23.8 | 25.0 | 0.0 | 90.9 | 95.5 | 59.1 | 59.1 | 90.9 | 90.9 | 72.7 |
| 25 | Yakut | NEA | 51 | 91.7 | 75.0 | 87.5 | 45.8 | 98.0 | 27.5 | 92.2 | 96.1 | 46.0 | 37.3 | 42.0 |
| 26 | Maya | NAM | 50 | 89.4 | 38.3 | 8.7 | 6.5 | 94.0 | 38.0 | 94.0 | 94.0 | 62.0 | 38.0 | 56.0 |
| 27 | Pima | NAM | 99 | 92.6 | 59.6 | 23.2 | 13.7 | 87.9 | 63.6 | 87.6 | 76.8 | 76.8 | 63.6 | 71.7 |
| 28 | Karitiana | SAM | 55 | 80.0 | 52.0 | 0.0 | 0.0 | 85.5 | 67.3 | 85.5 | 81.8 | 80.0 | 67.3 | 80.0 |
| 29 | Surui | SAM | 46 | 90.5 | 54.8 | 7.1 | 7.1 | 97.8 | 26.1 | 97.8 | 91.3 | 47.8 | 26.1 | 30.4 |
| 30 | Ticuna | SAM | 65 | 92.3 | 53.8 | 31.3 | 18.8 | 93.8 | 39.1 | 93.8 | 93.8 | 61.9 | 39.1 | 46.0 |
| | Total | | 1642 | | | | | | | | | | | |

[a]Details of the sampling and ethnographic information for these populations can be found at http://alfred.med.yale.edu/. [b]AFR = Africa; EUR = Europe; SWA = Southwest Asia; EAS = East Asia; OCE = Oceania; NEA = Northeast Asia; NAM = North America; SAM = South America.

© 2007 Nature Publishing Group http://www.nature.com/naturegenetics

encoding their corresponding ligands would support the hypothesis that these unlinked loci are coevolving.

To assess evidence for coevolution of *KIR* and *HLA* class I at the population level, we genotyped 1,642 unrelated individuals from 30 geographically distinct populations throughout the world (**Table 1** and **Fig. 1**). Each individual was tested for the presence or absence of specific *KIR* genes whose ligands have been defined previously. We grouped *HLA-B* and *HLA-C* alleles according to known ligand specificity for the following inhibitory receptors: (i) *KIR2DL2* and *KIR2DL3*, which segregate as alleles of a single locus, encode molecules that bind HLA-C allotypes with Asp80 (C-group1 alleles); (ii) KIR2DL1 binds HLA-C allotypes with Lys80 (C-group2 alleles) and (iii) KIR3DL1 binds HLA-B allotypes that contain the Bw4 epitope (*Bw4* alleles), which is defined by amino acid variation at positions 77–83 of the HLA-B molecules. Data from three studies indicate that Bw4 molecules with Ile80 (*Bw4-80I* alleles) serve as better ligands for KIR3DL1 than do Bw4 molecules with Thr80 (*Bw4-80T* alleles)[11–13]. Therefore, we considered these two subgroups in analyses involving *Bw4*. The activating KIR2DS2 and KIR2DS1 bind to HLA-C-group1 allotypes and HLA-C-group2 allotypes, respectively, with low affinity. The ligand for KIR3DS1 is not known definitively, but genetic epidemiological studies[14,15] point to Bw4-80I as its ligand, which is consistent with the strong similarity between the extracellular domains of KIR3DS1 and KIR3DL1. Similarly, the activating KIR2DS2 and KIR2DS1 show 97%–98% sequence similarity in their extracellular domains to the corresponding inhibitory KIRs (KIR2DL2 and KIR2DL1, respectively). Frequencies of the presence or absence (that is, phenotypic or carrier frequencies) of the *KIR* genes and the *HLA-B* and *HLA-C* ligand groups are shown in **Figure 2**.

We investigated population-level evidence for coevolution of the *KIR* and *HLA* loci by determining correlations between the frequencies of functionally relevant receptor-ligand pairs. For each pair, we assessed the significance of these correlations using empirical methods based on genomic comparisons to avoid issues of statistical non-independence among populations and to account for the effects of demographic history[16] (see Methods and **Supplementary Note** online). The phenotypic frequencies for the activating *KIR3DS1* and *Bw4* ligand group showed a strong negative correlation ($r = -0.63$; $P_{empirical} = 0.041$; **Fig. 3** and **Table 2**) that was enhanced slightly when we considered the *Bw4-80I* subset of *Bw4* alleles ($r = -0.66$; $P_{empirical} = 0.038$). In fact, *Bw4-80I* accounts almost completely for the correlation observed between *KIR3DS1* and *Bw4*, as there was only a minimal negative correlation between *KIR3DS1* and the alternative *Bw4-80T* subset ($r = -0.19$; *P* value not significant). For the activating *KIR2DS1*

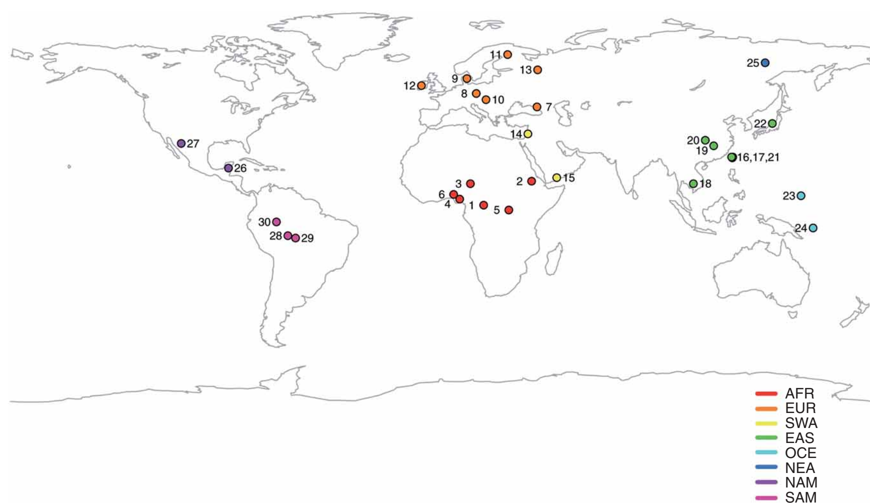


**Figure 1** Map of the world showing the location of sampled populations. Numbers correspond to populations in **Table 1**.

and *KIR2DS2* genes, the correlation with C-group2 and C-group1, respectively, was also negative ($r$ = –0.48 and –0.37, respectively; **Supplementary Fig. 1** online) but not significant. *KIR3DL1*, a common inhibitory gene, showed weak, nonsignificant positive correlations with its *Bw4* ligand group and the *Bw4-80I* subset (**Supplementary Fig. 1**). The population-level correlation between *KIR3DL1* and *Bw4-80T* frequencies was even weaker, corresponding with data indicating that Bw4-80I molecules serve as better ligands for KIR3DL1 (refs. 11–13). For the inhibitory *KIR2DL1* and *KIR2DL3* genes, the positive correlation with their ligands was very weak and not significant.

C-group2 alleles and *Bw4* alleles are in significant positive linkage disequilibrium (LD) with one another ($P$ < 0.001), as are *KIR3DS1* and *KIR2DS1* ($P$ < 0.001). Therefore, it is possible that only one receptor-ligand pair is the dominant factor in shaping the frequency distributions of these two activating *KIR-HLA* ligand pairs and that members of the other pair are simply hitchhiking. Owing to the degree of LD for both sets of genes, it is difficult to assess the individual effects of the two receptor-ligand pairs based on these population-level data. However, the relationship between *KIR3DS1* and *Bw4-80I* frequencies was stronger and more significant than that between any other receptor-ligand pairing, lending credence to the putative receptor-ligand relationship between KIR3DS1 and Bw4-80I molecules, as implicated in previous disease association studies[14,15,17].

Although the direction of the KIR-HLA associations was predicted *a priori*, we have reported two-sided tests throughout the paper. Even though the negative correlation between *KIR3DS1* and *Bw4* or *Bw4-80I* was the only one to reach statistical significance, the general pattern of negative and positive correlations between activating and inhibitory receptor-ligand pairs, respectively, was very consistent. The single exception is the negative correlation between *KIR2DL2* and C-group1, which is in contrast with the positive correlations between other inhibitory genes tested and their respective ligands. *KIR2DS2* and *KIR2DL2* were virtually in complete LD in all populations tested (except for the Yakut), explaining the identical C-group1 correlations with *KIR2DS2* and *KIR2DL2*. Several lines of evidence suggest that *KIR2DL2* is simply tracking an interaction between *KIR2DS2* and C-group1: (i) the fairly weak correlations between other inhibitory receptors and their ligand groups, (ii) the positive correlations between the other inhibitory receptors and their ligand groups, (iii) the negative correlation between *KIR2DS2* and its ligand group, consistent with respect to the other activating *KIR* tested and (iv) the fact that *KIR2DL3* recognizes the same ligand group as *KIR2DL2*, yet the correlations with C-group1 are diametrically opposed.

In order to investigate the geographical distribution of gene frequencies, we tested *KIR* and *HLA* ligand group frequencies for correlations with distance from East Africa based on previous results[18] and the latitude and longitude values for the population samples listed in the ALFRED database (**Table 3**). We observed significant positive clines in the frequencies of *KIR3DS1* and *KIR2DS1* with increasing distance from East Africa ($R^2$ > 45%, $P_{empirical}$ = 0.01); in contrast, we observed more subtle declines in the frequencies of their corresponding ligands, *Bw4-80I* ($R^2$ > 35%, $P_{empirical}$ = 0.004) and C-group2 ($R^2$ > 20%; $P$ value not significant), with increasing distance from East Africa. These data are consistent with the negative correlations that we observed between these activating receptor-ligand pairs (**Fig. 3b** and **Supplementary Fig. 1g**) and indicate an inverse correlation with respect to geography. We have necessarily restricted our analyses to certain *KIR-HLA* pairs with known or suspected receptor-ligand relationships. Although this subset may not represent the entire pattern of coevolution at the *HLA* class I and *KIR* loci, overall, the data suggest the maintenance of a balance between the frequencies of the activating receptors and their corresponding (putative) ligands across populations.

We found it intriguing that some pairs of neighboring populations shared similar *HLA* ligand group frequencies but had highly distinct phenotypic *KIR* gene frequencies (particularly in East Asia and the Americas; **Fig. 2**), as this is consistent with a situation in which the *KIR* genes are evolving at a more rapid rate than the *HLA* class I ligand groupings. A particularly notable example is the two tribal populations from Taiwan: the Atayal from the inland mountain ranges and the Ami from the east coast, which presumably stem from common ancestors settling Taiwan roughly 4,000 to 6,000 years ago[19]. The Atayal were completely missing *KIR2DS2* and *KIR2DL2*. In contrast, the Ami sample had some of the highest levels of *KIR2DL2* and *KIR2DS2* worldwide. However, both the Ami and Atayal had very low frequencies of *Bw4*. Additional geographic regions with several neighboring but distinct populations will need to be surveyed for both *KIR* and *HLA* to corroborate and extend these results.

Signatures of balancing selection are evident not only at the *HLA* class I allele level[20] but also at the amino acid level[21], similar to reports on HLA class II data[22]. Notably, recent work has shown that amino acids at residues 77 and 80 in HLA-C, which are in strong LD, were among those with the strongest signals for balancing selection relative to other polymorphic HLA-C residues (S.J. Mack and H.A. Erlich, personal communication). The importance of position 80 in determining alternative KIR2D ligand specificities, along with its involvement in peptide binding, raises the possibility of selection driven by both pathogens and *KIR-HLA* interactions. A very similar situation has been found for position 80 of HLA-B, which appears in a cluster of high LD (residues 80–83) showing a signature of balancing selection (S.J. Mack and H.A. Erlich, personal communication) and which determines KIR3DL1 and KIR3DS1 specificities. Overall, these data

© 2007 Nature Publishing Group http://www.nature.com/naturegenetics

agree nicely with the balanced relationship between activating receptors and their ligands (or putative ligands) shown in **Figure 3** and give further credibility to the coevolution of the *KIR* and *HLA* loci.

Taken together, the data presented here provide population-level evidence for the evolution of the *KIR* gene cluster owing to selection pressure favoring frequencies of activating *KIR* that suit the locale-specific *HLA* repertoire. Although the only significant correlation between receptor-ligand pairs involved KIR3DS1 and Bw4-80I using our empirical method ($r = -0.66$, $P_{\mathrm{empirical}} = 0.038$), we observed clear negative correlations between each of the other activating KIRs and their corresponding ligand groups. The strong correlation between activating, as opposed to inhibitory, *KIR* with *HLA* ligand makes sense, as NK cell inhibition is essential for preventing harmful responses to self under healthy conditions and is therefore a fixed trait. Thus, it is important that genes encoding inhibitory receptors are common members of *KIR* haplotypes, and indeed there is always at least one known inhibitory *KIR-HLA* ligand pairing present across virtually all humans. The activating *KIR* genes, on the other hand, are present on some, but not all, *KIR* haplotypes, and there are a fair number of haplotypes that do not contain any activating *KIR* for which a ligand is known. Activating KIRs probably serve as a fine-tuning mechanism: they enhance protection against certain diseases

© 2007 Nature Publishing Group http://www.nature.com/naturegenetics
npg

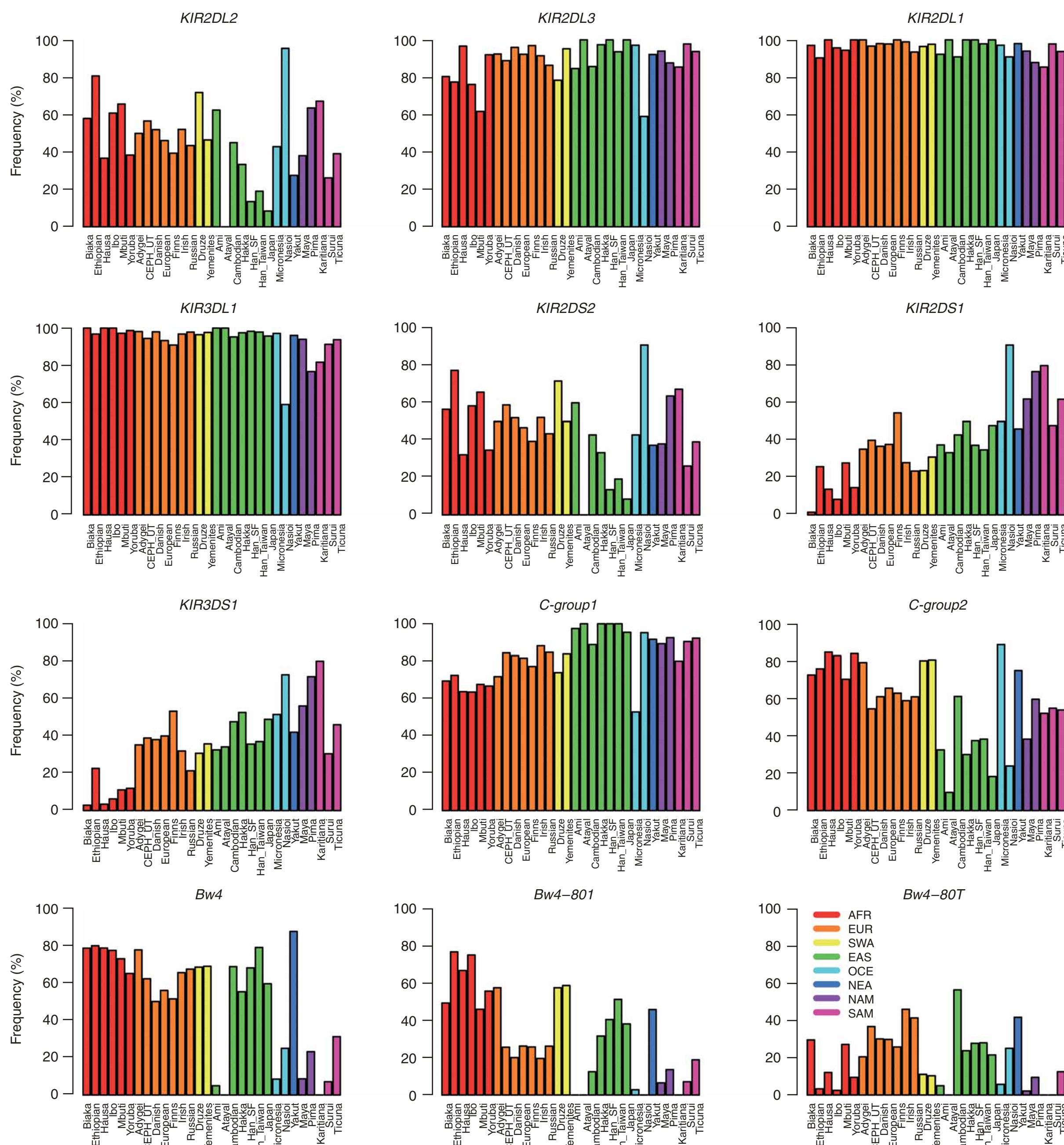


**Figure 2** *KIR* and *HLA* ligand carrier frequencies across populations. Populations are grouped by geographic region, and regions are ordered by increasing distance from East Africa. Two activating *KIR* loci (*KIR2DS1* and *KIR3DS1*) show an increasing trend in phenotypic frequency with increasing geographic distance from East Africa. An inverse relationship is seen for the *Bw4 HLA* ligand group. This trend is stronger for the *Bw4-80I* subset of *Bw4* ligands.

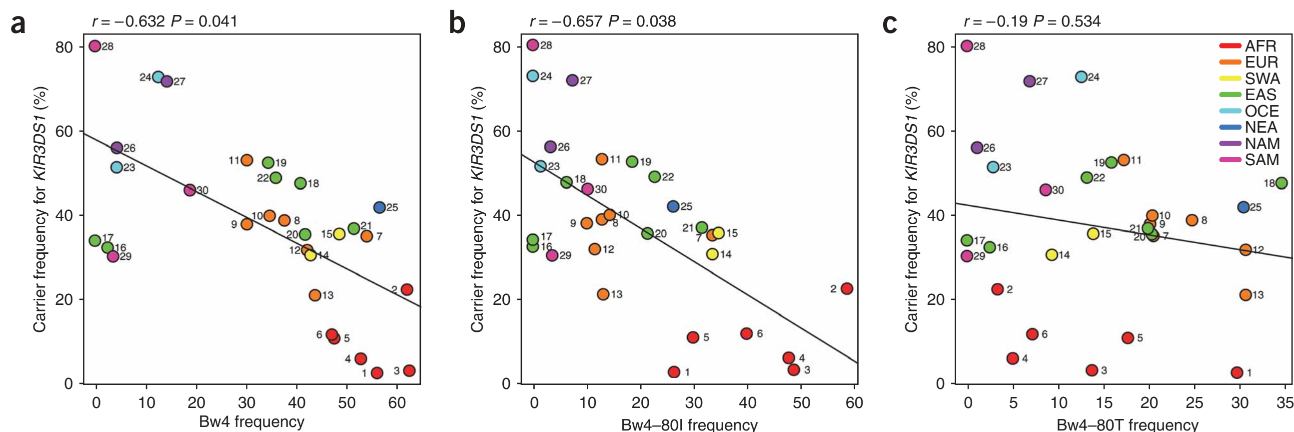


**Figure 3** Correlation between *KIR* and *HLA* ligand frequencies. (**a**) There is a strong negative correlation between the carrier frequencies for activating *KIR3DS1* and gene frequencies for its putative *HLA Bw4* ligand group. (**b**,**c**) The subset of *Bw4-80I* alleles (**b**) accounts for the majority of the relationship of *KIR3DS1* with *Bw4*, as there is little correlation between the frequency of this *KIR* gene and that of *Bw4-80T* alleles (**c**). There is also a negative, but nonsignificant, correlation for other activating KIRs (**Supplementary Figure 1g**,**h**) and a weaker nonsignificant positive correlation for inhibitory KIRs (**Supplementary Figure 1a**,**b**,**d**,**f**). Numbers correspond to populations in **Table 1**, and the solid line represents the estimated regression line.

but may actually confer susceptibility to others. This range in *KIR-HLA* genotypes, conferring greater activation to greater inhibition, may correlate well with the level of resistance to different types of complex diseases, where the underlying mechanisms of pathogenesis vary as a result of the level of immune responsiveness.

## METHODS

**Populations.** In this study, we genotyped 29 populations from the Allele Frequency Database (ALFRED), a freely accessible collection of allele frequency data in anthropologically defined human populations[23]. The populations were classified as belonging to one of the following eight geographic regions: Africa (AFR), Europe (EUR), Southwest Asia (SWA), East Asia (EAS), Oceania (OCE), Northeast Asia (NEA), North America (NAM) or South America (SAM). In addition, 90 Centre d'Etude du Polymorphisme Humain (CEPH) individuals were included in this study. These CEPH individuals from Utah were classified as belonging to the European geographic region based on ancestry. Population names and sample sizes are given in **Table 1**, and the geographic distribution of these populations is given in **Figure 1**. This study was approved by the Protocol Review Office of the National Cancer Institute institutional review board. Informed consent was obtained at the study sites from all participants.

**Table 2 Correlation between KIR and HLA ligand frequencies**

| Locus pair | Correlation (*r*)[a] | Empirical *P* value[b] |
|---|---|---|
| *KIR3DL1*, *Bw4* | 0.426 | 0.218 |
| *KIR3DL1*, *Bw480i* | 0.416 | 0.191 |
| *KIR3DL1*, *Bw480t* | 0.171 | 0.758 |
| *KIR3DS1*, *Bw4* | –0.632 | 0.041 |
| *KIR3DS1*, *Bw480i* | –0.657 | 0.038 |
| *KIR3DS1*, *Bw480t* | –0.190 | 0.534 |
| *KIR2DL1*, C-group2 | 0.046 | 0.924 |
| *KIR2DL2*, C-group1 | –0.366 | 0.542 |
| *KIR2DL3*, C-group1 | 0.184 | 0.328 |
| *KIR2DS1*, C-group2 | –0.478 | 0.149 |
| *KIR2DS2*, C-group1 | –0.371 | 0.479 |

[a]Pearson product-moment correlation between KIR carrier frequency (CF) and HLA ligand gene frequency (GF). [b]The empirical *P* value is the proportion of times that a larger absolute correlation than the true correlation between KIR GF and HLA GF was observed in the empirical distribution, generated by computing correlations between GFs for 10,000 pairs of unlinked sites typed in the same populations as the present study from the ALFRED database (see Methods).

***KIR* genotyping.** Genomic DNA was genotyped for presence or absence of the following KIR genes: *KIR2DL1*, *KIR2DL2*, *KIR2DL3*, *KIR2DL4*, *KIR2DL5*, *KIR2DS1*, *KIR2DS2*, *KIR2DS3*, *KIR2DS4*, *KIR2DS5*, *KIR3DL1*, *KIR3DL2*, *KIR3DL3*, *KIR3DS1*, *KIR2DP1* and *KIR3DP1*. Genotyping was performed using PCR amplification with two pairs of primers specific for each locus (PCR-SSP) as previously described[24,25]. The KIR genes are named based on their structure, including the number of extracellular domains they possess (for example, 2D or 3D) and the length of their intracytoplasmic tail (L for long, or S for short). KIR genes with a long intracytoplasmic tail send inhibitory signals, and those with a short intracytoplasmic tail send activating signals.

***HLA* class I genotyping.** Genomic DNA was amplified using locus-specific primers flanking exons 2 and 3. The PCR products were blotted on nylon membranes and hybridized with a panel of sequence-specific oligonucleotide (SSO) probes. Alleles were assigned by the reaction patterns of the SSO probes. Ambiguous SSOP typing results were resolved by sequencing analysis.

**Table 3 Relationship between distance from East Africa and *KIR/HLA* frequency**

| | Distance from East Africa | | | | |
|---|---|---|---|---|---|
| Locus | Slope[a] | $R^2$ [b] | *P* value[c] (regression) | *P* value[d] (empirical) |
| *KIR2DL1* | –0.000164 | 5.1% | 0.247 | 0.098 |
| *KIR2DL2* | –0.000018 | 0.5% | 0.731 | 0.747 |
| *KIR2DL3* | 0.000072 | 1.8% | 0.497 | 0.327 |
| *KIR3DL1* | –0.000248 | 17.7% | 0.026 | 0.019 |
| *KIR2DS1* | 0.000127 | 47.7% | <0.001 | 0.010 |
| *KIR2DS2* | –0.000020 | 0.6% | 0.691 | 0.745 |
| *KIR3DS1* | 0.000118 | 45.6% | <0.001 | 0.010 |
| C-group2 | –0.000052 | 20.3% | 0.016 | 0.286 |
| *Bw4* | –0.000197 | 32.4% | 0.002 | 0.022 |
| *Bw4-80I* | –0.000314 | 35.3% | 0.001 | 0.004 |
| *Bw4-80T* | –0.000163 | 18.7% | 0.022 | 0.083 |

[a]Slope of the relationship between GF* = log(GF/(1 – GF)) on the distance from East Africa, where GF is the gene frequency. [b]Proportion of variation explained ($R^2$) values for the regression of GF* on the distance from East Africa. [c]The ordinary regression-based *P* value. [d]The empirical *P* value is the proportion of times that a stronger relationship between GF* and distance from East Africa was observed in the empirical distribution, generated by computing the above regression for 538 sites from the ALFRED database typed in the same populations as the present study (see Methods).

© 2007 Nature Publishing Group http://www.nature.com/naturegenetics

**Statistical methods.** For *HLA-B* and *HLA-C*, analyses of allele frequencies (gene frequencies) were carried out using the PyPop software package[26]. The methods implemented in PyPop are given in ref. 27. *KIR* carrier frequencies (CF) were determined for each of the 30 population samples. Gene frequencies (GF), assuming HWP, were computed via Bernstein's formula[28] as $GF_i = 1 - \sqrt{(1 - CF_i)}$.

The significance of the correlation between *KIR* and *HLA* frequencies was assessed using both empirical and permutation-based approaches to avoid issues of non-independence among populations and to account for the effects of demographic history. In each case, this was accomplished by comparing the observed statistic to a null distribution of values in the same set of populations to account for any correlation induced by shared history. The ordinary *P* value for the Pearson product moment correlation does not account for this non-independence. Although the direction of the *KIR-HLA* associations was predicted *a priori*, we have reported more conservative two-sided tests throughout the paper based on the methods described below.

In the empirical approach, we used gene frequencies for 538 sites in 202 genes, typed in the same set of populations as the present study and provided in the ALFRED database, to generate an empirical distribution of correlations between gene frequencies for pairs of unlinked sites. The empirical *P* value ($P_{empirical}$) represents the proportion of times that the correlation between two randomly chosen sites on different chromosomes was larger in absolute value than the correlation between *KIR* gene frequencies and *HLA* gene frequencies. Each of the sites that we used had a heterozygosity of at least 0.25.

The empirical distribution used to assess the strength of the relationship between the distance from East Africa and the *KIR* gene frequency (or *HLA* gene frequency) was based on the results relating distance from East Africa and each of the 538 sites from ALFRED. The dependent variable used in the regressions was $GF^* = \log(GF / (1 - GF))$, where GF is the gene frequency. The empirical *P* value represents the proportion of times that transformed gene frequencies (GF*) in the empirical distribution showed a stronger relationship with distance from East Africa than did *KIR* (or the *HLA* ligand grouping). A weight of $1/k$ was used when there were *k* sites in the same gene.

Fisher's exact test was used to test for nonrandom association between the presence of pairs of *KIR* genes. The strength of the association was measured using Cramer's *V* statistic[29]. This measure of association, sometimes referred to as $W_n$, is equivalent to the correlation coefficient ($r = \sqrt{D_{11}^2/p_1p_2q_1q_2}$) between two biallelic loci but has a range from –1 to +1. Fisher's method for combining *P* values[30] was used to test for significant LD across a set of populations for a locus pair.

**URLs.** Details of the sampling and ethnographic information for the studied populations can be found at http://alfred.med.yale.edu/. SSO protocols are available at http://www.ihwg.org/protocols/protocol.htm. PyPop is available at http://www.pypop.org/.

*Note: Supplementary information is available on the Nature Genetics website.*

**ACKNOWLEDGMENTS**
We thank H. Rajeevan and M. Osier for their assistance with the ALFRED database. This project has been funded in whole or in part with federal funds from the US National Cancer Institute (National Institutes of Health) under contract N01-CO-12400. The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products or organizations imply endorsement by the US Government. This research was supported in part by the Intramural Research Program of the US National Institutes of Health (NIH), National Cancer Institute, Center for Cancer Research. R.M.S. was supported by NIH grant GM35326, Department of Energy grant DE-FG02-00ER45828 and the Vermont Advanced Computing Center under NASA grant NNG-05GO96G. D.M. was supported by FAPESP grant 03/08973-6.

**AUTHOR CONTRIBUTIONS**
M.C. designed and supervised the project and prepared the manuscript; R.M.S. contributed to the design of the project, developed the statistical methods and prepared the manuscript; M.P.M. performed *KIR* genotyping and prepared the manuscript; X.G. performed HLA genotyping; D.M. and M.Y. provided intellectual input; J.R.K. and K.K.K. provided samples and intellectual input; all authors discussed the results and commented on the manuscript.

**COMPETING INTERESTS STATEMENT**
The authors declare no competing financial interests.

Published online at http://www.nature.com/naturegenetics
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions

1. Carrington, M. & Martin, M.P. The impact of variation at the KIR gene cluster on human disease. *Curr. Top. Microbiol. Immunol.* **298**, 225–257 (2006).
2. Parham, P. MHC class I molecules and KIRs in human history, health and survival. *Nat. Rev. Immunol.* **5**, 201–214 (2005).
3. Carrington, M. & Cullen, M. Justified chauvinism: Advances in defining meiotic recombination through sperm typing. *Trends Genet.* **20**, 196–205 (2004).
4. Khakoo, S.I. & Carrington, M. KIR and disease: A model system or system of models? *Immunol. Rev.* **214**, 186–201 (2006).
5. Apanius, V., Penn, D., Slev, P.R., Ruff, L.R. & Potts, W.K. The nature of selection on the major histocompatibility complex. *Crit. Rev. Immunol.* **17**, 179–224 (1997).
6. Borghans, J.A., Beltman, J.B. & De Boer, R.J. MHC polymorphism under host-pathogen coevolution. *Immunogenetics* **55**, 732–739 (2004).
7. Hedrick, P.W. Pathogen resistance and genetic variation at MHC loci. *Evolution Int. J. Org. Evolution* **56**, 1902–1908 (2002).
8. Khakoo, S.I. *et al.* Rapid evolution of NK cell receptor systems demonstrated by comparison of chimpanzees and humans. *Immunity* **12**, 687–698 (2000).
9. Vilches, C. & Parham, P. KIR: Diverse, Rapidly Evolving Receptors of Innate and Adaptive Immunity. *Annu. Rev. Immunol.* **20**, 217–251 (2002).
10. Hiby, S.E. *et al.* Combinations of maternal KIR and fetal HLA-C genes influence the risk of preeclampsia and reproductive success. *J. Exp. Med.* **200**, 957–965 (2004).
11. Carr, W.H., Pando, M.J. & Parham, P. KIR3DL1 polymorphisms that affect NK cell inhibition by HLA-Bw4 ligand. *J. Immunol.* **175**, 5222–5229 (2005).
12. Cella, M., Longo, A., Ferrara, G.B., Strominger, J.L. & Colonna, M. NK3-specific natural killer cells are selectively inhibited by Bw4- positive HLA alleles with isoleucine 80. *J. Exp. Med.* **180**, 1235–1242 (1994).
13. Gumperz, J.E. *et al.* Conserved and variable residues within the Bw4 motif of HLA-B make separable contributions to recognition by the NKB1 killer cell- inhibitory receptor. *J. Immunol.* **158**, 5237–5241 (1997).
14. Martin, M.P. *et al.* Epistatic interaction between KIR3DS1 and HLA-B delays the progression to AIDS. *Nat. Genet.* **31**, 429–434 (2002).
15. Qi, Y. *et al.* KIR/HLA pleiotropism: Protection against both HIV and opportunistic infections. *PLoS Pathog.* **2**, e79 (2006)(doi:10.1371/journal.ppat.0020079).
16. Bamshad, M. & Wooding, S.P. Signatures of natural selection in the human genome. *Nat. Rev. Genet.* **4**, 99–111 (2003).
17. Lopez-Vazquez, A. *et al.* Protective effect of the HLA-Bw4I80 epitope and the killer cell immunoglobulin-like receptor 3DS1 gene against the development of hepatocellular carcinoma in patients with hepatitis C virus infection. *J. Infect. Dis.* **192**, 162–165 (2005).
18. Prugnolle, F. *et al.* Pathogen-driven selection and worldwide HLA class I diversity. *Curr. Biol.* **15**, 1022–1027 (2005).
19. Su, B. *et al.* Polynesian origins: insights from the Y chromosome. *Proc. Natl. Acad. Sci. USA* **97**, 8225–8228 (2000).
20. Meyer, D., Single, R.M., Mack, S.J., Erlich, H.A. & Thomson, G. Signatures of demographic history and natural selection in the human major histocompatibility complex Loci. *Genetics* **173**, 2121–2142 (2006).
21. Hedrick, P.W., Whittam, T.S. & Parham, P. Heterozygosity at individual amino acid sites: extremely high levels for HLA-A and -B genes. *Proc. Natl. Acad. Sci. USA* **88**, 5897–5901 (1991).
22. Salamon, H. *et al.* Evolution of HLA class II molecules: allelic and amino acid site variability across populations. *Genetics* **152**, 393–400 (1999).
23. Osier, M.V. *et al.* ALFRED: An allele frequency database for anthropology. *Am. J. Phys. Anthropol.* **119**, 77–83 (2002).
24. Martin, M.P. & Carrington, M. KIR locus polymorphisms: Genotyping and disease association analysis. in *Methods in Molecular Biology: Innate Immunity* (eds. Ewbank, J. & Vivier, E.) (Humana Press, Totowa, New Jersey, in the press).
25. Martin, M.P. *et al.* Cutting edge: Susceptibility to psoriatic arthritis: Influence of activating killer Ig-like receptor genes in the absence of specific HLA-C alleles. *J. Immunol.* **169**, 2818–2822 (2002).
26. Lancaster, A., Nelson, M.P., Meyer, D., Thomson, G. & Single, R.M. PyPop: a software framework for population genomics: analyzing large-scale multi-locus genotype data. *Pac. Symp. Biocomput.* 514–525 (2003).
27. Single, R.M., Meyer, D. & Thomson, G. Statistical methods for analysis of population genetic data. in *Immunobiology of the Human MHC: Proceedings of the 13th International Histocompatibility Workshop and Conference* Vol. 1 (ed. Hansen, J.A.) (IHWG Press, Seattle, 2007).
28. Williams, R.C., Steinberg, A.G., Knowler, W.C. & Pettitt, D.J. Gm 3;5,13,14 and stated-admixture: independent estimates of admixture in American Indians. *Am. J. Hum. Genet.* **39**, 409–413 (1986).
29. Cramer, H. *Mathematical Methods of Statistics* (Princeton Univ. Press, Princeton, New Jersey, 1946).
30. Fisher, R. *Statistical Methods for Research Workers* (Oliver & Boyd, Edinburgh, 1970).

© 2007 Nature Publishing Group http://www.nature.com/naturegenetics