Published in final edited form as:

Immunogenetics. 2012 February ; 64(2): 97–109. doi:10.1007/s00251-011-0569-x.

KIR Genotypic Diversity Can Track Ancestries in Heterogeneous Populations: A Potential Confounder for Disease Association Studies

Komal Manpreet Singh¹, Yume T. Phung¹, Mohamed S. Kohla^{1,6}, Billy Y-A Lan⁴, Sharon Chan⁴, Diana L. Suen¹, Sahar Murad^{1,3}, Shana Rheault¹, Peter Davidson⁵, Jennifer Evans⁵, Manpreet Singh⁷, Sofie Dohil¹, Robert W. Osorio³, Adil E. Wakil^{1,2,3}, Kimberly Page⁵, Sandy Feng⁴, and Stewart L. Cooper^{1,2,3,§}

¹Kalmanovitz Liver Immunology Laboratory, California Pacific Medical Center & Research Institute, San Francisco, California, USA

²Division of Hepatology, California Pacific Medical Center & Sutter Pacific Medical Foundation, San Francisco, California, USA

³Department of Transplantation, California Pacific Medical Center, San Francisco, California, USA

⁴Department of Transplantation, University of California, San Francisco, San Francisco, California, USA

⁵Department of Epidemiology, University of California, San Francisco, San Francisco, California, USA

⁶The National Liver Institute, Menoufiya, Egypt

Abstract

Killer cell immunoglobulin-like receptors (KIR) are encoded by highly polymorphic genes that regulate the activation of natural killer (NK) cells and other lymphocyte subsets, and likely play key roles in innate and adaptive immunity. Association studies increasingly implicate KIR in disease predisposition and outcome but could be confounded by unknown KIR genetic structure in heterogeneous populations. To examine this we characterized the diversity of 16 KIR genes in 712 Northern Californians (NC) stratified by selfassigned ethnicities, and compared the profiles of KIR polymorphism with other US and global populations using a reference database. Sixty-eight distinct KIR genotypes were characterized: 58 in 457 Caucasians (NCC); 17 in 47 African Americans (NCAA); 21 in 80 Asians (NCA); 20 in 74 Hispanics (NCH) and 18 in 54 “other”

[§]To whom correspondence should be addressed 2340 Clay Street, 3rd floor, San Francisco, CA 94115, Telephone: 415.600.1530 (office); 415.577.2710 (mobile), Fax: 415-776-0292, cooperstl@sutterhealth.org.

⁷<https://sourceforge.net/users/manpsing>

Article Note The study was conducted in a multi-ethnic cohort from Northern California. Our analysis shows the potential of KIR genotype frequency polymorphism to trace populations to their native ancestries and to foil disease association studies if ethnicity is ignored. The present study focuses the importance of defining underlying population structure in conducting KIR association studies and highlights KIR genotype polymorphism itself as a useful tool in population structure analysis. The work describes a new dataset and novel analysis. None of the material has been published or is under consideration for publication elsewhere.

Conflict of Interest

The authors declare that they have no conflicts of interest with the organizations that funded the research.

ethnicities (NCO). KIR genotype patterns and frequencies in the 4 defined ethnicities were compared with each other and with 34 global populations by phylogenetic analysis. Although there were no population-specific genotypes, the KIR genotype frequency patterns faithfully traced the ancestry of NCC, NCAA and NCA but not of NCH whose ancestries are known to be more heterogeneous. KIR genotype frequencies can therefore track ethnic ancestries in modern urban populations. Our data emphasize the importance of selecting ethnically matched controls in KIR based studies to avert spurious associations.

Keywords

KIR genotype frequency; Polymorphism; Northern California population; Ethnicity

Introduction

Human killer cell immunoglobulin-like receptors (KIR) are encoded within a dynamically evolving gene cluster located at chromosome 19q13.4, in the leukocyte receptor complex (Trowsdale and Parham 2004). Profound polymorphism has been generated by multiple mechanisms including gene duplication, deletion, inversion and non-allelic homologous recombination (NAHR) (Rajalingam et al. 2004; Vilches and Parham 2002). The human KIR locus has been shown to contain between 6 and 15 expressed genes and 2 pseudogenes. KIR polymorphism therefore exhibits marked haplotypic variation whereby an individual can possess up to 17, and as few as 8, KIR genes in distinct diploid combinations or “genotypes”. Two major haplotype groups, “A” and “B” were originally defined (Uhrberg et al. 1997) and still retain useful systematic classification. Group A haplotypes comprise a core set of 7 KIR genes that encode predominantly inhibitory receptors, and group B possess variable complements of genes that encode a greater proportion of activating receptors (Uhrberg et al. 1997). Certain genes are haplodiagnostic for each group. Generally, group A haplotypes are characterized by the presence of KIR2DL3 and KIR3DL1, which both encode inhibitory receptors, and by KIR2DS4, which can encode an activating receptor but is commonly null (Maxwell et al. 2002). Correspondingly, group B KIR haplotypes are distinguished by variable combinations of KIR2DL5, KIR2DL2, KIR2DS1-3, KIR2DS5 and KIR3DS1. KIR genotypic polymorphisms have been studied and reported in different human populations. Through March 2011, 396 distinct KIR genotypes have been reported from 12,741 individuals within 108 distinct populations (Gonzalez-Galarza et al. 2011; Middleton D 2003). New KIR genotypes are still being identified, even in Western populations where scrutiny has been most intense, emphasizing the still incompletely defined extent of haplotypic KIR variation.

The receptors expressed by KIR genes regulate the function of natural killer (NK) cells and other lymphocyte subsets (Lanier 2005). Now well established is that the ligands for 5 inhibitory receptors, KIR2DL1-3 and KIR3DL1-2, are subsets of classical HLA class I molecules, whose KIR specificity is defined by oligomorphic epitopes within the alpha 1 domain. KIR2DL1 engages HLA-C molecules possessing the key epitopic residue Lysine (K)80, whereas KIR2DL2 and KIR2DL3 bind HLA-C molecules harboring residue Asparagine (N)80. The dimorphism at residue 80 is conserved among hominid MHC class I

C types and defines the “HLAC2” (K80) and “HLAC1” (N80) KIR2DL ligands respectively. KIR3DL1 binds HLA class I molecules expressing the Bw4 motif at residues 77–83, with similar fine specificity for residue 80, and KIR3DL2 has potentially peptide-dependent specificity for HLA-A3 and HLA-A11 allotypes (Hansasuta et al. 2004). Relatively little is known about the natural ligands for activating KIR but there is evidence that at least some may also bind MHC class I molecules (Chewning et al. 2007). The *in vivo* function of KIR and the clinical implications of KIR polymorphism are still under intense scrutiny.

Genetic association studies increasingly implicate KIR in influencing the outcome of virus infections (Khakoo et al. 2004; Martin et al. 2002), tumors (Carrington et al. 2005), transplantation (Ruggeri et al. 2002), pregnancy success (Hiby et al. 2004) and autoimmunity predisposition (Nelson et al. 2004). The underlying theme is that genetic differences influence KIR-ligand interactions and tilt a subtle balance between activating and inhibitory signaling, and result in NK cell-mediated activation or dampening (Lanier 2005). Such disease-outcome associations are being conducted within and between various human populations (Denis et al. 2005; Du et al. 2007; Frassati et al. 2006; Jiang et al. 2005; Lee et al. 2008; Martin et al. 2004; Norman et al. 2001; Santin et al. 2006; Single et al. 2008; Uhrberg et al. 1997; Whang et al. 2005). However, the evolutionary pressures exerted across the KIR locus and the rate at which the locus can be shaped are unknown, generating concern that undetected KIR genetic structure, including within ethnic subpopulations, could confound the accuracy of inferences made in KIR association studies. To begin investigating this we set out to examine KIR polymorphism in a local, multi-ethnic North Californian population comprising 712 individuals.

Materials and methods

Study subjects

712 Northern Californians residing mainly in the greater San Francisco Bay area were characterized for KIR gene content (Table 1). Subjects were identified through clinical and/or research contact at 2 institutions in San Francisco (SF): California Pacific Medical Center (CPMC) and the University of California, San Francisco (UCSF). IRB approval for the current study was obtained at each institution. All persons gave their informed consent prior to their inclusion in the study. An initial cohort of 487 hospital clinic-based subjects are, or have been, infected with viral hepatitis B and/or C. A second cohort of 225 NC individuals was comprised by young injection drug users (IDU) and their associates: some were not infected with HCV, HBV or HIV, others were chronically infected and some resolved acute HCV infection. The ethnicity of subjects was self-assigned and stratified into the 4 major ethnic groups designated by the latest U.S. census (U.S. Census 2010): Caucasian, African American, Asian and Hispanic. A fifth group designated, “others” comprised individuals who either did not provide ethnicity information or represented other minority groups and those who indicated more than 1 ethnicity. For extended comparison, KIR data from 4206 multiethnic individuals was extracted from a central database: www.allelefrequencies.net, that captures 34 worldwide populations including from European, North American, South American, African and Asian countries (Gonzalez-

Galarza et al. 2011; Middleton D 2003). The selection of comparator populations was based on ethnic similarity with the 4 major ethnic groups within Northern California and on the availability of data for the same 16 KIR genes characterized in our study, with the exception of 3 genotypes in Native populations from North and South America, where the information for pseudogenes was missing. KIR genotype information was tabulated for each of the 34 comparator populations and our North Californian residents were abbreviated for convenient analysis and tabulated (Table 2).

DNA Isolation

Peripheral blood mononuclear cells (PBMC) were isolated using a Ficoll (GE Healthcare) density gradient method and genomic DNA (gDNA) was extracted using the QIAamp™ DNA Blood Mini kit (Qiagen® Valencia, CA). Samples were eluted in DNase/RNase-free water and measured for concentration using a NanoDrop 8000 spectrophotometer.

KIR genotyping

A sequence-specific priming polymerase chain reaction (PCR-SSP) genotyping method was established and optimized in our laboratory to determine the presence of 16 KIR genes in the diploid chromosomal complement of each subject. The PCR-SSP method used 50–100 ng of gDNA template, 10 pmol of each primer and GoTag® Green Master Mix (Promega). To ensure optimal conditions for accurate typing of 16 KIR genes, each primer pair was tested using established control samples. Our first round reactions incorporated the following primer pairs: for KIR2DL1 as described in (Vilches and Parham 2002).; for KIR2DL2-5, KIR3DL1-3, KIR2DS1-2, KIR2DS4, KIR3DS1, KIR2DP1, KIR3DP1, and the KIR2DS3-reverse primer as described in (Uhrberg et al. 1997). We utilized the KIR2DS3-forward primer originally described in (Norman et al. 2001) and a primer pair for KIR2DS5 as described in (IHWG 13th International Histocompatibility Workshop, Seattle, WA 2002). Internal controls included primers specific for the β -actin gene. The cycling conditions used for all PCR reactions were as follows: denaturation at 95°C for 2 min, 5 cycles (97°C 20 sec, 62°C 45 sec, 72°C 90 sec) then 30 cycles (95°C 20sec, 60°C 45 sec, 72°C 90 sec) and a final extension at 72°C for 5 min. Amplicons were electrophoresed through 1% agarose gels and visualized using ethidium bromide. When faint bands were observed, the reactions were repeated, initially using the same primers. If bands remained faint or absent, reactions were again repeated using primers described in (Vilches et al. 2007). As a further quality control, a spectrum of samples were also typed using Luminex® methodology at a different laboratory and results were 97% concordant. Sequence comparisons between primers used in this study and the latest IMGT KIR sequence database (July 2011) suggests that 15/522 (2.7%) KIR alleles might not amplify efficiently in our typing system (Supplementary Table 2). Two of the 15 are alleles of KIR2DL4, which is present in all 68 of our NC KIR genotypes (Figure 1), implying that there may be 13/552 (2.36%) alleles that have not been captured by our typing primers. In some genotypes, putatively missed alleles would have likely been present in combination with typed alleles at the same loci. Our genotyping error rate is therefore anticipated to have been low and likely comparable with the reference genotypes used in our Californian and global analysis. The KIR genotype of each individual was recorded using binary code to indicate the absence or presence of each KIR gene respectively. KIR2DL5A and KIR2DL5B, were not resolved in the current study.

Data analysis

A new KIR analysis application—KIR genotype patterns were tabulated using a simple and flexible in-house program (“*KGP*”) written in Perl script (Singh et al. manuscript in preparation). *KGP* parses the individual, binary-encoded KIR genotype data such that individuals or populations sharing the same genotype pattern are grouped together. The program output associates each KIR genotype pattern with the codified identities and ethnicities of individuals sharing the same genotype. By tabulating the exact number of distinct genotypes in this manner, it facilitates analysis of KIR genotypic diversity in large datasets of individuals and populations, and allows examination of potential ethnic differences. KIR genotypes can thus be grouped, quantified and ethnically stratified, permitting convenient comparison of KIR genotype frequency distributions within and between respective populations. The flexibility of *KGP* also enables convenient assimilation of KIR genotype frequencies downloaded from other databases, including from the central international KIR database: www.allelefrequencies.net. Therefore, *KGP* was also used to analyze KIR genotype data from our cohort with that extracted from 34 comparator populations.

Phylogenetic analysis—Phylogenetic analyses were conducted using the individual KIR genotype frequencies in respective study populations. The phylogenetic analysis was performed using PHYLIP version 3.6 (Felsenstein 1989) based on the restricted maximum likelihood estimation method, and displayed using Hypertree (Bingham and Sudarsanam 2000).

Results

KIR and Population Diversity in Northern California

We analyzed the diversity and frequency of KIR genes in a Northern Californian (NC) population comprising 712 subjects, selected by their enrolment in viral hepatitis research studies at 2 Medical Centers in San Francisco. PCR-SSP typing was conducted to determine the presence of 16 KIR genes in the diploid chromosomal complement of each individual and identified 68 distinct genotypes in the NC cohort (Figure 1). In order to seek evidence for KIR genetic substructure the NC study population was stratified into the 5 ethnic groups prevalent in the State (U.S. Census 2010), based upon self-assigned ethnicity (Table 1). Because KIR genotypes display a wide range of polymorphism within and between populations, a computer program (*KGP*) was written and deployed to display their variation and frequency and thus facilitate analysis of potential ethnic signatures. The KIR genotype content and frequency in the major NC ethnic groups: Caucasian (NCC, n=457), African American (NCAA, n=47), Asian (NCA, n=80), Hispanic (NCH, n=74), and in the ethnically composite group, “Others” (NCO, n=54), is shown in Figure 1. The NCO subgroup encapsulates subjects with undeclared ethnicity (n=42) and subjects whose self-assigned ethnicity did not belong to any of the 4 specified categories (n=12).

KIR genotype diversity among Caucasians in Northern California

Caucasians comprised the majority population in our cohort (Table 1), consistent with contemporary State demographics (U.S. Census 2010). Our initial analysis therefore focused

on this subgroup. The NCC subgroup comprised 457 individuals, from which 58 distinct KIR genotypes were characterized (Figure 1). Respective NCC genotypes were ranked by population frequency, numbered (NC1-58) and clustered. The range of KIR gene content varied from 8 to 16 genes between all NCC subjects and the most common genotype (NC1) was shared by 29.5% of this population (Figure 1). NC1 is recognized as the most common KIR genotype in different human populations, comprised by AA haplotypes (Uhrberg et al. 1997) constituted by 9 KIR genes. In contrast, genotypes NC38 to NC58 were *singletons*, each found in only 1 individual within the NCC subgroup. Of the 58 NCC genotypes, NC28 was unshared among all of the populations examined (Figure 1) but has been described in Iranian (Ashouri et al. 2009), Brazilian (Ewerton et al. 2007) and Indian populations (Kulkarni et al. 2008; Rajalingam et al. 2002) with frequencies varying from 0.8 to 3%. These populations were not included in our analysis either because their migration histories into California have been historically light and/or due to the lack of data for all 16 KIR genes typed in our study. Only 3.7% of NCC individuals possessed all 16 KIR genes, and 2.4% possessed the minimum of 8 genes; the latter in 2 different genotypes (NC9 and NC26). Emphasizing an asymptotic KIR genotype frequency distribution, the top 3 (5.2%) of the 58 NCC KIR genotypes were present in 52.3% of the NCC population (Figure 1). With the exception of NC1, the frequency of the remaining NCC genotypes ranked differently in the other ethnically defined NC populations studied.

KIR genotype diversity among African Americans, Asians, Hispanics and Others in Northern California

Our non-Caucasian study population comprising NCAA (6.6%), NCA (11.2%), NCH (10.4%), and NCO (7.6%) (Table 1), respectively possessed 17, 21, 20 and 18 distinct KIR genotypes (Figure 1). Ten KIR genotypes (NC59 to NC68) were identified in these groups that were not found in the NCC group. Five of these genotypes (NC59-63) occurred in NCAA with only 1 found in a non-NCAA ethnicity (NC59 in NCA). Among these 5 genotypes, NC61 was present at highest frequency (8.5%), ranking third in NCAA. Of the remaining 5 KIR genotypes, NC64 and NC65 were found only in NCA; NC66 and NC67 only in NCH, and NC68 only in NCO, each possessed by single individuals in respective ethnicities. While NC1 was ubiquitously the most frequent genotype, the frequency of other KIR genotypes varied among ethnicities. For example, genotype NC2 was the second most frequent in NCC (12.9%) and NCO (11.1%), but ranked third in NCH (9.5%), seventh in NCAA (4.3%) and eighth in NCA (2.5%) (Figure 1). Among the 10 most frequent NCC KIR genotypes, NC5 (4.4%), NC8 (1.8%) and NC10 (1.5%) were not present in NCAA, and only NC8 was absent in NCA. KIR genotype frequency was therefore the most striking population variable, which we interrogated further by phylogenetic analysis. As shown in Supplementary Figure 1, the NC KIR genotype frequencies of the 4 major ethnicities segregated into distinct clades.

Comparison of KIR genotype frequency and diversity with other populations

Our NC dataset provided opportunity to compare KIR gene structure with the same ethnic groups studied by others in different geographical locations. The primary objectives were to examine the extent to which our NC population captured known KIR polymorphism, to seek

ethnic signatures and to expose potential sampling errors associated with the observed KIR patterns in our NC population.

(i) Northern versus Southern California—Initially we extended the analysis to a healthy Southern California (SC) population of similar size using a reference 16-KIR gene dataset generated by Du and colleagues (Du et al. 2007). We compared all 68 KIR genotypes from the 712 NC individuals with 70 KIR genotypes reported from 759 SC subjects. This identified 84 distinct KIR genotypes in California, with 17 NC and 13 SC genotypes that were unshared between the respective populations. We then analyzed the KIR genotypes and their frequencies in each of the 4 major ethnicities, excluding NCO. In Californian Caucasians, 65 distinct KIR genotypes were found, with 29 and 7 unshared genotypes in NCC and SCC respectively (Supplementary Table 1). Similar analyses were performed for Californian African Americans, Asians and Hispanics. Twenty-seven KIR genotypes were observed among Californian African Americans, with 5 and 10 unshared genotypes in NCAA and SCAA respectively. Two hundred and thirty Californian Asians exhibited 39 distinct KIR genotypes, with 5 NCA and 18 SCA unshared genotypes. Among 202 California Hispanics 35 distinct genotypes were identified with 8 NCH and 15 SCH unshared genotypes.

Phylogenetic analyses using ethnic KIR genotype frequencies indicated that despite differences in subject selection methods, respective Caucasian, Asian and African American populations in SC and NC are genetically similar (Figure 2). Therefore, KIR genotype frequencies retained the distinct ethnic profiles observed in our NC analyses. In contrast, the KIR genotype frequencies in NCH and SCH segregated into two distinct clades (Figure 2) suggesting that they may represent genetically different populations, which is in fact plausible given that Hispanics are known to be markedly heterogeneous (Shtir et al. 2009).

(ii) Global populations—To overcome the potentially obscuring effects of admixture in contemporary U.S. populations, and to address the potential for local sampling error, we expanded the comparative KIR genotype analysis to global populations. The global comparison used data extracted in March 2011 from the major reference KIR genotype database (Gonzalez-Galarza et al. 2011; Middleton D 2003), which contains 396 distinct KIR genotypes reported from 12,741 individuals in 108 populations. We extracted KIR genotype data from various native European, Asian and African populations in an attempt to capture the principal ancestral demographics of modern North American immigrants. Native American populations from South and North America were also included to further scrutinize the KIR genetic make-up of the Hispanic population in California. Information for the same 16 KIR genes was available from 7 Native American, 1 Mestizo and 22 non-American populations (Table 2). The comparator groups included the 4 major NC ethnicities in our study (excluding NCO) and the 4 ethnically matched SC populations. In total, 162 KIR genotypes were present in the 38 comparator populations for which complete KIR genotype information was available. In the selected populations there were 4,206 individuals, of which 712 (~17%) were provided by the current NC study. To analyze KIR population genetic structure, we constructed progressively inclusive population cohorts starting with Caucasians, which were the largest ethnic group (Table 3). NCC were

sequentially combined with other Caucasian populations (Cohorts 1–3), and ultimately with non-Caucasians (Cohort 4), encompassing all 39 global populations (Table 3).

The Caucasian KIR genotype comparison, as anticipated, showed that the number of unshared KIR genotypes in NCC was inversely related to the number of individuals sampled. The number of unshared NCC KIR genotypes reduced from 29 to 10 as NCC was sequentially compared to larger Caucasian cohorts, and reduced further to a single KIR genotype, NCC28 (Figure 1), when the comparison was extended to a multi-ethnic global population (Table 3). Similar analyses were performed for American African, Asian and Hispanic populations and we observed complete resolution of unshared genotypes within respective ethnicities when comparisons were extended to global populations (data not shown). These results emphasized that KIR gene content is not associated with any specific population or ethnicity.

Phylogenetic analysis of the frequencies of the 162 KIR genotypes in Cohort 4 (Table 3), clustered populations into 3 broad ancestries: European, African and Asian (Figure 3). Consistent with the KIR genotype frequency analysis conducted solely in California, NCC and SCC populations grouped closely within the European Caucasian clade, and NCAA and NCA clustered within native African and Asian populations respectively. Although the Californian Hispanic populations grouped within the European clade, as we have already shown NCH and SCH had formed separate branches when only the Californian KIR genotype frequencies were analyzed (Figure 2). The latter suggests more complex KIR genotype profiles in Hispanics. Therefore, while no single KIR genotype is population specific, as previously known, the KIR genotype frequency profiles broadly tracked ancestries.

Ethnic KIR genotype diversity influences KIR gene frequencies

Because disease association studies typically examine individual KIR genes, we investigated whether the observed ethnic KIR genotype frequency differences might influence the frequencies of individual KIR. Comparison of individual KIR gene frequencies in the Californian (SC+NC) population, which exhibited similar genotype frequencies respectively in Caucasians, Asians and African Americans (Figure 2), showed that ethnicity significantly influenced the frequency of KIR2DL2-3 KIR3DL1, KIR2DS1-2, KIR2DS4 and KIR3DS1 respectively (Supplementary Figure 2).

Discussion

The 712 NC residents characterized here for KIR gene content exhibited 68 KIR genotypes, sampled the State's 4 major ethnic groups (U.S. Census 2010) and allowed us to investigate ethno-geographical features of KIR gene structure. Our results indicate that KIR genotype frequencies can effectively illuminate ancestry, highlighting an under appreciated utility of KIR genotype polymorphism in broadly defining the ethnic structure of modern heterogeneous populations. This finding had not been intuitive *a priori*, especially given that KIR genotypes each represent the diploid state. We anticipated that ethnic admixture would likely homogenize and obscure ancestral influences on KIR genotype frequency. Our data, therefore, support that admixture remains fairly light, including among African Americans,

as recently reported (Halder et al. 2009; Tishkoff et al. 2009), and indicate that the frequency dependent and balancing selection influencing individual KIR genotypes in native populations remains evident in modern Americans. We infer that the effect must be quite strong for it to be observed in comparatively few individuals and suspect that it has likely been facilitated by relatively recent migrations combined with sociocultural patterns. KIR genotype frequencies can thus retain ethnic signatures in extant US populations, as recently reported for individual gene frequencies in African Americans (Jiang et al. 2010). Our parallel analysis (Supplementary Figure 2) further supports that ethnic KIR genotype profiles in California are linked to variations in individual KIR gene frequencies between the ethnic groups.

Although the NC study population was recruited from individuals exposed to hepatitis B and C viruses, which could potentially skew KIR genotype frequencies, several factors mitigate the composition of our cohorts with respect to our principal findings and conclusions. Not all study members had chronic viral hepatitis, and the cohorts were variable in disease status (Table 1). The IDU cohort comprised 225 individuals who were selected mainly by predisposition to drug addiction. They are too young to have selection bias created by disease outcome and are effectively a group of young NC individuals who use or are exposed to drugs: some are not infected (with HCV, HBV or HIV), others are chronically infected and some resolved acute HCV infection. The chronic HBV and HCV subjects in our other cohorts could, however, conceivably select for certain KIR genotypes. That Caucasian, Asian and African ethnicity still shapes the KIR genotype frequency profiles when our NC data is compared with local and with global datasets makes it unlikely that there is a strong viral hepatitis influence on the KIR genotype frequencies in our NC population. Most notably, the geographically immediate comparator population in SC is comprised by unrelated blood donors who are likely to have been predominantly healthy, but as shown in Figure 2, phylogenetic analysis discretely groups Caucasians, Asians and African Americans respectively. This argues for a KIR genotype effect that is likely independent of HCV and HBV, which do not exert any obvious impact on host gene flow.

Of the 65 KIR genotypes in Californian Caucasians, 29 (44.62%) (Table 3) were specific to either NC or to SC. Although initially surprising, we suspect that different Caucasian sample sizes in the NC and SC populations contributed to this NCC vs. SCC difference. The first 195 NCC individuals sampled yielded 39 KIR genotypes, which is a similar number to the 36 KIR genotypes characterized among the comparator population of 195 SCC individuals (Du et al. 2007). Taking a very simplified approach we therefore note that, on average, in the NCC population 1 distinct KIR genotype was identified for every 5 NCC individuals and for every 5.4 SCC individuals sampled. The ratio in NCC increases to 7.9 when we extend the number of KIR genotypes observed (N=58) among our entire Caucasian (N=457) cohort. Thus, 39 (67%) of our 58 NCC genotypes were identified after sampling 43% (195/457) of individuals in NCC. This implies that in the NC and SC Caucasian populations the rate of identification of KIR genotypes is similar and that sampling the same number of Caucasians in SC might yield a similar total number of distinct KIR genotypes. In our NC population, the rate of KIR genotype identification falls as more people are sampled but the total number of genotypes continues to increase. The quantitative NCC-SCC KIR genotype difference is therefore likely due, in significant part, to sample size differences. Potential qualitative

differences in ancestral demography within the sampled populations in NC and SC could conceivably contribute also. A focused NCC and SCC population KIR analysis will be required to more confidently resolve these questions.

The complex patterns of KIR gene presence or absence generates a high degree of heterogeneity between individuals, as reflected by the 396 KIR genotypes found in 12,741 individuals from 108 populations reported through March 2011 (Gonzalez-Galarza et al. 2011; Middleton D 2003). Our NC population therefore exhibits ~17% (68/396) of the KIR genotype polymorphism reported so far. The NC population is composed predominantly by immigrants from different parts of the world at different times over the past ~400 years (Diner 2008). Unlike KIR genotype NC1, (Figure 1), which is known to be present at highest frequency in most ethnicities (Norman et al. 2001; Single et al. 2007; Uhrberg et al. 2002; Yawata et al. 2002), our data show other genotypes that are shared between populations at variable frequencies and are phylogenetically informative. When the frequencies of 162 distinct, 16-KIR genotypes in 38 global populations were compared (Table 2) (Middleton D 2003), the analysis accurately segregated African, European and Asian ancestries on phylogenetic trees (Figure 3). Notably also, when Native American populations were included in the analyses, members of all indigenous groups clustered discretely.

While KIR genotype frequency variability between populations is thus a striking feature in our dataset, its degree of finer resolution in discriminating population substructure merits further investigation. It was notable that the Finn and Basque populations fell outside of the main European cluster (Figure 3). European populations have been found to be unusually homogeneous with respect to gene frequencies (Cavalli-Sforza 1994), mitochondrial lineages (Pult et al. 1994; Sajantila et al. 1995) and a spectrum of genetic diseases. Fennoscandian populations, however, which include Finns, are distinct compared with other European populations at nuclear (Cavalli-Sforza 1994) and mitochondrial (Sajantila et al. 1995) genetic markers. A previous European genetic study further showed that Finns differ from surrounding populations by exhibiting markedly reduced Y-chromosomal and mitochondrial diversity (Sajantila et al. 1996). The same study showed that, like Finns, Basques have lower genetic diversity and also appear to be genetically distinct, as further reflected by a relatively high frequency of Rhesus-negative genes and a distinctive non Indo-European language. Whether such distinctive features are captured in the KIR frequency profiles shown in our study would require greater sampling and further evaluation. It is unclear why the French population from Reims (FRA-RE) (Levinson et al. 2008), clusters with Basques or why the Portuguese sample does not cluster firmly within the European group. We were also interested to note that Mestizos, who by definition are admixed (Gutierrez-Rodriguez et al. 2006), still fall within the Native American clade. KIR genotype frequency variability between populations is thus a striking feature in our dataset and its degree of finer resolution merits further investigation.

These data argue that environmental factors and natural selection are likely shaping and apparently *harmonizing* population KIR genotype frequencies. The ubiquitously high frequency “NC1” KIR gene combination appears to confer uniform advantage while subordinate genotypes are subjected to more ethno-geographically variable influences. This

will need further consideration and analysis, and is beyond the scope of this manuscript. Currently, it would be difficult to invoke a concise mechanism, particularly since specific factors driving KIR polymorphism remain largely obscure. Founder effects and linkage disequilibrium among individual genes and alleles, and selection for KIR combinations that deliver net functional consequences (the extent of activation or inhibition) could most conceivably play a role.

While we found that KIR genotype frequencies among NCC, NCAA and NCA retraced respective European, African and Asian ancestries and established discrete clades, the KIR genotype frequency profile of Hispanics suggested greater complexity. When analyzed only against other major Californian ethnicities, NCH, like SCH, established phylogenies that were distinct from the other ethnic clades (Supplementary Figure 1 and Figure 2). However, when Californian Hispanic KIR genotype frequencies were compared with global populations, NCH and SCH fell within the dominant European cluster (Figure 3). A less conclusive Hispanic KIR phylogeny is, however, anticipated. “Hispanics” (also called “Latinos”) possess shared culture and usually a common language, but varied recent ancestry. Genome-wide admixture mapping supports prior studies (Chakraborty et al. 1999; Sans 2000) indicating that Hispanics encapsulate varying proportions of European, Native American and (predominantly West) African origins (Mao et al. 2007). Using a panel of ancestry informative markers (AIMs), an elegant study of the fourfold difference in asthma predisposition in Mexicans (M) versus Puerto Ricans (PR) found European, Native American and African ancestries in 45%M vs. 66%PR, 52%M vs. 18%PR and 3%M vs. 16%PR respectively (Salari et al. 2005). The 74 Hispanic individuals in our NC study therefore undoubtedly constitute a heterogeneously admixed group and as already noted, our trees pointed to a possible genetic difference within California between NCH and SCH. As anticipated from the AIMs study, the KIR genotypic frequencies of Hispanics also influenced the topology of Caucasians in our phylogenetic trees, which, however, suggested dominant influence of European over African and Native American genotype frequency signatures (Figure 3). Future studies of KIR structure in the Californian Hispanic population will be conducted, including in healthy individuals, to examine the extent to which ancestral KIR frequency signatures are retained in the respective ethnic subgroups. We recognize that the Hispanic subjects in our study may have been most obviously influenced by disease selection bias given that 63/74 (85%) were chronically infected with hepatitis C (Table 1).

Because we have shown that KIR genotype frequency can be a molecular marker for ethnicity, and influences the frequency of individual KIR genes, our data hold implications for disease association studies involving KIR. Case-control studies require thorough scrutiny of a population's genetic structure to minimize sampling error (Pritchard and Donnelly 2001) because hidden population structure can introduce bias and misleading conclusions (Barbujani and Colonna 2010; Lander and Schork 1994; Risch et al. 2002; Rosenberg et al. 2002; Thomas and Witte 2002). The complexity of KIR genetics in modern populations and the retention of ethnic KIR signatures, as highlighted here, increase the risk for falsely ascribing genetic association, particularly in studies with small sample sizes or in settings where a minor population is folded into a dominant one due to a shared disease, such as hepatitis C or HIV. Reliable associations that provide insight into KIR involvement in disease outcome will only be achieved by careful selection of ethnically matched controls.

Our data more deeply elucidate the structure of KIR genotype polymorphism, especially in California, and set a firmer foundation for KIR-related studies. By cataloguing the KIR genetic constitution of the NC population, in combination with other regional datasets, these data facilitate selection of appropriate controls for future association studies. In fact, we show that KIR genotypic polymorphism itself can be a useful tool in helping to define population substructure.

In conclusion, this study emphasizes the importance of incorporating population structure and stratification in conducting KIR association studies. It also highlights that stratification among Hispanics remains problematic and will require finer resolution of KIR polymorphism in combination with AIM panels in local populations. Our study provides deeper insight into KIR genotypic structure in California that should facilitate future disease association studies in this region.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank the subjects who participated. The work was supported by the National Institutes of Health (NIH) grant R01-DK064051 (to SLC). We also thank the “The Ibrahim El Hefni Liver Biorepository” at CPMC, especially Adrian Wadley and Dr. Daniel Stone for providing samples and for valuable input in the manuscript, and Catherine Carr for enthusiastically recruiting patients at CPMC liver clinics. We greatly appreciate the support of the other surgeons and hepatologists who dedicated time for patient recruitment, and of the exceptional staff in the Department of Transplantation and Division of Hepatology at CPMC and UCSF and those involved with The UFO Study. Our sincere thanks to Dr. Martin Brotman for facilitating a translational research program at CPMC. We also thank Dr. Marcelo Pando, Ms Susan Perkin and Dr. Dolly Tyan at Stanford Blood Center, for testing duplicate samples. We are grateful to Dr. Jason D. Barbour for critically reading the manuscript and to the reviewers, whose suggestions and comments enriched the scientific content. We also extend profound gratitude to The Ibrahim El-Hefni Technical Training Foundation and the El-Hefni family, The Kalmanovitz Foundation, The Raab Foundation, Mr. Harry Litke, The Unbroken Chain Foundation and Mr and Mrs John Mumford for supporting multiple aspects of this study; it could not have been accomplished otherwise.

References

1. Ashouri E, Farjadian S, Reed EF, Ghaderi A, Rajalingam R. KIR gene content diversity in four Iranian populations. *Immunogenetics*. 2009; 61:483–492. [PubMed: 19521696]
2. Barbujani G, Colonna V. Human genome diversity: frequently asked questions. *Trends Genet*. 2010; 26:285–295. [PubMed: 20471132]
3. Bingham J, Sudarsanam S. Visualizing large hierarchical clusters in hyperbolic space. *Bioinformatics*. 2000; 16:660–661. [PubMed: 11038340]
4. Carrington M, Wang S, Martin MP, Gao X, Schiffman M, Cheng J, Herrero R, Rodriguez AC, Kurman R, Mortel R, Schwartz P, Glass A, Hildesheim A. Hierarchy of resistance to cervical neoplasia mediated by combinations of killer immunoglobulin-like receptor and human leukocyte antigen loci. *J Exp Med*. 2005; 201:1069–1075. [PubMed: 15809352]
5. Cavalli-Sforza, LL.; Menozzi, P.; Piazza, A. *The History and Geography of Human Genes*. Princeton University Press; New Jersey: 1994. Europe; p. 270-276.
6. Chakraborty BM, Fernandez-Esquer ME, Chakraborty R. Is being Hispanic a risk factor for noninsulin dependent diabetes mellitus (NIDDM)? *Ethn Dis*. 1999; 9:278–283. [PubMed: 10421091]
7. Chewning JH, Gudme CN, Hsu KC, Selvakumar A, Dupont B. KIR2DS1-positive NK cells mediate alloresponse against the C2 HLA-KIR ligand group in vitro. *J Immunol*. 2007; 179:854–868. [PubMed: 17617576]

8. Denis L, Sivula J, Gourraud PA, Kerdudou N, Chout R, Ricard C, Moisan JP, Gagne K, Partanen J, Bignon JD. Genetic diversity of KIR natural killer cell markers in populations from France, Guadeloupe, Finland, Senegal and Reunion. *Tissue Antigens*. 2005; 66:267–276. [PubMed: 16185321]
9. Diner H. Immigrants Joining The Mainstream. e- Journal USA. 2008; 13
10. Djulejic E, Petlichkovski A, Trajkov D, Hristomanova S, Middleton D, Spiroski M. Distribution of killer cell immunoglobulinlike receptors in the Macedonian population. *Hum Immunol*. 2010; 71:281–288. [PubMed: 20004695]
11. Du Z, Gjertson DW, Reed EF, Rajalingam R. Receptor-ligand analyses define minimal killer cell Iglike receptor (KIR) in humans. *Immunogenetics*. 2007; 59:1–15. [PubMed: 17103212]
12. Ewerton PD, Leite Mde M, Magalhaes M, Sena L, Melo dos Santos EJ. Amazonian Amerindians exhibit high variability of KIR profiles. *Immunogenetics*. 2007; 59:625–630. [PubMed: 17551723]
13. Felsenstein J. PHYLIP - Phylogeny Inference Package (Version 3.68). *Cladistics*. 1989:164–166.
14. Frassati C, Touinssi M, Picard C, Segura M, Galicher V, Papa K, Gagne K, Vivier E, Degioanni A, Boetsch G, Mercier P, Vely F, de Micco P, Reviron D, Chiaroni J. Distribution of killer-cell immunoglobulin-like receptor (KIR) in Comoros and Southeast France. *Tissue Antigens*. 2006; 67:356–367. [PubMed: 16671942]
15. Gendzekhadze K, Norman PJ, Abi-Rached L, Layrisse Z, Parham P. High KIR diversity in Amerindians is maintained using few gene-content haplotypes. *Immunogenetics*. 2006; 58:474–480. [PubMed: 16738943]
16. Gonzalez-Galarza FF, Christmas S, Middleton D, Jones AR. Allele frequency net: a database and online repository for immune gene frequencies in worldwide populations. *Nucleic Acids Res*. 2011; 39:D913–D919. [PubMed: 21062830]
17. Gutierrez-Rodriguez ME, Sandoval-Ramirez L, Diaz-Flores M, Marsh SG, Valladares-Salgado A, Madrigal JA, Mejia-Arangure JM, Garcia CA, Huerta-Zepeda A, Ibarra-Cortes B, Ortega-Camarillo C, Cruz M. KIR gene in ethnic and Mestizo populations from Mexico. *Hum Immunol*. 2006; 67:85–93. [PubMed: 16698429]
18. Halder I, Yang BZ, Kranzler HR, Stein MB, Shriver MD, Gelernter J. Measurement of admixture proportions and description of admixture structure in different U.S. populations. *Hum Mutat*. 2009; 30:1299–1309. [PubMed: 19572378]
19. Hansasuta P, Dong T, Thananchai H, Weekes M, Willberg C, Aldemir H, Rowland-Jones S, Braud VM. Recognition of HLA-A3 and HLA-A11 by KIR3DL2 is peptide-specific. *Eur J Immunol*. 2004; 34:1673–1679. [PubMed: 15162437]
20. Hiby SE, Walker JJ, O'Shaughnessy KM, Redman CW, Carrington M, Trowsdale J, Moffett A. Combinations of maternal KIR and fetal HLA-C genes influence the risk of preeclampsia and reproductive success. *J Exp Med*. 2004; 200:957–965. [PubMed: 15477349]
21. Jiang B, Hou L, Chen M, Ng J, Hurley CK. The profile of KIR3DL1 and KIR3DS1 alleles in an African American population resembles that found in African populations. *Tissue Antigens*. 2010; 76:64–66. [PubMed: 20230527]
22. Jiang K, Zhu FM, Lv QF, Yan LX. Distribution of killer cell immunoglobulin-like receptor genes in the Chinese Han population. *Tissue Antigens*. 2005; 65:556–563. [PubMed: 15896204]
23. Khakoo SI, Thio CL, Martin MP, Brooks CR, Gao X, Astemborski J, Cheng J, Goedert JJ, Vlahov D, Hilgartner M, Cox S, Little AM, Alexander GJ, Cramp ME, O'Brien SJ, Rosenberg WM, Thomas DL, Carrington M. HLA and NK cell inhibitory receptor genes in resolving hepatitis C virus infection. *Science*. 2004; 305:872–874. [PubMed: 15297676]
24. Kulkarni S, Single RM, Martin MP, Rajalingam R, Badwe R, Joshi N, Carrington M. Comparison of the rapidly evolving KIR locus in Parsis and natives of India. *Immunogenetics*. 2008; 60:121–129. [PubMed: 18351333]
25. Lander ES, Schork NJ. Genetic dissection of complex traits. *Science*. 1994; 265:2037–2048. [PubMed: 8091226]
26. Lanier LL. NK cell recognition. *Annu Rev Immunol*. 2005; 23:225–274. [PubMed: 15771571]
27. Lee YC, Chan SH, Ren EC. Asian population frequencies and haplotype distribution of killer cell immunoglobulin-like receptor (KIR) genes among Chinese, Malay, and Indian in Singapore. *Immunogenetics*. 2008; 60:645–654. [PubMed: 18668235]

28. Levinson RD, Du Z, Luo L, Monnet D, Tabary T, Brezin AP, Zhao L, Gjertson DW, Holland GN, Reed EF, Cohen JH, Rajalingam R. Combination of KIR and HLA gene variants augments the risk of developing birdshot chorioretinopathy in HLA-A*29-positive individuals. *Genes Immun*. 2008; 9:249–258. [PubMed: 18340360]
29. Mao X, Bigham AW, Mei R, Gutierrez G, Weiss KM, Brutsaert TD, Leon-Velarde F, Moore LG, Vargas E, McKeigue PM, Shriver MD, Parra EJ. A genomewide admixture mapping panel for Hispanic/Latino populations. *Am J Hum Genet*. 2007; 80:1171–1178. [PubMed: 17503334]
30. Martin AM, Kulski JK, Gaudieri S, Witt CS, Freitas EM, Trowsdale J, Christiansen FT. Comparative genomic analysis, diversity and evolution of two KIR haplotypes A and B. *Gene*. 2004; 335:121–131. [PubMed: 15194195]
31. Martin MP, Gao X, Lee JH, Nelson GW, Detels R, Goedert JJ, Buchbinder S, Hoots K, Vlahov D, Trowsdale J, Wilson M, O'Brien SJ, Carrington M. Epistatic interaction between KIR3DS1 and HLA-B delays the progression to AIDS. *Nat Genet*. 2002; 31:429–434. [PubMed: 12134147]
32. Maxwell LD, Wallace A, Middleton D, Curran MD. A common KIR2DS4 deletion variant in the human that predicts a soluble KIR molecule analogous to the KIR1D molecule observed in the rhesus monkey. *Tissue Antigens*. 2002; 60:254–258. [PubMed: 12445308]
33. Middleton D, Meenagh A, Gourraud PA. KIR haplotype content at the allele level in 77 Northern Irish families. *Immunogenetics*. 2007; 59:145–158. [PubMed: 17200871]
34. Middleton D ML, Rood H, Komerofsky R. New Allele Frequency Database: <http://www.allelefrequencies.net>. *Tissue Antigens*. 2003; 61:403–407. [PubMed: 12753660]
35. Nelson GW, Martin MP, Gladman D, Wade J, Trowsdale J, Carrington M. Cutting edge: heterozygote advantage in autoimmune disease: hierarchy of protection/susceptibility conferred by HLA and killer Ig-like receptor combinations in psoriatic arthritis. *J Immunol*. 2004; 173:4273–4276. [PubMed: 15383555]
36. Norman PJ, Stephens HA, Verity DH, Chandanayingyong D, Vaughan RW. Distribution of natural killer cell immunoglobulin-like receptor sequences in three ethnic groups. *Immunogenetics*. 2001; 52:195–205. [PubMed: 11220621]
37. Pavlova Y, Kolesar L, Striz I, Jabor A, Slavcev A. Distribution of KIR genes in the Czech population. *Int J Immunogenet*. 2008; 35:57–61. [PubMed: 18186800]
38. Pritchard JK, Donnelly P. Case-control studies of association in structured or admixed populations. *Theor Popul Biol*. 2001; 60:227–237. [PubMed: 11855957]
39. Pult I, Sajantila A, Simanainen J, Georgiev O, Schaffner W, Paabo S. Mitochondrial DNA sequences from Switzerland reveal striking homogeneity of European populations. *Biol Chem Hoppe Seyler*. 1994; 375:837–840. [PubMed: 7710701]
40. Rajalingam R, Krausa P, Shilling HG, Stein JB, Balamurugan A, McGinnis MD, Cheng NW, Mehra NK, Parham P. Distinctive KIR and HLA diversity in a panel of north Indian Hindus. *Immunogenetics*. 2002; 53:1009–1019. [PubMed: 11904677]
41. Rajalingam R, Parham P, Abi-Rached L. Domain shuffling has been the main mechanism forming new hominoid killer cell Ig-like receptors. *J Immunol*. 2004; 172:356–369. [PubMed: 14688344]
42. Risch N, Burchard E, Ziv E, Tang H. Categorization of humans in biomedical research: genes, race and disease. *Genome Biol*. 2002; 3 comment2007.
43. Rosenberg NA, Pritchard JK, Weber JL, Cann HM, Kidd KK, Zhivotovsky LA, Feldman MW. Genetic structure of human populations. *Science*. 2002; 298:2381–2385. [PubMed: 12493913]
44. Ruggeri L, Capanni M, Urbani E, Perruccio K, Shlomchik WD, Tosti A, Posati S, Rogaia D, Frassonni F, Aversa F, Martelli MF, Velardi A. Effectiveness of donor natural killer cell alloreactivity in mismatched hematopoietic transplants. *Science*. 2002; 295:2097–2100. [PubMed: 11896281]
45. Sajantila A, Lahermo P, Anttinen T, Lukka M, Sistonen P, Savontaus ML, Aula P, Beckman L, Tranebjaerg L, Gedde-Dahl T, Issel-Tarver L, Di Rienzo A, Paabo S. Genes and languages in Europe: an analysis of mitochondrial lineages. *Genome Res*. 1995; 5:42–52. [PubMed: 8717054]
46. Sajantila A, Salem AH, Savolainen P, Bauer K, Gierig C, Paabo S. Paternal and maternal DNA lineages reveal a bottleneck in the founding of the Finnish population. *Proc Natl Acad Sci U S A*. 1996; 93:12035–12039. [PubMed: 8876258]

47. Salari K, Choudhry S, Tang H, Naqvi M, Lind D, Avila PC, Coyle NE, Ung N, Nazario S, Casal J, Torres-Palacios A, Clark S, Phong A, Gomez I, Matallana H, Perez-Stable EJ, Shriver MD, Kwok PY, Sheppard D, Rodriguez-Cintron W, Risch NJ, Burchard EG, Ziv E. Genetic admixture and asthma-related phenotypes in Mexican American and Puerto Rican asthmatics. *Genet Epidemiol*. 2005; 29:76–86. [PubMed: 15918156]
48. Sans M. Admixture studies in Latin America: from the 20th to the 21st century. *Hum Biol*. 2000; 72:155–177. [PubMed: 10721616]
49. Santin I, de Nancrales GP, Calvo B, Gaafar A, Castano L, Bilbao JR. Killer cell immunoglobulinlike receptor (KIR) genes in the Basque population: association study of KIR gene contents with type 1 diabetes mellitus. *Hum Immunol*. 2006; 67:118–124. [PubMed: 16698433]
50. Shtir CJ, Marjoram P, Azen S, Conti DV, Le Marchand L, Haiman CA, Varma R. Variation in genetic admixture and population structure among Latinos: the Los Angeles Latino eye study (LALES). *BMC Genet*. 2009; 10:71. [PubMed: 19903357]
51. Single RM, Martin MP, Gao X, Meyer D, Yeager M, Kidd JR, Kidd KK, Carrington M. Global diversity and evidence for coevolution of KIR and HLA. *Nat Genet*. 2007; 39:1114–1119. [PubMed: 17694058]
52. Single RM, Martin MP, Meyer D, Gao X, Carrington M. Methods for assessing gene content diversity of KIR with examples from a global set of populations. *Immunogenetics*. 2008; 60:711–725. [PubMed: 18797862]
53. Thomas DC, Witte JS. Point: population stratification: a problem for case-control studies of candidate gene associations? *Cancer Epidemiol Biomarkers Prev*. 2002; 11:505–512. [PubMed: 12050090]
54. Tishkoff SA, Reed FA, Friedlaender FR, Ehret C, Ranciaro A, Froment A, Hirbo JB, Awomoyi AA, Bodo JM, Doumbo O, Ibrahim M, Juma AT, Kotze MJ, Lema G, Moore JH, Mortensen H, Nyambo TB, Omar SA, Powell K, Pretorius GS, Smith MW, Thera MA, Wambebe C, Weber JL, Williams SM. The genetic structure and history of Africans and African Americans. *Science*. 2009; 324:1035–1044. [PubMed: 19407144]
55. Trowsdale J, Parham P. Mini-review: defense strategies and immunity-related genes. *Eur J Immunol*. 2004; 34:7–17. [PubMed: 14971025]
56. U.S.Census. U.S. Census Bureau. 2010
57. Uhrberg M, Parham P, Wernet P. Definition of gene content for nine common group B haplotypes of the Caucasoid population: KIR haplotypes contain between seven and eleven KIR genes. *Immunogenetics*. 2002; 54:221–229. [PubMed: 12136333]
58. Uhrberg M, Valiante NM, Shum BP, Shilling HG, Lienert-Weidenbach K, Corliss B, Tyan D, Lanier LL, Parham P. Human diversity in killer cell inhibitory receptor genes. *Immunity*. 1997; 7:753–763. [PubMed: 9430221]
59. Vilches C, Castano J, Gomez-Lozano N, Estefania E. Facilitation of KIR genotyping by a PCR-SSP method that amplifies short DNA fragments. *Tissue Antigens*. 2007; 70:415–422. [PubMed: 17854430]
60. Vilches C, Parham P. KIR: diverse, rapidly evolving receptors of innate and adaptive immunity. *Annu Rev Immunol*. 2002; 20:217–251. [PubMed: 11861603]
61. Whang DH, Park H, Yoon JA, Park MH. Haplotype analysis of killer cell immunoglobulin-like receptor genes in 77 Korean families. *Hum Immunol*. 2005; 66:146–154. [PubMed: 15695000]
62. Yawata M, Yawata N, Draghi M, Little AM, Partheniou F, Parham P. Roles for HLA and KIR polymorphisms in natural killer cell repertoire selection and modulation of effector function. *J Exp Med*. 2006; 203:633–645. [PubMed: 16533882]
63. Yawata M, Yawata N, McQueen KL, Cheng NW, Guethlein LA, Rajalingam R, Shilling HG, Parham P. Predominance of group A KIR haplotypes in Japanese associated with diverse NK cell repertoires of KIR expression. *Immunogenetics*. 2002; 54:543–550. [PubMed: 12439616]

NC KIR GENOTYPE ID	INHIBITORY GENE					ACTIVATING GENE					PSEUDO GENE		GENOTYPE FREQUENCY						GENOTYPE RANKING						
	KIR2DL1	KIR2DL2	KIR2DL3	KIR3DL1	KIR3DL2	KIR3DL3	KIR2DS1	KIR2DS2	KIR2DS3	KIR2DS4	KIR2DS5	KIR3DS1	KIR2DL4	KIR3DP1	KIR2DP1	NCC [45/78]	NCAA [47/17]	NCA [30/21]	NCH [74/20]	NCO [34/16]	NCC	NCAA	NCA	NCH	NCO
1																135 (29.5)	18 (38.3)	30 (37.5)	25 (33.8)	24 (44.4)	1	1	1	1	1
2																59 (12.9)	2 (4.3)	2 (2.5)	7 (9.5)	6 (11.1)	2	7	8	3	2
3																45 (9.8)	3 (6.4)	12 (15.0)	8 (10.8)	4 (7.4)	3	5	2	2	3
4																25 (5.5)	4 (8.5)	1 (1.3)	6 (8.1)	3 (5.6)	4	2	19	5	4
5																20 (4.4)		3 (3.8)	2 (2.7)		5		4	7	
6																19 (4.2)	2 (4.3)	7 (8.8)	7 (9.5)		6	8	3	4	
7																17 (3.7)	1 (2.1)	3 (3.8)	2 (2.7)	2 (3.7)	7	12	6	9	5
8																8 (1.8)			1 (1.4)	1 (1.9)	8			12	10
9																8 (1.8)	1 (2.1)				9	16			
10																7 (1.5)		2 (2.5)			10		11		
11																7 (1.5)	1 (2.1)		1 (1.4)		11	17		19	
12																6 (1.3)		1 (1.3)	3 (4.1)		12		18	6	
13																6 (1.3)	1 (2.1)			1 (1.9)	13	13			11
14																5 (1.1)		3 (3.8)			14		5		
15																5 (1.1)	1 (2.1)			2 (3.7)	15	15			6
16																5 (1.1)			1 (1.4)		16			18	
17																5 (1.1)			1 (1.4)		17			13	
18																4 (0.9)		1 (1.3)			18		20		
19																4 (0.9)		3 (3.8)		1 (1.9)	19		7		14
20																4 (0.9)					20				
21																4 (0.9)					21				
22																3 (0.7)			2 (2.7)		22			8	
23																3 (0.7)			1 (1.4)		23			15	
24																3 (0.7)		2 (2.5)		1 (1.9)	24		9		13
25																3 (0.7)					25				
26																3 (0.7)				1 (1.9)	26				9
27																3 (0.7)			1 (1.4)	1 (1.9)	27		20		18
28																2 (0.4)					28				
29																2 (0.4)		1 (1.3)			29		17		
30																2 (0.4)			2 (2.7)		30			10	
31																2 (0.4)			1 (1.4)	1 (1.9)	31			17	16
32																2 (0.4)					32				
33																2 (0.4)					33				
34																2 (0.4)				2 (3.7)	34				7
35																2 (0.4)					35				
36																2 (0.4)					36				
37																2 (0.4)					37				
38																1 (0.2)		1 (1.3)		1 (1.9)	38		14		8
39																1 (0.2)					39				
40																1 (0.2)					40				
41																1 (0.2)					41				
42																1 (0.2)	1 (2.1)				42		14		
43																1 (0.2)	3 (6.4)		1 (1.4)	1 (1.9)	43		4	16	15
44																1 (0.2)					44				
45																1 (0.2)				1 (1.9)	45				17
46																1 (0.2)		2 (2.5)			46		12		
47																1 (0.2)					47				
48																1 (0.2)					48				
49																1 (0.2)					49				
50																1 (0.2)					50				
51																1 (0.2)		1 (1.3)			51			16	
52																1 (0.2)					52				
53																1 (0.2)					53				
54																1 (0.2)					54				
55																1 (0.2)					55				
56																1 (0.2)					56				
57																1 (0.2)					57				
58																1 (0.2)					58				
59																	1 (2.1)	1 (1.3)				10		13	
60																	1 (2.1)					11			
61																	4 (8.5)					3			
62																	2 (4.3)					6			
63																	1 (2.1)					9			
64																		1 (1.3)					15		
65																		1 (1.3)				21			
66																			1 (1.4)					11	
67																			1 (1.4)					14	
68																				1 (1.9)					12

Fig. 1.

KIR genotypes observed among North Californian (NC) ethnicities; Caucasian (NCC), African American (NCAA), Asian (NCA), Hispanic (NCH) and “others” (NCO). The first column labels and stratifies by frequency the distinct KIR genotypes generated from 712 subjects by *KGP*. The KIR genotypes are depicted in rows. The respective KIR inhibitory, activating and pseudo genes are ordered in columns, from left to right. By convention, black cells designate the presence and white cells the absence, of respective genes. The ethnic breakdown of the dataset is shown in the next 5 columns, NCC-NCO. Adjacent to the

respective ethnicity headings, within square brackets [], are the total number of individuals separated by (/) from the total number of KIR genotypes in each ethnic group. In each column, the number and percent (in parentheses) of individuals possessing respective KIR genotypes is shown. The frequency rank of respective KIR genotypes in each ethnic group is shown in the 5 right columns. NC1, alone, can be seen as the most frequent genotype in all ethnicities. The genotype shaded in grey (NC28) occurred only in NCC. The 10 *non-NCC genotypes* (NC59 to NC68) were found only among non-Caucasians in our NC cohort.

Fig. 2.

Phylogenetic tree constructed using KIR genotype frequencies of 4 ethnicities sampled from NC (this study) and SC. Ethnic abbreviations are shown in Table 2. Respective Caucasian, Asian and African American clades are indicated.

Fig. 3.

Phylogenetic trees based on frequencies of 162 KIR genotypes in respective populations (Cohort 4 in Table 3). Ethnic abbreviations are shown in Table 2. In clockwise order, African, European, Native American and Asian clusters are respectively enclosed within distinct outlines, as shown. The NC and SC populations are clustered in proximity to their native ancestries.

Table 1

Ethnic and clinical composition of the study cohort

		Caucasian (NCC)	African American (NCAA)	Asian (NCA)	Hispanic (NCH)	Others (NCO)
HCV Cirrhotic (n=329)	Total HCV Cirrhotic (n=329)	215/329 (65.35%)	26/329 (7.90%)	20/329 (6.08%)	58/329 (17.63%)	10/329 (3.04%)
	*CPMC (n=166)	113/166 (68.08%)	11/166 (6.62%)	10/166 (6.02%)	27/166 (16.27%)	5/166 (3.01%)
	**UCSF (n=163)	102/163 (62.58%)	15/163 (9.20%)	10/163 (6.13%)	31/163 (19.02%)	5/163 (3.07%)
HBV Chronic (n=57)		4/57 (7.02%)	-	50/57 (87.72%)	-	3/57 (5.26%)
HCV Chronic (n=101)		68/101 (67.33%)	15/101 (14.85%)	7/101 (6.93%)	5/101 (4.50%)	6/101 (5.94%)
IDU (n=225)		170/225 (75.55%)	6/225 (2.67%)	3/225 (1.33%)	11/225 (4.89%)	35/225 (15.56%)
Total NC Population (n=712)		457/712 (64.19%)	47/712 (6.60%)	80/712 (11.24%)	74/712 (10.39%)	74/712 (7.58%)

Ethnic and clinical composition of the 712 study subjects. Individuals were categorized in the 5 major ethnicities that compose the California population (US Census 2010): Caucasian, African American, Asian American, Hispanic and "Others". The fifth category, "Others" (n=54), refers to individuals who lack ethnicity information in our record or cannot be categorized under the four principal ethnic groups. Clinical classification into "HCV Cirrhotic" (n=329), non-cirrhotic "HCV Chronic" (n=101) and HBV Chronic (n=57) reflects the patient groups who attend local liver and transplant centers (i)*CPMC (n=166) and (ii) **UCSF (n=163). An additional study group of 225 local injection drug users (IDU) was included. The bottom row shows the total number of individuals from respective ethnicities within the cohort.

Table 2

Populations considered for the analysis

Population	Demography	Abbreviations	Geographical Location	Total # of Individuals
1	Northern California Caucasian ^(Present Study)	NCC	North America	457
2	Northern California Asian American ^(Present Study)	NCA	North America	80
3	Northern California African American ^(Present Study)	NCAA	North America	47
4	Northern California Hispanic ^(Present Study)	NCH	North America	74
5	Northern Californian "Others" ^(Present Study)	NCO	North America	54
6	Southern California Asian ^(Du et al. 2007)	SCA	North America	150
7	Southern California African American ^(Du et al. 2007)	SCAA	North America	58
8	Southern California Caucasian ^(Du et al. 2007)	SCC	North America	195
9	Southern California Hispanic ^(Du et al. 2007)	SCH	North America	128
10	England ^{(Middleton D 2003) *}	ENG 4	Europe	99
11	Finland (Helsinki) ^(Denis et al. 2005)	FIN	Europe	101
12	France (Reims) ^(Levinson et al. 2008)	FRA-RE	Europe	102
13	France Southeast pop2 ^(Frassati et al. 2006)	FRA-SE	Europe	38
14	France West ^(Denis et al. 2005)	FRA-W	Europe	108
15	Spain (Granada) ^{(Middleton D 2003) **}	SPA-GR	Europe	100
16	Spain (Basque) ^(Santin et al. 2006)	SPA-BA	Europe	71
17	Scotland (Orkney) ^{(Middleton D 2003) ***}	SCO	Europe	90
18	Portugal ^{(Middleton D 2003) *}	POR	Europe	65
19	Italy (Lombardy) ^{(Middleton D 2003) ***}	ITA	Europe	50
20	Ireland (Northern) ^(Middleton et al. 2007)	IRE	Europe	200
21	Macedonia ^(Djuljic et al. 2010)	MAC	Europe	214
22	Czech Republic ^(Pavlova et al. 2008)	CZE	Europe	125
23	South Africa Xhosa ^{(Middleton D 2003) ****}	SAX	Africa	50
24	South Africa San ^{(Middleton D 2003) ***}	SAS	Africa	91
25	Senegal ^(Denis et al. 2005)	SEN	Africa	118
26	Ghana ^{(Middleton D 2003) ****}	GHA	Africa	41
27	China Zhejiang Han ^(Jiang et al. 2005)	ZHC	Asia	104
28	Hong Kong Chinese ^{(Middleton D 2003) *}	HKC	Asia	100
29	South Korea ^(Whang et al. 2005)	SK	Asia	154
30	Singapore Chinese2 ^(Lee et al. 2008)	SC	Asia	210
31	Japan ^(Yawata et al. 2006)	JPN	Asia	132
32	Argentina (Wichis) ^{(Middleton D 2003) ***}	WIC	South America	82
33	Argentina (Chiriguano) ^{(Middleton D 2003) ***}	CHI	South America	54

Population	Demography	Abbreviations	Geographical Location	Total # of Individuals
34	Mexico (Mestizo) ^(Gutierrez-Rodriguez et al. 2006)	MES	North America	86
35	Mexico (Huichol) ^(Gutierrez-Rodriguez et al. 2006)	HUI	North America	73
36	Mexico (Taramumara) ^(Gutierrez-Rodriguez et al. 2006)	TAR	North America	65
37	Mexico (Purepecha) ^(Gutierrez-Rodriguez et al. 2006)	PUR	North America	53
38	Venezuela (Bari) ^(Gendzekhadze et al. 2006)	BAR	South America	80
39	Venezuela (Warao) ^(Gendzekhadze et al. 2006)	WAR	South America	89

Geographical information for 39 populations including 5 Northern California and 34 global comparator populations extracted from www.allelefrequencies.net (Middleton D 2003) [* Bone Marrow Registry, ** Controls for disease study, *** Anthropology Study, **** Blood Donor].

Abbreviations for each population were formulated for labeling used in the present study.

Table 3

KIR genotypes observed within the Caucasian and global populations

Study Cohort (No. of Individuals)	No. of Observed KIR Genotypes	No of Unshared NCC Genotypes (%)
Cohort 1: Northern California Caucasian + Southern California Caucasian (652)	65	29 (44.62)
Cohort 2: Northern California Caucasian + European Caucasian (1820)	105	14 (13.33)
Cohort 3: Northern California Caucasian + Southern California Caucasian + European Caucasian (2015)	110	10 (9.09)
Cohort 4: All Global populations (4206)	162	1 (0.61) *

The qualitative (ethnic) and quantitative properties of Cohorts 1–4 are indicated in the left column. Fifty-eight genotypes from 457 NCC individuals were compared with progressively larger and more inclusive cohorts. The total numbers of distinct KIR genotypes observed in respective cohorts are shown in the middle column. The unshared NCC genotypes not seen in the respective non-NCC populations are indicated in the right column. The previously unreported NCC genotype denoted with an asterisk (*) in cohort 4 refers to NC28 (see Figure 1).