Plasticity in the organization and sequences of human KIR/ILT gene families

Michael J. Wilson*†, Michaela Torkar*†, Anja Haude*, Sarah Milne‡, Tania Jones§, Denise Sheer§, Stephan Beck‡, and John Trowsdale*¶

*Immunology Division, Department of Pathology, Tennis Court Road, Cambridge CB2 1QP, United Kingdom; ‡Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, United Kingdom; and §Imperial Cancer Research Fund, Lincolns Inn Fields, Holborn, London WC2A 3PX, United Kingdom

Edited by Johannes van Rood, Leiden University, Leiden, The Netherlands, and approved February 14, 2000 (received for review January 3, 2000)

The ≈1-Mb leukocyte receptor complex at 19q13.4 is a key polymorphic immunoregion containing all of the natural killer-receptor *KIR* and related *ILT* genes. When the organization of the leukocyte receptor complex was compared from two haplotypes, the gene content in the *KIR* region varied dramatically, with framework loci flanking regions of widely variable gene content. The *ILT* genes were more stable in number except for *ILT6*, which was present only in one haplotype. Analysis of *Alu* repeats and comparison of *KIR* gene sequences, which are over 90% identical, are consistent with a recent origin. *KIR* genesis was followed by extensive duplication/deletion as well as intergenic sequence exchange, reminiscent of MHC class I genes, which provide KIR ligands.

Natural Killer (NK) cells are regulated by interaction of surface receptors with MHC class I molecules (1, 2). In humans, these interactions are mediated by two structurally distinct receptor families, the C-type lectin-like (CD94/NKG2) heterodimers and Ig-like receptors, generically known as *KIR* for killer Ig receptor (3–5). The *KIR* loci are located on chromosome 19q13.4 near genes encoding related molecules such as the Ig-like transcripts (*ILT*s), also known as monocyte inhibitory receptors (*MIR*s) or leukocyte Ig-like receptors (*LIR*), and the leukocyte associated inhibitory receptors (*LAIR*) (6–12). The region has been called the leukocyte receptor complex or cluster (LRC) (10, 13).

Over 100 highly homologous *KIR* sequences have been deposited in databases (2). Although most *KIR*s show 90–95% identity, some cDNAs may represent different alleles (14, 15). Similarly, a vast number of transcripts belonging to the *ILT* family show high identity (8, 9). There are about 10 expressed *ILT* loci in addition to *ILT9* and *10* genes (12) for which no functional transcripts have been found.

Different cells within an individual may each express a subset of the available *KIR* repertoire. Haploid genomes may encode different numbers of *KIR* genes (16). *ILT* polymorphism may not be so extensive except in certain loci, such as *ILT5* (17).

Published data suggested that patterns of *KIR* expression relate to isotypic and allotypic variation in addition to differential regulation of gene expression. To clarify the nature of polymorphism and variation in gene number in the polygenic LRC, we undertook a complete analysis of two different haplotypes.

Methods

P1 Artificial Chromosome (PAC) Clones. PACs were identified from the RPCI-1 library (18) with a *KIR* probe for exon 3 and an *ILT2* cDNA, from the Human Genome Mapping Project Resource Center.

DNA Sequencing and Analysis. PAC DNA was randomly subcloned into M13mp18 and pUC18 and amplified in 96-well microtiter plates (19). The sequence was determined by using chain termination chemistry (20). The reads were assembled into contigs (21) and analyzed (http://www.sanger.ac.uk/Teams/HGP/Humana).

Dot matrix comparisons were done by using DOTTER (22), http://www.sanger.ac.uk/software/. Amino acid motifs were identified by using the PSORT (23) program and the PROSITE database (24). The genomic sequences are available from ftp://ftp.sanger.ac.uk/pub/human/sequences/Chr_19/unfinished_sequence).

Directed Sequencing. Directed sequencing to detect genes on the first haplotype was done with locus-specific primers (16). For intergenic regions, a redundant *KIR* exon 1 primer was used with an upstream exon 7 primer. *ILT* cDNAs were aligned, and primers were designed by using differences in the 3′ ends (Table 1). Long-range PCR was used (Boehringer Mannheim).

Results

Strategy. To gain insight into the genomic diversity of the *KIR* and *ILT* loci, we collected extensive mapping and sequence data from the LRC. The genomic organization of the *ILT* and *KIR* clusters was determined on two different haplotypes by using a PAC genomic library made from a single individual (18). Contigs were assembled by conventional mapping. They fell into two distinct sets, and the PACs were divided into haplotypes on the basis of additional polymorphic markers or partial sequencing (Fig. 1). PAC clones were designated haplotype 1 or 2 by a supercript number. Selected clones were sequenced to study gene arrangement after comparison by Southern blot and *KIR*/*ILT*-specific PCR to determine that no rearrangements or deletions had occurred. We obtained 270 kb of complete sequence of two clones (52N12[1] and 1060P11[2]) spanning part of the *ILT* region and all of the *KIR* region (Fig. 1).

Gene Content. The *FcαR* locus was telomeric of the *KIR* cluster close to another Ig-like NK gene, *NKp46*. Strikingly, all *ILT* and *KIR* genes in the sequenced section were in head-to-tail orientation from centromere to telomere. This, together with the conservation of gene structures and sequence homologies between the different receptor families, indicates that the LRC has evolved as a result of extensive duplication. Comparison with databases revealed a genomic clone extending telomeric of the KIR region. A gene X was identified telomeric of *NKp46* but in the opposite orientation. This gene was expressed in testis and not in tissue of lymphoid origin by Northern blot analysis (data not shown), suggesting that it marks the telomeric boundary of the LRC. In total, there are at least 24 structurally and func-

This paper was submitted directly (Track II) to the PNAS office.

Abbreviations: LRC, leukocyte receptor complex; NK, natural killer; PAC, P1 artificial chromosome.

†M.J.W. and M.T. contributed equally to this work.

¶To whom reprint requests should be addressed. E-mail: jt233@mole.bio.cam.ac.uk.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Article published online before print: *Proc. Natl. Acad. Sci. USA*, 10.1073/pnas.080588597. Article and publication date are at www.pnas.org/cgi/doi/10.1073/pnas.080588597

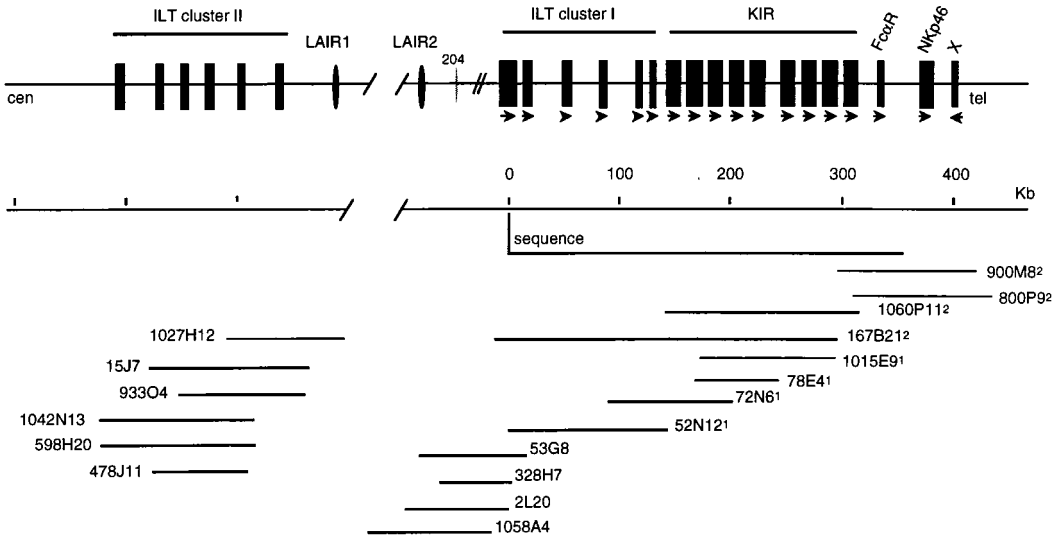

Fig. 1. Organization of the leukocyte receptor complex on human chromosome 19q13.4. Two contigs were deduced from PACs positive with combinations of individual probes. The complete sequences of PACs 1060P11[2] and 52N12[1] are displayed in Figs. 2 and 4, respectively. A sequence of a section of the *KIRs* from a haplotype similar to that of 72N6[1] is available (GenBank database accession no. AC011501). The *KIR* gene family is flanked by the *ILT* cluster I and the *FcαR* gene. The numbers of *KIR* genes differ depending on the haplotype. The *ILT* genes are grouped in two clusters separated by the *LAIR* genes. A link between the two clusters remains to be established. Marker 204 is based on a PCR product designed for other sequences in the database. Although all PAC sizes are to scale, distances between genes outside the sequenced region are approximate. At the telomeric end, the gene designated X is not related structurally to the Ig-like receptors. 52N12[1] and 72N6[1] represent one haplotype, and the contig containing 167B21[2] is part of the second haplotype. This was based on sequence polymorphism in the PCR-amplified *ILT*3 gene and other differences. Arrows point 5′ to 3′.

tionally related Ig-like receptors in the LRC region, spanning approximately 1 Mb of human chromosome 19q13.4.

Sequences of *KIR* Genes. Sequencing of the ≈160-kb PAC clone 1060P11[2], which spans the *KIR* region, revealed numerous tightly clustered loci, most of which were less than 3 kb apart (Fig. 2*A*). Analysis of the sequence data from 1060P11[2] revealed 10 *KIR* genes, most of which have matching cDNA sequences. The exceptions are the recently described centromeric *KIR* locus *KIR3DL3* [aka *KIRCI* (12), named according to convention (25)], the gene

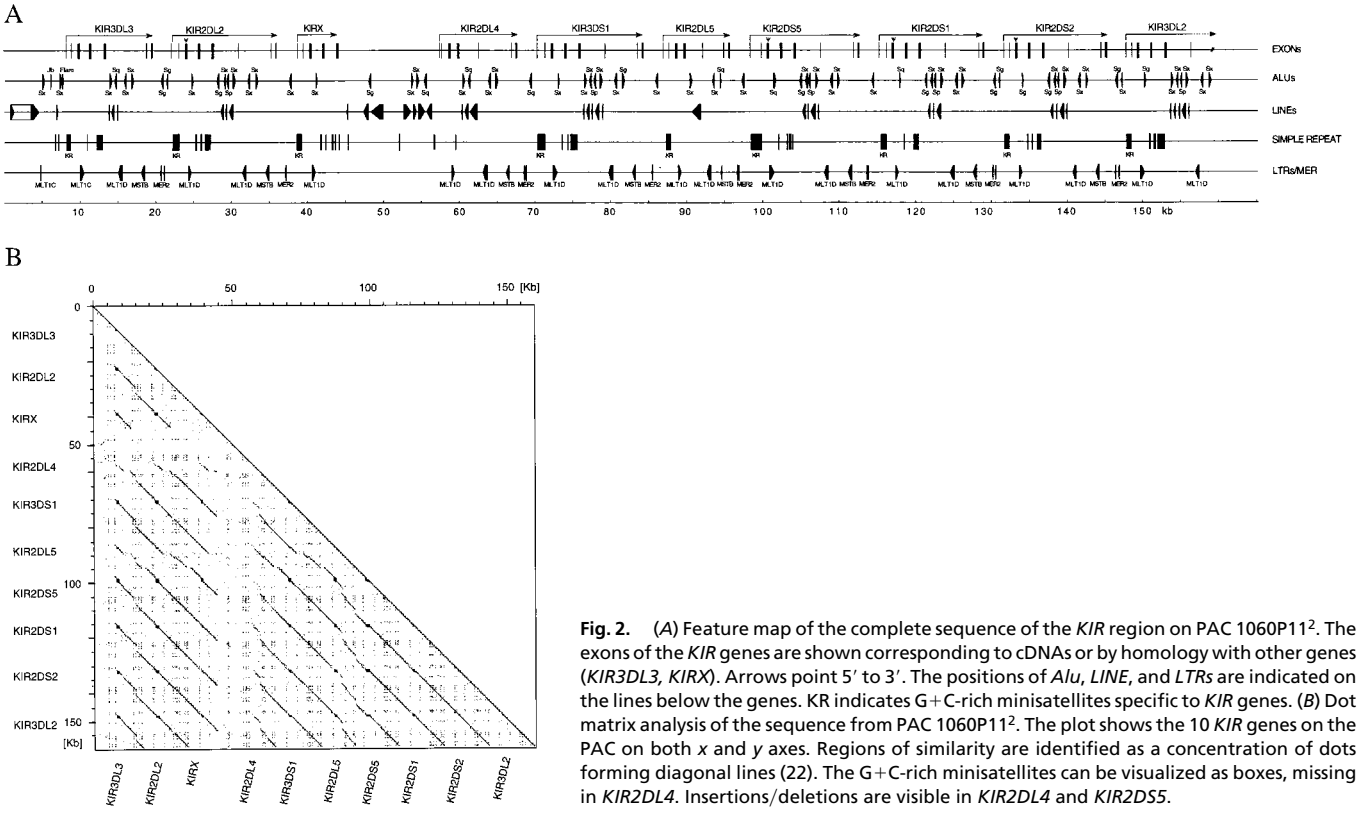

Fig. 2. (*A*) Feature map of the complete sequence of the *KIR* region on PAC 1060P11[2]. The exons of the *KIR* genes are shown corresponding to cDNAs or by homology with other genes (*KIR3DL3, KIRX*). Arrows point 5′ to 3′. The positions of *Alu*, *LINE*, and *LTRs* are indicated on the lines below the genes. KR indicates G+C-rich minisatellites specific to *KIR* genes. (*B*) Dot matrix analysis of the sequence from PAC 1060P11[2]. The plot shows the 10 *KIR* genes on the PAC on both *x* and *y* axes. Regions of similarity are identified as a concentration of dots forming diagonal lines (22). The G+C-rich minisatellites can be visualized as boxes, missing in *KIR2DL4*. Insertions/deletions are visible in *KIR2DL4* and *KIR2DS5*.

Table 1. Oligonucleotides used for detection of *ILT*, *LAIR*, and *KIR* genes by PCR and determination of polymorphic sites by sequence analysis

	Sequence	Specificity
1	5′ GCCACAATCACTCATCAGAGTA 3′	*ILT1* specific fwd
2	5′ GTATCGCTGTTACTATGGTAGCG 3′	*ILT2* specific fwd
3	5′ ATGATCCCCACCTTCACGGCT 3′	*ILT3* specific fwd
4	5′ CAGCTTCCATGCCTTCTGGG 3′	*ILT3* specific rev
5	5′ TCAGTATTACAGCCGCGCTCGG 3′	*ILT4* specific fwd
6	5′ GGTGCTATGGTTATGACTCGCGCG 3′	*ILT6* specific fwd
7	5′ ATGACCGCCGCCCTCACAGCCT 3′	*ILT9* specific fwd
8	5′ GCAGAGCAGGGCATCATGGTGT 3′	*ILT10* specific fwd
9	5′ AGCCTCCGAGTGTCCACACTG 3′	*LIR6* specific fwd
10	5′ TTGGCAGACAGTCCAGATAACATC 3′	*LIR6* specific rev
11	5′ TGTCGTGGCCCGCGGAGGC 3′	*LIR8* specific fwd
12	5′ TGACTGACACAGCAGGGTCACG 3′	*ILT* redundant rev
13	5′ CGTGACCCTGCTGTGTCAGTCA 3′	*ILT4* sequencing primer
14	5′ CGAAGCCATGAGTTGCACACTG 3′	Marker 204 fwd
15	5′ CAACCACAGCATCTGTAGGCTCC 3′	Marker 204 rev
16	5′ GGGGAGCCATGTGACTTTCGTG 3′	*LAIR* fwd
17	5′ GTCACTCTGCTCAGACCATTTAG 3′	*LAIR* rev
18	5′ TGATTGGGACCTCAGTGGTCA 3′	*KIR* exon 7 redundant fwd
19	5′ CCCAACRCAYRCCATGMTGA 3′	*KIR* exon 1 redundant rev

The KIR oligonucleotides were used for linking adjoining genes on haplotype 1. KIR genes were detected by using primer sets as described (16).

*KIR*X, and a gene telomeric of *KIR3DS1*, which we have named *KIR2DL5*. *KIR*X has only exons 1–5, i.e., those encoding the matching extracellular domains of other *KIR*s. *KIR2DL5* has a similar gene structure and is most homologous to *KIR2DL4*. *KIR2DL5* has a predicted ORF that would encode potentially for a protein with two ITIM motifs in its cytoplasmic tail. We found transcripts corresponding to *2DL5* by reverse transcription–PCR using cDNA derived from PBLs. All of the two-Ig-domain *KIR*s, with the exception of *KIR2DL4* and *KIR2DL5*, have a pseudo exon 3 that has remarkable similarity to the first Ig domain of the three Ig domain *KIR*s but are all marked by a 3-bp deletion and a nucleotide change leading to an in-frame stop codon at the same position as described for the *KIR2DL3* gene (26).

Dot-Matrix Analysis of *KIR* Genes. Dot matrix analysis of the 1060P11[2] sequence (Fig. 2*B*) revealed a remarkable organization of reiterated sequences. The only unique sequences over 100 bp were: (*i*) upstream of *KIR3DL3*, (*ii*) outside the reiterated region, and (*iii*) upstream of *KIR2DL4*. The *KIR* gene sequences, including intergenic regions, are highly conserved. The sequences comprise a continuous loop that extends seamlessly from gene to gene. The reiterativeness of the loop is broken only by 14 kb upstream of the *KIR*2DL4 locus, which displays some unique features, characterized by L1 repeats. The high level of homology could facilitate exchange of exons between different *KIR* loci, by some form of illegitimate crossing over or gene conversion (27). Such mechanisms may be behind the variation in number of Ig exons in some members of the extended *KIR* family (see below).

Comparison of Two *KIR* Haplotypes. Using the sequence from 1060P11[2], we analyzed the other haplotype (haplotype 1) by PCR using locus-specific (16) and intergenic primer sets (Table 1). All PCR products were subcloned and completely sequenced to determine the gene arrangement of this haplotype (Fig. 3*A*). Genomic databases concurred with our data on this haplotype, which is the most common (16). The two haplotypes revealed a remarkable difference in organization. Certain framework loci, such as *KIR3DL3* at the centromeric end, *KIR*X-*KIR2DL4* in the middle, and *KIR3DL2* at the telomeric end, are present on both haplotypes, consistent with their genotype frequencies of 100% in the populations studied (12, 16). The position of the *KIR3DL1* locus was occupied by an activating gene, *KIR3DS1* on the second haplotype.

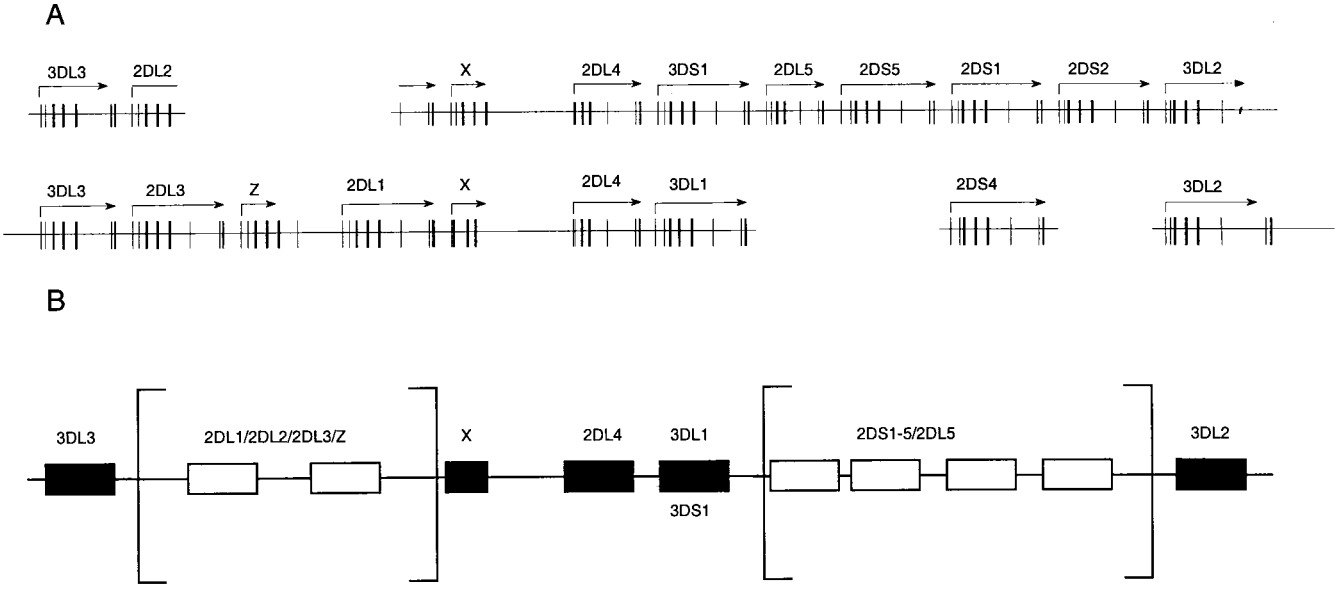

Fig. 3. (*A*) Differences in the organization of the *KIR* region in two different haplotypes. On the top line is the plot of the *KIR* region from 1060P11[2] (Fig. 2*A*). Partial sequence was obtained from PACs 72N6[1], 78E4[1], 1015E9[1], and 1015 M9[1] for the other haplotype (Fig. 1) by using PCR as well as comparison with other data (see Fig. 1). The gene distances are not to scale and have been exaggerated to display contiguity, suggested by sequence comparisons. Thus, *2DL2* is shown as a composite of two genes on the second haplotype, namely *2DL3* and *2DL1*. Overall, the genes are so similar that the precise lineup of the alleles remains uncertain, but *3DS1* and *3DL1* are shown as alleles, as supported by other data (see below and ref. 16). (*B*) Framework genes and variable bubbles in the *KIR* region. Comparison of the gene organization data in Fig. 2*A*, from the two haplotypes, with other data (16) is consistent with invariant framework loci, flanking regions where the gene number shows marked flexibility, indicated as open boxes flanked by square boxes. A similar model has been proposed to explain (mostly interspecies) gene expansion/contraction in the MHC, proposing independent expansion of class I genes within a framework of ancestral loci (48).

The two genes are highly homologous in both the exon and intron sequences, consistent with an allelic relationship. On haplotype 2, a set of three activating *KIR* genes (*KIR2DS5, DS1*, or *DS2*) was present between *KIR3DS1* and *3DL2*, whereas only one activating locus, *2DS4*, was found at the corresponding position in haplotype 1. The sequences of these genes are very similar to each other. It was therefore difficult to determine to which locus on haplotype 2 *KIR2DS4* is allelic, and it may represent a distinct locus.

Further Centromeric, Haplotype 2 Displayed a Single *KIR* Gene, *KIR2DL2*. The cDNA sequence for this locus shows similarities to the extracellular domain of 2DL3 and the transmembrane and cytoplasmic part of *2DL1.* Interestingly, haplotype 1 contained *KIR2DL3* and *2DL1* at the position corresponding to *2DL2* (Fig. 3*A*). Detailed comparison of the genomic sequences revealed the precise relationship between the two haplotypes in this region, as shown in Fig. 3*A*. A deletion extending from exon 5 of *KIR2DL3* and exon 6 of *KIR2DL1* would result in the formation of a composite gene, *2DL2,* with loss of the intervening *KIRZ* locus. This is the simplest scenario to account for the difference in organization, but such is the similarity of the sequences of all of the loci in this region, a more complex rearrangement should not be ruled out. The data suggest that there is flexibility in the presence/absence of certain *KIR* genes, whereas certain "framework" genes are always present, as depicted on Fig. 3*B*. A similar situation prevails in other variable regions, such as the MHC, where genes like *C4* or *DRB* may be present in variable numbers on different haplotypes (28, 29), and other genes such as *DRA* are invariably present as singletons. In the case of the MHC framework, "orthologous" regions are conserved across species such as mice (30). *KIR* genes have not been identified in rodents. As mentioned above, two of the framework genes, present in all or most haplotypes, are positioned at the end of the complex (*KIR3DL3, KIR3DL2*) adjacent to unreiterated sequence. Similarly, the internal framework locus, *KIR2DL4*, also sits next to unique sequence. This property may help prevent loss of these genes by DNA looping, to which the other genes with variable presence may be prone.

Minisatellites in *KIR* Genes. A feature of all of the *KIR* genes, with the exception of *KIR2DL4*, is the presence of a sequence resembling a classic moderately G+C-rich minisatellite. The repeat unit of 19–20 residues is typically GGGCCTGGAGGGAGATAT. Taking into account all of the repeats, none of the nucleotides is totally invariant at any of the positions, although the first section is generally more conserved. The number of repeat units varies from around 30–60 (≈600–1,200 bp). Apart from the residues immediately flanking the consensus-splicing signals, the first introns of the *KIR* genes are wholly taken up by the minisatellites. They could be connected in some way to the variation in numbers of different *KIR* loci. Other G+C-rich minisatellites are associated with instability via meiotic recombination processes such as gene conversion or unequal crossing over (31). Another feature of minisatellites is their association with recombination hotspots (32). Recombination in the KIR region has not been studied so far. In view of the linkage disequilibrium of various combinations of *KIR* alleles, this topic deserves to be explored.

Repeat Analysis Is Consistent with the Recent Origin of the *KIR* Region. In addition to the microsatellites referred to above, the sequences were analyzed for repeats, including *Alus, LINEs*, and *SINEs*. Together they account for over 30% of the sequence exceeding the coding sequence ≈5-fold. The *Alu* family is the most frequently represented at a density of 0.49 *Alu* per kilobase, which is not atypical. Outstanding is the highly significant *Alu* S/J ratio. This ratio can be used as a measure of the age or plasticity of a given sequence. According to their evolutionary origin, *Alu*s can be divided into two main classes, J-*Alu* (old) and S-*Alu* (new), and various subclasses of S (33). Random *Alu*s in GenBank are at a universal S/J ratio of 3.00 (34). The S/J ratio over the *KIR* region approaches 70! This is consistent with a recent origin of the region since J-*Alu*s retrotransposed between ≈55–31 million years ago. The lack of a *KIR* region in mice is consistent with its recent conception. The S-*Alu*s are similarly located in all *KIR* genes, suggesting that all *KIR* loci were derived by duplication of a single primordial locus. The simplest explanation for development of the region is that of the emergence of a *KIR* gene in human ancestors after the mouse/human divide, followed by multiple duplications of the gene or its derivatives. This could have been followed by sequence exchange by gene conversion or nonreciprocal crossovers. In other words, the *KIR* region is a young region that has undergone considerable genetic turbulence. Indeed, the specificity of KIR for subsets of HLA-A, -B, or -C allotypes requires that these receptor/ligand combinations developed subsequent to the divergence of HLA loci from each other (35). Analysis of the *KIR* region in other primates will enable more precise dating of the origin of the region (36).

***ILT* Genes.** The *ILT* genes fell into two clusters. We sequenced a set of six *ILT* loci in the region proximal to the *KIR* loci on a 148-kb PAC clone, 52N12[1]. We called this group, in close proximity to the *LAIR*2 gene, *ILT* cluster I. We performed three color fluorescence *in situ* hybridization analyses (not shown) using PAC clones corresponding to the *FcαR/KIR* border (800P9), the second *ILT* region (598H20), and *NKG7*, a marker centromeric of the *ILTs*, demonstrating that the order of these groups of genes, centromere to telomere, is *NKG7-ILT* cluster II-*ILT* cluster I-*KIR*. The FISH results taken together with the molecular data and information from databases show the *ILT* cluster II to be within 150 kb of *ILT* cluster I (Fig. 1). The *LAIR1* locus was grouped together with the *ILT* cluster II (Fig. 1). *ILT* cluster I encoded on PAC 52N12[1] included the *ILT1, LIR6, ILT2,* and *ILT3* genes (9, 37). The *ILT9* and *10* genes and exons 1–4 of *KIR3DL3* were present on this PAC clone (Fig. 1*A*). Thus, 52N12[1] overlaps with clone 1060P11[2] (Fig. 4*A*).

A comparison of the exon/intron organization of the *ILT* genes revealed that *ILT2* conformed to the prototypic structure of inhibitory *ILT* genes as described recently for *ILT3* (12). The activating *ILT*s encoded in cluster I showed some variation. Although *ILT9* and *10* have an exon/intron organization characteristic of activating *ILTs*, an extra exon was found in the *LIR6* gene encoding for part of a stalk region that is extended in the *LIR6* transcript in comparison with any other activating *ILT. ILT1* had a largely extended intron 3′ of the last Ig domain exon, which results in the large gene size (15 kb). Upstream of the predicted translation start site in the *LIR6* and in the *ILT2* gene was an extra 5′ untranslated region exon present in matching transcripts (*LIR6a* and *MIRcl7*, respectively). Corresponding exons may be predicted in the *ILT1* and *ILT9* genes but not for *ILT3* and *ILT10*.

Comparison of the *ILT* genes (Fig. 4*A*) revealed that, unlike the *KIRs*, their introns are not highly homologous and do not contain many repeat elements with some exceptions, such as *ILT1.* These data suggest an older origin of the *ILT* region than the *KIRs*, consistent with the existence of rodent counterparts, the paired Ig-like receptors (*PIRs*), which are in a syntenic region on mouse chromosome 7 (38–40). Analysis of the repeat composition of the *ILT* sequence reveals a modern–ancient S/J *Alu* ratio of ≈5, a value that supports the contention of a greater maturity than the *KIR* complex. Sequences from the two haplotypes revealed a degree of variation in the *ILT* coding regions, for example: *ILT3* exon 12 A/G and *ILT4* Ig3 T/C, both of which correspond to known polymorphic cDNA for these genes.

Haplotype-Specific Variation in *ILT* Cluster II. A contig of six clones was formed outside the sequenced region described above that all encoded the *ILT4* locus. On the basis of a sequence polymorphism identified in the Ig3 domain of the *ILT4* gene, they

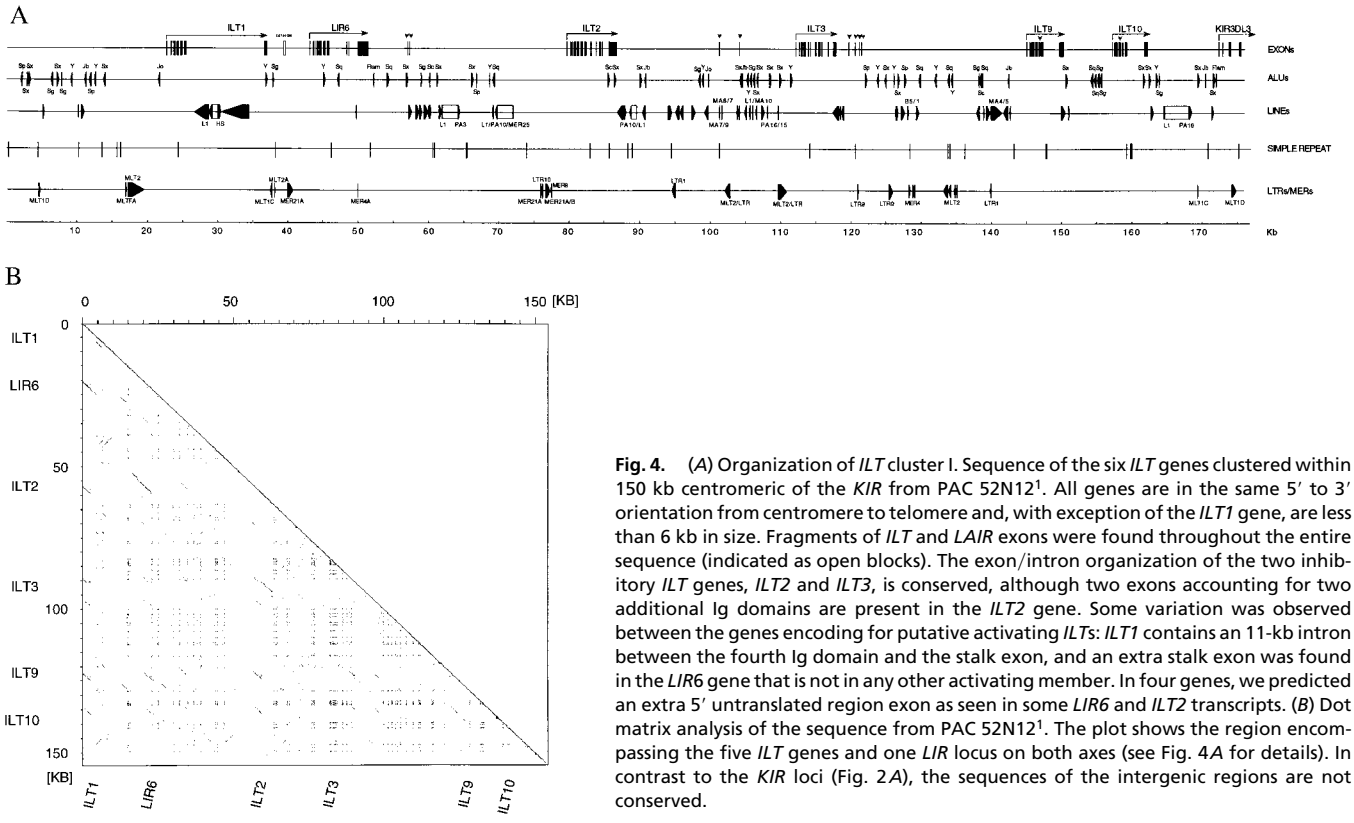

Fig. 4. (*A*) Organization of *ILT* cluster I. Sequence of the six *ILT* genes clustered within 150 kb centromeric of the *KIR* from PAC 52N12[1]. All genes are in the same 5′ to 3′ orientation from centromere to telomere and, with exception of the *ILT1* gene, are less than 6 kb in size. Fragments of *ILT* and *LAIR* exons were found throughout the entire sequence (indicated as open blocks). The exon/intron organization of the two inhibitory *ILT* genes, *ILT2* and *ILT3*, is conserved, although two exons accounting for two additional Ig domains are present in the *ILT2* gene. Some variation was observed between the genes encoding for putative activating *ILT*s: *ILT1* contains an 11-kb intron between the fourth Ig domain and the stalk exon, and an extra stalk exon was found in the *LIR*6 gene that is not in any other activating member. In four genes, we predicted an extra 5′ untranslated region exon as seen in some *LIR6* and *ILT2* transcripts. (*B*) Dot matrix analysis of the sequence from PAC 52N12[1]. The plot shows the region encompassing the five *ILT* genes and one *LIR* locus on both axes (see Fig. 4*A* for details). In contrast to the *KIR* loci (Fig. 2*A*), the sequences of the intergenic regions are not conserved.

were grouped according to their haplotype (Fig. 5*A*). The link between this cluster and the two haplotypes identified in the *ILT*/*KIR* region remains to be determined. The *ILT5* gene was at the centromeric end of this cluster on two PAC clones from each haplotype. *ILT8* was close to *ILT5*. Telomeric of *ILT8* is

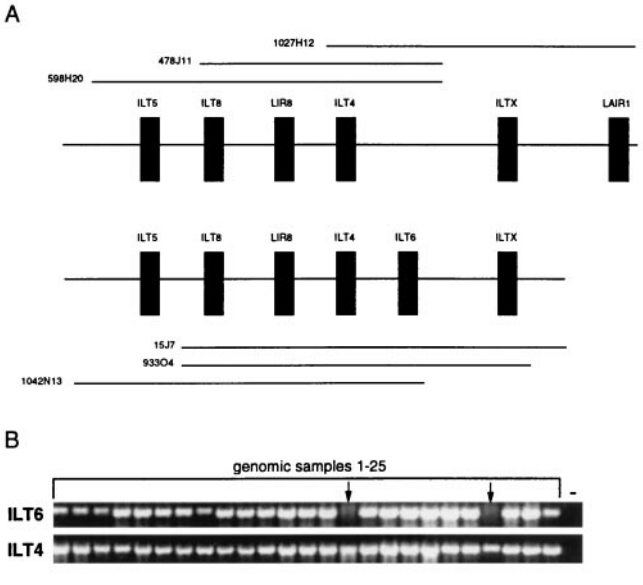

Fig. 5. (*A*) Haplotypic variation for presence of *ILT6* in *ILT* cluster II. The centromeric cluster of *ILT* genes (Fig. 1) contained six genes in most haplotypes. On the haplotype shown (*Top*), the *ILT6* gene is missing. (*B*) PCR analysis for the presence of *ILT6*. The genotypes of 25 individuals were analyzed by PCR by using *ILT6*- and *ILT4*-specific primers. The two samples that were negative for *ILT6* are indicated by arrows.

LIR8, which is present on PAC 478J11 and 933O4. On three PAC clones, we identified a *Bgl*II (2 kb) and a *Xba*I fragment (3 kb) by Southern blot analysis that could not be accounted for by any of the identified *ILT* genes. We mapped this fragment to the telomeric end of cluster II. This locus could be the *ILT7* gene, which we have not been able to amplify by PCR.

We amplified *ILT6* from three clones (15J7, 933O4, 1042N13) that correspond to a single haplotype (Fig. 5*A*). Because the only overlap between these clones is between *LIR8* and *ILT4*, the *ILT6* gene must be located either between these two loci or in close proximity to one of them. Two clones of the second haplotype also span this region but do not contain the *ILT6* gene. We identified a haplotype-specific 3.5-kb *Xba*I fragment by Southern blot analysis that is present only on the three *ILT6*-positive PAC clones. We postulate that the *ILT6* gene is absent on one haplotype in this PAC library and therefore shows presence/absence variation similar to what has been observed for some of the *KIR* loci. To examine this hypothesis further, we amplified *ILT6* from 25 genomic samples and found 2/25 homozygous negative for this locus (Fig. 5*B*). This is consistent with a gene frequency for the presence of the *ILT6* gene of about 0.72, (0.51 homozygotes, 0.41 heterozygotes, and 0.08 homozygote negative, applying the Hardy–Weinberg equation).

Discussion

All genes encoded in the LRC on human chromosome 19q13.4 are members of the Ig superfamily. Parts of their gene structures are remarkably conserved, and all are in the same head-to-tail orientation, with the possible exception of *ILT* complex II. The many *ILT*/*LAIR*/*KIR* gene fragments we found in the *ILT* region (Fig. 4) could be the fallout of abortive rearrangements at repeated duplications throughout evolution.

Rodents have genes equivalent to *NKG7* (41) and *NKp46* (42) as well as the probable *ILT* orthologues, the paired Ig-like

receptors (*PIR*s), which are located on the syntenic mouse chromosome 7 (40). However, no rodent *KIR* genes have been identified. There are at least 14 mouse *Ly-49* genes, which fulfill a function similar to the *KIR*s (43, 44). These are encoded on mouse chromosome 6 in the NK complex. Only a single human *LY49* pseudogene has been identified in the syntenic region on human chromosome 12 (45). Multiple duplication events could have led to rapid expansion of the *KIR* gene family in primates (2) and, conversely, the *Ly49* genes in rodents. This argues for convergent evolution of function of these receptors.

Recently, a "hybrid" cDNA molecule, *KIR2DL1v*, was identified that appeared to comprise a *2DL1* sequence with the proximal part of the second Ig domain to the TM region being replaced by a sequence resembling *KIR2DS1* (46). Our data show how the two putative parent genes are unlikely to be alleles. They are separated by over 50 kb of intervening DNA and are in different variable *KIR* regions. These data could be explained by some form of nonreciprocal recombination such as gene conversion, which is known to operate in the MHC (27). The arrangement of *KIR* genes consisting of highly related sequences in the same orientation may provide the ideal substrate for gene conversion.

Close examination of the *KIR* region shows that the sequences upstream of the transcribed region are remarkably similar, with the exception of the *2DL4* gene, suggesting common promoters. Different groups of 2–9 *KIR* genes are expressed in NK clones, but the sequences of the *KIR* promoter sequences are homogeneous, except for that of the *KIR2DL4*. Therefore, it seems likely that regulation of expression of *KIR* transcripts is facilitated by a stochastic process. The sequence upstream of *KIR2DL4* may be significant because this gene is unique among its peers in being expressed in 100% of NK clones (15). *2DL4* is the only *KIR* gene lacking the repeat region in intron 1.

Examination of the LRC from two different haplotypes revealed variation in the *KIR* cluster. *KIR* genes present on all haplotypes represent framework loci that flank regions of variability. In these heterogeneous regions, there are at least 11 possible *KIR* loci. If we accept this model as a first approximation of the different arrangements of *KIR* genes, there are at least two main positions where haplotypes may differ (Fig. 3). This hypothesis would account for a large number of different haplotypes, and it concurs with the variation in the number of *KIR* genes observed in different genotypes (ref. 16; unpublished data). The frequencies with which certain combinations of *KIR* loci are found in different individuals exceeded levels expected from random association (16), indicative of linkage disequilibrium of alleles on haplotypes. The independent segregation of *KIR3DL1* and *KIR3DS1* suggested that these two specificities were alleles. This is the case for the two haplotypes on Fig. 3.

Taking into account further variation in sequences, particularly *ILT4* and *ILT5* (17), as well as the presence/absence of the *ILT6* gene, the LRC is clearly highly variable. Another region of the genome that exhibits extensive variation is the MHC, the products of which are ligands for some of the *KIR* and *ILT* molecules. The MHC *class I*, *class II,* and *C4* genes exhibit high levels of variability for presence/absence (28, 47). The common functional link to both sets of loci is resistance to pathogens. Like the MHC, the LRC has all the hallmarks of a dynamic genomic region. Selection for variation in *KIR* gene arrangement could be infection, in which case we may expect to find some haplotype frequencies skewed in different diseases.

We thank the Medical Research Council, the Wellcome Foundation, the European Economic Community (CT961105), and the Imperial Cancer Research Fund for support, and A. Ziegler, A. Volz, and A. Jeffries for helpful discussions, as well as Prof. Pieter de Jong for the genomic DNA (BACPAC Resources, Oakland, CA).

1. Karre, K. & Colonna, M., eds. (1998) *Curr. Top. Microbiol. Immunol.* **230**.
2. Parham, P. (1997) *Immunol. Rev.* **155**.
3. Lanier, L. L. (1998) *Annu. Rev. Immunol.* **16,** 359–394.
4. Brown, M., Scalzo, A., Matsumoto, K. & Yokoyama, W. (1997) *Immunol. Rev.* **155,** 53–65.
5. Plougastel, B. & Trowsdale, J. (1998) *Genomics* **49,** 193–199.
6. D'Andrea, A., Chang, C., Franz-bacon, K., McClanahan, T., Phillips, J. H. & Lanier, L. L. (1995) *J. Immunol.* **155,** 2306–2310.
7. Suto, Y., Maenaka, K., Yabe, T., Hirai, M., Tokunaga, K., Tadokoro, K. & Juji, T. (1996) *Genomics* **35,** 270–272.
8. Borges, L. (1997) *J. Immunol.* **159,** 5192–5196.
9. Colonna, M., Nakajima, H., Navarro, F. & Lopez-Botet, M. (1999) *J. Leukocyte Biol.* **66,** 375–381.
10. Wagtmann, N., Rojo, S., Eichler, E., Mohrenweiser, H. & Long, E. O. (1997) *Curr. Biol.* **7,** 615–618.
11. Meyaard, L., Adema, G. J., Chang, C., Lanier, L. L. & Phillips, J. H. (1997) *Immunity* **7,** 283–290.
12. Torkar, M., Norgate, Z., Colonna, M., Trowsdale, J. & Wilson, M. (1998) *Eur. J. Immunol.* **28,** 3959–3967.
13. Wende, H., Colonna, M., Ziegler, A. & Volz, A. (1998) *Mamm. Genome* **10,** 154–160.
14. Selvakumar, A., Steffens, U. & Dupont, B. (1997) *Immunol. Rev.* **155,** 183–195.
15. Valiante, N., Lienert, K., Shilling, H., Smits, B. & Parnham, P. (1997) *Immunol. Rev.* **155,** 155–164.
16. Uhrberg, M., Valiante, N. M., Shum, B. P., Shilling, H. G., Lienert-Weidenbach, K., Corliss, B., Tyan, D., Lanier, L. L. & Parham, P. (1997) *Immunity* **7,** 753–763.
17. Colonna, M., Navarro, F., Bellon, T., Llano, M., Garcia, P., Samaridis, J., Angman, J., Cella, M. & Lopez-Botet, M. (1997) *J. Exp. Med.* **186,** 1809–1818.
18. Ioannou, P. A., Amemiya, C. T., Garnes, J., Kroisel, P. M., Shizuya, H., Batzer, M. A. & de Jong, P. J. (1994) *Nat. Genet.* **6,** 84–89.
19. Beck, S. & Alderton, R. P. (1993) *Anal. Biochem.* **212,** 498–505.
20. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 5463–5467.
21. Sanger, F. (1998) *Genome Res.* **8,** 1097–1108.
22. Sonnhammer, E. L. & Durbin, R. (1995) *Gene* **167,** 1–10.
23. Nakai, K. & Horton, P. (1999) *Trends Biochem. Sci.* **24,** 34–36.
24. Bairoch, A. (1990) *Prosite: A Dictionary of Protein Sites and Patterns,* 5th Ed. (Département de Biochimie Médicale, Université de Genève, Geneva).
25. Long, E. O., Colonna, M. & Lanier, L. L. (1996) *Immunol. Today* **17,** 100.
26. Wilson, M. J., Torkar, M. & Trowsdale, J. (1997) *Tissue Ant.* **49,** 574–579.
27. Hogstrand, K. & Bohme, J. (1999) *Immunol. Rev.* **167,** 305–317.
28. Trowsdale, J., Ragoussis, J. & Campbell, R. D. (1991) *Immunol. Today* **12,** 443–446.
29. Campbell, R. D. & Trowsdale, J. (1993) *Immunol. Today* **14,** 349–352.
30. Amadou, C., Kumanovics, A., Jones, E. P., Lambracht-Washington, D., Yoshino, M. & Lindahl, K. F. (1999) *Immunol. Rev.* **167,** 211–222.
31. Jeffreys, A. J., Neil, D. L. & Neumann, R. (1998) *EMBO J.* **17,** 4147–4157.
32. Jeffreys, A. J., Murray, J. & Neumann, R. (1998) *Mol. Cell* **2,** 267–273.
33. Jurka, J. & Milosavljevic, A. (1991) *J. Mol. Evol.* **32,** 105–121.
34. Jurka, J. & Smith, T. (1988) *Proc. Natl. Acad. Sci. USA* **85,** 4775–4778.
35. Parham, P. (1994) *Semin. Immunol.* **6,** 373–382.
36. Zietkiewicz, E., Richter, C., Malalowski, W., Jurka, J. & Labuda, D. (1994) *Nucleic Acids Res.* **22,** 5608–5612.
37. Cosman, D., Fanger, N., Borges, L., Kubin, M., Chin, W., Peterson, L. & Hsu, M.-L. (1997) *Immunity* **7,** 273–282.
38. Yamashita, Y., Fukuta, D., Tsuji, A., Nagabukuro, A., Matsuda, Y., Nishikawa, Y., Ohyama, Y., Ohmori, H., Ono, M. & Takai, T. (1998) *J. Biochem.* **123,** 358–368.
39. Hayami, K., Fukuta, D., Nishikawa, Y., Yamashita, Y., Inui, M., Ohyama, Y., Hikida, M., Ohmori, H. & Takai, T. (1997) *J. Biol. Chem.* **272,** 7320–7327.
40. Kubagawa, H., Burrows, P. D. & Coopers, M. D. (1997) *Proc. Natl. Acad. Sci. USA* **94,** 5261–5266.
41. Berg, S. F., Westgaard, I. H., Fossum, S. & Dissen, E. (1999) *Immunogenetics* **49,** 815–818.
42. Falco, M., Cantoni, C., Bottino, C., Moretta, A. & Biassoni, R. (1999) *Immunol. Lett.* **68,** 411–414.
43. McQueen, K. L., Freeman, J. D., Takei, F. & Mager, D. L. (1998) *Immunogenetics* **48,** 174–183.
44. Brown, M. G., Fulmek, S., Matsumoto, K., Cho, R., Lyons, P. A., Levy, E. R., Scalzo, A. A. & Yokoyama, W. M. (1997) *Genomics* **42,** 16–25.
45. Barten, R. & Trowsdale, J. (1999) *Immunogenetics* **49,** 731–734.
46. Shilling, H. G., Lienert-Weidenbach, K., Valiante, N. M., Uhrberg, M. & Parham, P. (1998) *Immunogenetics* **48,** 413–416.
47. Campbell, R. D. & Trowsdale, J. (1997) *Immunol. Today* **18** (Suppl.).
48. Amadou, C. (1999) *Immunogenetics* **49,** 362–367.