Genes and Immunity (2000) 1, 280–287
© 2000 Macmillan Publishers Ltd All rights reserved 1466-4879/00 $15.00
www.nature.com/gene

The centromeric part of the human NK gene complex: linkage of LOX-1 and LY49L with the CD94/NKG2 region

C Bull[1], Y Sobanov[2], B Röhrdanz[1], J O'Brien[1], H Lehrach[1] and E Hofer[2]

[1]*Max-Planck-Institute for Molecular Genetics, Ihnestrasse 73, D-14195 Berlin, Germany;* [2]*Department of Vascular Biology and Thrombosis Research, Vienna International Research Cooperation Center, University of Vienna, Brunnerstrasse 59, A-1235 Vienna, Austria*

The natural killer (NK) gene complex is a genomic region containing lectin-type receptor genes. We have established a contig of PAC and BAC clones comprising about 1 Mb of the centromeric part of the NK gene complex. This region extends from the LOX-1 gene, which encodes a receptor for oxidized LDL and was found within 100 kb telomeric of the STS marker D12S77, contains the CD94 and NKG2 NK receptor genes and reaches beyond D12S852 on the proximal side. In this part we have mapped the human LY49L gene, a homologue of the rodent Ly49 genes, which encode important MHC class I receptors for the regulation of NK cell activity in rodents. The LY49L gene is localized 100 to 200 kb centromeric of the NKG2 gene cluster and 300 to 400 kb telomeric of the STS marker D12S841. Genomic sequencing of the complete gene including promoter and intron sequences confirmed that the structure is similar to the mouse Ly49 genes. Screening of several cDNA libraries did not detect any transcripts of putative additional human LY49 genes. In addition, in the course of these studies several EST sequences were localized in the region, one immediately upstream of the LY49L gene. Genes and Immunity (2000) **1,** 280–287.

Keywords: *natural killer cells; NK gene complex; chromosome 12; LOX, CD94, NKG2 and LY-49 genes*

Introduction

Cytotoxicity of natural killer (NK) cells towards tumor or virally infected target cells[1] is regulated by inhibitory and activating signals from cell surface receptors. NK cell receptors that recognize MHC class I or related molecules on target cells[2–6] can be grouped into two structurally distinct classes, namely the immunoglobulin-(Ig)-like and the C-type lectin receptors. The killer immunoglobulin-like receptors (KIR) have been found to be expressed on human NK cells and different variants display either repressive or triggering functions. They are encoded by a cluster of genes in a region on chromosome 19 named the leukocyte receptor complex (LRC).[7,8] A gene distantly related to the Ig-like receptor gene family has recently also been detected in mice, but its relevance for rodent NK cell function has not been firmly established.[9]

C-type lectin NK receptors, which occur in both inhibitory and activating forms and are encoded within a syntenic region on human chromosome 12,[10–13] mouse chromosome 6[14,15] and rat chromosome 4,[16] called the NK gene complex, can be sub-grouped into different families. Among these are the CD94/NKG2, the Ly49 and the NKR-P1 families. Whereas members of the CD94/NKG2 and the NKR-P1 families have been found to be expressed in mice,[17–19] rats[20] and humans,[5,21–24] expression of functional Ly49 proteins has so far only been detected in rodents.[15,19]

Recently, an aberrantly spliced cDNA coding for a human LY49 molecule (named Ly49L) has been detected[25] in the expressed sequence tags (EST) database and in cDNA libraries derived from human NK cells. Therefore genes related to all different NK cell receptor classes have now been found in rodents and in humans. However, it remains unclear whether intact gene families related to the human KIR receptors and the rodent Ly49 receptors coexist and fulfill similar functions in human as well as in rodent NK cells or whether these two receptor classes, which bind classical MHC class I molecules, have evolved differently from common ancestor genes in humans and rodents.

Furthermore, there are additional lectin-type receptors encoded in the NK gene complex which possess different functions in various cell types. The CD69 or AICL receptors[26,27] are widely expressed in the hematopoietic lineage and their functions and ligands are not clearly defined. Recently, the lectin-type LLT1 receptor was reported to be broadly expressed in lymphocytes.[28] Furthermore, a member of another lectin-type receptor class, LOX-1, has been implicated as a receptor for oxidized low-density lipoprotein on endothelial cells and monocytes.[29]

In order to establish the basis to search for novel forms of lectin-type receptor genes, to link the *NKG2* and *CD94* NK cell receptor genes[13,23] with the region containing the

Correspondence: Erhard Hofer, Department of Vascular Biology and Thrombosis Research, Vienna International Research Cooperation Center, University of Vienna, Brunnerstrasse 59, A-1235 Vienna, Austria
This work was supported in part by a grant from the Austrian National Bank (Jubiläumsfond No. 7567) to E.H.
Supplementary material to this paper is available from the author.
Received 28 January 2000; revised and accepted 22 February 2000

LY49L and other potentially related genes and to get some clues to the evolution of the locus, we have now constructed a physical map based on PAC and BAC clones of the centromeric part of the NK complex region. We have focused here on the centromeric part of the NK gene complex since the available data indicate that the syntenic region of rodents contains more than nine *Ly49* genes and resistance loci for cytomegalovirus and pox virus.[15] In addition we find the *LOX-1* gene to be located 200 kb telomeric of *CD94*. Sequencing of the complete *LY49L* gene confirmed previous results concerning the exon/intron structure.[30] Furthermore, we have identified several genes represented as EST sequences in the analyzed region, one EST is encoded within a very short distance to the 5′-end of the *LY49L* gene.

Results

A contig of PAC and BAC clones comprising the centromeric part of the NK gene complex

Starting from the *NKG2/CD94* gene region[13,23] and the *LY49L* gene we have established contigs of PAC and BAC clones that extend over about 1 Mb and contain the centromeric part of the NK gene complex. Most parts of the region are covered by several clones (see Figure 1). The *LOX-1* gene[29] was detected in clone L21228, the first clone on the telomeric side of this contig, and this gene is therefore located about 100 kb telomeric of the STS marker *D12S77*, which flanks the *NKG2/CD94* region. The *CD94* gene and all five *NKG2* genes are placed between 100 to 200 kb centromeric of *D12S77* and their relative orientation and linkage has been described previously.[13] The *LY49L* gene is 100 to 200 kb proximal of the *NKG2A* gene and has been linked to the STS marker *D12S841* which is found 200 to 300 kb centromeric of *LY49L*. The contig region extends further beyond the STS markers *D12S112* and *D12S852* and therefore covers about 600 kb of the region proximal of *LY49L*.

Additional genes and ESTs in the region

By end-sequencing of individual PAC clones several sequences identical or closely similar to database EST entries were detected in the region. Thirty to 50 kb telomeric of *LY49L*, a sequence highly similar or identical to the human translation initiation factor eIF-2 gamma subunit, was found (L19161, HUMIEF2G). Immediately upstream of the *LY49L* gene (559 bp from the translation start site) we mapped the EST AA005053/AA005005. On the T7 end of P23267 and on the SP6 end of G21796 sequences identical and highly similar to GenBank EST AA749426 and AA846955, respectively, have been detected.

Features of the *LY49L* gene

To reveal any unusual features of the *LY49L* gene in comparison to the rodent Ly49 genes, it was completely sequenced including the introns and promoter region using a shotgun library that was produced from the clone RCPIp704I2437. The complete gene covers a chromosomal region of 10.7 kb (EMBL/GenBank accession numbers BankIt308184 and AF213453). A genomic organisation of seven exons (329 bp, 167 bp, 90 bp, 237 bp, 143 bp [+82 bp intron], 107 bp and 378 bp) interrupted by six introns (1312 bp, 1057 bp, 964 bp, 1683 bp, 1480 bp [1398 bp] and 2729 bp) was confirmed as has been previously published.[30] The first exon, however, containing only untranslated sequence extends at least 285 bp beyond the 5′-end described by others.[25,30] We also noticed that the exon-intron boundaries of the human *LY49L* gene (Figure 2a) differ for all exons from the sequences that have been found most frequently at mammalian splice sites.[31,32] The cDNAs isolated by us and others always contained a part of intron 5. This is due to the use of a cryptic splice site in intron 5. An alignment of all known *Ly49* genes from mouse, rat and human (Figure 2b) shows that the very 3′ base of exon 5 is a G in all rodent genes, whereas it is a C in the human *LY49L* gene. The G is most frequently found at the 3′ end of mammalian exons.

The potential promoter region of the *LY49L* gene

The approximate position of the potential transcription start site was inferred from a cDNA clone that contained a 381 bp sequence stretch upstream from the predicted translation initiation codon. This region was not functionally tested but as another gene (EST AA005053) was already located only 559 bp upstream of this site, it seems unlikely that transcription initiates much further upstream. The 559 bp sequence was analyzed in order to find potential promoter binding sites by the SIGSCAN algorithm (Version 4.05, http://bimas.dcrt.nih.gov:80/molbio/signal/) developed by Dan S Prestridge. This analysis suggested some potential TATA boxes and transcription factor consensus binding sites based on the TRANSFAC database[33] (see Figure 3).

Screening for additional *LY49L* genes

After screening two natural killer cell cDNA libraries as well as human fetal liver, fetal lung, fetal spleen and spleen cDNA libraries we detected another incompletely spliced cDNA in a human fetal lung library in addition to the one described previously.[25] This cDNA lacked exon 4 and contained part of intron 5 due to the use of a cryptic splice site in the intron. However, we did not identify any clone that would encode an LY49L molecule with a complete carbohydrate recognition domain. Taking into account the possibility that other human *LY49* genes might have diverged quite far from the rodent genes, we performed homology screening under reduced stringency, but we could not isolate any different *LY49* variant by this approach.

Discussion

The NK gene complex region has presumably evolved by consecutive gene duplications and contains several lectin-type genes, which display selective expression in NK and NK/T cells and function as MHC class I receptors, such as the Ly49 receptors in rodents and the CD94/NKG2 receptors in rodents and humans.[3,5,18] Other forms, such as the *CD69*, the *AICL*, and the *LLT1* genes,[10,13,15,34] are more widely expressed and may possess different functions. Furthermore, we have now located the lectin-type receptor for oxidized low-density lipoprotein, which is expressed on endothelial cells and macrophages,[35] just 200 kb telomeric of the *CD94* gene. Thus this region contains lectin-type receptors important for various functions of hemopoetic cells and endothelial cells which are capable of binding ligands as varied as MHC class I/peptide complexes, oxidized LDL or carbohydrates.

Figure 1 Physical map of the centromeric part of the human NK gene complex. A contig of PAC and BAC clones was established starting from the *LY49L* gene. In addition, the region around the *NKG2* genes was extended. The whole region covers about 1 Mb. Four clones (marked *) proximal to *D12S77* containing the *CD94* and the *NKG2* genes have been reported previously.[13] Additional clones are from the RPCI-1,3,4,5 PAC library and the RPCI-11 BAC library of the RZPD or from the CEPH BAC library (marked +). BAC clones are indicated by a b in front of the clone designation. Overlapping clones were identified using labeled end-probes and the presence of the individual sequences of the clone ends, genes and STS markers in the different clones were determined by PCR as described in materials and methods. The line at the top displays the genes, PAC-ends and STS markers detected in the clones shown below. In addition, some EST's detected by partial sequencing are shown.

a

Number	Exon Size	Intron size	3′ Intron sequence (splice donor site)		Exon sequence	5′ Intron sequence (splice acceptor site)		Intron phase
			yyyy-cag	G-----	Consenus	---MAG	gtragt	
I	329 bp	1312 bp		ACAAAA	Exon 1	ATACAG	gtaata	
II	167 bp	1057 bp	ttcaacag	AAACCA	Exon 2	ACAAAG AspLysG	gtatgc	1
III	90 bp	964 bp	atttaaag	AATTTT luPhe	Exon 3	CAAATA ThrAsnI	gtgagt	1
IV	237 bp	1683 bp	aattttag	TCTTTC lePhe	Exon 4	ATACAG AsnThrG	gtacaa	1
V	143 bp (potential)	1480 bp			Exon 5 (potential)	GAACTC GluLeu	gtattt	0
V	225 bp (observed)	1398 bp	tattgtag	GCAAAT lyLys	Exon 5 (observed)	TCCAGA	gtgata	
VI	107 bp	2729 bp	tattatag	GCCTTC AlaPhe	Exon 6	TGGAAT GlyIl	gtaagt	2
VII	373 bp		ttctacag	TAATTCT eAsnSer	Exon 7			

b

```
             1001                                                 1050
hLY-49L      GACAAAGATG AACTCGTATT TTACATTCAC TTTTATTCTC TTGGACTCTG
mly-49E-GE   GATGAAGATG AACTG..... .......... .......... ..........
mly-49F-GE   GATGAAAATG AACTG..... .......... .......... ..........
mly-49C      GATGAAGATG AACTG..... .......... .......... ..........
mly-49H      GATGAAGATG AACTG..... .......... .......... ..........
mly-49G1     AATGAGGATG AACTG..... .......... .......... ..........
mly-49G3     AATGAGGATG AACTG..... .......... .......... ..........
mly-49G2     AATGAGGATG AACTG..... .......... .......... ..........
mly-49G4     AATGAGGATG AACTG..... .......... .......... ..........
mly-49A      GATGAGGATG AACTG..... .......... .......... ..........
mly-49D-GE   GATGAGGATG AACTG..... .......... .......... ..........
mly-49B      GATGAGGATG AATTG..... .......... .......... ..........
rly-49-12    GATGAGGATG AACTG..... .......... .......... ..........
rly-49-19    GATGAGGATG AACTG..... .......... .......... ..........
rly-49-29    GATGAGGACG AACGG..... .......... .......... ..........
rly-49-9     GATGAGGACG AACGG..... .......... .......... ..........

             1051                                                 1100
hLY-49L      TTTCTCAATG TTGGACCTAA GATATTGAAG ACAGGCTGGA GTCCAGAGCC
mly-49E-GE   .......... .......... .......... .......... .......AAA
mly-49F-GE   .......... .......... .......... .......... .......AAA
mly-49C      .......... .......... .......... .......... .......AAA
mly-49H      .......... .......... .......... .......... .......AAA
mly-49G1     .......... .......... .......... .......... .......AAG
mly-49G3     .......... .......... .......... .......... .......AAG
mly-49G2     .......... .......... .......... .......... .......AAG
mly-49G4     .......... .......... .......... .......... .......AAG
mly-49A      .......... .......... .......... .......... .......AAG
mly-49D-GE   .......... .......... .......... .......... .......AAG
mly-49B      .......... .......... .......... .......... .......AAG
rly-49-12    .......... .......... .......... .......... .......ATG
rly-49-19    .......... .......... .......... .......... .......CTG
rly-49-29    .......... .......... .......... .......... .......AAG
rly-49-9     .......... .......... .......... .......... .......AAG

             1101                                                 1150
hLY-49L      TTCATTCAAT CTCAGATTTA TGAAAATAAT TACTGGATCG GATTATCATA
mly-49E-GE   TTCCTTCAGT TCCAGGTTAT TTCAGACAGT TACTGGATTG GATTGTCATA
mly-49F-GE   TTCCTTCAAT TCCAGGTTAT TCCAGACAGT TACTGGATTG GATTGTCATA
mly-49C      TTCCTTCAAC GCCATGTTAT TCCAGAGAAT TACTGGATTG GATTGTCTTA
mly-49H      TTCCTTCAAC GCCATGTTAT TCTAGAGAGT TACTGGATTG GATTGTCATA
mly-49G1     TTCCTTCAGA ACCTGGCTCC TTCAGACATT TCCTGGATTG GATTGTCATA
mly-49G3     TTCCTTCAGA ACCTGGCTCC TTCAGACATT TCCTGGATTG GATTGTCATA
mly-49G2     TTCCTTCAGA ACCTGGCTCC TTCAGACATT TCCTGGATTG GATTGTCATA
mly-49G4     TTCCTTCAGA ACCTGGCTCC TTCAGACATT TCCTGGATTG GATTGTCATA
mly-49A      TTCCTTCATC TCGTGGTTCC TTCAGACAGT TGCTGGGTTG GATTGTCATA
mly-49D-GE   TTCCTTCAGC TCGTGGTTCC TTCAGACAGT TGCTGGATTG GATTGTCATA
mly-49B      TTCCTTAAAT CCCAACTTCA AAGAAACACA TACTGGATTT CACTGACACA
rly-49-12    TTCCTTCATC TCCTTGTTAC TCCAGACAGT TACTGGATTG GATTGTCTTA
rly-49-19    TTCCTTCATC TCCTGGTTAC TCCAGACAGT TACTGGATTG GATTATTTTA
rly-49-29    TTCCTTCAGC AACAGCTTAT TACAGACAAT TACTGGATTG GATTTTCATA
rly-49-9     TTCCTTCAGC AACAGCTTAT TCCAGACAAT TACTGGATTG GATTTTCATA
```

Figure 2 Sequence at the exon-intron boundary of the *LY49L* gene. Exon-Intron boundary sequences of the human *LY49L* gene do not always match the known splice site consensus sequences. (**a**) The information on exon size, intron size, splice donor site and splice acceptor site, intronic and exonic sequences and the intron phase are given across the lines for all exons and introns (numbers from I to VII are shown down the left column). Nucleotides that match the splice consensus sequences given in the second line are underlined. (**b**) For comparison the corresponding sequences of the 3′-border of exon 5 and the 5′-border of exon 6 are shown for 15 *Ly49* genes. An alignment of the cDNA sequences of all fully sequenced *Ly49* genes from mouse (11 sequences), rat (4 sequences) and human (1 sequence) that has been performed using PileUp algorithm from the GCG10 software is shown. The alignment shows that the 3′ terminal base of exon 5 in all known rodent *Ly49* genes is a G whereas in the human *LY49L* gene a C was found. This may have lead to the activation of a cryptic splice site in intron 5 which leads to incomplete splicing of this intron.

```
                              TATA element                       c-Myb
  1  TAAAAAAAAA AAGTACAGAC CAAATATATT TTACAATAGA AAAAATTGAA
                    CCAAT
 51  AACAATTTAA GTATCCAATA GTAGGGGATT GCCTAAATAA ATGATGGTTT
                           F2F
101  ATCCCTGTGA GTAATATTTT TAAAATAACA TTTCAGATGA ATTTCTTTGA

151  AATAATGGAA TACATATTAT TTATTAGAAA ATAATAAATA AAAAATACTG

201  TGTAAGGAAT GATACCAACC TTGATAGTAA GCACAACATA CAGAGAAAAA

251  AAACTGGAAG AATATATGCC ACAATTTTAC AGAGGAATTA CGGATCTTTT
                TATA element
301  AAAGTAATTT AAAAAATACA TTTACTTGAA ATAAACAGAT AGAAAAAGGA
           TATA element
351  ACTTTTAAAA AATAATGACA TATAGGAGAA CATTTCCAAC AGAGAGACTG

401  TTATTTAAAC TAAGAGCTAA GTAGTTCAAT CATTACACTT CTTTTATATG
      TATA element
451  TCTATATTTT TATACATAGA CATTTCAGTA ATGTCTTTTA CCAGCAGTTT
                                        TATA element  Hox4
501  CTCTAAAGTA CGTTAGTTTT TTGTTAATGA TATTAGTTTT ATTTTCTCCA

551  AAGATGTGT
```

Figure 3 Predicted TATA boxes and transcription factor binding sites in the potential promoter region of the *LY49L* gene. The region 559 bp upstream of the area of the presumptive transcription start site has been analyzed in order to characterize the promoter region using the SIGSCAN algorithm (Version 4.05, http://bimas.dcrt.nih.gov:80/molbio/signal/) developed by Dan S Prestridge. The sites are marked with boxes and the transcription factors are given above.

Concerning the genes and gene families with specific expression pattern and function in cells with natural killer activity, a single *CD94* gene and several *NKG2* genes have been described for humans,[13,36] and have more recently also been detected in rodents.[17,20,37] In mice the NK gene complex adjacent to the *CD94* and *NKG2* genes contains a cluster of at least nine *Ly49* genes spread over a region of about 500 kb.[15] Genetic mapping data in the mouse have further revealed the presence of resistance loci to mouse cytomegalovirus (*Cmv1*), to mouse pox virus (*Rmp1*) and a locus conferring the capacity to kill Chinese hamster ovary cells (*Chok*) close to the region containing the *Ly49* genes.[15] The corresponding syntenic region of human chromosome 12 should be contained in the PAC contig established in this work reaching from the *NKG2A* gene beyond *D12S852* (Figure 1). The human *Ly49L* gene has been precisely mapped in this region and it remains to be seen whether additional genes and features of this region are shared between the mouse and human genome.

Since the previously isolated Ly49L cDNA was the product of an aberrant splicing event, we have searched several cDNA libraries including NK cell and fetal cDNA libraries. This screening detected only cDNAs spliced by using a cryptic splice site in intron 5 of the *LY49L* gene. These cDNAs encode proteins which have the second and third carbohydrate recognition domain of the lectin part replaced by a different sequence. This results from a part of intron 5 retained in the cDNA and an out of frame fusion with the exon 6 sequences leading to a premature stop codon. A closer inspection and comparison of the sequences of the exon 5/intron 5 boundaries of the *LY49L* gene shows that in all cloned Ly49 cDNA sequences a G nucleotide is absolutely conserved at the very 3′-end of exon 5 among mouse and rat molecules (Figure 2b). In accordance with the exon theory of splicing, an involvement of the 5′ and 3′ exon sequences in the regulation of splicing is postulated. It can therefore be speculated that the CTG to CTC change at the 3′ end of exon 5, which represents a silent mutation (leucine in both cases), interferes with proper 5′-splice site recognition and as a consequence a cryptic splice site in intron 5 is activated.

It is intriguing that splice variants and truncated or altered C-terminal portions of the lectin type receptors are frequently observed. For example, several cDNA variants for the mouse Ly49G molecule have been isolated.[38] The Ly49G1 sequence has a prolonged exon 4 with a sequence stretch that could represent a splice donor site from the intron. Ly49G3 also lacks a part of exon 4 plus most of exon 5. From the human lectin-type receptors two members of the NKG2 family, NKG2E and NKG2F have altered C-terminal portions. In the case of the NKG2E transcript a cryptic splice site in the coding region of exon 6 is fused to an Alu type sequence several kb downstream of exon 6.[23] This results in the exchange of several C-ter-

minal amino acids of the CRD domain and a somewhat longer C-terminal part of the NKG2E protein. In the NKG2F transcript the insertion of two nucleotides in exon 4 results in a frame-shift leading to a premature stop codon.[23,39] Therefore the NKG2F mRNA encodes only part of the first CRD domain and an altered short C-terminus. It is not clear whether these variants are functionally relevant. It is possible that the C-terminal end is not important for the binding properties of the receptors and some minor changes, such as in NKG2E, can be tolerated and give functional proteins. This is less likely for molecules like NKG2F and LY49L, where CRD domains 2 and 3 are missing. However, in view of the fact that these mRNAs are produced in NK cells, these alterations presumably occurred relatively recently in evolution and these proteins could possess some functions, eg, mediated by their functional cytoplasmic signaling domain. A clarification will have to await functional analysis of these molecules.

An analysis of the sequence similarities between the various rodent and human lectin NK receptors showed that the human *LY49L* gene is of comparable similarity to the rodent *Ly49* genes as the human *NKG2* genes are to the rodent *NKG2* genes (data not shown). This is in line with the assumption that the *LY49* gene(s) have been functional over long periods of the evolution of humans. However, our attempts to detect products of additional *LY49* genes by screening under reduced stringency did not provide any further evidence of additional closely related human *LY49* genes expressed in NK cells. It is not clear at the moment whether this means that human lectin-type genes with the function of the murine *Ly49* genes have diverged so far in sequence that they cannot be detected by homology screening, even under reduced stringency. This is conceivable as they had to co-evolve with their potential ligands, human MHC class I molecules. Alternatively, the *Ly49* region in humans could have been deleted or replaced or never developed to the complexity seen in rodents. It is possible that the functions of the rodent *Ly49* genes are taken over in humans by the KIR receptor genes in the LRC complex. This is supported by the current lack of described functional KIR family genes in NK cells of rodents and that, based on the capacity of the human KIR receptors as well as the murine Ly49 receptors to bind a wide variety of classical MHC class I molecules, functional equivalence is possible. A definite answer to this question will have to await a closer analysis by gene mapping and sequencing of the human genomic region syntenic to the rodent Ly49-encoding region. In any case these questions promise to provide interesting clues to the evolution of the murine and human genome and will give an example of how coevolution between receptors and their ligands can progress.

Materials and methods

Libraries and primers

High-density colony filters and clones from a human PAC library[40] were obtained from the Resource Center and Primary Database (RZPD) of the German Human Genome Project. Clone pools for PCR screening and clones from the human CEPH or RPCI BAC library (http://www.cephb.fr/) were kindly provided by Ralf Sudbrak or ordered from the RZPD, respectively. For cDNA screening the following colony filters from the RZPD were used: human fetal thymus (HTE cDNA, ICRFp508, ~84,864 clones) human fetal liver (HFL cDNA, ICRFp512, ~96,768 clones), human fetal lung (HPO cDNA, ICRFp515, ~49,920 clones), human NK cells (HNK cDNA, CIMLp544, ~22,272 clones), a human NK cell line (NKL, ~34,000 clones) and human spleen (DKFZp597, ~27,648 clones).

Oligonucleotide primers specific for PAC ends and used for PCR were obtained by end-sequencing of the isolated PAC clones and are available from the author as supplementary material. Primers specific for the *LY49L* gene and the *LOX-1* gene were taken from the sequence established in this work (accession number BankIt308184 and AF213453) and the work of Yamanaka *et al*,[35] respectively. Probes for the CD94 and NKG2 genes have been described previously,[13,23] primers for STS markers were from the Human Genome Database.

Screening of human PAC, BAC, cosmid, cDNA and shotgun libraries with nonradioactive probes

Probes from the PAC or BAC DNA were generated either in a PCR using DIG-11-dUTP (STSs) or synthesized by *in vitro* transcription with SP6- or T7-RNA polymerase to produce riboprobes according to the manufacturer's instructions (Roche Molecular Systems, Mannheim, Germany). For hybridization laminated filters were prehybridized in a modified Church buffer (250 mM Na_2HPO_4 pH 7.2, 5% SDS, 1 mM EDTA) for 2 h at 65°C, probes were denatured at 100°C for 10 min and in the case of riboprobes pre-anealed with sonicated human placental DNA (Sigma, St Louis, MO, USA) at a final concentration of 2 mg/ml in 4× SSC, then diluted and sprayed onto each individual filter as described previously.[41] Filters were sealed in plastic bags, hybridized for 16 h at 65°C and washed consecutively for 30 min in 40 mM Na_2HPO_4 pH 7.2, 0,2% SDS at 20°C and then 65°C. Positives were detected by alkaline phosphatase linked anti-DIG antibodies and attophos as a substrate. An image was captured with a CCD camera and analyzed by the Xdigitise software (Steinfath *et al*, in preparation).

Establishing a contig of PAC and BAC clones

To establish a contig spanning the centromeric region of the NK gene complex, PAC clones that map to this region were isolated by library screening with PCR probes from the STS markers *D12S841*, *D12S852*, *D12S1123* and the *LY49L* gene. Primer sequences and PCR conditions were according to information provided in the Genome Database (http://www.hgmp.mrc.ac.uk/gdb/). The contig was then extended by screening libraries with end-probes obtained from available PACs, including those of the *CD94*/*NKG2* region,[13] using an *in vitro* transcription system. Alternatively, the inserts of clones were end-sequenced and oligonucleotides based on these sequences were used in PCR to obtain probes for further screening or to verify overlapping regions of the clones. The probes were hybridized onto the human PAC library using four-fold coverage in the first round and if no clones could be identified, 15-fold coverage in a second round. Positive clones were purchased from the RZPD and the markers were confirmed by PCR using non-purified whole colonies as a template. Additionally, clone pools and colony filters from the CEPH and RPCI-11 BAC

libraries were screened in a PCR with primers derived from clone end sequences.

Cloning of the *LY49L* genomic region
The PAC library with 15-fold coverage of the human genome was screened with a *LY49L* probe derived from the partial insert of EST AA464841 and the complete insert of a human *LY49L* cDNA clone (kindly provided by Eric Dissen/Sigbjörn Fossum, Dept of Anatomy, University of Oslo, Norway). All clones that showed a hybridization signal with the probe were obtained from the RZPD and re-analyzed in a PCR using the *LY49L* specific primer pairs Ly49F/Ly49R and hLy49–5′/hLy49Ex2-R which amplify exonic and intronic sequence of the 5′ or the 3′ region of the gene. Initially 17 positive PAC clones were identified.

Construction of a shotgun library
PAC DNA from clone RPCIp704I2437 was purified by the alkaline lysis procedure followed by cesium chloride density centrifugation. After purification 10 μg of DNA was sonicated, end-repaired and sized on a preparative agarose gel. Three fractions ranging from 1 to 2 kb were cut and blunt end-cloned into pUC18 vector. The DNA was transformed by electroporation into competent DH10B cells. The shotgun library was plated onto 2 × YT agarose plates containing 50 μg/ml Ampicillin. After incubation overnight blue/white selection was performed and 3000 positively selected clones were picked and arrayed into 384 well microtitre plates.

Preparation of PAC DNA, size determination by PFGE and restriction mapping
Bacteria of the strain DH10B harboring the PAC were grown in 2 × YT containing 30 μg/ml Kanamycin. Minipreparations were performed using guanidinium-thyocyanate.[42] For sizing 500 ng PAC DNA was digested with *NotI* (New England Biolabs, Schwalbach, Germany) for 2 h and then subjected to pulsed field gel electrophoresis (PFGE) on a CHEF Biorad apparatus for 20 h at 14°C using a switch time of 4–20 sec at 6V/cm.

Restriction mapping was performed by *EcoRI* or *EcoRI/HindIII* (New England Biolabs) digestion of PAC DNA and subsequent gel ectrophoresis on 0.8% agarose gels for 16 h. The gel was stained with Vistra Green (Amersham, Braunschweig, Germany). An image was captured with a FluorImager (Molecular Dynamics, Freiburg, Germany) and analyzed with the Image[43] and FPC[44] software packages.

Plating, gridding and screening of the shotgun library
The library was robotically spotted onto nylon membranes. Colonies were grown overnight, then lysed and the DNA fixed to the membrane by UV crosslinking. The membranes were then hybridized with DIG-labeled probes that were generated by PCR from the cDNA clone. For the initial screening the insert of a *LY49L* cDNA clone was amplified and labeled using the hLy49–5′/pL397 primer pair. To obtain complete coverage of the genomic region two additional rounds of shotgun library re-screening had to be performed using the Ly49UTR5/pL397 and Contig44F/Contig44R primer pairs to produce probes from the cDNA clone. Positive shotgun clones were sequenced.

Sequencing
For sequencing the shotgun clones were grown in microtitre plates. The insert was amplified by PCR from the clones with SP6 and T7 primers and directly sequenced on ABI machines. Sequences were assembled into a contig using the Staden-Package software.[45] Finishing was done by PAC sequencing with the gene specific primers I2437A3F/H2F1, I2437B2F1, I2437F1F1 and I2437H1R1.

Cloning and analysis of Ly49 cDNAs
To identify a completely spliced *LY49L* sequence or other *LY49* variants, several cDNA libraries derived from the following tissues were screened with *LY49L* DIG-dUTP-labeled PCR probe which was produced with the pL263/pU60 primer pair first under stringent conditions and then under non-stringent conditions at 40°C: human fetal thymus, human fetal liver, human fetal lung, human NK cells and a human NK cell line. Positive clones were purchased from the RZPD, tag sequenced and if confirmed positive, fully sequenced by primer walking.

Acknowledgements

We want to emphasize that CB and YS contributed equally to this work. We would like to thank Ralf Sudbrak for helpful discussions and providing the CEPH human BAC library (http://www.cephb.fr/), Karl Rak and Sabine Thamm for technical help and Heinz Himmelbauer for critically reading the manuscript. We acknowledge the Resource Center Primary Database of the German Human Genome Project in Berlin (http://www.rzpd.de) for providing filter spotted PAC, BAC and cDNA libraries. The nucleotide sequence data reported have been submitted to the EMBL/GenBank nucleotide sequence databases and have been assigned the accession numbers BankIt308184 and AF213453.

Nomenclature
According to the Human Gene Nomenclature Database (www.gene.ucl.ac.uk/nomenclature) the various loci described have obtained the following symbols: *OLR1* (*LOX-1*), *KLRA1* (*LY49L*), *KLRC1* (*NKG2A/B*), *KLRC2* (*NKG2C*), *KLRC3* (*NKG2E*), *KLRC4* (*NKG2F*), *KLRD1* (*CD94*).

References

1 Trinchieri G. Biology of natural killer cells. *Adv Immunol* 1989; **47**: 187–376.
2 Moretta A, Biassoni R, Bottino C *et al*. Major histocompatibility complex class I-specific receptors on human natural killer and T lymphocytes. *Immunol Rev* 1997; **155**: 105–117.
3 Hoglund P, Sundback J, Olsson Alheim MY *et al*. Host MHC class I gene control of NK-cell specificity in the mouse. *Immunol Rev* 1997; **155**: 11–28.
4 Lanier LL. NK cell receptors. *Annu Rev Immunol* 1998; **16**: 359–393.
5 Lopez-Botet M, Bellon T. Natural killer cell activation and inhibition by receptors for MHC class I. *Curr Opin Immunol* 1999; **11**: 301–307.
6 Bauer S, Groh V, Wu J *et al*. Activation of NK cells and T cells by NKG2D, a receptor for stress- inducible MICA. *Science* 1999; **285**: 727–729.
7 Wilson MJ, Torkar M, Trowsdale J. Genomic organization of a human killer cell inhibitory receptor gene. *Tissue Antigens* 1997; **49**: 574–579.

8 Wagtmann N, Rojo S, Eichler E, Mohrenweiser H, Long EO. A new human gene complex encoding the killer cell inhibitory receptors and related monocyte/macrophage receptors. *Curr Biol* 1997; **7**: 615–618.
9 Wang LL, Mehta IK, LeBlanc PA, Yokoyama WM. Mouse natural killer cells express gp49B1, a structural homologue of human killer inhibitory receptors. *J Immunol* 1997; **158**: 13–17.
10 Renedo M, Arce I, Rodriguez A *et al*. The human natural killer gene complex is located on chromosome 12p12-p13. *Immunogenetics* 1997; **46**: 307–311.
11 Yabe T, McSherry C, Bach FH *et al*. A multigene family on human chromosome 12 encodes natural killer-cell lectins. *Immunogenetics* 1993; **37**: 455–460.
12 Suto Y, Yabe T, Maenaka K *et al*. The human natural killer gene complex (NKC) is located on chromosome 12p13.1-p13.2. *Immunogenetics* 1997; **46**: 159–162.
13 Sobanov Y, Glienke J, Brostjan C *et al*. Linkage of the NKG2 and CD94 receptor genes to D12S77 in the human natural killer gene complex. *Immunogenetics* 1999; **49**: 99–105.
14 Brown MG, Fulmek S, Matsumoto K *et al*. A 2-Mb YAC contig and physical map of the natural killer gene complex on mouse chromosome 6. *Genomics* 1997; **42**: 16–25.
15 Brown MG, Scalzo AA, Matsumoto K, Yokoyama WM. The natural killer gene complex: a genetic basis for understanding natural killer cell function and innate immunity. *Immunol Rev* 1997; **155**: 53–65.
16 Dissen E, Ryan JC, Seaman WE, Fossum S. An autosomal dominant locus, Nka, mapping to the Ly-49 region of a rat natural killer (NK) gene complex, controls NK cell lysis of allogeneic lymphocytes. *J Exp Med* 1996; **183**: 2197–2207.
17 Vance RE, Tanamachi DM, Hanke T, Raulet DH. Cloning of a mouse homolog of CD94 extends the family of C-type lectins on murine natural killer cells. *Eur J Immunol* 1997; **27**: 3236–3241.
18 Vance RE, Kraft JR, Altman JD, Jensen PE, Raulet DH. Mouse CD94/NKG2A is a natural killer cell receptor for the nonclassical major histocompatibility complex (MHC) class I molecule Qa-1(b). *J Exp Med* 1998; **188**: 1841–1848.
19 Ryan JC, Seaman WE. Divergent functions of lectin-like receptors on NK cells. *Immunol Rev* 1997; **155**: 79–89.
20 Dissen E, Berg SF, Westgaard IH, Fossum S. Molecular characterization of a gene in the rat homologous to human CD94. *Eur J Immunol* 1997; **27**: 2080–2086.
21 Houchins JP, Yabe T, McSherry C, Bach FH. DNA sequence analysis of NKG2, a family of related cDNA clones encoding type II integral membrane proteins on human natural killer cells. *J Exp Med* 1991; **173**: 1017–1020.
22 Hofer E, Duchler M, Fuad SA *et al*. Candidate natural killer cell receptors. *Immunol Today* 1992; **13**: 429–430.
23 Glienke J, Sobanov Y, Brostjan C *et al*. The genomic organization of NKG2 C, E, F and D receptor genes in the human natural killer gene complex. *Immunogenetics* 1998; **48**: 163–173.
24 Chang C, Rodríguez A, Carretero M *et al*. Molecular characterization of human CD94: a type II membrane glycoprotein related to the C-type lectin superfamily. *Eur J Immunol* 1995; **25**: 2433–2437.
25 Westgaard IH, Berg SF, Orstavik S, Fossum S, Dissen E. Identification of a human member of the Ly-49 multigene family. *Eur J Immunol* 1998; **28**: 1839–1846.
26 Hamann J, Fiebig H, Strauss M. Expression cloning of the early activation antigen CD69, a type II integral membrane protein with a C-type lectin domain. *J Immunol* 1993; **150**: 4920–4927.
27 Hamann J, Montgomery KT, Lau S, Kucherlapati R, van Lier RA. AICL: a new activation-induced antigen encoded by the human NK gene complex. *Immunogenetics* 1997; **45**: 295–300.
28 Bakker AB, Baker E, Sutherland GR, Phillips JH, Lanier LL. Myeloid DAP12-associating lectin (MDL)-1 is a cell surface receptor involved in the activation of myeloid cells. *Proc Natl Acad Sci USA* 1999; **96**: 9792–9796.
29 Sawamura T, Kume N, Aoyama T *et al*. An endothelial receptor for oxidized low-density lipoprotein. *Nature* 1997; **386**: 73–77.
30 Barten R, Trowsdale J. The human Ly-49L gene. *Immunogenetics* 1999; **49**: 731–734.
31 Senapathy P, Shapiro MB, Harris NL. Splice junctions, branch point sites, and exons: sequence statistics, identification, and applications to genome project. *Methods Enzymol* 1990; **183**: 252–278.
32 Shapiro MB, Senapathy P. RNA splice junctions of different classes of eukaryotes: sequence statistics and functional implications in gene expression. *Nucleic Acids Res* 1987; **15**: 7155–7174.
33 Wingender E, Dietze P, Karas H, Knüppel R. TRANSFAC: a database on transcription factors and their DNA binding sites. *Nucleic Acids Res* 1996; **24**: 238–241.
34 Boles KS, Barten R, Kumaresan PR, Trowsdale J, Mathew PA. Cloning of a new lectin-like receptor expressed on human NK cells. *Immunogenetics* 1999; **50**: 1–7.
35 Yamanaka S, Zhang XY, Miura K, Kim S, Iwao H. The human gene encoding the lectin-type oxidized LDL receptor (OLR1) is a novel member of the natural killer gene complex with a unique expression profile. *Genomics* 1998; **54**: 191–199.
36 Plougastel B, Trowsdale J. Sequence analysis of a 62-kb region overlapping the human KLRC cluster of genes. *Genomics* 1998; **49**: 193–199.
37 Lohwasser S, Hande P, Mager DL, Takei F. Cloning of murine NKG2A, B and C: second family of C-type lectin receptors on murine NK cells. *Eur J Immunol* 1999; **29**: 755–761.
38 Smith HR, Karlhofer FM, Yokoyama WM. Ly-49 multigene family expressed by IL-2-activated NK cells. *J Immunol* 1994; **153**: 1068–1079.
39 Plougastel B, Trowsdale J. Cloning of NKG2-F, a new member of the NKG2 family of human natural killer cell receptor genes. *Eur J Immunol* 1997; **27**: 2835–2839.
40 Ioannou PA, Amemiya CT, Garnes J *et al*. A new bacteriophage P1-derived vector for the propagation of large human DNA fragments. *Nat Genet* 1994; **6**: 84–89.
41 Bancroft DR, O'Brien JK, Guerasimova A, Lehrach H. Simplified handling of high-density genetic filters using rigid plastic laminates. *Nucleic Acids Res* 1997; **25**: 4160–4161.
42 Nelson JE, Krawetz SA. Purification of cloned and genomic DNA by guanidine thiocyanate/isobutyl alcohol fractionation. *Anal Biochem* 1992; **207**: 197–201.
43 Sulston J, Mallett F, Durbin R, Horsnell T. Image analysis of restriction enzyme fingerprint autoradiograms. Comput *Appl Biosci* 1989; **5**: 101–106.
44 Soderlund C, Longden I, Mott R. FPC: a system for building contigs from restriction fingerprinted clones. *Comput Appl Biosci* 1997; **13**: 523–535.
45 Bonfield JK, Smith KF, Staden R. A new DNA sequence assembly program. *Nucleic Acids Res* 1995; **23**: 4992–4999.