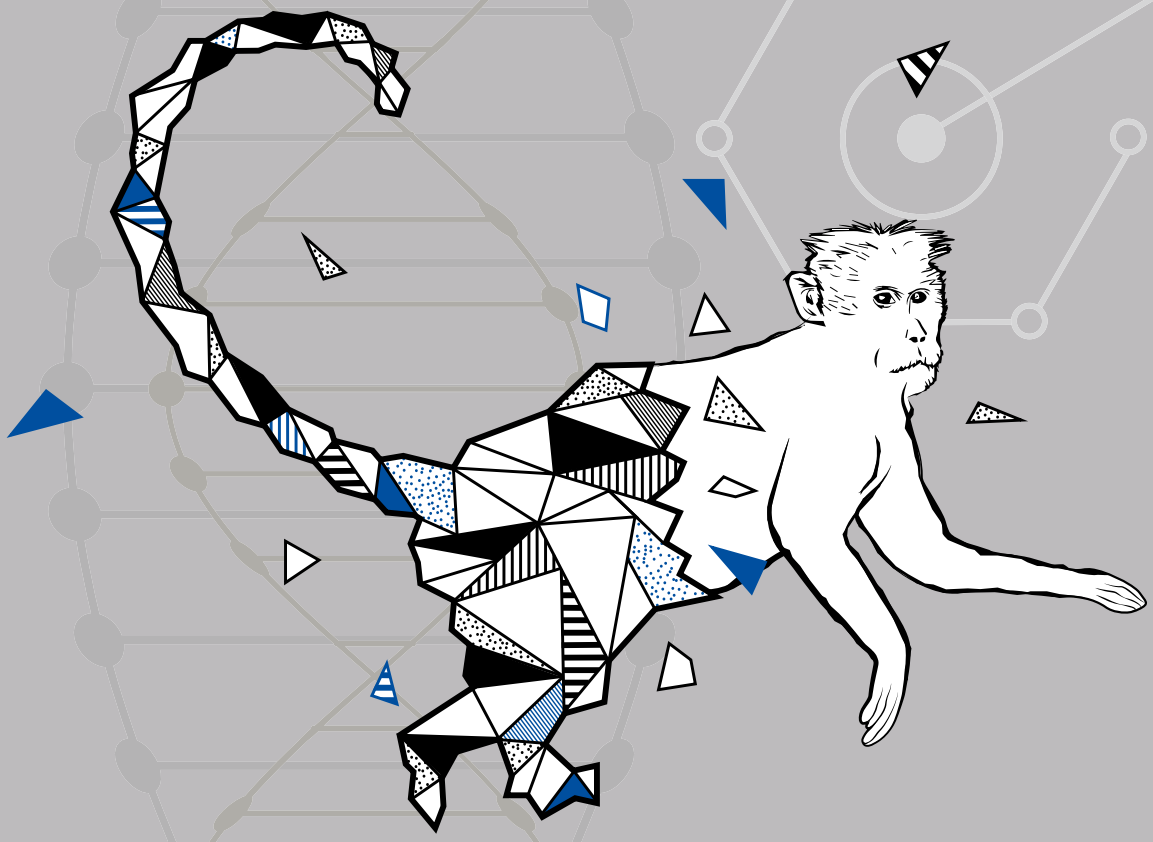
KIR gene complexity in primates

A genetic arsenal equipped to arm and control a killer



Jesse Bruijnesteijn

KIR gene complexity in primates: a genetic arsenal equipped to arm and control a killer

Jesse Bruijnesteijn

KIR gene complexity in primates: a genetic arsenal equipped to arm and control a killer

KIR gen complexiteit in primaten: een genetisch arsenaal om NK cellen te activeren en controleren

(met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van de graad van doctor aan de
Universiteit Utrecht
op gezag van de
rector magnificus, prof.dr. H.R.B.M. Kummeling,
ingevolge het besluit van het college voor promoties
in het openbaar te verdedigen op

woensdag 3 november 2021 des ochtends te 10.15 uur

door

Jesse Bruijnesteijn

geboren op 28 februari 1991 te Hoorn

Promotor
Prof.dr. R. E. Bontrop

Copromotor
Dr. N. G. de Groot

Assessment Committee
Prof. P. Parham
Prof. L. Walter
Prof. J. Kaufman
Prof. E. P. G. J. Cuppen
Prof. R. J. de Boer

"Encumbered forever by desire and ambition,
There's a hunger still unsatisfied."

David Gilmour (Pink Floyd – High hopes)

The research described in this thesis was performed at the Dept. of Comparative Genetics and Refinement at the Biomedical Primate Research Centre, Rijswijk, the Netherlands

Financial support was provided by the Biomedical Primate Research Centre.

Illustrations: F. van Hassel
Lay-out: F. van Hassel
Cover design: Demi Bruijnesteijn
Printed by: Ridderprint BV
ISBN: 978-94-6416-837-2

Table of Contents

Chapter 1

General Introduction

Jesse Bruijnesteijn

General introduction

The killer-cell immunoglobulin-like receptors (KIR) are type I transmembrane glycoproteins that are expressed on NK cells and subsets of T cells. The KIR receptors can be categorized into activating and inhibitory forms that recognize different epitopes on MHC class I molecules. Through these interactions, the development and functional activity of NK cells are regulated. The functional capacity of NK cells is first acquired in a process termed NK cell education, which involves the recognition of self-MHC molecules by inhibitory KIR. The activity of mature NK cells is then modulated by the balanced integration of simultaneous activating and inhibitory signals. Cells with aberrant MHC class I expression, which might occur during infection or tumor formation, are recognized and subsequently lysed by the NK cells. Non-cytotoxic NK cells mainly secrete cytokines upon activation, which modulates consecutive immune responses or regulates vascular remodeling during pregnancy.

The human KIR family comprises 17 members, which vary in their structure, function, ligand specificity, expression status, and cellular localization. The genes that encode the receptors are located on chromosome 19 and are arranged in a head-to-tail manner. Genomic and transcriptomic characterization studies illustrated allelic polymorphism and gene copy number variation, which is further complicated by recombination events, differential haplotype configurations and alternative splicing. The characterization of the *KIR* gene system contributes to a better understanding of the abundant disease association studies, which demonstrated the importance of KIR in health and disease.

The characterization of *KIR* genes in other primate species indicates a rapid and species-specific co-evolution with their ligand-encoding *MHC class I* genes. The KIR receptors in different primate species share a similar structure and function, but might genetically vary in their lineage expansion and haplotype configurations. An initial characterization of the *KIR* gene cluster has been performed in rhesus macaques, a species that is commonly used as model in biomedical research to test new vaccines and drugs to treat human and animal diseases. In line with the extensively expanded *MHC-A* and-*B* genes in macaques, the macaque *KIR* genes illustrate extensive diversity. Taking into consideration the many health and disease associations with KIR (and their MHC ligands) in humans, and the relevance of macaques as preclinical model, this manuscript is focused on understanding the *KIR* gene system in different macaque species.

Primate taxonomy

The definition of species, or taxonomy, is the complex study to distinguish groups of organisms, and affects the fields of genetics, population biology, ecology, and ethology [1]. At present, over 300 primate species (*Primata*) are recognized, which share anatomic, functional, and behavioral features that reflect common ancestry, and are

generally divided into wet- and dry-nosed primates (**Figure 1**). The latter suborder includes the tarsier, monkey, and ape clades, which are further distinguished by two parvorders that separates the New World monkeys from the Old World monkeys, apes and humans. The Old World monkey superfamily (*Cercopithecoidea*), including the macaques, share a common ancestor with the humans and apes (*Hominoidea*) approximately 25-33 million years ago. The macaque genera (*Macaca*) comprise over 20 species, which evolutionary diverged in the past 1-3 million years (**Figure 1**). Even more, populations within macaque species, such as the Indian, Burmese and Chinese rhesus macaques, and the insular cynomolgus macaques, emerged by natural barriers and drove intra-species evolution.

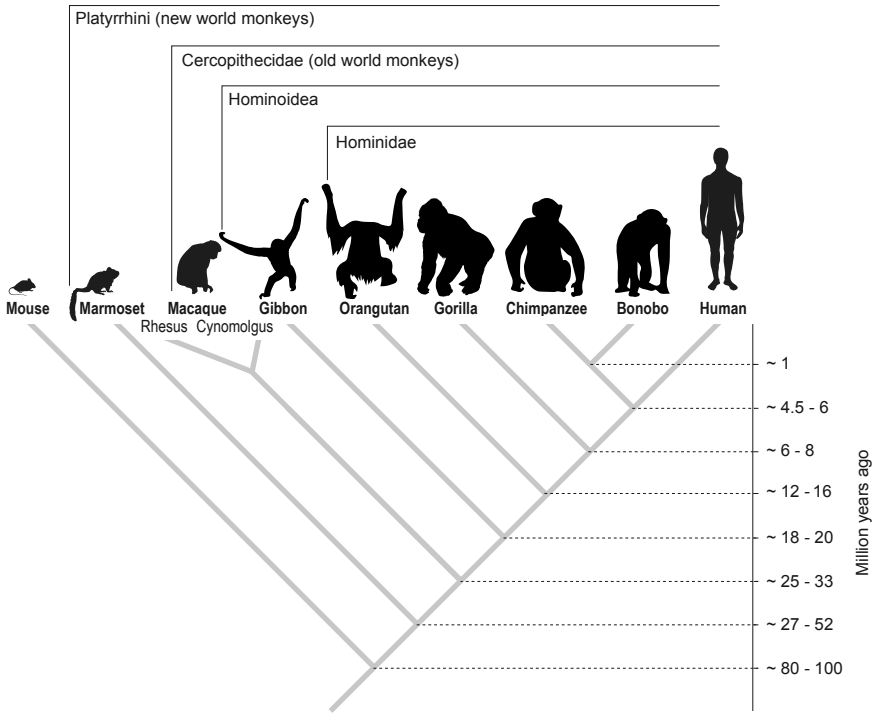


Figure 1. Evolutionary relationships and taxonomy of primates. The pedigree illustrates the estimated genetic distance to a common ancestor with humans for the different non-human primate species. Mouse has been added to display the divergence of primates from other mammals. On top, the orders and families are defined for the different primate species.

The bridge between innate and adaptive immunity

During the evolution of species, a protective immune system was acquired that fights potentially lethal pathogens. The adaptive arm of this defense mechanism is able to specifically recognize and eliminate viruses, bacteria, fungi, and parasites, and owns a memory function to launch a more rapid response upon recurrent infection. These adaptive responses are mainly mediated by B- and T-lymphocytes that rely on a highly diverse repertoire of receptors, which are generated by somatic chromosomal rearrangements [2]. The B-lymphocytes mainly harbor a humoral function, by secreting antibodies and cytokines, whereas different subsets of T-lymphocytes perform regulatory and cytotoxic tasks. The evolutionary older innate arm of the immune system plays a crucial role in the initiation and activation of the adaptive immune response, and also participates in the killing and removal of pathogens by a-specific priming through interactions of generic receptors. The effectors of the innate immunity are a diverse group of lymphoid cells, which are grouped based on their cytokine production profiles and cytotoxicity [3].

Among the innate lymphoid cells are the Natural Killer (NK) cells, which were discovered in mice and were initially described for their capability to eliminate cancer cells [4-6]. This cytotoxic role is analogous to that of the $CD8^+$ T-lymphocyte subset, although the response in NK cells is more rapid and in absence of antigen-specific priming. The lysis of target cells by activated NK cells is mediated by the engagement of death receptors, such as Fas/CD95, or by the release of small granules, which result in the secretion of perforins and granzymes. In addition to the killing of tumor and pathogen-infected cells, NK cells have immunoregulatory functions. Upon activation, they facilitate the fast production and secretion of numerous cytokines and chemokines, such as interferon (IFN)-γ, and thereby modulate both the innate and adaptive immune responses [7-9]. The NK cell response can be enhanced in the presence of different cytokines, like IL-12, IL-15, and IL-18, which are secreted by other innate immune cells. Recently, expansion of long-lasting subsets of NK cells are described in response to infectious pathogens, such as the mouse and human cytomegalovirus, which indicate memory-like characteristics for NK cells [10, 11]. These regulatory and adaptive features, although initiated by a-specific pathogen priming, contribute to adaptive immune responses, and suggests NK cells to bridge the innate and adaptive immune system.

NK cell development and subset differentiation

In additional to shared functional features, both the NK cells and the B- and T-lymphocytes primarily originate from the common lymphoid progenitor (CLP) in the bone marrow (**Figure 2**). Whereas T-lymphocytes migrate and continue to develop in the thymus, the NK cells expand through the lineage-restricted NK cell progenitor (NKP), from which further differentiation is dependent on CD127 expression. Influenced

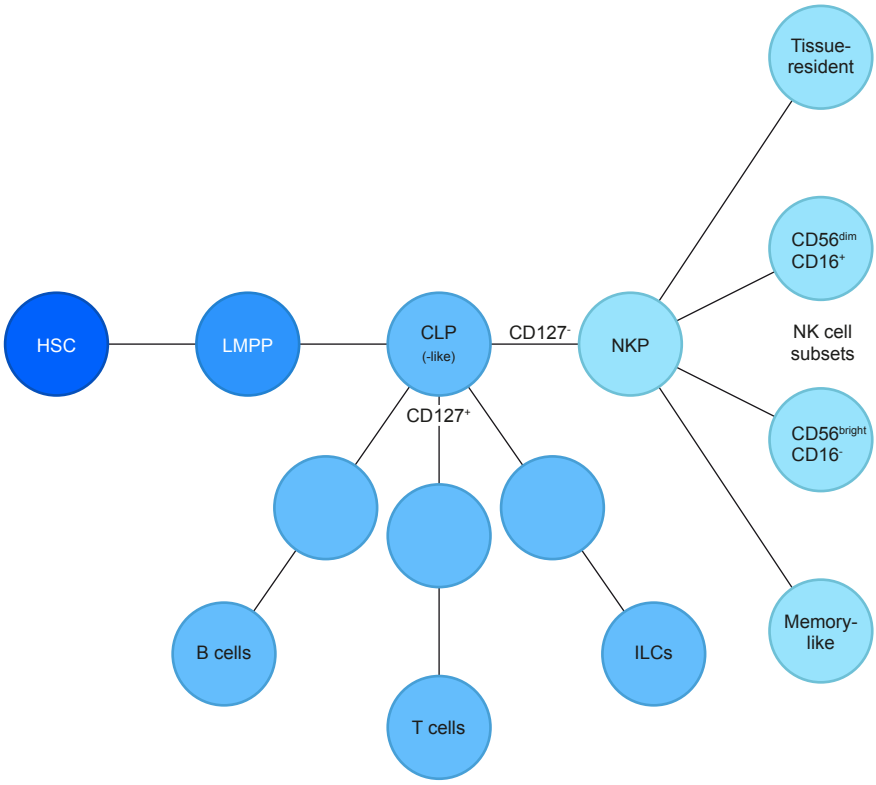


Figure 2. Development pathway of lymphocytes and NK cell subsets. Activation of hematopoietic stem cells (HSC) leads to the development of lymphoid-primed multipotent precursors (LMPP), which then differentiate into common lymphoid progenitor (CLP) cells in the bone marrow. Subsequent commitment that depends on CD127 expression generates either B- and T-cell populations and ILCs through different development stages, or a NK precursor stage (NKP). Eventually, these NKPs then differentiate into subsets of NK cells.

Table I. Four human NK cell subsets that are distinguished by specifc markers.

Subset	Markers	Main function
$CD56^{dim}CD16^{+}$	$CD56^{+/-}$, $CD16^{+}$, KIR^{+}, $CD117^{-}$, $CD127^{-}$	Highly cytotoxic NK cells, death receptors and cytolytic granules
$CD56^{bright}CD16^{-/+}$	$CD56^{+}$, $CD16^{-}$, KIR^{-}, $CD117^{+/-}$, $CD127^{+/-}$	Regulatory NK cells, secrete cytokines and chemokines
Tissue-resident $CD56^{bright}$	$CD56^{+}$, KIR^{low}, $CXCR6^{+}$, $CCR5^{+}$, $CD49a^{+/-}$, $CD69^{+}$	Still elusive, but might shape the local microenvoriment
Adaptive NK cells	$CD56^{+/-}$, KIR^{+}, $CD2^{+}$, $CD57^{+}$, $NKG2C^{+}$	Memory-like expansion of cytotoxic NK cells

by multiple transcription factors, such as GATA3, the $CD127^+$ NKPs differentiate into diverse ILC populations, whereas the $CD127^-$ NKPs develop through different stages to the mature peripheral blood NK cells and NK cell subsets in tissues (**Figure 2**). The NK cells form a heterogenous cell population that represent 10-15% of the circulating lymphoid cells. Based on surface phenotype and cytokine profiles, NK cell subsets are defined, which – according to the prevailing linear model – represent distinct development stages. The most widely used phenotypic markers are $CD3^-CD56^+$ and $CD3^-CD16^+$ expression, which together define two of the four major subsets in humans: $CD56^{dim}CD16^+$ and $CD56^{bright}CD16^{-/+}$ (**Table I**). The $CD56^{dim}CD16^+$ subset, which represents 90% of the circulating NK cells, exerts higher natural cytotoxicity and expresses increased levels of Ig-like receptors compared to the $CD56^{bright}CD16^{-/+}$ subset. Although circulating $CD56^{bright}CD16^{-/+}$ NK cells can mediate cytotoxicity, the main function is the production and secretion of cytokines, such as IFN-γ, TNF-α, and IL-10 [12, 13]. The linear development model states that the $CD56^{bright}CD16^{-/+}$ NK cells are precursors of the terminally differentiated $CD56^{dim}CD16^+$ subset [14-18]. Although this naturally cytotoxic subset predominates in blood, the majority of NK cells in the body are represented by the $CD56^{bright}CD16^-$ cells that reside and differentiate in second lymphoid organs, such as the thymus, lymph nodes, liver, and uterus, but also in inflamed and tumor tissues [19-24]. This tissue-resident $CD56^{bright}$ NK cell subset mediates specialized and tissue-specific functions, and can be distinguished from circulating NK cells by the expression of CD69, chemokine receptors (CXCR6 and CCR5), and adhesion molecules (CD49a) [25-28]. A fourth major NK cell subset can be distinguished subsequent to pathogenic encounters, such as cytomegalovirus (CMV) infection [29, 30], and comprise NK cells with memory-like or adaptive features, including robust responses upon recurrent infection and clonal expansion. These adaptive $CD56^{dim}CD16^+$ NK cells express multiple activating receptors and contain elevated levels of perforin and granzymes, which indicates high cytolytic capacity. The adaptive NK cells, which are thought to reside in the liver, poorly respond to activated immune cells and several activating cytokines, such as IL-12. Epigenetic modifications might explain the altered functional profile of this NK cell subset [31].

The broad spectrum of NK cell subsets in humans is currently thought to arise from different development stages, which hypothesize linear differentiation from a common progenitor. However, *in vitro* studies demonstrated NK cell differentiation from common myeloid progenitors (CMP) in the presence of NK cell-supporting cytokines, and, although artificial culture systems are used, it does speculate a branched development [32, 33]. This is further supported by macaque studies, in which the $CD56^{bright}$ and $CD56^{dim}$ NK cell subsets were acquired from different lineage progenitors [34, 35]. Although there is still debate on the contrasting development models, the extensive subset diversity is well illustrated and hint for a broad phenotypic and functional spectrum of NK cells.

NK cell education and 'missing-self' activation

To recognize pathogenic or infected cells and to maintain self-tolerance, NK cells are educated through a process of receptor engagements. These 'educative' interactions involve a variety of inhibitory surface receptors on the NK cells and major histocompatibility complex (MHC) class I molecules, also known in humans as human leukocyte antigens (HLA), which are ubiquitously expressed by nucleated cells and present cytosolic peptides. In a fluid and ongoing process, NK cells that express at least one inhibitory receptor that recognizes a host MHC class I molecule acquire effector functions; only these NK cells are licensed to kill [36-38]. The importance of NK cell education is demonstrated in *B2m*$^{-/-}$ mice, which lack expression of MHC class I molecules. The NK cells from these mice showed significantly decreased killing of MHC class I deficient target cells and abrogated cytokine production compared to NK cells from wild type mice [39, 40]. Similarly, decreased effector functions were described for NK cells that lack MHC-specific inhibitory receptors [41]. In addition, the number of inhibitory receptors that recognize host MHC molecules positively correlates with the NK cell responsiveness [41, 42].

During viral infection and tumor formation, stressed cells are recognized through the presentation of pathogenic antigens by MHC class I molecules, which primes cytotoxic T-lymphocytes. This defense mechanism can be evaded by the chronic downregulation of MHC molecules from the surface of aberrant cells, which is a common evolutionary adaption for viruses and tumors [43-45]. Educated NK cells, however, detect the absence of their deterrents, and become activated to lyse the target cell. This process of NK cell activation, which has been proposed in 1981, is termed the 'missing-self' hypothesis and is thought to be the main mechanism of NK cell activation (**Figure 3**) [46].

The pool of NK cells in an individual is highly diverse, which is not only reflected by different subsets, but also on single-cell level. NK cells show a variegated expression of different activating and inhibitory receptors, which are encoded by genes that segregate independently from the MHC-encoding genes. As a result, a significant NK cell population is unable to interact with self-MHC molecules, and, therefore, do not acquire effector functions and remain unlicensed. These uneducated cells persist in the circulation, but are hyporesponsive.

Two models were initially proposed for the differential responsiveness states of NK cells. On one hand, the 'arming' model states that the interaction of inhibitory receptors with self-MHC molecules is required for functional maturation. On the other hand, the 'disarming' model suggests that all NK cells are initially responsive, but continuous activating engagements, in the absence of inhibitory interactions with the host MHC molecules, desensitize the NK cells and result in hyporesponsiveness. A third and more supported model is the Rheostat model that describes the education procedure as a continuous process, in which the NK cell responsiveness is dependent on the quantity of environmental factors, such as the presence of absence of cytokines and MHC class I ligands. This last model would explain the persistence of unlicensed

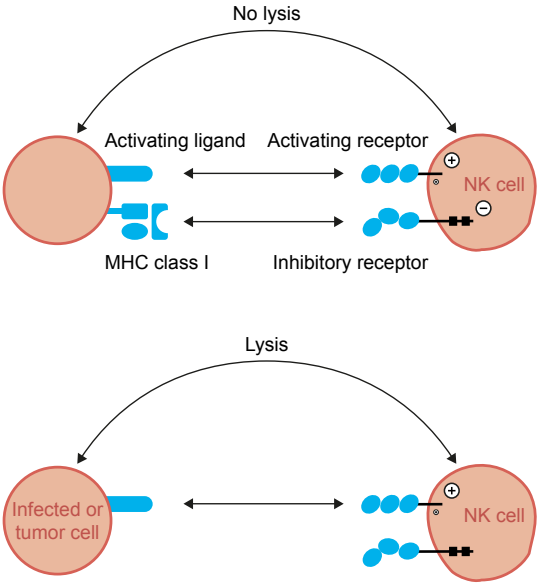


Figure 3. A schematic representation of the missing-self hypothesis. A balance of activating and inhibitory signals maintain tolerance in the upper panel through interactions of regulatory receptors expressed by NK cells and their ligands, including MHC class I. Upon infection or tumor cell formation, target cells might downregulate their MHC class I expression and the total balance of signals the NK cell receives might shift towards a more activating potential. This triggers NK cell activation and lysis of the target cell.

cells, which are thought to have effector function under specific conditions, such as viral infection and tumor formation [47, 48], and can be 're-educated' upon environmental changes [37, 49].

NK cell receptors

A broad spectrum of activating and inhibitory receptors is expressed on the surface of the NK cell and cooperate with each other to regulate the cytotoxicity. The transmembrane and cytoplasmic domains determine the signaling potential of these receptors. In general, inhibitory receptors contain one or multiple immunoreceptor tyrosine-based inhibitory motifs (ITIM) in their cytoplasmic tail, whereas activating receptors possess a positively charged residue in their transmembrane region that can interact adaptor molecules, such as DAP10, DAP12, and FcR-γ. These adaptors contain immunoreceptor tyrosine-based activation motifs (ITAM), which transduce activating signals.

The majority of activating and inhibitory receptors belong to three receptor families that contribute to NK cell regulation. The Natural Cytotoxicity Receptor (NCR) family comprise only activating receptors, namely NKp46, NKp44, and NKp30, and are known to interact with several non-MHC molecules, such as heparan sulfate glycosaminoglycans [50]. NKp46 and NKp30 are constitutively expressed on NK cells, but are also identified on subsets of ILCs and T cells. Another family involved in NK cell regulation is the calcium-dependent lectin-like receptor family that includes CD94-NKG2-A/C/E/F/H heterodimers and NKG2D homodimers, which are inhibitory and activating receptors, respectively. These inhibitory receptors interact with several non-classical MHC molecules, whereas multiple other ligands are documented for NKG2D, including the major histocompatibility complex class I chain-related molecules (MIC-A/B) [51, 52]. The third receptor family comprise the killer cell immunoglobulin-like receptors (KIR), which form a highly diverse and expanded group of inhibitory and activating receptors expressed on NK cells and subsets of T cells.

KIR discovery, structure and nomenclature

After the 'missing-self' hypothesis was proposed in 1981, researchers started to work on the question how cells with an aberrant MHC expression were recognized and killed by the NK cells. Transfection studies demonstrated that the loss of HLA-A,-B, and -C expression resulted in susceptibility to NK cell-mediated lysis [53, 54]. This susceptibility was reversed by the transfection of certain *HLA* class I genes, indicating a molecular interaction of HLA molecules with proteins present on NK cells. Subsequently, multiple immunoglobulin-superfamily receptors were cloned to demonstrate putative interactions, including the receptors now known as KIR [55-57]. However, before standard nomenclature was introduced in 1996 [58], many different names were given to these receptors, including p50, p58 and p70, which referred to their molecular weight determined by immunoprecipitation, and the NK-associated transcripts (NKat). Even more, the acronym KIR first referred to killer inhibitor receptors, and was later changed to killer cell immunoglobulin-like receptors when the activating family members were discovered [59, 60].

The KIR family comprises receptors that consist of up to three extracellular domains, followed by a stem region, a transmembrane region, and a cytoplasmic tail (**Figure 4**). The nomenclature of the KIR receptors helps to distinguish differential structures. The first digit following the KIR abbreviation indicates the number of extracellular domains (denoted as 'D'). In humans, KIR2D and KIR3D structures are reported, whereas in several Old World monkeys KIR1D structures are documented as well. The length of the cytoplasmic can either be long or short, denoted with an 'L' or 'S', and characterizes inhibitory or activating KIR, respectively, whereas a 'P' denotes a pseudogene. A sequential two-digit numbering distinguish the different *KIR* genes. Non-synonymous *KIR* alleles are distinguished by three-digit numbers that are separated from the gene digits by an asterisk, whereas synonymous polymorphisms in the coding sequence of a

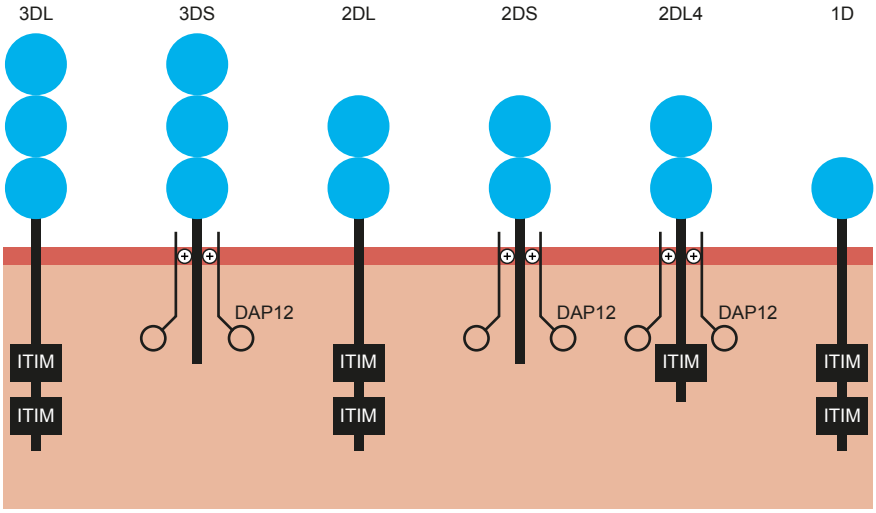


Figure 4. A schematic overview of the KIR receptor structure and nomenclature. KIR receptors display differential structures, with 2 or 3 extracellular domains, to which is referred to with KIR2D or KIR3D in nomenclature, respectively. In some primate species, receptors with a single extracellular domain are documented (KIR1D). A long (L) cytoplasmic tail can contain two immunoreceptor tyrosine-based inhibitory motifs (ITIM) and represent inhibitory receptors (e.g. KIR3DL). Short (S) cytoplasmic tails have a positive residue present in their transmembrane region (e.g. KIR3DS), which can interact with adaptor molecules, such as DAP12, that contain immunoreceptor tyrosine-based activation motifs (ITAM). An exception is KIR2DL4, which has both activating and inhibitory signaling potential.

KIR gene are distinguished by a second set of two digits (e.g., *KIR3DL1*001:02*). Optional suffixes can be provided to denote the expression and recombination status of a receptor [61, 62]. These nomenclature guidelines apply to the *KIR* gene system in all primates, with a few species-specific exceptions.

The genetic cluster

The genes encoding the human KIR receptors are located on chromosome 19q13.4 within the leukocyte receptor complex (LRC) and span a 100-200 kb region. The length of a single *KIR* gene is approximately 10-14 kb and the different genes are separated by intergenic regions of about 2-3 kb. The receptor is encoded by up to 9 exons, of which the first 2 encode the leader peptide, and exons 3-5 encode the extracellular domains (D0, D1 and D2). The stem region is encoded by exon 6, whereas exons 7-9 encode the transmembrane region and the cytoplasmic tail.

In addition to the *KIR* cluster, the LRC harbors more genes belonging to the immunoglobulin-like superfamily. For example, at the centromeric side the *KIR* gene

cluster is flanked by genes that encode the leukocyte immunoglobulin-like receptor (LILR) family, which are predominantly expressed in different lymphoid and myeloid cell types. The LILR family seems to be involved in the regulation of a diverse set of immune cells through interactions with MHC and non-MHC ligands [63]. Other structurally related members located in the LRC are the leukocyte-associated immunoglobulin-like receptors (LAIR), the sialic acid-binding immunoglobulin-type lectins (SIGLEC), the NCR1 (also referred to as NKp46), and the Fc-α receptor (FCAR), the latter of which is flanking the *KIR* genes at the telomeric side [64, 65]. In the extended LRC region, genes are identified that encode the ITAM-containing adapter molecules, which interact with activating immunoglobulin-like receptors, such as DAP10 and DAP12.

The *KIR* gene region is subjected to expansion and contraction, which results in a variable number of genes present within the cluster. The *KIR* genes display a high degree of similarity and are tandemly arranged in a head-to-tail manner. In primates, however, one conserved *KIR* gene is located outside the *KIR* cluster, namely *KIR3DX1*, which is identified within the *LILR* region. Phylogenetic analysis illustrates that *KIR3DX1* belongs to a different ancient lineage then the other *KIR* genes, and share a common progenitor 135.5 ± 10.5 million years ago [66]. Since then, the *KIR3DL/S* genes expanded and diverged by recombination, deletion and inclusion events, whereas the *KIR3DX1* remained a conserved pseudogene. In contrast, the KIR3DX lineage is expanded in cattle [66, 67]. The expansion of the *KIR3DL/S* genes resulted in four defined phylogenetic lineages that distinguish the receptors in their structure and MHC class I specificity. This lineage diversification probably predates the primate speciation, indicating that the *KIR* genes in different primate species represent the same lineages. However, after speciation, the *KIR* genes have diverged, which resulted in separately clustered and species-specific genes.

The human KIR repertoire and haplotypes

In humans, 17 *KIR* genes are documented, of which two are identified as pseudogenes, namely *KIR2DP1* and *KIR3DP1*. The expressed *KIR* genes are divided into four different lineages; lineage I includes *KIR2DL4* and *KIR2DL5(A/B)*, linage II includes *KIR3DL1*, *KIR3DL2* and *KIR3DS1*, lineage III includes *KIR2DL1*, *KIR2DL2*, *KIR2DL3*, *KIR2DS1*, *KIR2DS2*, *KIR2DS3*, *KIR2DS4*, *KIR2DS5*, and the pseudogenes, and lineage V includes *KIR3DL3*.

Both lineages I and III include receptors with two extracellular domains, but these are differently encoded. The main difference is within the extracellular domains: *KIR2DL4* and *KIR2DL5* (lineage I) contain the D0 and D2 domains, whereas the lineage III *KIR* genes express the D0 and D1 domains [68]. The KIR2DL4 further differs from the remaining receptors, as the long cytoplasmic tail only contains a single ITIM, in contrast to two ITIMs in the tail of all other inhibitory KIR. In addition, KIR2DL4 possess a positively charged residue in the transmembrane region, like in all activating KIR, which

can interact with ITAM-containing adaptor molecules. This provides KIR2DL4 with both inhibitory and activating signaling potentials [68-70].

Another atypical receptor is KIR3DL3, which is likely to encode an inhibitory receptor with 3 Ig-like domains, but lack the stem region (exon 6) and contain only a single ITIM in the cytoplasmic tail [71]. This protein is only expressed at low levels, but can be upregulated by treatment with methyltransferase inhibitors, which indicates that the promotor activity is regulated by methylation [72, 73]. A recent study also demonstrated regulation of KIR3DL3 expression levels via different miRNAs [74]. The receptor seems to have emerged from an ancient recombination event, as the domain-encoding exons cluster into lineage V, whereas the cytoplasmic tail is more similar to that observed in lineage III KIR. This structure is conserved in primates, including the Old World monkeys. In addition, KIR3DL3 is identified in all individuals from different geographically and ancestrally human populations that were studied for their KIR repertoire [75], which indicates an essential function for this receptor. The function and ligands of KIR3DL3 are, however, not discovered yet.

The human KIR haplotypes can be categorized into two groups, termed A and B haplotypes. Four framework genes form the basis of most haplotype configurations and include *KIR3DL3*, *KIR3DL2*, *KIR3DP1*, and *KIR2DL4*. A recombination hotspot in the center of the haplotype distinguishes a centromeric and telomeric region. In humans, the *KIR* genes expanded in both of these haplotype regions. The two haplotype groups have different *KIR* gene content: group A haplotypes contain a fixed set of 7 *KIR* genes, of which only a single gene encodes an activating receptor (*KIR2DS4*), whereas group B haplotypes contain a variable number of genes, including multiple activating receptors. Haplotype configurations that contain both group A and B haplotype content can result from chromosomal recombination events. The two standard A and B configurations are documented in all populations, but with different relative frequencies that correlate with the frequencies of the KIR ligands [76].

KIR ligands

As mentioned, interactions of NK cell receptors with proteins on target cells are involved in NK cell education, pathogen recognition and subsequent immune response initiation. The KIR receptors play a central role in these processes, as the individual KIR receptors recognize distinct MHC/HLA allotypes. In humans, the *HLA* genes are located on chromosome 6 and are classified as class I, II, and III. The *HLA class II* genes encode molecules that are present on specialized immune cells, such as dendritic cells and present extracellular proteins. The structure and function of HLA class III molecules are less defined, but involve cell signaling. The HLA class I molecules, which include ligands for KIR, are encoded by six genes: *HLA-A*, *-B*, *-C*, *-E*, *-F*, and *-G*. The first three genes (classical) are expressed on almost all nucleated cells and are characterized by extensive polymorphism, whereas the latter three genes (non-classical) show more restricted expression and encode conserved receptors. Some of the non-classical HLA molecules

are described as ligand for KIR. For example, KIR2DL4 expressed in endosomes of uterine NK cells might interact with soluble HLA-G [77], whereas open conformers of HLA-F are suggested to interact with KIR3DS1 [78].

However, most ligands of KIR belong to the classical HLA class I molecules (**Table II**) [57, 77-92]. The KIR receptors interact with specific epitopes that are present on HLA-A, -B and -C molecules, which are termed A3/A11, Bw4, C1, and C2. The A3 and A11 epitopes are identified on approximately 45% of the HLA-A molecules, whereas about 36% of the HLA-B molecules, and a few HLA-A alleles, contain a Bw4 epitope [93]. The majority of the HLA-A and-B molecules do not function as ligand for KIR, but rather evolve under the pressure of the T-cell response. In contrast, all HLA-C molecules facilitate interactions with human KIR as they possess either an C1 or an C2 epitope. *HLA-C* is the most recently evolved *HLA* gene and probably resulted from a duplication of an *HLA-B*-like entity that encoded an C1 epitope [94]. A general overview of the HLA and KIR interactions are provided in table II.

Table II. Ligand specificity of human KIR for different HLA molecules. The open conformer status of HLA-F is abbriviated as OC.

Lineage	KIR	Signal	HLA ligand	Epitope	Reference
Lineage I	KIR2DL4	Inhibitory/Activating	HLA-G	-	[77]
	KIR2DL5A	Inhibitory	Unknown	-	[88]
	KIR2DL5B	Inhibitory	Unknown	-	[88]
Lineage III	KIR2DL1	Inhibitory	HLA-C (some HLA-B)	C2	[85-87, 92]
	KIR2DL2	Inhibitory	HLA-C (some HLA-B)	C1	[83-85, 87]
	KIR2DL3	Inhibitory	HLA-C (some HLA-B)	C1	[83-85, 87]
	KIR2DS1	Activating	HLA-C	C2	[85, 92]
	KIR2DS2	Activating	HLA-A and-C	Peptide-dependent	[82]
	KIR2DS3	Activating	Unknown	-	[84]
	KIR2DS4	Activating	HLA-A (some HLA-C)	A11	[79, 81]
	KIR2DS5	Activating	Unknown	-	[80, 84]
	KIR2DP1	Pseudogene	-	-	[89]
	KIR3DP1	Pseudogene	-	-	[90]
Lineage II	KIR3DL1	Inhibitory	HLA-A and-B	Bw4	[57]
	KIR3DL2	Inhibitory	HLA-A	A3/A11 peptide-dependent	[91]
	KIR3DS1	Inhibitory	HLA-B and-F	HLA-F: OC	[78]
Lineage V	KIR3DL3	Inhibitory	Unknown	-	[73]

The specificity and avidity of the KIR-HLA interactions can, however, vary for different alleles, and might also be dependent on the peptide bound by the HLA molecule. For example, KIR2DL1 molecules are generally able to interact with HLA-C2 epitopes, but the KIR2DL1*022 allele switched its specificity to the C1 epitope [95]. In addition, different *KIR2DL1* alleles are reported to interact with variable avidity to C2 epitopes. The importance of HLA-bound peptides is demonstrated by differential specificity and avidity of the interactions with KIR, and provide a mechanism to recognize changes in the peptide content, which may occur during viral infection or tumor formation [96, 97].

For several KIR receptors the ligand is not identified. They might utilize specialized interactions under specific conditions, like infection and tumor formation, or during development. Also, non-MHC molecules might interact with KIR, like is demonstrated for heparan sulfate and KIR2DL4 [98]. This interaction directly modules the cellular localization of KIR2DL4 and might also affect its function.

KIR in disease and pregnancy

Given the independent segregation of the highly polymorphic *KIR* and *HLA* gene families, and their important role in modulating the innate and adaptive immune responses, the diversity of KIR-HLA interactions might influence the health and disease of an individual. However, the extensive diversity of KIR-HLA interactions and the incomplete knowledge of the KIR ligands make the interpretation of disease association studies difficult. Nevertheless, abundant genomic studies have provided insights in the role of KIR in disease.

Associations with a variety of infectious diseases are reported [99-101], of which most infection studies involved the effect of KIR genetics on the progression of HIV. These studies mainly determined the control of HIV infection in the presence or absence of specific *KIR* genes, with special interest for *KIR3DL1* and *KIR3DS1*. For example, interactions of certain *KIR3DL1* and *HLA-B* allotypes seem to protect from HIV progression [102], and the presence of both *KIR3DL1* and *KIR3DS1* further suppress the viral load [103]. Even more, the presence of *KIR2DL2*, *KIR2DL5*, *KIR2DS5*, and *KIR2DS2* in the maternal KIR repertoire is associated with a decreased HIV transmission to the child [103]. An meta-analysis of 13 cohort studies, however, demonstrated that only the presence of *KIR2DL3* and *KIR3DS1* (homozygous) in the Caucasian subpopulation was associated with protection against HIV progression [104].

The role of KIR in different forms of cancer is extensively studied [105-109]. For example, an increased frequency of *KIR2DL2* and *KIR2DS2*, both linked to KIR haplotype B, was reported in patients with leukemia. Therefore, it was suggested that the inhibitory KIR profile of haplotype A was protective against leukemia [110]. In a different cohort of leukemia patients it was shown that the presence of *KIR2DL2* was not significantly higher in patients, but instead *KIR2DS*2 and *KIR2DS4* showed a positive correlation with disease [111]. In opposition, a model has been proposed that suggests

a protective role of *KIR2DL2* in leukemia patients [112]. The contradicting results illustrate the difficulty to associate KIR with different forms of cancer [113]. These association studies are mainly based on the presence or absence of *KIR* genes and the corresponding disease progression, but might miss information on the tumor phenotype, HLA ligands, KIR haplotype configurations, and *KIR* alleles, which might influence the disease status. However, a more tailored approach might provide opportunities to target the KIR-HLA interactions in certain types of cancer [105].

Other disease association studies involved inflammatory and autoimmune disorders [114-117], or the effect of KIR in organ and stem cell transplantation [118-122].

During pregnancy, the local immune system is involved in vascular remodeling and in avoiding the rejection of the foetus, which possesses paternal antigens. Specialized NK cells that reside in the uterine (uNK cells) might play an important role by secreting cytokines and chemokines upon binding of KIR with HLA-C molecules, which are present on trophoblasts. Association studies illustrated an increased risk of pregnancy disorders for certain maternal inhibitory KIR and fetal HLA-C combinations [123-126]. Also, the absence of activating KIR, like *KIR2DS1* and *KIR2DS4*, might result in poor trophoblast invasion, placental stress, and growth restriction due to insufficient NK cell activation and subsequent cytokine release [127, 128]. In contrast, however, the presence of activating KIR, including *KIR2DS1*, was also associated with increased risk of recurrent miscarriage [129, 130]. In addition to the lineage III KIR receptors, KIR2DL4 might also play a role in early pregnancy through interactions with soluble HLA-G, which is predominantly expressed in fetal trophoblast cells [77, 131].

The majority of the association studies are based on the presence or absence of *KIR* and *HLA* genes. The outcome of the studies might be influenced by additional factors, such as allelic variation, disease phenotype, missed recombinant genes, potential other ligands, and genetic variation of populations and individuals.

NK cells in animal models

The involvement in essential immunological processes, such as the killing of aberrant cells, but also the associations with pregnancy, therapies and diseases, underlines the need for a better understanding of NK cells and their receptors. Studies in humans are mainly performed on peripheral blood lymphocytes, which represents a single subset of NK cells. More specialized NK cells that reside in tissues are less accessible. In addition, controlled disease experiments in humans are challenging and unethical, and are therefore limited to patient cohort studies. The contradicting results in the association studies and the restrictions in human studies illustrate the urgency for models to study NK cells in health and disease.

Studies in murine models elucidated several basic principles of the NK cells, such as the ‘missing self’ hypothesis and memory-like functions [46, 132]. These findings demonstrate that mouse models can be a powerful model to study fundamental processes. However, hardly any outcomes of mouse studies that involved NK cells were

of medical relevance, which might be explained by the substantial differences in human and mouse NK cell biology. Significant variance is observed for the cytotoxicity, subsets, localization, and cytokine production profiles of the NK cells in both species [133]. The major difference is, however, within the receptors that are expressed. The MHC class I molecules are recognized by Ly49 molecules in mice, whereas in humans KIR facilitate these interactions [134, 135]. Although Ly49 and KIR have similar functions (analogs), their structure, genetics, and MHC binding sites are different and are not comparable. In addition, the structure and function of the inhibitory NKG2A and activating NKG2D receptors are highly conserved, but the ligands diverged in humans and mice, indicating a diverged immune response [136]. Humanized mouse models might improve immunological studies, although the applicability may be limited for lymphocyte studies [137, 138]. In overview, mouse models provide insights into the basic principles of NK cell immunology, but, considering the broad differences, might be less useful in disease-related studies.

Translational biomedical research with non-human primate species has played an important role in the advances of the medical, pharmaceutical, and scientific field. Macaque species share a relatively close common ancestor with humans that lived approximately 25 million years ago, which explains the similarity to humans in their physiology, reproduction, development, and immunology. Rhesus and cynomolgus macaques, two species that diverged from each other 1-3 million years ago, are widely used as model in infectious and autoimmune diseases, including models for HIV/SIV, tuberculosis, and zika infection [139-141]. Although differential properties are reported for several immune cells in primate species [142], the macaque and human NK cells share abundant characteristics [143]. Three NK cells subsets are defined in macaques, of which the $CD56^{+}CD16^{-}$ and $CD16^{+}$ populations resemble the $CD56^{bright}$ and $CD56^{dim}$ human NK cell subsets, respectively [144]. The third macaque subset is double negative ($CD56^{-}CD16^{-}$) and might be an intermediate NK cell population. In addition, like in humans, tissue-resident NK cells are described in macaques, although data is limited [145, 146]. The activity of macaque NK cells is regulated by inhibitory and activating receptors, including a highly diverse set of KIR receptors that recognizes MHC class I molecules. These characteristics indicate that macaque species are suitable biomedical models in studies that involve NK cells.

Co-evolution of the MHC and KIR genes in hominids

The *KIR* gene region show considerable variation and expansion across different primate species, as do the genes that encode their MHC class I ligands. Of these ligands, MHC-C is the most recently evolved member and share orthologs in humans, chimpanzees, bonobos, and orangutans [147]. In humans, the HLA-C molecules are fixed and their C1 and C2 epitopes act as dominant ligands for the lineage III KIR, whereas the more ancient HLA-A and -B are only recognized by a few KIR [148]. In contrast, in orangutans, MHC-C is present on half of the MHC haplotypes and only have

the C1 epitope [149]. The orangutan KIR show specificity for the C1 epitopes, but not C2 epitopes. This indicates that MHC-C emerged in a common ancestor of humans and orangutans, and that HLA-C and KIR further evolved in humans to gain C2 epitopes and specific KIR ligands. This rapid and species-specific co-evolution is substantiated in chimpanzees by the presence of KIR receptors specific for both C1 and C2 epitopes [150].

The rapid evolution of the *KIR* genes is further illustrated by the presence of only a few orthologs in hominids. Chimpanzees and humans diverged from a common ancestor relatively recently, but share only four orthologs, namely *KIR2DL4*, *KIR2DL5*, *KIR2DS4*, and *KIR3DL3* [93]. In orangutans, similar orthologs are identified, with the exception of *KIR2DS4*. The majority of the *KIR* genes are, however, species-specific. In humans, both the telomeric and centromeric haplotype regions show an expansion of genes [151], whereas the chimpanzee and orangutan *KIR* genes are mainly confined to the centromeric region [93]. Even more, the standard A and B configurations observed in humans is absent in other primates, which might indicate an evolutionary selective factor that drive the differentially distributed haplotype groups [76].

A comparable framework is described for the KIR haplotypes in hominids, which have similar genes flanking the centromeric and telomeric regions. This suggests an ancestral haplotype that was generated by duplications before speciation [149, 152]. On one side, the centromeric region is flanked by *KIR3DL3* in all hominids, which was probably generated by a recombination event [153], whereas the other side is marked by a pseudogene. The sequence of this pseudogene encodes the first 5 exons in all great apes and is followed by repetitive elements. The recombination events that resulted in the pseudogene are, however, species-specific, with genes from distinct lineages involved. This framework pseudogene is therefore not considered as an ortholog in primate species. The telomeric region is flanked by *KIR2DL4* and *KIR3DL2*/*KIR3DL1*. The former gene is largely conserved in primates, with little species-specific variation. The number of ITIM motifs in the cytoplasmic tail of KIR2DL4 might, however, vary in different species, with either one or two motifs present [154]. The boundary of the *KIR* gene cluster, which is adjacent to the *FCAR* gene, is marked by a lineage II *KIR* gene in all great apes, although this gene is not conserved.

KIR genes in macaques

Comparative analysis of the *KIR* gene cluster in different higher primate species highlighted shared characteristics, although species-specific evolution was evident. At least in part, the diversification of the *KIR* gene system can be explained by co-evolution with the MHC class I ligands. In contrast to higher primates, however, macaques do not have a homolog of MHC-C, and instead have an expanded repertoire of *MHC-A* and-*B* genes [155-157]. In agreement with the absence of *MHC-C*, just a single lineage III *KIR* gene is described in macaques, which is transcribed as a receptor with a single extracellular domain (KIR1D) [158, 159]. Macaque MHC haplotypes can contain up to

3 *MHC-A* genes and even more copies of *MHC-B*, both of which are characterized by abundant polymorphism. Several rhesus macaque lineage II KIR are identified to interact with MHC-A, and a few MHC-B (**Table III**), which is in line with an extensive expansion reported for this *KIR* lineage in macaques [160-163]. In total, 22 rhesus macaque *KIR* genes were described before this manuscript was published, of which 18 clustered into lineage II [61, 164, 165]. These lineage II *KIR* genes encode receptors with three extra cellular domains (KIR3DL and KIR3DS). Genomic and transcriptomic segregation studies defined several rhesus macaque *KIR* haplotypes that contained 6 to 13 *KIR* genes, indicating copy number variation [166]. *KIR3DL20* was reported on the majority of the defined haplotypes, suggesting a framework gene status. This gene probably shares a common progenitor with human *KIR3DL3*, as the exons that encode the extracellular domains are similar [153]. The stem region and cytoplasmic tail are, however, encoded by exons that cluster into a different lineage. The pseudogene that

Table III. Rhesus macaque KIR interactions with MHC molecules. This overview represents a generalization, as the specificity for epitopes seems to be allele specific, which are described in more detailed in the corresponding literature.

Lineage	KIR	Signal	MHC ligand	Epitope	Reference
Lineage II	KIR3DL01	Inhibitory	Mamu-B	Bw4	[161]
	KIR3DLW03	Inhibitory	Mamu-A1	Bw4	[160]
	KIR3DL05	Inhibitory	Mamu-A1,-A3,-B,-AG	Bw4, Bw6	[160, 162, 163]
	KIR3DL06	Inhibitory	Mamu-B	Bw4	[163]
	KIR3DL08	Inhibitory	Mamu-B	Bw4	[163]
	KIR3DL11	Inhibitory	Mamu-A1	-	[160]
	KIR3DS05	Activating	Mamu-A1 (very weak)	-	[160]
	KIR3DSW08	Activating	Mamu-A3 (very weak),-B	Bw4	[160, 163]

might be present on the ancestral *KIR* haplotype is also reported in rhesus macaques, although its framework status is not confirmed. The documented rhesus macaque pseudogene (*KIRDP*) is not orthologous to human *KIR3DP1*, as the first three exons associate with different rhesus macaque sequences [149]. On approximately half of the haplotypes, *KIR2DL04* is identified, which represents the only macaque *KIR* ortholog that is shared with humans and other hominids [166]. All macaques KIR2DL04 allotypes contain two ITIMs in contrast to the single ITIM observed for human KIR2DL4. In addition, soluble HLA-G is identified as KIR2DL4 ligand in humans, but in macaques the gene encoding this non-classical MHC diverged into a pseudogene. However, a similar gene, *Mamu-AG*, has been identified that is expressed on trophoblasts and might facilitate interactions with macaque KIR2DL04, although binding is not

demonstrated [167]. A framework gene that marks the other side of the telomeric *KIR* haplotype in macaques has not been reported, but is most likely a lineage II *KIR* gene.

In 2005, and more recently in 2019, BAC clones that covered complete rhesus macaque *KIR* haplotypes were completely sequenced and constructed [159, 168]. These two haplotypes demonstrated that the physical location of *KIR3DL20* resembles the *KIR3DL3* locus in humans, substantiating that these genes share a common progenitor. *KIR1D* has been identified next to the *KIR3DL20*. Although all 9 exons are identified for *KIR1D* at the genomic level, exons 3 and 5, which contain deletions that would shift the reading frame, are not incorporated at the transcription level. In addition, a pseudogene, *KIR2DL04*, and lineage II *KIR* genes are documented on these relatively short *KIR* haplotypes. In contrast to humans, chimpanzees, and orangutans, the macaque *KIR* genes are expanded in the telomeric region, whereas the centromeric region may only contain *KIR3DL20*, *KIR1D*, and a putative pseudogene. Although informative, these two haplotypes do not provide complete information on the haplotype configurations in macaques.

Most studies were performed in Indian rhesus macaques, whereas only a few individuals from the Chinese and Burmese populations are genomically characterized for their *KIR* repertoire [166, 169-175]. Less well characterized is the *KIR* gene repertoire in cynomolgus macaques. Only the isolated and relatively recently established Mauritian population is analyzed for their *KIR* gene content and shows limited variation due to the small number of founders [176, 177]. Despite that the majority of the macaque studies involved the characterization of *KIR* genotypes and haplotypes by determining the presence or absence of *KIR* genes, species-specific variation is documented for both macaque species. A handful of studies involved the transcriptomic characterization by Sanger sequencing and illustrated extensive allelic polymorphism and putative recombinant genes.

Evolution of sequencing techniques

Over the years, sequencing techniques have been improved, which enables a higher throughput for the characterization of complex genetic regions, like the *KIR* gene region. Up to now, however, only the conventional sequencing approaches have been used for the macaque *KIR* gene cluster, although new techniques are available that might improve the characterization.

It was in 1953 that Watson and Crick solved the structure of DNA [178], which is composed of four deoxyribonucleic acids (adenine, cytosine, thymine, guanine). Upon this discovery, it took over 20 years to develop sufficient techniques that were able to decode the DNA, which harbors the essential information about our phylogeny, ancestry, and susceptibility to disease. Early attempts to sequence DNA involved two-dimensional separation of radioactively labeled DNA fragments and reported 24 bases of a lactose-repressor binding site, which took two years [179]. In around 1976, the

development of first-generation sequencing techniques accelerated with approaches reported by Sanger and Gilbert, which enabled the separation of DNA fragments on a polyacrylamide gel at a single-base resolution [180, 181]. Especially Sanger's sequencing approach was improved over the years by the introduction of enhanced polymerases and chemicals, and this led to the release of the first fluorescent-based sequencing machines in 1987, which enabled the sequencing of 1.000 bases per day [182]. These improvements paved to way for the Human Genome Project (HGP) that started in 1990 and aimed to sequence the complete human genome. Propelled by a string of breakthroughs, the efforts of labs around the world, and billions of dollars, the human genome sequence was completed in 2004 [183]. One year later, second-generation sequencing instruments became commercially available, such as the 454 (Roche), the Solexa (Illumina), and the IonTorrent (Rothberg), the fast developments of which enhanced the quality and quantity of sequenced DNA and lowered the costs. To illustrate, the number of bases that could be sequenced within two days by a second-generation sequencing approach exceeds the size of the HGP by a factor of 40. Nowadays, the Illumina platform is the only second-generation technique that is still widely used and supported. Third-generation sequencing platforms, which enable real-time and single-molecule (SMRT) sequencing, include Pacific Biosciences (PacBio) and Oxford Nanopore sequencing. The PacBio platform measures the incorporation of a single fluorescently labelled nucleotide into a growing DNA chain using a zero-mode waveguide, and enables relatively high accuracy sequencing of long reads (10-100 kb) [184, 185]. The Nanopore platform, of which the technique was already hypothesized in the 1990s [186], pulls complete strains of DNA through a narrow pore. This causes shifts in ion currents that reveals the sequence of the DNA [187]. The highest read-lengths that are currently obtained by Nanopore sequencing goes up to 900 kb, but the relatively low accuracy requires correction by a high coverage. A major advantage of the Nanopore platform is the portability of the pocket-sized device. The two third-generation sequencing platforms do not only enable a high sequencing throughput of large DNA fragments, which allows the sequencing of a complete human genome within two days at the cost of approximately thousand dollars, but are also capable of DNA modification sequencing, such as methylation (**Table IV**) [188-190].

The efforts made in the last half century resulted in the development and improvement of sequencing approaches, which are still not plateaued in accuracy, throughput and cost. By combining the golden standard of Sanger sequencing with both third-generation sequencing platforms, PacBio and Nanopore, complex immunogenic regions in the macaque genome, such as the *KIR* gene cluster that exceeds the complexity of similar regions in humans, can be elucidated.

Table IV. Comparison of Sanger sequencing and third-generation sequencing platforms. Table is adapted from [190].

Technique	Platform	Cost	Avg. Output (Gb)	Avg./Max. read length	Run time	Error rate
Sanger	-	$	0.00006	650 bp	0.5-1h	0.1%
PacBio	RS II	$$$$	2	5-15 kb / >60 kb	0.5-6h	10-15%
	Sequel	$$$	8-10	5-30 kb / >200 kb	0.5-20h	10-15%
	Sequel II	$$	56-70	5-30 kb / >200 kb	0.5-30h	10-15%
Nanopore	Flongle	$	1	5-35 kb / >2 Mb	0.5- 48h	5-15%
	MinION	$	15	5-35 kb / >2 Mb	0.5- 48h	5-15%
	GridION	$$	75	5-35 kb / >2 Mb	0.5- 48h	5-15%
	PromethION	$$$	4-6 Tb	5-35 kb / >2 Mb	0.5- 72h	5-15%

Thesis aim and overview

This manuscript is aimed to provide more insights into the *KIR* gene system of macaques. Previous characterization studies already hinted on the complexity of this system, but these experiments relied on conventional sequencing techniques, such as Sanger sequencing and Roche/454 pyrosequencing, or on the determination of the presence or absence of *KIR* genes at the genomic DNA level. Although these studies were informative, a comprehensive overview on the macaque *KIR* gene system is still lacking. Furthermore, it has been demonstrated that the geographical origin of macaques that are used as biomedical model influences experimental outcomes, which is most probably related to immunogenetic differences. An improved characterization of the macaque *KIR* genes might refine the selection of animals that are used for biomedical studies based on genetic knowledge in different macaque species, populations and individuals. Such an approach has already shown beneficial in case of the *MHC* genes [191, 192]. In addition, this manuscript is intended to provide a comparison of the macaque and human *KIR* gene clusters, and how these clusters relate with previous published data on the *KIR* genes in other primate species. The immunogenetic comparison might elucidate differential and conserved characteristics in the *KIR* gene evolution in different species, which further the understanding on receptor function, ligands, and evolutionary selective forces. At last, novel methods, techniques, and sequencing platforms that are described in this manuscript to characterize the highly complex macaque *KIR* gene system can be applied for the characterization of plastic multigenic regions in other species, including humans.

The sequencing of macaque and human *KIR* genes was mainly performed by Sanger sequencing, or by the profiling of *KIR* genes at the genomic DNA level, which are time

consuming methods or do not provide allele level resolution. Allele level specificity and functionality is, however, documented for multiple KIR and might be of importance to interpret disease association studies. Therefore, a method was established that enabled a high-throughput characterization of both macaque and human *KIR* gene transcripts by sequencing on a PacBio platform. Abundant novel KIR alleles and gene entities were confirmed in these transcriptome datasets. Even more, multiple recombinant *KIR* genes were identified that would be missed by most conventional sequencing techniques. Transcriptomic macaque and human *KIR* haplotypes were subsequently defined by segregation studies, which illustrated the plasticity of the macaque *KIR* gene organization, and confirmed novel and previously reported human haplotype configurations (**Chapter 2**).

The divergence of the human and macaque *KIR* genes is in line with the evolutionary distance to their common ancestor that lived approximately 25 million years ago. There is, however, also significant diversity reported in the closely related rhesus and cynomolgus macaque species that diverged only 1-3 million years ago, which indicates an unparalleled rapid evolution of the macaque *KIR* gene system. This is further substantiated by allele and gene level diversity in different geographical macaque populations, such as the Indian, Chinese, and Burmese rhesus macaques, and the mainland, insular, and Mauritian cynomolgus macaques (**Chapter 3**). The *KIR* gene characterization of the cynomolgus macaques and the large number of recombinant *KIR* genes also stressed the need for a revised macaque *KIR* nomenclature report [62].

Allelic polymorphism, recombinations, and copy number variations contribute to the diversity of the *KIR* genes, but less well defined is the variation that is generated by alternative splicing. The impact of this posttranscriptional process is evident for the *KIR* genes in both macaques and humans, as abundant exon skipping events, alternative splice sites, and cryptic exons are described. These isoforms might differ in their expression status, function, cellular localization, and binding properties, and thereby significantly contribute to the variation of the *KIR* gene system (**Chapter 4**).

Whereas human *KIR* haplotypes mostly follow standard configurations with some differential recombinations, the knowledge on macaque KIR haplotypes is restricted to two completely sequenced rhesus macaque haplotypes. Macaque *KIR* haplotypes that were deduced from transcriptome studies have shown variable gene content and a less strict framework, but the exact configuration remains elusive. The combination of targeted enrichment using Cas9 nuclease followed by sequencing on an Oxford Nanopore MinION platform, enabled the construction of complete macaque *KIR* haplotypes at an allele level resolution. This technique was further validated by the assembly of human *KIR* haplotypes. Overall, the targeted enrichment and long-read sequencing provides a cost- and time-efficient method to characterize complete haplotypes of complex multigenic regions in different species (**Chapter 5**).

In conclusion, a comprehensive overview of the macaque *KIR* gene region is provided that include a summary and discussion on the extensive diversity that has been reported in rhesus and cynomolgus macaques. Furthermore, a comparison of the *KIR* cluster in other primate species illustrates the differences and similarities that

have emerged during the rapid evolution of the *KIR* genes, and shows the relevance for non-human primate models in translatable biomedical research (**Chapter 6**).

With the new sequencing techniques available, the genetic organization of the *KIR* gene cluster in macaque species is more and more elucidated. These insights pave the way to examine the role of KIR in different biological processes (**Chapter 7**). The expression of KIR on tissue-resident NK cells, for instance, might display more specialized functions. Even more, a highly regulated alternative splicing machinery might generate specific KIR isoforms during different processes. As such, the functional impact of the broad KIR receptor diversity might stretch beyond the current findings and assumptions.

References

1. Groves C: Species concept in primates. *American journal of primatology* 2012, 74(8):687-691.
2. Bassing CH, Swat W, Alt FW: The Mechanism and Regulation of Chromosomal V(D)J Recombination. *Cell* 2002, 109(2, Supplement 1):S45-S55.
3. Hazenberg MD, Spits H: Human innate lymphoid cells. *Blood* 2014, 124(5):700-709.
4. Kiessling R, Klein E, Pross H, Wigzell H: "Natural" killer cells in the mouse. II. Cytotoxic cells with specificity for mouse Moloney leukemia cells. Characteristics of the killer cell. *Eur J Immunol* 1975, 5(2):117-121.
5. Takasugi M, Mickey MR, Terasaki PI: Reactivity of lymphocytes from normal persons on cultured tumor cells. *Cancer research* 1973, 33(11):2898-2902.
6. Herberman RB, Nunn ME, Holden HT, Lavrin DH: Natural cytotoxic reactivity of mouse lymphoid cells against syngeneic and allogeneic tumors. II. Characterization of effector cells. *International journal of cancer* 1975, 16(2):230-239.
7. Kim S, Iizuka K, Aguila HL, Weissman IL, Yokoyama WM: In vivo natural killer cell activities revealed by natural killer cell-deficient mice. *Proceedings of the National Academy of Sciences of the United States of America* 2000, 97(6):2731-2736.
8. Smyth MJ, Cretney E, Kelly JM, Westwood JA, Street SE, Yagita H, Takeda K, van Dommelen SL, Degli-Esposti MA, Hayakawa Y: Activation of NK cell cytotoxicity. *Molecular immunology* 2005, 42(4):501-510.
9. Tripp CS, Wolf SF, Unanue ER: Interleukin 12 and tumor necrosis factor alpha are costimulators of interferon gamma production by natural killer cells in severe combined immunodeficiency mice with listeriosis, and interleukin 10 is a physiologic antagonist. *Proceedings of the National Academy of Sciences of the United States of America* 1993, 90(8):3725-3729.
10. Sun JC, Beilke JN, Lanier LL: Adaptive immune features of natural killer cells. *Nature* 2009, 457(7229):557-561.
11. Guma M, Angulo A, Vilches C, Gomez-Lozano N, Malats N, Lopez-Botet M: Imprint of human cytomegalovirus infection on the NK cell receptor repertoire. *Blood* 2004, 104(12):3664-3671.
12. Fauriat C, Long EO, Ljunggren HG, Bryceson YT: Regulation of human NK-cell cytokine and chemokine production by target cell recognition. *Blood* 2010, 115(11):2167-2176.
13. De Maria A, Bozzano F, Cantoni C, Moretta L: Revisiting human natural killer cell subset function revealed cytolytic CD56(dim)CD16+ NK cells as rapid producers of abundant IFN-gamma on activation. *Proceedings of the National Academy of Sciences of the United States of America* 2011, 108(2):728-732.
14. Dulphy N, Haas P, Busson M, Belhadj S, Peffault de Latour R, Robin M, Carmagnat M, Loiseau P, Tamouza R, Scieux C *et al*: An unusual CD56(bright) CD16(low) NK cell subset dominates the early posttransplant period following HLA-matched hematopoietic stem cell transplantation. *Journal of immunology (Baltimore, Md : 1950)* 2008, 181(3):2227-2237.
15. Yu J, Mao HC, Wei M, Hughes T, Zhang J, Park I-k, Liu S, McClory S, Marcucci G, Trotta R *et al*: CD94 surface density identifies a functional intermediary between the CD56bright and CD56dim human NK-cell subsets. *Blood* 2010, 115(2):274-281.
16. Chan A, Hong DL, Atzberger A, Kollnberger S, Filer AD, Buckley CD, McMichael A, Enver T, Bowness P: CD56bright human NK cells differentiate into CD56dim cells: role of contact with peripheral fibroblasts. *Journal of immunology (Baltimore, Md : 1950)* 2007, 179(1):89-94.
17. Romagnani C, Juelke K, Falco M, Morandi B, D'Agostino A, Costa R, Ratto G, Forte G, Carrega P, Lui G *et al*: CD56brightCD16- killer Ig-like receptor- NK cells display longer telomeres and acquire features of CD56dim NK cells upon activation. *Journal of immunology (Baltimore, Md : 1950)* 2007, 178(8):4947-4955.
18. Nguyen S, Dhedin N, Vernant JP, Kuentz M, Al Jijakli A, Rouas-Freiss N, Carosella ED, Boudifa A, Debre P, Vieillard V: NK-cell reconstitution after haploidentical hematopoietic stem-cell transplantations: immaturity of NK cells and inhibitory effect of NKG2A override GvL effect. *Blood* 2005, 105(10):4135-4142.
19. Andrews DM, Smyth MJ: A potential role for RAG-1 in NK cell development revealed by analysis of NK cells during ontogeny. *Immunology and cell biology* 2010, 88(2):107-116.
20. Freud AG, Becknell B, Roychowdhury S, Mao HC, Ferketich AK, Nuovo GJ, Hughes TL, Marburger TB, Sung J, Baiocchi RA *et al*: A human CD34(+) subset resides in lymph nodes and differentiates into CD56bright natural killer cells. *Immunity* 2005, 22(3):295-304.

21. Kim S, Iizuka K, Kang HS, Dokun A, French AR, Greco S, Yokoyama WM: In vivo developmental stages in murine natural killer cell maturation. *Nature immunology* 2002, 3(6):523-528.

22. Santoni A, Carlino C, Gismondi A: Uterine NK cell development, migration and function. *Reproductive biomedicine online* 2008, 16(2):202-210.

23. Ferlazzo G, Thomas D, Lin SL, Goodman K, Morandi B, Muller WA, Moretta A, Munz C: The abundant NK cells in human secondary lymphoid tissues require activation to express killer cell Ig-like receptors and become cytolytic. *Journal of immunology (Baltimore, Md : 1950)* 2004, 172(3):1455-1462.

24. Dalbeth N, Gundle R, Davies RJ, Lee YC, McMichael AJ, Callan MF: CD56bright NK cells are enriched at inflammatory sites and can engage with monocytes in a reciprocal program of activation. *Journal of immunology (Baltimore, Md : 1950)* 2004, 173(10):6418-6426.

25. Shiow LR, Rosen DB, Brdickova N, Xu Y, An J, Lanier LL, Cyster JG, Matloubian M: CD69 acts downstream of interferon-alpha/beta to inhibit S1P1 and lymphocyte egress from lymphoid organs. *Nature* 2006, 440(7083):540-544.

26. Jenne CN, Enders A, Rivera R, Watson SR, Bankovich AJ, Pereira JP, Xu Y, Roots CM, Beilke JN, Banerjee A *et al*: T-bet-dependent S1P5 expression in NK cells promotes egress from lymph nodes and bone marrow. *The Journal of experimental medicine* 2009, 206(11):2469-2481.

27. Campbell JJ, Qin S, Unutmaz D, Soler D, Murphy KE, Hodge MR, Wu L, Butcher EC: Unique subpopulations of CD56+ NK and NK-T peripheral blood lymphocytes identified by chemokine receptor expression repertoire. *Journal of immunology (Baltimore, Md : 1950)* 2001, 166(11):6477-6482.

28. Melsen JE, Lugthart G, Lankester AC, Schilham MW: Human Circulating and Tissue-Resident CD56(bright) Natural Killer Cell Populations. *Front Immunol* 2016, 7:262-262.

29. Sun JC, Beilke JN, Lanier LL: Adaptive immune features of natural killer cells. *Nature* 2009, 457(7229):557-561.

30. Halenius A, Gerke C, Hengel H: Classical and non-classical MHC I molecule manipulation by human cytomegalovirus: so many targets-but how many arrows in the quiver? *Cellular & molecular immunology* 2015, 12(2):139-153.

31. Schlums H, Cichocki F, Tesi B, Theorell J, Beziat V, Holmes TD, Han H, Chiang SC, Foley B, Mattsson K *et al*: Cytomegalovirus infection drives adaptive epigenetic diversification of NK cells with altered signaling and effector function. *Immunity* 2015, 42(3):443-456.

32. Grzywacz B, Kataria N, Kataria N, Blazar BR, Miller JS, Verneris MR: Natural killer–cell differentiation by myeloid progenitors. *Blood* 2011, 117(13):3548-3558.

33. Chen Q, Ye W, Jian Tan W, Mei Yong KS, Liu M, Qi Tan S, Loh E, Te Chang K, Chye Tan T, Preiser PR *et al*: Delineation of Natural Killer Cell Differentiation from Myeloid Progenitors in Human. *Scientific Reports* 2015, 5:15118.

34. Wu C, Espinoza DA, Koelle SJ, Yang D, Truitt L, Schlums H, Lafont BA, Davidson-Moncada JK, Lu R, Kaur A *et al*: Clonal expansion and compartmentalized maintenance of rhesus macaque NK cell subsets. *Science immunology* 2018, 3(29).

35. Wu C, Li B, Lu R, Koelle SJ, Yang Y, Jares A, Krouse AE, Metzger M, Liang F, Lore K *et al*: Clonal tracking of rhesus macaque hematopoiesis highlights a distinct lineage origin for natural killer cells. *Cell stem cell* 2014, 14(4):486-499.

36. Elliott JM, Wahle JA, Yokoyama WM: MHC class I-deficient natural killer cells acquire a licensed phenotype after transfer into an MHC class I-sufficient environment. *The Journal of experimental medicine* 2010, 207(10):2073-2079.

37. Joncker NT, Shifrin N, Delebecque F, Raulet DH: Mature natural killer cells reset their responsiveness when exposed to an altered MHC environment. *The Journal of experimental medicine* 2010, 207(10):2065-2072.

38. Kim S, Poursine-Laurent J, Truscott SM, Lybarger L, Song YJ, Yang L, French AR, Sunwoo JB, Lemieux S, Hansen TH *et al*: Licensing of natural killer cells by host major histocompatibility complex class I molecules. *Nature* 2005, 436(7051):709-713.

39. Bix M, Liao NS, Zijlstra M, Loring J, Jaenisch R, Raulet D: Rejection of class I MHC-deficient haemopoietic cells by irradiated MHC-matched mice. *Nature* 1991, 349(6307):329-331.

40. Liao NS, Bix M, Zijlstra M, Jaenisch R, Raulet D: MHC class I deficiency: susceptibility to natural killer (NK) cells and impaired NK activity. *Science (New York, NY)* 1991, 253(5016):199-202.

41. Belanger S, Tu MM, Rahim MM, Mahmoud AB, Patel R, Tai LH, Troke AD, Wilhelm BT, Landry JR, Zhu Q *et al*: Impaired natural killer cell self-education and "missing-self" responses in Ly49-deficient mice. *Blood* 2012, 120(3):592-602.

42. Joncker NT, Fernandez NC, Treiner E, Vivier E, Raulet DH: NK cell responsiveness is tuned commensurate with the number of inhibitory receptors for self-MHC class I: the rheostat model. *Journal of immunology (Baltimore, Md : 1950)* 2009, 182(8):4572-4580.

43. Restifo NP, Esquivel F, Kawakami Y, Yewdell JW, Mule JJ, Rosenberg SA, Bennink JR: Identification of human cancers deficient in antigen processing. *The Journal of experimental medicine* 1993, 177(2):265-272.

44. Restifo NP, Marincola FM, Kawakami Y, Taubenberger J, Yannelli JR, Rosenberg SA: Loss of functional beta 2-microglobulin in metastatic melanomas from five patients receiving immunotherapy. *Journal of the National Cancer Institute* 1996, 88(2):100-108.

45. Collins KL, Chen BK, Kalams SA, Walker BD, Baltimore D: HIV-1 Nef protein protects infected primary cells against killing by cytotoxic T lymphocytes. *Nature* 1998, 391(6665):397-401.

46. Ljunggren HG, Karre K: In search of the 'missing self': MHC molecules and NK cell recognition. *Immunology today* 1990, 11(7):237-244.

47. Tarek N, Le Luduec JB, Gallagher MM, Zheng J, Venstrom JM, Chamberlain E, Modak S, Heller G, Dupont B, Cheung NK *et al*: Unlicensed NK cells target neuroblastoma following anti-GD2 antibody treatment. *The Journal of clinical investigation* 2012, 122(9):3260-3270.

48. Tu MM, Mahmoud AB, Makrigiannis AP: Licensed and Unlicensed NK Cells: Differential Roles in Cancer and Viral Control. *Front Immunol* 2016, 7:166.

49. Orr MT, Murphy WJ, Lanier LL: 'Unlicensed' natural killer cells dominate the response to cytomegalovirus infection. *Nature immunology* 2010, 11(4):321-327.

50. Barrow AD, Martin CJ, Colonna M: The Natural Cytotoxicity Receptors in Health and Disease. *Front Immunol* 2019, 10(909).

51. Braud VM, Allan DS, O'Callaghan CA, Soderstrom K, D'Andrea A, Ogg GS, Lazetic S, Young NT, Bell JI, Phillips JH *et al*: HLA-E binds to natural killer cell receptors CD94/NKG2A, B and C. *Nature* 1998, 391(6669):795-799.

52. Lopez-Larrea C, Suarez-Alvarez B, Lopez-Soto A, Lopez-Vazquez A, Gonzalez S: The NKG2D receptor: sensing stressed cells. *Trends in molecular medicine* 2008, 14(4):179-189.

53. Storkus WJ, Alexander J, Payne JA, Dawson JR, Cresswell P: Reversal of natural killing susceptibility in target cells expressing transfected class I HLA genes. *Proceedings of the National Academy of Sciences of the United States of America* 1989, 86(7):2361-2364.

54. Shimizu Y, DeMars R: Demonstration by class I gene transfer that reduced susceptibility of human cells to natural killer cell-mediated lysis is inversely correlated with HLA class I antigen expression. *Eur J Immunol* 1989, 19(3):447-451.

55. Colonna M, Samaridis J: Cloning of immunoglobulin-superfamily members associated with HLA-C and HLA-B recognition by human natural killer cells. *Science (New York, NY)* 1995, 268(5209):405-408.

56. Wagtmann N, Biassoni R, Cantoni C, Verdiani S, Malnati MS, Vitale M, Bottino C, Moretta L, Moretta A, Long EO: Molecular clones of the p58 NK cell receptor reveal immunoglobulin-related molecules with diversity in both the extra- and intracellular domains. *Immunity* 1995, 2(5):439-449.

57. D'Andrea A, Chang C, Franz-Bacon K, McClanahan T, Phillips JH, Lanier LL: Molecular cloning of NKB1. A natural killer cell receptor for HLA-B allotypes. *Journal of immunology (Baltimore, Md : 1950)* 1995, 155(5):2306-2310.

58. Long EO, Colonna M, Lanier LL: Inhibitory MHC class I receptors on NK and T cells: a standard nomenclature. *Immunology today* 1996, 17(2):100.

59. Andre P, Biassoni R, Colonna M, Cosman D, Lanier LL, Long EO, Lopez-Botet M, Moretta A, Moretta L, Parham P *et al*: New nomenclature for MHC receptors. *Nature immunology* 2001, 2(8):661.

60. Marsh SG, Parham P, Dupont B, Geraghty DE, Trowsdale J, Middleton D, Vilches C, Carrington M, Witt C, Guethlein LA *et al*: Killer-cell immunoglobulin-like receptor (KIR) nomenclature report, 2002. *Human immunology* 2003, 64(6):648-654.

61. Robinson J, Guethlein LA, Maccari G, Blokhuis J, Bimber BN, de Groot NG, Sanderson ND, Abi-Rached L, Walter L, Bontrop RE *et al*: Nomenclature for the KIR of non-human species. *Immunogenetics* 2018, 70(9):571-583.

62. Bruijnesteijn J, de Groot NG, Otting N, Maccari G, Guethlein LA, Robinson J, Marsh SGE, Walter L, O'Connor DH, Hammond JA *et al*: Nomenclature report for killer-cell immunoglobulin-like receptors (KIR) in macaque species: new genes/alleles, renaming recombinant entities and IPD-NHKIR updates. *Immunogenetics* 2019.

63. Trowsdale J, Jones DC, Barrow AD, Traherne JA: Surveillance of cell and tissue perturbation by receptors in the LRC. *Immunological reviews* 2015, 267(1):117-136.

64. Wende H, Colonna M, Ziegler A, Volz A: Organization of the leukocyte receptor cluster (LRC) on human chromosome 19q13.4. *Mammalian genome : official journal of the International Mammalian Genome Society* 1999, 10(2):154-160.

65. Barrow AD, Trowsdale J: The extended human leukocyte receptor complex: diverse ways of modulating immune responses. *Immunological reviews* 2008, 224:98-123.

66. Guethlein LA, Abi-Rached L, Hammond JA, Parham P: The expanded cattle KIR genes are orthologous to the conserved single-copy KIR3DX1 gene of primates. *Immunogenetics* 2007, 59(6):517-522.

67. Sanderson ND, Norman PJ, Guethlein LA, Ellis SA, Williams C, Breen M, Park SD, Magee DA, Babrzadeh F, Warry A *et al*: Definition of the cattle killer cell Ig-like receptor gene family: comparison with aurochs and human counterparts. *Journal of immunology (Baltimore, Md : 1950)* 2014, 193(12):6016-6030.

68. Moradi S, Berry R, Pymm P, Hitchen C, Beckham SA, Wilce MC, Walpole NG, Clements CS, Reid HH, Perugini MA *et al*: The structure of the atypical killer cell immunoglobulin-like receptor, KIR2DL4. *The Journal of biological chemistry* 2015, 290(16):10460-10471.

69. Rajagopalan S, Long EO: KIR2DL4 (CD158d): An activation receptor for HLA-G. *Front Immunol* 2012, 3:258-258.

70. Faure M, Long EO: KIR2DL4 (CD158d), an NK Cell-Activating Receptor with Inhibitory Potential. *The Journal of Immunology* 2002, 168(12):6208.

71. Trundley AE, Hiby SE, Chang C, Sharkey AM, Santourlidis S, Uhrberg M, Trowsdale J, Moffett A: Molecular characterization of KIR3DL3. *Immunogenetics* 2006, 57(12):904-916.

72. Trompeter HI, Gomez-Lozano N, Santourlidis S, Eisermann B, Wernet P, Vilches C, Uhrberg M: Three structurally and functionally divergent kinds of promoters regulate expression of clonally distributed killer cell Ig-like receptors (KIR), of KIR2DL4, and of KIR3DL3. *Journal of immunology (Baltimore, Md : 1950)* 2005, 174(7):4135-4143.

73. Leaton LA, Shortt J, Kichula KM, Tao S, Nemat-Gorgani N, Mentzer AJ, Oppenheimer SJ, Deng Z, Hollenbach JA, Gignoux CR *et al*: Conservation, Extensive Heterozygosity, and Convergence of Signaling Potential All Indicate a Critical Role for KIR3DL3 in Higher Primates. *Front Immunol* 2019, 10:24.

74. Nutalai R, Gaudieri S, Jumnainsong A, Leelayuwat C: Regulation of KIR3DL3 Expression via Mirna. *Genes* 2019, 10(8).

75. Hollenbach JA, Nocedal I, Ladner MB, Single RM, Trachtenberg EA: Killer cell immunoglobulin-like receptor (KIR) gene content variation in the HGDP-CEPH populations. *Immunogenetics* 2012, 64(10):719-737.

76. Hilton HG, Guethlein LA, Goyos A, Nemat-Gorgani N, Bushnell DA, Norman PJ, Parham P: Polymorphic HLA-C Receptors Balance the Functional Characteristics of KIR Haplotypes. *Journal of immunology (Baltimore, Md : 1950)* 2015, 195(7):3160-3170.

77. Rajagopalan S, Long EO: KIR2DL4 (CD158d): An activation receptor for HLA-G. *Front Immunol* 2012, 3:258.

78. Garcia-Beltran WF, Hölzemer A, Martrus G, Chung AW, Pacheco Y, Simoneau CR, Rucevic M, Lamothe-Molina PA, Pertel T, Kim T-E *et al*: Open conformers of HLA-F are high-affinity ligands of the activating NK-cell receptor KIR3DS1. *Nature immunology* 2016, 17(9):1067-1074.

79. Graef T, Moesta AK, Norman PJ, Abi-Rached L, Vago L, Older Aguilar AM, Gleimer M, Hammond JA, Guethlein LA, Bushnell DA *et al*: KIR2DS4 is a product of gene conversion with KIR3DL2 that introduced specificity for HLA-A*11 while diminishing avidity for HLA-C. *The Journal of experimental medicine* 2009, 206(11):2557-2572.

80. Della Chiesa M, Romeo E, Falco M, Balsamo M, Augugliaro R, Moretta L, Bottino C, Moretta A, Vitale M: Evidence that the KIR2DS5 gene codes for a surface receptor triggering natural killer cell function. *Eur J Immunol* 2008, 38(8):2284-2289.

81. Katz G, Markel G, Mizrahi S, Arnon TI, Mandelboim O: Recognition of HLA-Cw4 but not HLA-Cw6 by the NK cell receptor killer cell Ig-like receptor two-domain short tail number 4. *Journal of immunology (Baltimore, Md : 1950)* 2001, 166(12):7260-7267.

82. Liu J, Xiao Z, Ko HL, Shen M, Ren EC: Activating killer cell immunoglobulin-like receptor 2DS2 binds to HLA-A*11. *Proceedings of the National Academy of Sciences of the United States of America* 2014, 111(7):2662-2667.

83. Colonna M, Spies T, Strominger JL, Ciccone E, Moretta A, Moretta L, Pende D, Viale O: Alloantigen recognition by two human natural killer cell clones is associated with HLA-C or a closely linked gene. *Proceedings of the National Academy of Sciences of the United States of America* 1992, 89(17):7983-7985.

84. Moesta AK, Norman PJ, Yawata M, Yawata N, Gleimer M, Parham P: Synergistic polymorphism at two positions distal to the ligand-binding site makes KIR2DL2 a stronger receptor for HLA-C than KIR2DL3. *Journal of immunology (Baltimore, Md : 1950)* 2008, 180(6):3969-3979.

85. Parham P: MHC class I molecules and kirs in human history, health and survival. *Nature Reviews Immunology* 2005, 5(3):201-214.

86. Winter CC, Gumperz JE, Parham P, Long EO, Wagtmann N: Direct binding and functional transfer of NK cell inhibitory receptors reveal novel patterns of HLA-C allotype recognition. *Journal of immunology (Baltimore, Md : 1950)* 1998, 161(2):571-577.

87. Wagtmann N, Rajagopalan S, Winter CC, Peruzzi M, Long EO: Killer cell inhibitory receptors specific for HLA-C and HLA-B identified by direct binding and by functional transfer. *Immunity* 1995, 3(6):801-809.

88. Vilches C, Rajalingam R, Uhrberg M, Gardiner CM, Young NT, Parham P: KIR2DL5, a novel killer-cell receptor with a D0-D2 configuration of Ig-like domains. *Journal of immunology (Baltimore, Md : 1950)* 2000, 164(11):5797-5804.

89. Hilton HG, Blokhuis JH, Guethlein LA, Norman PJ, Parham P: Resurrecting KIR2DP1: A Key Intermediate in the Evolution of Human Inhibitory NK Cell Receptors That Recognize HLA-C. *Journal of immunology (Baltimore, Md : 1950)* 2017, 198(5):1961-1973.

90. Bono M, Pende D, Bertaina A, Moretta A, Della Chiesa M, Sivori S, Zecca M, Locatelli F, Moretta L, Bottino C *et al*: Analysis of KIR3DP1 Polymorphism Provides Relevant Information on Centromeric KIR Gene Content. *The Journal of Immunology* 2018:ji1800564.

91. Shaw J, Kollnberger S: New perspectives on the ligands and function of the killer cell immunoglobulin-like receptor KIR3DL2 in health and disease. *Front Immunol* 2012, 3(339).

92. Chapel A, Garcia-Beltran WF, Hölzemer A, Ziegler M, Lunemann S, Martrus G, Altfeld M: Peptide-specific engagement of the activating NK cell receptor KIR2DS1. *Scientific Reports* 2017, 7(1):2414.

93. Parham P, Norman PJ, Abi-Rached L, Guethlein LA: Human-specific evolution of killer cell immunoglobulin-like receptor recognition of major histocompatibility complex class I molecules. *Philosophical transactions of the Royal Society of London Series B, Biological sciences* 2012, 367(1590):800-811.

94. Older Aguilar AM, Guethlein LA, Adams EJ, Abi-Rached L, Moesta AK, Parham P: Coevolution of killer cell Ig-like receptors with HLA-C to become the major variable regulators of human NK cells. *Journal of immunology (Baltimore, Md : 1950)* 2010, 185(7):4238-4251.

95. Hilton HG, Norman PJ, Nemat-Gorgani N, Goyos A, Hollenbach JA, Henn BM, Gignoux CR, Guethlein LA, Parham P: Loss and Gain of Natural Killer Cell Receptor Function in an African Hunter-Gatherer Population. *PLoS genetics* 2015, 11(8):e1005439.

96. Cassidy SA, Cheent KS, Khakoo SI: Effects of Peptide on NK cell-mediated MHC I recognition. *Front Immunol* 2014, 5:133.

97. Cassidy S, Mukherjee S, Myint TM, Mbiribindi B, North H, Traherne J, Mulder A, Claas FH, Purbhoo MA, Das J *et al*: Peptide selectivity discriminates NK cells from KIR2DL2- and KIR2DL3-positive individuals. *European journal of immunology* 2015, 45(2):492-500.

98. Brusilovsky M, Cordoba M, Rosental B, Hershkovitz O, Andrake MD, Pecherskaya A, Einarson MB, Zhou Y, Braiman A, Campbell KS *et al*: Genome-wide siRNA screen reveals a new cellular partner of NK cell receptor KIR2DL4: heparan sulfate directly modulates KIR2DL4-mediated responses. *Journal of immunology (Baltimore, Md : 1950)* 2013, 191(10):5256-5267.

99. Mele D, Pasi A, Cacciatore R, Mantovani S, Oliviero B, Mondelli MU, Varchetta S: Decreased interferon-γ production by NK cells from KIR haplotype B carriers in hepatitis C virus infection. *Liver International* 2019, 39(7):1237-1245.

100. Wauquier N, Padilla C, Becquart P, Leroy E, Vieillard V: Association of KIR2DS1 and KIR2DS3 with fatal outcome in Ebola virus infection. *Immunogenetics* 2010, 62(11):767-771.

101. Aranda-Romo S, Garcia-Sepulveda CA, Comas-García A, Lovato-Salas F, Salgado-Bustamante M, Gómez-Gómez A, Noyola DE: Killer-cell immunoglobulin-like receptors (KIR) in severe A (H1N1) 2009 influenza infections. *Immunogenetics* 2012, 64(9):653-662.

102. Martin MP, Qi Y, Gao X, Yamada E, Martin JN, Pereyra F, Colombo S, Brown EE, Shupert WL, Phair J *et al*: Innate partnership of HLA-B and KIR3DL1 subtypes against HIV-1. *Nature genetics* 2007, 39(6):733-740.

103. Pelak K, Need AC, Fellay J, Shianna KV, Feng S, Urban TJ, Ge D, De Luca A, Martinez-Picado J, Wolinsky SM *et al*: Copy number variation of KIR genes influences HIV-1 control. *PLoS Biol* 2011, 9(11):e1001208-e1001208.

104. Chaisri S, Pabalan N, Tabunhan S, Tharabenjasin P, Sankuntaw N, Leelayuwat C: Effects of the killer immunoglobulin-like receptor (KIR) polymorphisms on HIV acquisition: A meta-analysis. *PloS one* 2019, 14(12):e0225151.

105. Leone P, De Re V, Vacca A, Dammacco F, Racanelli V: Cancer treatment and the KIR-HLA system: an overview. *Clinical and experimental medicine* 2017, 17(4):419-429.

106. He Y, Bunn PA, Zhou C, Chan D: KIR 2D (L1, L3, L4, S4) and KIR 3DL1 protein expression in non-small cell lung cancer. *Oncotarget* 2016, 7(50):82104-82111.

107. Vineretsky KA, Karagas MR, Christensen BC, Kuriger-Laber JK, Perry AE, Storm CA, Nelson HH: Skin Cancer Risk Is Modified by KIR/HLA Interactions That Influence the Activation of Natural Killer Immune Cells. *Cancer research* 2016, 76(2):370-376.

108. Ewen EM, Pahl JHW, Miller M, Watzl C, Cerwenka A: KIR downregulation by IL-12/15/18 unleashes human NK cells from KIR/HLA-I inhibition and enhances killing of tumor cells. *Eur J Immunol* 2018, 48(2):355-365.

109. Mahaweni NM, Ehlers FAI, Bos GMJ, Wieten L: Tuning Natural Killer Cell Anti-multiple Myeloma Reactivity by Targeting Inhibitory Signaling via KIR and NKG2A. *Front Immunol* 2018, 9:2848.

110. Verheyden S, Bernier M, Demanet C: Identification of natural killer cell receptor phenotypes associated with leukemia. *Leukemia* 2004, 18(12):2002-2007.

111. Zhang Y, Wang B, Ye S, Liu S, Liu M, Shen C, Teng Y, Qi J: Killer cell immunoglobulin-like receptor gene polymorphisms in patients with leukemia: possible association with susceptibility to the disease. *Leukemia research* 2010, 34(1):55-58.

112. Middleton D, Diler AS, Meenagh A, Sleator C, Gourraud PA: Killer immunoglobulin-like receptors (KIR2DL2 and/or KIR2DS2) in presence of their ligand (HLA-C1 group) protect against chronic myeloid leukaemia. *Tissue antigens* 2009, 73(6):553-560.

113. Augusto DG: The Impact of KIR Polymorphism on the Risk of Developing Cancer: Not as Strong as Imagined? *Frontiers in Genetics* 2016, 7(121).

114. Niepieklo-Miniewska W, Majorczyk E, Matusiak L, Gendzekhadze K, Nowak I, Narbutt J, Lesiak A, Kuna P, Poninska J, Pietkiewicz-Sworowska A *et al*: Protective effect of the KIR2DS1 gene in atopic dermatitis. *Gene* 2013, 527(2):594-600.

115. Saito H, Hirayama A, Umemura T, Joshita S, Mukawa K, Suga T, Tanaka E, Ota M: Association between KIR-HLA combination and ulcerative colitis and Crohn's disease in a Japanese population. *PloS one* 2018, 13(4):e0195778-e0195778.

116. Liang H-L, Ma S-J, Tan H-Z: Association between killer cell immunoglobulin-like receptor (KIR) polymorphisms and systemic lupus erythematosus (SLE) in populations: A PRISMA-compliant meta-analysis. *Medicine (Baltimore)* 2017, 96(10):e6166-e6166.

117. Mohammad-Ebrahim H, Kamali-Sarvestani E, Mahmoudi M, Beigy M, Karami J, Ahmadzadeh N, Shahram F: Association of killer cell immunoglobulin-like receptor (KIR) genes and their HLA ligands with susceptibility to Behcet's disease. *Scandinavian journal of rheumatology* 2018, 47(2):155-163.

118. van Duin D, Avery RK, Hemachandra S, Yen-Lieberman B, Zhang A, Jain A, Butler RS, Barnard J, Schold JD, Fung J *et al*: KIR and HLA interactions are associated with control of primary CMV infection in solid organ transplant recipients. *American journal of transplantation : official journal of the American Society of Transplantation and the American Society of Transplant Surgeons* 2014, 14(1):156-162.

119. Littera R, Piredda G, Argiolas D, Lai S, Congeddu E, Ragatzu P, Melis M, Carta E, Michittu MB, Valentini D *et al*: KIR and their HLA Class I ligands: Two more pieces towards completing the puzzle of chronic rejection and graft loss in kidney transplantation. *PloS one* 2017, 12(7):e0180831.

120. Rajalingam R: The Impact of HLA Class I-Specific Killer Cell Immunoglobulin-Like Receptors on Antibody-Dependent Natural Killer Cell-Mediated Cytotoxicity and Organ Allograft Rejection. *Front Immunol* 2016, 7(585).

121. Neuchel C, Fürst D, Niederwieser D, Bunjes D, Tsamadou C, Wulf G, Pfreundschuh M, Wagner E, Stuhler G, Einsele H *et al*: Impact of Donor Activating KIR Genes on HSCT Outcome in C1-Ligand Negative Myeloid Disease Patients Transplanted with Unrelated Donors-A Retrospective Study. *PloS one* 2017, 12(1):e0169512-e0169512.

122. Shaffer BC, Hsu KC: How important is NK alloreactivity and KIR in allogeneic transplantation? *Best Pract Res Clin Haematol* 2016, 29(4):351-358.

123. Hiby SE, Apps R, Sharkey AM, Farrell LE, Gardner L, Mulder A, Claas FH, Walker JJ, Redman CW, Morgan L *et al*: Maternal activating KIRs protect against human reproductive failure mediated by fetal HLA-C2. *The Journal of clinical investigation* 2010, 120(11):4102-4110.

124. Hiby SE, Walker JJ, O'Shaughnessy K M, 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. *The Journal of experimental medicine* 2004, 200(8):957-965.

125. Hiby SE, Regan L, Lo W, Farrell L, Carrington M, Moffett A: Association of maternal killer-cell immunoglobulin-like receptors and parental HLA-C genotypes with recurrent miscarriage. *Human reproduction (Oxford, England)* 2008, 23(4):972-976.

126. Chazara O, Xiong S, Moffett A: Maternal KIR and fetal HLA-C: a fine balance. *Journal of leukocyte biology* 2011, 90(4):703-716.

127. Xiong S, Sharkey AM, Kennedy PR, Gardner L, Farrell LE, Chazara O, Bauer J, Hiby SE, Colucci F, Moffett A: Maternal uterine NK cell-activating receptor KIR2DS1 enhances placentation. *The Journal of clinical investigation* 2013, 123(10):4264-4272.

128. Kennedy PR, Chazara O, Gardner L, Ivarsson MA, Farrell LE, Xiong S, Hiby SE, Colucci F, Sharkey AM, Moffett A: Activating KIR2DS4 Is Expressed by Uterine NK Cells and Contributes to Successful Pregnancy. *The Journal of Immunology* 2016, 197(11):4292.

129. Wang S, Zhao YR, Jiao YL, Wang LC, Li JF, Cui B, Xu CY, Shi YH, Chen ZJ: Increased activating killer immunoglobulin-like receptor genes and decreased specific HLA-C alleles in couples with recurrent spontaneous abortion. *Biochemical and biophysical research communications* 2007, 360(3):696-701.

130. Faridi RM, Agrawal S: Killer immunoglobulin-like receptors (KIRs) and HLA-C allorecognition patterns implicative of dominant activation of natural killer cells contribute to recurrent miscarriages. *Human reproduction (Oxford, England)* 2011, 26(2):491-497.

131. Witt CS, Goodridge J, Gerbase-Delima MG, Daher S, Christiansen FT: Maternal KIR repertoire is not associated with recurrent spontaneous abortion. *Human reproduction (Oxford, England)* 2004, 19(11):2653-2657.

132. Cooper MA, Elliott JM, Keyel PA, Yang L, Carrero JA, Yokoyama WM: Cytokine-induced memory-like natural killer cells. *Proceedings of the National Academy of Sciences of the United States of America* 2009, 106(6):1915-1919.

133. Sungur CM, Murphy WJ: Utilization of mouse models to decipher natural killer cell biology and potential clinical applications. *Hematology American Society of Hematology Education Program* 2013, 2013:227-233.

134. Carlyle JR, Mesci A, Fine JH, Chen P, Belanger S, Tai LH, Makrigiannis AP: Evolution of the Ly49 and Nkrp1 recognition systems. *Seminars in immunology* 2008, 20(6):321-330.

135. Hanke T, Takizawa H, McMahon CW, Busch DH, Pamer EG, Miller JD, Altman JD, Liu Y, Cado D, Lemonnier FA *et al*: Direct assessment of MHC class I binding by seven Ly49 inhibitory NK cell receptors. *Immunity* 1999, 11(1):67-77.

136. Eagle RA, Trowsdale J: Promiscuity and the single receptor: NKG2D. *Nature reviews Immunology* 2007, 7(9):737-744.

137. Herndler-Brandstetter D, Shan L, Yao Y, Stecher C, Plajer V, Lietzenmayer M, Strowig T, de Zoete MR, Palm NW, Chen J *et al*: Humanized mouse model supports development, function, and tissue residency of human natural killer cells. *Proceedings of the National Academy of Sciences of the United States of America* 2017, 114(45):E9626-e9634.

138. Douam F, Ziegler CGK, Hrebikova G, Fant B, Leach R, Parsons L, Wang W, Gaska JM, Winer BY, Heller B *et al*: Selective expansion of myeloid and NK cells in humanized mice yields human-like vaccine responses. *Nature Communications* 2018, 9(1):5031.

139. Dudley DM, Aliota MT, Mohr EL, Weiler AM, Lehrer-Brey G, Weisgrau KL, Mohns MS, Breitbach ME, Rasheed MN, Newman CM *et al*: A rhesus macaque model of Asian-lineage Zika virus infection. *Nature Communications* 2016, 7(1):12204.

140. Dijkman K, Sombroek CC, Vervenne RAW, Hofman SO, Boot C, Remarque EJ, Kocken CHM, Ottenhoff THM, Kondova I, Khayum MA *et al*: Prevention of tuberculosis infection and disease by local BCG in repeatedly exposed rhesus macaques. *Nature Medicine* 2019, 25(2):255-262.

141. Hatziioannou T, Ambrose Z, Chung NPY, Piatak M, Jr., Yuan F, Trubey CM, Coalter V, Kiser R, Schneider D, Smedley J *et al*: A macaque model of HIV-1 infection. *Proceedings of the National Academy of Sciences of the United States of America* 2009, 106(11):4425-4429.

142. Bjornson-Hooper ZB, Fragiadakis GK, Spitzer MH, Madhireddy D, McIlwain D, Nolan GP: A comprehensive atlas of immunological differences between humans, mice and non-human primates. *bioRxiv* 2019:574160.

143. Hong H, Rajakumar P, Billingsley J, Reeves RK, Johnson R: No monkey business: why studying NK cells in non-human primates pays off. *Front Immunol* 2013, 4(32).

144. Reeves RK, Gillis J, Wong FE, Yu Y, Connole M, Johnson RP: CD16- natural killer cells: enrichment in mucosal and secondary lymphoid tissues and altered function during chronic SIV infection. *Blood* 2010, 115(22):4439-4446.

145. Okoye AA, DeGottardi MQ, Fukazawa Y, Vaidya M, Abana CO, Konfe AL, Fachko DN, Duell DM, Li H, Lum R *et al*: Role of IL-15 Signaling in the Pathogenesis of Simian Immunodeficiency Virus Infection in Rhesus Macaques. *The Journal of Immunology* 2019:ji1900792.

146. Vargas-Inchaustegui DA, Helmold Hait S, Chung HK, Narola J, Hoang T, Robert-Guroff M: Phenotypic and Functional Characterization of Circulatory, Splenic, and Hepatic NK Cells in Simian Immunodeficiency Virus–Controlling Macaques. *The Journal of Immunology* 2017, 199(9):3202.

147. Adams EJ, Parham P: Species-specific evolution of MHC class I genes in the higher primates. *Immunological reviews* 2001, 183:41-64.

148. Valiante NM, Uhrberg M, Shilling HG, Lienert-Weidenbach K, Arnett KL, D'Andrea A, Phillips JH, Lanier LL, Parham P: Functionally and structurally distinct NK cell receptor repertoires in the peripheral blood of two human donors. *Immunity* 1997, 7(6):739-751.

149. Guethlein LA, Older Aguilar AM, Abi-Rached L, Parham P: Evolution of killer cell Ig-like receptor (KIR) genes: definition of an orangutan KIR haplotype reveals expansion of lineage III KIR associated with the emergence of MHC-C. *Journal of immunology (Baltimore, Md : 1950)* 2007, 179(1):491-504.

150. Moesta AK, Abi-Rached L, Norman PJ, Parham P: Chimpanzees use more varied receptors and ligands than humans for inhibitory killer cell Ig-like receptor recognition of the MHC-C1 and MHC-C2 epitopes. *Journal of immunology (Baltimore, Md : 1950)* 2009, 182(6):3628-3637.

151. Abi-Rached L, Moesta AK, Rajalingam R, Guethlein LA, Parham P: Human-specific evolution and adaptation led to major qualitative differences in the variable receptors of human and chimpanzee natural killer cells. *PLoS genetics* 2010, 6(11):e1001192.

152. 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.

153. Leaton LA, Shortt J, Kichula KM, Tao S, Nemat-Gorgani N, Mentzer AJ, Oppenheimer SJ, Deng Z, Hollenbach JA, Gignoux CR *et al*: Conservation, Extensive Heterozygosity, and Convergence of Signaling Potential All Indicate a Critical Role for KIR3DL3 in Higher Primates. *Front Immunol* 2019, 10(24).

154. Hershberger KL, Shyam R, Miura A, Letvin NL: Diversity of the killer cell Ig-like receptors of rhesus monkeys. *Journal of immunology (Baltimore, Md : 1950)* 2001, 166(7):4380-4390.

155. Boyson JE, Shufflebotham C, Cadavid LF, Urvater JA, Knapp LA, Hughes AL, Watkins DI: The MHC class I genes of the rhesus monkey. Different evolutionary histories of MHC class I and II genes in primates. *Journal of immunology (Baltimore, Md : 1950)* 1996, 156(12):4656-4665.

156. Doxiadis GGM, de Groot N, Otting N, Blokhuis JH, Bontrop RE: Genomic plasticity of the MHC class I A region in rhesus macaques: extensive haplotype diversity at the population level as revealed by microsatellites. *Immunogenetics* 2011, 63(2):73-83.

157. Otting N, Heijmans CM, Noort RC, de Groot NG, Doxiadis GG, van Rood JJ, Watkins DI, Bontrop RE: Unparalleled complexity of the MHC class I region in rhesus macaques. *Proceedings of the National Academy of Sciences of the United States of America* 2005, 102(5):1626-1631.

158. 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(3):254-258.

159. Sambrook JG, Bashirova A, Palmer S, Sims S, Trowsdale J, Abi-Rached L, Parham P, Carrington M, Beck S: Single haplotype analysis demonstrates rapid evolution of the killer immunoglobulin-like receptor (KIR) loci in primates. *Genome research* 2005, 15(1):25-35.

160. Rosner C, Kruse PH, Hermes M, Otto N, Walter L: Rhesus macaque inhibitory and activating KIR3D interact with Mamu-A-encoded ligands. *Journal of immunology (Baltimore, Md : 1950)* 2011, 186(4):2156-2163.

161. Schafer JL, Colantonio AD, Neidermyer WJ, Dudley DM, Connole M, O'Connor DH, Evans DT: KIR3DL01 recognition of Bw4 ligands in the rhesus macaque: maintenance of Bw4 specificity since the divergence of apes and Old World monkeys. *Journal of immunology (Baltimore, Md : 1950)* 2014, 192(4):1907-1917.

162. Colantonio AD, Bimber BN, Neidermyer WJ, Jr., Reeves RK, Alter G, Altfeld M, Johnson RP, Carrington M, O'Connor DH, Evans DT: KIR polymorphisms modulate peptide-dependent binding to an MHC class I ligand with a Bw6 motif. *PLoS Pathog* 2011, 7(3):e1001316.

163. Banerjee P, Ries M, Janaka SK, Grandea AG, Wiseman R, O'Connor DH, Golos TG, Evans DT: Diversification of Bw4 Specificity and Recognition of a Nonclassical MHC Class I Molecule Implicated in Maternal-Fetal Tolerance by Killer Cell Ig-like Receptors of the Rhesus Macaque. *Journal of immunology (Baltimore, Md : 1950)* 2018, 201(9):2776-2786.

164. Bimber BN, Evans DT: The killer-cell immunoglobulin-like receptors of macaques. *Immunological reviews* 2015, 267(1):246-258.

165. de Groot NG, Blokhuis JH, Otting N, Doxiadis GG, Bontrop RE: Co-evolution of the MHC class I and KIR gene families in rhesus macaques: ancestry and plasticity. *Immunological reviews* 2015, 267(1):228-245.

166. Blokhuis JH, van der Wiel MK, Doxiadis GG, Bontrop RE: The mosaic of KIR haplotypes in rhesus macaques. *Immunogenetics* 2010, 62(5):295-306.

167. Boyson JE, Iwanaga KK, Golos TG, Watkins DI: Identification of a novel MHC class I gene, Mamu-AG, expressed in the placenta of a primate with an inactivated G locus. *Journal of immunology (Baltimore, Md : 1950)* 1997, 159(7):3311-3321.

168. Graves T, Eichler E.E. and Wilson R.K. (2019) Macaca mulatta isolate AG07107 chromosome 19, whole genome shotgun sequence. Published by NCBI. https://www.ncbi.nlm.nih.gov/nuccore/ CM014354. Accessed 21 November 2019.

169. Blokhuis JH, van der Wiel MK, Doxiadis GG, Bontrop RE: Evidence for balancing selection acting on KIR2DL4 genotypes in rhesus macaques of Indian origin. *Immunogenetics* 2009, 61(7):503-512.

170. Blokhuis JH, van der Wiel MK, Doxiadis GG, Bontrop RE: The extreme plasticity of killer cell Ig-like receptor (KIR) haplotypes differentiates rhesus macaques from humans. *Eur J Immunol* 2011, 41(9):2719-2728.

171. Moreland AJ, Guethlein LA, Reeves RK, Broman KW, Johnson RP, Parham P, O'Connor DH, Bimber BN: Characterization of killer immunoglobulin-like receptor genetics and comprehensive genotyping by pyrosequencing in rhesus macaques. *BMC genomics* 2011, 12:295.

172. Hellmann I, Lim S-Y, Gelman RS, Letvin NL: Association of activating KIR copy number variation of NK cells with containment of SIV replication in rhesus monkeys. *PLoS Pathog* 2011, 7(12):e1002436-e1002436.

173. Hershberger KL, Shyam R, Miura A, Letvin NL: Diversity of the Killer Cell Ig-Like Receptors of Rhesus Monkeys. *The Journal of Immunology* 2001, 166(7):4380.

174. Chaichompoo P, Bostik P, Stephenson S, Udompunturuk S, Kobkitjaroen J, Pattanapanyasat K, Ansari AA: Multiple KIR gene polymorphisms are associated with plasma viral loads in SIV-infected rhesus macaques. *Cellular immunology* 2010, 263(2):176-187.

175. Kruse PH, Rosner C, Walter L: Characterization of rhesus macaque KIR genotypes and haplotypes. *Immunogenetics* 2010, 62(5):281-293.

176. Bimber BN, Moreland AJ, Wiseman RW, Hughes AL, O'Connor DH: Complete characterization of killer Ig-like receptor (KIR) haplotypes in Mauritian cynomolgus macaques: novel insights into nonhuman primate KIR gene content and organization. *Journal of immunology (Baltimore, Md : 1950)* 2008, 181(9):6301-6308.

177. Prall TM, Graham ME, Karl JA, Wiseman RW, Ericsen AJ, Raveendran M, Alan Harris R, Muzny DM, Gibbs RA, Rogers J *et al*: Improved full-length killer cell immunoglobulin-like receptor transcript discovery in Mauritian cynomolgus macaques. *Immunogenetics* 2017, 69(5):325-339.

178. Watson JD, Crick FHC: Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. *Nature* 1953, 171(4356):737-738.

179. Gilbert W, Maxam A: The nucleotide sequence of the lac operator. *Proceedings of the National Academy of Sciences of the United States of America* 1973, 70(12):3581-3584.

180. Maxam AM, Gilbert W: A new method for sequencing DNA. *Proceedings of the National Academy of Sciences of the United States of America* 1977, 74(2):560-564.

181. Sanger F, Nicklen S, Coulson AR: DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences of the United States of America* 1977, 74(12):5463-5467.

182. Smith LM, Sanders JZ, Kaiser RJ, Hughes P, Dodd C, Connell CR, Heiner C, Kent SB, Hood LE: Fluorescence detection in automated DNA sequence analysis. *Nature* 1986, 321(6071):674-679.

183. Finishing the euchromatic sequence of the human genome. *Nature* 2004, 431(7011):931-945.

184. Eid J, Fehr A, Gray J, Luong K, Lyle J, Otto G, Peluso P, Rank D, Baybayan P, Bettman B *et al*: Real-time DNA sequencing from single polymerase molecules. *Science (New York, NY)* 2009, 323(5910):133-138.

185. Levene MJ, Korlach J, Turner SW, Foquet M, Craighead HG, Webb WW: Zero-mode waveguides for single-molecule analysis at high concentrations. *Science (New York, NY)* 2003, 299(5607):682-686.

186. Church G, Deamer David W, Branton D, Baldarelli R, Kasianowicz J: Characterization Of Individual Polymer Molecules Based On Monomer-interface Interactions. In. EP: Hardvard College Univ California; 2008.

187. Laszlo AH, Derrington IM, Ross BC, Brinkerhoff H, Adey A, Nova IC, Craig JM, Langford KW, Samson JM, Daza R *et al*: Decoding long nanopore sequencing reads of natural DNA. *Nature biotechnology* 2014, 32(8):829-833.

188. Flusberg BA, Webster DR, Lee JH, Travers KJ, Olivares EC, Clark TA, Korlach J, Turner SW: Direct detection of DNA methylation during single-molecule, real-time sequencing. *Nature methods* 2010, 7(6):461-465.

189. Ni P, Huang N, Zhang Z, Wang D-P, Liang F, Miao Y, Xiao C-L, Luo F, Wang J: DeepSignal: detecting DNA methylation state from Nanopore sequencing reads using deep-learning. *Bioinformatics* 2019.

190. Kraft F, Kurth I: Long-read sequencing in human genetics. *medizinische genetik* 2019, 31(2):198-204.

191. Joag SV, Stephens EB, Adams RJ, Foresman L, Narayan O: Pathogenesis of SIVmac Infection in Chinese and Indian Rhesus Macaques: Effects of Splenectomy on Virus Burden. *Virology* 1994, 200(2):436-446.

192. Zhou Y, Bao R, Haigwood NL, Persidsky Y, Ho W-z: SIV infection of rhesus macaques of Chinese origin: a suitable model for HIV infection in humans. *Retrovirology* 2013, 10(1):89.

Chapter 2

Human and rhesus macaque KIR haplotypes defined by their transcriptomes

Jesse Bruijnesteijn[1], Marit K. H. van der Wiel[1], Wendy T. N. Swelsen[2], Nel Otting[1], Annemiek J. M. de Vos-Rouweler[1], Diënne Elferink[3], Gaby G. Doxiadis[1], Frans H. J. Claas[3], Neubury M. Lardy[2], Natasja G. de Groot[1] and Ronald E. Bontrop[1,4,*]

[1]Comparative Genetics and Refinement, Biomedical Primate Research Centre, 2288 GJ Rijswijk, The Netherlands
[2]Sanquin, Dept. of Immunogenetics, 1066 CX Amsterdam, The Netherlands
[3]LUMC, Immunohematology and Blood Transfusion, 2300 RC Leiden, The Netherlands
[4]Theoretical Biology and Bioinformatics, Utrecht University, 3584 CH Utrecht, The Netherlands

*Corresponding author:
Prof. Dr. Ronald E. Bontrop
bontrop@bprc.nl
Tel. nr. +31-15-2842699
Fax nr. +31-15-2842600

Abstract

The killer-cell immunoglobulin-like receptors (KIR) play a central role in the immune recognition in infection, pregnancy and transplantation through their interactions with MHC class I molecules. *KIR* genes display abundant copy number variation as well as high levels of polymorphism. As a result, it is challenging to characterize this structurally dynamic region. *KIR* haplotypes have been analysed in different species using conventional characterization methods, such as Sanger sequencing and Roche/454 pyrosequencing. However, these methods are time-consuming and often failed to define complete haplotypes, or do not reach allele-level resolution. In addition, most analyses were performed on genomic DNA, and thus were lacking substantial information about transcription and its corresponding modifications. In this communication, we present a Single-Molecule, Real-Time (SMRT) sequencing approach, using Pacific Bioscience's (PacBio) Sequel platform to characterize the KIR transcriptomes in human and rhesus macaque (*Macaca mulatta*) families. This high-resolution approach allowed the identification of novel *Mamu-KIR* alleles, the extension of reported allele sequences, and the determination of human and macaque *KIR* haplotypes. In addition, multiple recombinant *KIR* genes were discovered, all located on contracted haplotypes, which were likely the result of chromosomal rearrangements. The relatively high number of contracted haplotypes discovered might be indicative of selection on small KIR repertoires and/or novel fusion gene products. This next-generation method provides an improved high-resolution characterization of the *KIR* cluster in humans and macaques, which eventually may aid in a better understanding and interpretation of *KIR* allele associated diseases, as well as the immune response in transplantation and reproduction.

Introduction

Killer-cell Immunoglobulin-like Receptors (KIR) are expressed on natural killer (NK) cells and subsets of T cells (1, 2), and play a key role in immune recognition through interactions with the highly polymorphic major histocompatibility complex (MHC) class I molecules (3, 4). For example, KIR may play an important role in the detection of aberrant MHC class I expression on tumor and virally infected cells, and their subsequent elimination (5, 6). KIR are type I transmembrane glycoproteins that consist of two or three extracellular Ig-like domains (2D or 3D) as well as a stem, transmembrane region and cytoplasmic tail. The length of the cytoplasmic tail can be either long (L), including two ITIM motifs, or short (S), and characterizes inhibitory or activating KIR, respectively. KIR with one extracellular Ig-like domain and a truncated cytoplasmic tail (KIR1D) are observed in some non-human primate species, and seem to have no counterpart in humans (7, 8). The nomenclature of the *KIR* genes is based on the functional (L, S) and structural characteristics (2D or 3D), and takes into account allelic variation as well (9).

In humans, the *KIR* gene cluster is located within the leukocyte receptor complex (LRC) on chromosome 19q13.4, and displays copy number variation (CNV) at the population level, as reflected by a variable number of tandemly arranged *KIR* genes (10, 11). A *KIR* haplotype contains seven to twelve expressed genes, three of which are considered framework genes: *KIR3DL3*, *KIR3DL2*, and *KIR2DL4* (12). A fourth framework gene is *KIR3DP1*, which is a pseudo-gene. Based on the genetic make up, the human *KIR* haplotypes can be categorized into two groups (13). Group A haplotypes are characterized by seven *KIR* genes, including the framework genes and only the activating *KIR2DS4* structure, whereas group B haplotypes can contain up to twelve genes including the framework genes and multiple activating receptors. The *KIR* genes can be further divided into phylogenetic lineages (I, II, III and V), each characterized by structure and MHC class I specificity; lineage I includes *KIR2DL4* and *KIR2DL5*, lineage II includes *KIR3DL1/S1* and *KIR3DL2*, lineage III includes *KIR2DS1/2*, *KIR2DL1/2/3*, *KIR2DS3/5*, *KIR2DS4* and the pseudo-genes, and lineage V includes *KIR3DL3*.

In addition to the CNV, allelic polymorphism is another important feature of the *KIR* gene system. In humans, the greatest expansion is observed for *KIR* lineage III genes in both the telomeric and centromeric region of the haplotype. A total of fifteen human *KIR* genes and two pseudo-genes have been characterized, and up to 907 unique full-length *KIR* alleles have been catalogued (14).

As various KIR may bind specific but differential structures on MHC class I molecules, the KIR repertoire influences, in part, the variability of the immune response. Since both the *MHC* and *KIR* gene systems display substantial levels of polymorphism and segregate as independent entities located on different chromosomes, the potential repertoire of MHC-KIR interactions may vary considerably, even among related individuals within a family. The presence or absence of certain *KIR* alleles and MHC-KIR interactions are associated with disease susceptibility and its progress, but may also play a role in transplantation and reproductive biology (5, 6, 15, 16).

In recent years, our understanding of the biology and evolution of the *KIR* gene system has greatly expanded, although some key questions remain to be answered. Suitable animal models to study KIR-related diseases are more or less confined to non-human primate species, since rodent species have another system executing similar tasks that arose as result of convergent evolution (17, 18). Macaques, for example, share a close evolutionary relationship with humans, which is evidenced by similar pathology and immune responses in models for infectious and autoimmune diseases (19-21). Initial genomic characterization of the *KIR* gene repertoire in rhesus macaques (*Macaca mulatta*, *Mamu*) highlighted substantial similarities along with some differences as compared to humans (8, 22-24). For example, the macaque *KIR* gene cluster shows an extreme expansion of lineage II *KIR* genes, mainly in the telomeric part of the haplotype, which might be associated with the multiplicity of the KIR-interacting *MHC-A* and *-B* genes (25, 26). This extensive gene copy number variation exceeds the lineage III *KIR* expansion observed in humans. Thus far, twenty-two *KIR* genes and 218 alleles have been reported in macaques (7, 22, 27-35). *KIR* haplotypes in macaques can contain up to eleven genes, some including *Mamu-KIR2DL04* and

Mamu-KIR1D, which belong to KIR lineages I and III, respectively. The human *KIR* haplotypic division differentiating between the more activating (B) and the more inhibitory (A) haplotypes is not as obvious for macaques (22, 26).

A limited number of macaque haplotypes has been characterized by studying segregation in families in combination with conventional sequencing methods, leading to haplotype definitions that were based on the presence of both partial and full-length cDNA sequences. Although these methods provided insights, they were either not always sufficient to resolve allele-level haplotypes, or to identify genes with low transcription levels. In contrast, most human *KIR* haplotypes were characterized by determining the presence or absence of *KIR* genes at the genomic DNA (gDNA) level. As a consequence, crucial information on the allele level, copy number variation, transcription level, and transcriptional modifications, such as splicing, may be missed. As particular *KIR* alleles are expected to be associated with health and disease, a comprehensive method is required to characterize the complete KIR transcriptome. Here we report a next-generation Single-Molecule, Real-Time (SMRT) sequencing method on the PacBio Sequel platform, which allowed us to obtain full-length KIR transcriptomes, as well as *KIR* haplotypes for both human and rhesus macaque families. This approach provides a significant step forward, which may aid in a better understanding and interpretation of *KIR*-allele associated diseases, as well as the immune reactivity in transplantation and reproductive biology.

Materials & method

Animals and cells

A large pedigree-based Indian rhesus macaque family, with a total of thirty animals, was selected from the self-sustaining colony housed at the Biomedical Primate Research Centre (BPRC) (**Figure 1**). EDTA or heparin whole blood samples were obtained during regular annual health checks, and peripheral blood mononuclear cells (PBMC) were isolated from heparin blood samples. PBMCs of fifteen related humans were provided by the immunohematology and blood transfusion department of the Leiden University Medical Center (**Figure 4**). Informed consent was obtained from all participants.

RNA / gDNA extraction and cDNA synthesis

Total RNA was extracted directly from EDTA whole blood samples or from ±15 x 106 PBMCs with RNeasy Mini Kit (Qiagen, Valencia, CA, USA) in accordance with the manufacturer's instructions. First-strand cDNA was synthesized with the RevertAid First Strand cDNA Synthesis Kit (Invitrogen, Carlsbad, CA, USA) using oligo$(dT)_{18}$ primers. Genomic DNA was extracted from EDTA whole blood samples by a standard salting-out procedure, or from ±15 x 10^6 PBMCs with an AllPrep RNA/DNA easy Mini Kit (Qiagen, Valencia, CA, USA) according to the manufactures instructions.

PacBio Sequel PCR amplification, sample purification and SMRTbell libraries

Polymerase chain reaction (PCR) with different primer sets (**Table I**) was performed on cDNA using Phusion Hot Start II DNA Polymerase (Thermo Fisher Scientific, Waltham, MA, USA) to obtain full-length macaque and human KIR amplicons. Each primer was tagged at the 5' end with a 16 bp barcode, designed for the PacBio platform, to identify samples by unique barcode combinations. Thermal cycling conditions were: denaturation at 98 °C for 2 min, followed by 32 cycles of 98 °C for 20 s, 66 °C for 45 s, and 72 °C for 2 min, except for the primer set derived from Moreland and colleagues (24): denaturation at 98 °C for 2 min; 5 cycles of 98 °C for 20 s, 68 °C for 5 s, 66 °C for 5 s, 63 °C for 30 s, 60 °C for 5 s, and 72 °C for 2 min; 29 cycles of 98 °C for 20 s, 63 °C for 30 s, and 72 °C for 2 min. Appropriately sized PCR products of approximately 1250 bp were selected by gel electrophoresis and purified using a GeneJet Gel extraction kit (Invitrogen, Carlsbad, CA, USA). The amplified KIR amplicons were pooled and purified twice, using AMPure XP beads (Beckman-Coulter, Woerden, The Netherlands) at a 1:1 bead to DNA volume ratio. The DNA concentration of purified pooled samples (> 1 µg total DNA) was measured using the Qubit dsDNA HS assay kit and Qubit 2.0 Fluorometer (Thermo Fisher Scientific, Waltham, MA, USA).

SMRTbell libraries were generated according to the PacBio Amplicon Template Preparation protocol for circular consensus sequences (CCS) (36). In brief, PCR product pools were end-repaired, and hairpin adapters were subsequently incorporated using the PacBio DNA Template Prep Kit 2.0. After the removal of failed ligation products, the SMRTbell templates were purified with 0.6x volume of AMPure XP beads. The volume of sequencing primer and polymerase was determined using the PacBio calculator. Polymerase-bound SMRTbells were MagBead loaded over zero-mode waveguides and immobilized. SMRTbell library generation and sequencing were performed by the Leiden Genome Technology Centre (LGTC) using a PacBio Sequel instrument with P6-C4 sequencing chemistry.

PacBio Sequel data analysis

Circular consensus sequences were obtained and selected for high read quality (rq-value of 0.99 or higher). The data was demultiplexed based on unique barcoding, and were used to type individual samples.

Geneious Pro R10 software (37) was used to map the reads to a reference database, including all reported full-length and partial human or macaque *KIR* cDNA sequences (7, 14, 22, 27-35), to identify 100% matched reads (100% overlap, 0% mismatch, maximum ambiguity = 1). The remaining reads were *De Novo* assembled, and the consensus of each contig was phylogenetically aligned with the human or macaque *KIR* sequence reference database. Official designations of novel sequences were determined by comparison with the reference sequence databases and by phylogenetic analysis, and were provided by the Immuno Polymorphism Database (IPD), which will shortly release a database with non-human primate KIR sequences. The novel sequences were confirmed when observed in two independent PCRs or shown to

segregate into families, and were subsequently submitted to the European Nucleotide Archive (ENA) and assigned an association number (https://www.ebi.ac.uk/ena/data/view/PRJEB22235).

Allele confirmation by Sanger sequencing

Alleles that could not be confirmed by segregation and had low PacBio read counts (< 3 reads), or alleles that seemed to be duplicated on a single haplotype, were confirmed by Sanger sequencing. To amplify products that distinguished the different genes or alleles, specific primers were designed within the exons (**Supp. Table I**). For all primers, the PCR conditions were: denaturation at 98 °C for 2 min, followed by 32 cycles of 98 °C for 20 s, 66 °C for 45 s, and 72 °C for 2 min. PCR products were subjected to gel electrophoresis, and bands of the appropriate size were purified with a GeneJet Gel extraction kit. For *Mamu-KIR3DL07*, *Mamu-KIR3DL05* and *Mamu-KIR3DS02* cloning was performed as previously described (28). Sequencing of the PCR products or isolated cloned amplicons was performed on a 3500XL Genetic Analyzer automatic sequencer (Applied Biosystems, Foster City, USA). Sequences were analyzed with SeqMan Pro (Dnastar, Inc., Madison, USA) and MacVector (MacVector, Inc., Cambridge, UK) software.

Genomic DNA KIR typing

In addition to typing at the transcription level, both human and macaque samples were also typed at the gDNA level for the presence or absence of *KIR* genes. Rhesus macaque DNA samples were typed using quantitative PCR (qPCR) and melt curve analysis as previously described (23). Human *KIR* genotyping was performed by Sanquin (Dept. of Immunogenetics, Netherlands) using the Olerup SSP® KIR typing kit (Olerup SSP AB, Stockholm, Sweden) in accordance with the manufacturer's instructions.

Results

Comparison of the conventional characterization strategies versus the Pacbio Sequel platform

The rhesus macaque *KIR* system was previously characterized by cloning and Sanger sequencing, Roche/454 pyrosequencing, and microsatellite analysis, which resulted in a database of 218 partial and full-length *Mamu-KIR* alleles (22-24, 27, 33, 38). Validation of these sequences was provided by independent amplifications and, whenever possible, segregation studies. However, these characterization strategies were time-consuming, and were often insufficient to resolve full-length allele sequences.

Experience taught that the Pacbio Sequel platform offers a substantial number of improvements compared to conventional sequencing strategies, such as higher throughput and longer reads with high accuracy by circular consensus sequencing (CCS) (39, 40). Taking these advantages into account, we set up a pipeline to characterize the *KIR* gene system using a PacBio Sequel platform. Initially, we calibrated the PacBio

platform by re-analyzing macaque blood samples that had been previously typed for *KIR* by a conventional methodology. With the current PacBio sequencing protocol, we confirmed in considerably less time the rhesus macaque *KIR* results that had been obtained earlier. Furthermore, additional *KIR* alleles and genes were identified, and partial sequences that had been missed by conventional methods could be extended. In the following set-ups, we used a family segregation concept, so that identical sequences could be obtained and confirmed by analyzing different but related individuals.

Full-length KIR allele discovery at the transcription level in rhesus macaques

The previously published macaque *KIR3D/1D* and *KIR2DL04* primers located in conserved regions of the 5' and 3' UTR were barcoded with PacBio sequence tags, and used to amplify full-length KIR cDNA transcripts (22, 24). A combination of two generic *Mamu-KIR3D/1D* and one specific *Mamu-KIR2DL04* primer sets was required to amplify the complete KIR transcriptome in macaques (**Table I**).

A family of thirty macaques was selected that originated from eight founders and covered four generations, which allowed an extensive segregation study (**Figure 1**). An average of approximately 9,000 *Mamu-KIR3D/1D* and 5,800 *Mamu-KIR2DL04* reads were obtained per animal (±100,000 reads/PacBio Sequel cell), of which 20-45% were perfectly matched with the *Mamu-KIR* allele library that consisted of 218 annotated *Mamu-KIR3D/1D* and *Mamu-KIR2DL04* sequences. The remaining reads that did not match with the *Mamu-KIR* library were novel alleles, partial sequences or sequences containing random single nucleotide gaps, which had been introduced by PacBio sequencing. A total of twenty-nine unreported *Mamu-KIR* alleles were identified in thirty related rhesus macaques (**Tables II** and **III**). Six new *KIR* alleles were identified for

Table I. Overview of forward (Fw) and reverse (Rv) primers used for PacBio PCR amplification of full lenght KIR in humans and macaques. To amplify different KIR genes, wobbles had to be introduced in primer sequences: cytosine or thymine (Y), adenine or guanine (R), guanine or thymine (K) and adenine or thymine (W). The human KIR2D/3D primers were designed for this study to amplify all KIR genes, except for KIR2DL4, KIR2DL5, and KIR3DL3. Human KIR2DL4 was specifically amplified by a seperate primer set.

Target	Fw / Rv	Primer sequence 5' - 3'	Source
Mamu-KIR3D/1D	Fw	CAGCACCATGTCGCTCAT	Moreland et al. [24]
	Rv	GGGGTCAAGTGAAGTGGAGA	
Mamu-KIR3D/1D	Fw	AACATCCTGTGYRCTGCKGAGCWGAG	Blokhuis et al. [22]
	Rv	TGGAKAATTGTGGGYTAAGCAARGGAG	
Mamu-KIR2DL4	Fw	CCACATCCTCTGCACCGGTCAGTC	Blokhuis et al. [22]
	Rv	GCAGGGGTCAAGTGAAGGGGAGAA	
Human *KIR2D/3D*	Fw	CGTCAYCCTCCCATGATGTGG	Designed for this study
	Rv	GTTGGAGAGGTGGGCAGG	
Human *KIR2DL4*	Fw	CCTCACCACATCCTCTGCAC	Designed for this study
	Rv	GGTGTGAGGAAGAGTGATGCT	

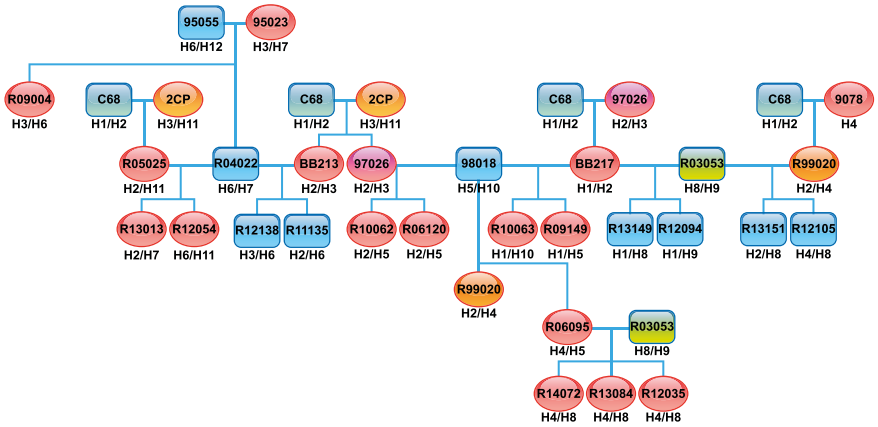


Figure 1. Pedigree of an Indian rhesus macaque family comprising thirty members. Depicted are the sires (blue squares) and dames (red ovals) with their offspring. Five animals (C68, 2CP, 97026, R99020, and R03053) are present in the pedigree multiple times, and are depicted by a different color. Different haplotypes are indicated by H1, H2, and so on. All haplotypes are found in at least two animals, except for H12, which did not segregate from sire 95055 into any of the offspring. Although no blood samples were obtained from dame 9078, H4 could be deduced based on the offspring's genetic content.

Mamu-KIR3DL07 and *Mamu-KIR3DL20*, whereas for the other *KIR* genes, two, one, or no new alleles were detected.

In two of the newly detected alleles, premature stop codons were identified, but both were located in exon 9. For example, an insertion of a single cytosine (C) was found at base pair position 1014 in *Mamu-KIR3DL11*009* (LT906600), which resulted in a frame shift and the introduction of an early stop codon at base pair position 1171 (exon 9). This cytosine insertion was confirmed using multiple independent PCRs, and the frame shift might suggest a *Mamu-KIR3DL11* isoform with a truncated cytoplasmic tail, thereby lacking the ability to signal via ITIMs. In *Mamu-KIR3DL05*013* (LT906588), a point mutation at base pair position 1294 (exon 9) introduced a stop codon, which resulted in a transcript likely to encode a KIR3DL protein with only one ITIM.

In addition to novel alleles, six previously reported *Mamu-KIR* sequences were extended, and another thirty alleles were confirmed. Overall, of the sixty-nine *KIR* genes/alleles identified in the rhesus macaque family **(Table II**), almost half of them were novel, illustrating the power of the platform and the extensive allelic polymorphism of the macaque *KIR* gene system.

Recombinant Mamu-KIR genes

In addition to allelic polymorphism and CNV, the plasticity of the *KIR* gene system is also reflected by recombination events, such as inter- and intrachromosomal rearrangements, which might result in the formation of recombinant in-frame *KIR* genes (**Figure 2A, B**) (41-44). This type of generation of novel gene entities, caused by the fusion of different genes, has been described in humans, but thus far has not been

Table II. Reported and novel alleles per *Mamu-KIR* gene. In total, 218 *Mamu-KIR* alleles are reported and used as reference database in this communication. Four recombinant and twenty-nine novel alleles were discovered and named by phylogenetic analysis. The recombinant genes consist of two segments from different *KIR* genes. Thirty-six alleles were confirmed or extended. In total, 69 *Mamu-KIR* alleles were reported in the studies rhesus macaque family.

Gene	R	C	N	Gene	R	C	N	Gene	R	C	N
1D	3	1	0	*3DL07*	13	1	6	*3DS04*	9	0	0
2DL4	28	5	0	*3DL08*	12	4	1	*3DS05*	5	2	1
3DL01	28	7	2	*3DL10*	8	1	2	*3DS06*	8	1	1
3DL02	9	1	0	*3DL11*	7	0	1	*3DSw07*	3	0	0
3DLw03	5	0	2	*3DL20*	15	4	6	*3DSw08*	12	2	1
3DL04	4	0	0	*3DS01*	5	1	1	*3DSw09*	6	1	1
3DL05	14	2	2	*3DS02*	17	2	2	Recombinant	0	0	4
3DL06	2	0	0	*3DS03*	5	1	0	Total	218	36	33

R= Reported alleles, **C**=Confirmed genes, **N**=Novel alleles

Table III. Overview of twenty-nine novel, four recombinant, and six extended *KIR* genes observed in thirty related rhesus macaques. Official designations of novel sequences were determined by comparison with the reference sequence database, containing 218 reported *Mamu-KI*R sequences, and were provided by the Immuno Polymorphism Database (IPD), which will shortly release a database with non-human primate KIR sequences.The number of animals that contained the allele and the European Nucleotide Archive (ENA) accession numbers are listed (https://www.ebi.ac.uk/ena/). E, extended KIR sequence; R, recombinant KIR sequence.

Gene name	Number of animals	ENA accession number
*Mamu-KIR3DL01*002:02*	11	LT906583
*Mamu-KIR3DL01*028*	3	LT906584
*Mamu-KIR3DLw03*006:01*	6	LT906585
*Mamu-KIR3DLw03*007*	2	LT906586
*Mamu-KIR3DL05*012*	6	LT906587
*Mamu-KIR3DL05*013*	2	LT906588
*Mamu-KIR3DL07*014*	11	LT906589
*Mamu-KIR3DL07*016*	5	LT906591
*Mamu-KIR3DL07*017*	6	LT906592
*Mamu-KIR3DL07*018*	2	LT906593
*Mamu-KIR3DL07*019*	2	LT906594
*Mamu-KIR3DL07*020*	2	LT906595
*Mamu-KIR3DL08*012*	6	LT906596
*Mamu-KIR3DL10*002:03*	6	LT906597
*Mamu-KIR3DL10*007*	11	LT906598
*Mamu-KIR3DL11*009*	2	LT906600
*Mamu-KIR3DL20*016*	6	LT906601
*Mamu-KIR3DL20*017*	3	LT906602
*Mamu-KIR3DL20*018*	17	LT906603
*Mamu-KIR3DL20*019*	6	LT906604
*Mamu-KIR3DL20*020*	2	LT906605
*Mamu-KIR3DL20*022*	3	LT906607
*Mamu-KIR3DS01*004*	2	LT906608
*Mamu-KIR3DS02*014*	2	LT906609
*Mamu-KIR3DS02*015:01*	6	LT906610
*Mamu-KIR3DS05*004*	3	LT906611
*Mamu-KIR3DS06*007*	7	LT906612
*Mamu-KIR3DSw08*010:02*	2	LT907836
*Mamu-KIR3DSw09*007*	3	LT907837
Mamu-KIR3DL02/3DL08A (R)	7	LT907838
Mamu-KIR3DL02/3DL08B (R)	2	LT907839
Mamu-KIR3DL10A/3DL02 (R)	7	LT907840
Mamu-KIR3DL10B/3DL02 (R)	1	LT907841
*Mamu-KIR3DL01*002* (E)	7	LT963634
*Mamu-KIR3DL02*004:01* (E)	5	LT963635
*Mamu-KIR3DL05*006:01* (E)	6	LT963636
*Mamu-KIR3DL07*009:01* (E)	3	LT963637
*Mamu-KIR3DL20*001* (E)	6	LT963638
*Mamu-KIR3DS05*002:02* (E)	12	LT963639

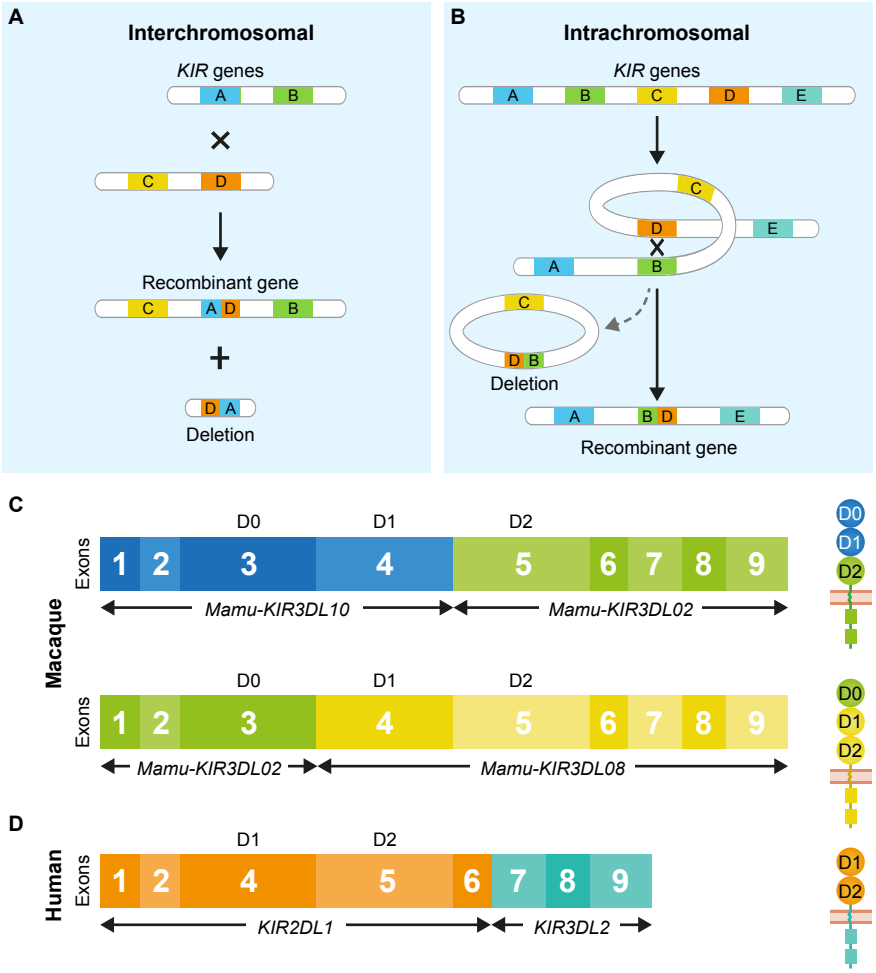


Figure 2. Chromosomal rearrangement mechanisms, inter- or intrachromosomal, which can result in recombinant *KIR* genes and the contraction of haplotypes. (A) Interchromosomal: recombination between two different (homologous) chromosomes can result in the generation of a novel recombinant gene and the deletion of gene segments. (B) Intrachromosomal: two genes on the same chromosome might align with each other, and subsequent rearrangements can result in the generation of a recombinant fusion gene and the deletion of a non-viable ring chromosome. The deletions described by (A) and (B), together with the generation of a recombinant gene, might explain the *KIR* haplotype contractions. (C) Two variants of recombinant gene transcripts observed in rhesus macaques are depicted. The 5' end to exon 4 of *Mamu-KIR3DL10* and exon 5 to the 3' end of *Mamu-KIR3DL02* encode one type of recombinant transcripts, with SNPs in the *Mamu-KIR3DL10* segment. The other recombinant genes found in this study are encoded by the 5' end to exon 3 of *Mamu-KIR3DL02* and exon 4 to the 3' end of *Mamu-KIR3DL08*, with variation in the *Mamu-KIR3DL08* segment. Predicted KIR protein structures are schematically illustrated adjacent to the respective gene, in which the colors of the exons and protein segments resemble each other. (D) Human recombinant KIR transcript with the 5' end to exon 6 of *KIR2DL1* and exon 7 to the 3' end of *KIR3DL2*. This transcript is likely to encode a KIR2D protein with the extracellular domains of KIR2DL1 and an intercellular tail of KIR3DL2, as schematically illustrated adjacent to the transcript.

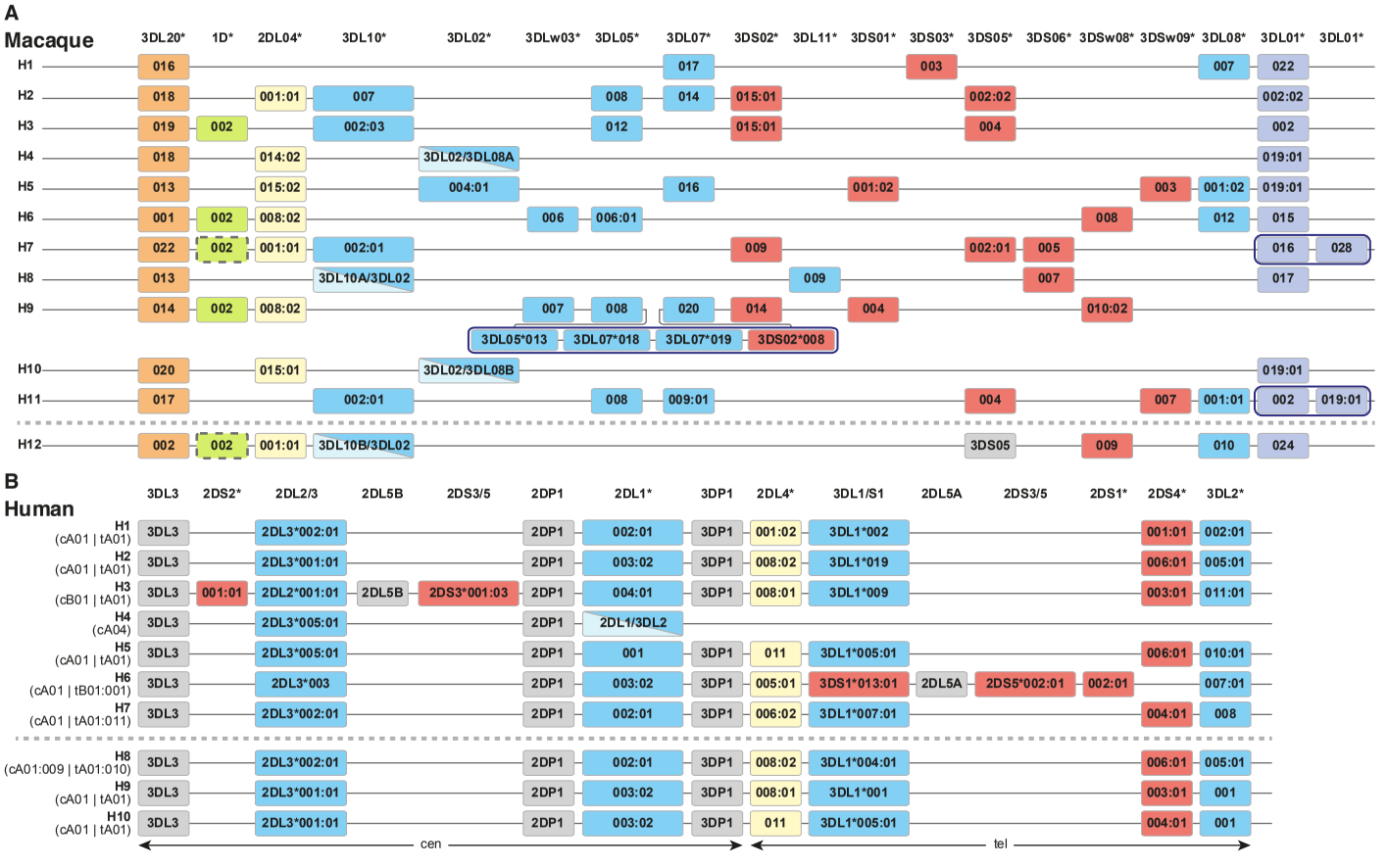
observed in rhesus macaques. In this study, comprising only one extended macaque family, four novel recombinant *Mamu-KIR* sequences were identified at the transcription level, composed of segments from two different *KIR* genes (**Figure 2C**). Two recombinant genes consisted of the 5' end to exon 4 of *Mamu-KIR3DL10* and exon 5 up to the 3' end of *Mamu-KIR3DL02*, but seemed to originate from two independent fusion events. In these two recombinant genes, most point mutations were observed in the *Mamu-KIR3DL10* segment, whereas the *Mamu-KIR3DL02* part was less variable; we therefore refer to these recombinant genes as *Mamu-KIR3DL10A/3DL02* and *Mamu-KIR3DL10B/3DL02*. The other two recombinant genes consisted of the 5' end up to exon 3 of *Mamu-KIR3DL02* and exon 4 up to the 3' end of *Mamu-KIR3DL08*. These recombinant sequences can be distinguished by a synonymous and a non-synonymous single nucleotide polymorphism (SNP) in the *Mamu-KIR3DL08* segment, and will be referred to as *Mamu-KIR3DL02/3DL08A* and *Mamu-KIR3DL02/3DL08B*.

The transcripts of the above-mentioned recombinant genes encoded three extracellular domains and a long cytoplasmic tail, suggesting an inhibitory function (**Figure 2C**). The recombinant genes were confirmed by independent PCRs and/or segregation into families. All four recombinant genes contained a segment of *Mamu-KIR3DL02*, suggesting that this gene is highly susceptible to engaging in fusion events. However, recombinant genes containing segments from other inhibitory, activating, or pseudo-genes might be discovered when larger populations are studied.

Rhesus macaque KIR haplotypes show extensive expansions and contractions

An analysis of a rhesus macaque family comprising thirty animals (**Figure 1**) revealed the segregation of *KIR* gene combinations. Twelve previously unreported *Mamu-KIR* haplotypes were deduced (**Figure 3A**). All haplotypes were confirmed in multiple animals, except for haplotype H12, for which material of informative offspring is missing (sire 95055). This deduced haplotype was, however, confirmed by multiple independent PCRs and, in addition, Sanger sequencing to certify low PacBio read counts for the *Mamu-KIR3DL20* and *KIR3DS05* genes (< 3 PacBio reads; Supp. **Figure 1A**). *KIR3DS05* on haplotype H12 could not be defined at the allelic level, but the presence was confirmed on gDNA. Furthermore, haplotype H12 was confirmed in part by a previous study in which the KIR repertoire of sire 95055 was analysed using Sanger sequencing (22).

The number of *KIR* genes per haplotype showed remarkable variability, and ranged from four to fourteen genes, which produced *bona fide* transcripts. *Mamu-KIR3DL20* was transcribed on all twelve haplotypes, whereas *Mamu-KIR3DL01* was present on all haplotypes except for haplotype H9. These two genes might be considered framework genes in rhesus macaques, although some previously reported macaque *KIR* haplotypes defined by conventional characterization methodology seem to lack those genes (22). Eight of the twelve *Mamu-KIR* haplotypes transcribed a *Mamu-KIR2DL04* gene, whereas in humans the orthologous gene is referred to as a framework gene (12). Other common lineage II *KIR* genes were *Mamu-KIR3DL05*, *Mamu-KIR3DL07*, *Mamu-KIR3DL08*, *Mamu-KIR3DL10*, *Mamu-KIR3DS02,* and *Mamu-KIR3DS05*, which were


A
Macaque
3DL20* 1D* 2DL04* 3DL10* 3DL02* 3DLw03* 3DL05* 3DL07* 3DS02* 3DL11* 3DS01* 3DS03* 3DS05* 3DS06* 3DSw08* 3DSw09* 3DL08* 3DL01* 3DL01*
H1 016 017 003 007 022
H2 018 001:01 007 008 014 015:01 002:02 002:02
H3 019 002 002:03 012 015:01 004 002
H4 018 014:02 3DL02/3DL08A 019:01
H5 013 015:02 004:01 016 001:02 003 001:02 019:01
H6 001 002 008:02 006 006:01 008 012 015
H7 022 002 001:01 002:01 009 002:01 005 016 028
H8 013 3DL10A/3DL02 009 007 017
H9 014 002 008:02 007 008 020 014 004 010:02
3DL05*013 3DL07*018 3DL07*019 3DS02*008
H10 020 015:01 3DL02/3DL08B 019:01
H11 017 002:01 008 009:01 004 007 001:01 002 019:01
H12 002 002 001:01 3DL10B/3DL02 3DS05 009 010 024
B
Human
3DL3 2DS2* 2DL2/3 2DL5B 2DS3/5 2DP1 2DL1* 3DP1 2DL4* 3DL1/S1 2DL5A 2DS3/5 2DS1* 2DS4* 3DL2*
H1 (cA01 | tA01) 3DL3 2DL3*002:01 2DP1 002:01 3DP1 001:02 3DL1*002 001:01 002:01
H2 (cA01 | tA01) 3DL3 2DL3*001:01 2DP1 003:02 3DP1 008:02 3DL1*019 006:01 005:01
H3 (cB01 | tA01) 3DL3 001:01 2DL2*001:01 2DL5B 2DS3*001:03 2DP1 004:01 3DP1 008:01 3DL1*009 003:01 011:01
H4 (cA04) 3DL3 2DL3*005:01 2DP1 2DL1/3DL2
H5 (cA01 | tA01) 3DL3 2DL3*005:01 2DP1 001 3DP1 011 3DL1*005:01 006:01 010:01
H6 (cA01 | tB01:001) 3DL3 2DL3*003 2DP1 003:02 3DP1 005:01 3DS1*013:01 2DL5A 2DS5*002:01 002:01 007:01
H7 (cA01 | tA01:011) 3DL3 2DL3*002:01 2DP1 002:01 3DP1 006:02 3DL1*007:01 004:01 008
H8 (cA01:009 | tA01:010) 3DL3 2DL3*002:01 2DP1 002:01 3DP1 008:02 3DL1*004:01 006:01 005:01
H9 (cA01 | tA01) 3DL3 2DL3*001:01 2DP1 003:02 3DP1 008:01 3DL1*001 003:01 001
H10 (cA01 | tA01) 3DL3 2DL3*001:01 2DP1 003:02 3DP1 011 3DL1*005:01 004:01 001
cen
tel

present in 55-70% of the studied animals. Most of the remaining *KIR* genes were present in approximately 10-30% of the animals, or were absent in this family. The more frequently present genes, like *Mamu-KIR3DL01, Mamu-KIR3DL05*, *Mamu-KIR3DL07*, *Mamu-KIR3DL20,* and *KIR3DS02,* showed the most allelic variation (**Table II**), which might indicate a selective pressure on relatively rapidly evolving genes, or that these *KIR* genes are old entities that accumulated variation over time.

Haplotypes H7, H9, and H11 showed an expansion of the *KIR* cluster that was characterized by two or three 'allelic' copies of a certain gene (**Figure 3A**). On haplotypes H7 and H11, *Mamu-KIR3DL01* was duplicated, resulting in two 'allelic' *Mamu-KIR3DL01* copies that differed at 22 and 12 base pair positions, respectively. Considering the number of SNPs between the *Mamu-KIR3DL01* copies and the fact that they are located on the same chromosome, designating them as different genes could be considered in the future. Macaque haplotype H9 showed a more extensive expansion, with two 'allelic' copies of *Mamu-KIR3DL05* and *Mamu-KIR3DS02*, and even three copies of *Mamu-KIR3DL07*. Each 'allelic' *Mamu-KIR3DL07* copy varied from the other by at least six SNPs, suggesting that these copies are old entities that accumulated variation over time, and are not the result of recent duplications. The duplicated *Mamu-KIR3DL05*, *Mamu-KIR3DL07*, and *Mamu-KIR3DS02* copies on haplotype H9 were also confirmed by Sanger sequencing to exclude potential *in vitro* artefacts generated by the PacBio platform (**Supp. Table I** and **Supp. Figure 1B, C, D**).

In contrast to extended haplotypes, macaque haplotypes H4, H8, H10, and H12 seemed to be contracted haplotypes, with only four or seven/eight *KIR* genes present, including a recombinant gene generated by an in-frame fusion event. All of these contracted haplotypes also contained a *Mamu-KIR3DL01* and *Mamu-KIR3DL20* allele along with a few additional lineage I and/or II *KIR* alleles. The presence of a fusion gene

◀ **Figure 3. Rhesus macaque and human *KIR* haplotypes defined on the transcription level.** (A) Twelve different *KIR* haplotypes were characterized by a segregation study involving a large rhesus macaque family. All haplotypes segregate except for haplotype H12, which was only found in sire 95055 and was confirmed by multiple independent PCRs. The physical localization of *Mamu-KIR3DL20*, *Mamu-KIR1D*, and *Mamu-KIR2DL04* is based on a completely sequenced macaque *KIR* haplotype, and these genes are indicated by different colors (27). All other *Mamu-KIR3DL* genes are indicated in blue boxes and all *Mamu-KIR3DS* genes in red boxes. Since *Mamu-KIR3DL01* is present on all haplotypes except haplotype H9, it might be considered a framework gene, and is therefore indicated by a different blue color. *KIR3DS05* on haplotype H12 (grey box) was determined as present on gDNA, but lacked allele-level typing due to low PacBio read counts (< 3 PacBio reads). Gene duplications, or genes that are likely to be inserted as an entity, are indicated by a dark blue outline around the boxes. A recombinant gene, which might be the result of a fusion event, is depicted by a two-colored box (see for instance haplotype H4). On haplotypes H7 and H12, it was undeterminable whether *KIR1D*002* was present, and these boxes are indicated by a dashed outline. The physical localizations of the *Mamu-KIR3DL* and the *Mamu-KIR3DS* genes are predicted based on information from recombinant genes or gene insertions. (B) Ten human *KIR* haplotypes were characterized. Haplotypes H1 to H7 did segregate, whereas haplotypes H8 to H10 were only found in a single individual. The inhibitory *KIR* (*KIR2/3DL*) are depicted by blue boxes, whereas the activating *KIR* (*KIR2/3DS*) are depicted by red boxes. *KIR2DL4* is illustrated by yellow boxes. *KIR3DL3* and *KIR2DL5* were only determined for their presence or absence on gDNA, and are depicted in grey boxes. The recombinant gene on haplotype H4 is illustrated by a two-colored box. The gene order and haplotype nomenclature are based on the haplotype conformations as described by Pyo and colleagues (52).

on all contracted haplotypes might be indicative of recombination events, such as unequal crossing-over and intrachromosomal recombination (**Figure 2A, 2B**), which might have caused the deletion of genes that were present on the original *KIR* haplotypes. At least one copy of a recombinant gene was found in 40% of the animals (**Supp. Table II**), which might indicate positive selection for contracted haplotypes containing recombinant genes.

The recombination events in macaques may be indicative of the physical position of *KIR* genes and, in combination with the previously sequenced macaque *KIR* haplotype by Sambrook and colleagues, the physical locations of the *KIR* genes were predicted and illustrated in **Figure 3A** (27). For example, haplotype H8 and H12 contain a recombinant gene with a 5' segment of *Mamu-KIR3DL10* and a 3' segment of *Mamu-KIR3DL02,* suggesting that the latter is located downstream of *Mamu-KIR3DL10*. The recombinant genes on haplotypes H4 and H10 suggest that *Mamu-KIR3DL02* should be localized in front of *Mamu-KIR3DL08*. Furthermore, *Mamu-KIR3DL05*, *Mamu-KIR3DL07*, and *Mamu-KIR3DS02* seemed to have been introduced on macaque haplotype H9 as a single entity, indicating that these genes were located next to each other, or are at least in close proximity. Nevertheless, to elucidate the precise *KIR* gene positions, additional genomic haplotype sequencing and phasing needs to be performed.

To confirm that no *KIR* genes were missed by PCR amplification at the transcription level, the absence or presence of several frequent KIR genes was also determined at the gDNA level (**Supp. Table II**). This approach suggested that no frequent *KIR* genes had been missed by the three primer sets used for amplification on cDNA. In addition, although typing for the presence or absence of these genes at the gDNA level might be less informative, in a few cases it provided extra information that was needed to assign them to a certain haplotype.

Identification of full-length cDNA KIR transcripts and a fusion gene in humans

In humans, one generic primer set, mapping to the UTRs, was able to amplify most *KIR2D/3D* genes, except for *KIR3DL3*, *KIR2DL4*, and *KIR2DL5*. For *KIR2DL4*, we designed an additional specific primer set to facilitate analysis of this gene at the transcriptional level (**Table I**). UTR-specific primers for *KIR3DL3* and *KIR2DL5* could not be designed, and therefore these genes were only analyzed for the presence or absence at the gDNA level. It is known from the literature that *KIR3DL3* is a framework gene, and should be present on all *KIR* haplotypes, whereas for *KIR2DL5*, two, one, or no copies can be present only on group B haplotypes (12, 45).

A human family consisting of fifteen members was selected, which allowed segregation analysis (**Figure 4**). The family comprised three generations, and had been founded by five individuals. Approximately 7,900 *KIR2D/3D* reads and 7,350 *KIR2DL4* reads per individual were obtained, and an average of 10-22,5% of the reads mapped 100% to the human *KIR2D/3D* allele library containing 907 reported *KIR* alleles (14). The remaining reads contained single nucleotide gaps, or were partial sequences introduced by PacBio sequencing. No unreported human *KIR* alleles were discovered,

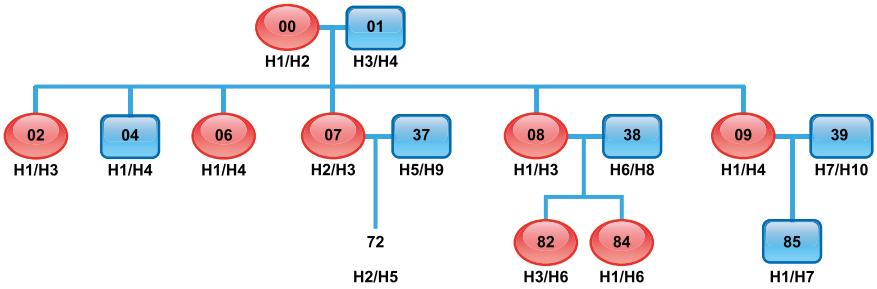


Figure 4. Pedigree of a human family that comprises fifteen individuals. Depicted are the males (blue squares) and the females (red ovals) with their children. The gender of individual 72 was unknown. Different haplotypes are indicated by H1, H2, and so on. In total, ten *KIR* haplotypes were identified in this family, of which seven did segregate (H1 – H7) and three were found in a single individual (H8 – H10).

although one known partial sequence was extended (*KIR3DL2*011:01,* LT934502). Furthermore, a recently reported fusion gene was confirmed, consisting of the 5' end up to exon 6 of *KIR2DL1* and exon 7 to the 3' end of *KIR3DL2* (**Figure 2D**, LT963640) (42). The fusion transcript encoded a D1-D2 extracellular segment and a long intercellular tail, suggesting an inhibitory KIR2D receptor. Standard *KIR* typing kits, which are commonly used for *KIR* characterization in humans, readily miss recombinant *KIR* genes as they only type for the presence or absence of gene segments at the gDNA level.

On average, less copy number variation for human versus rhesus macaque KIR haplotypes

Ten *KIR* haplotypes from fifteen related human individuals were thoroughly defined based on full-length cDNA transcripts (**Figure 3B**). Each haplotype encoded from three to ten *KIR* gene transcripts, including representatives of the framework genes *KIR3DL3*, *KIR2DL4*, and *KIR3DL2*, except for haplotype H4, which lacked a copy of the *KIR2DL4* gene. The fourth human framework gene, *KIR3DP1*, is a pseudo-gene, and is therefore not amplified at the transcription level.

In this communication, we followed the human *KIR* haplotype nomenclature conventions of Pyo and colleagues, which was later adapted by Vierra-Green and colleagues (46, 47). Human haplotypes H1, H2, H5, and H7-H10 represent so-called non-variable A haplotypes (cA01|tA01), characterized by six *KIR* genes (on cDNA level), including the framework genes *KIR2DL4* and *KIR3DL2*, and the activating receptor *KIR2DS4*. Haplotype H8 was identical to the previously reported cA01:009|tA01:010 haplotype, whereas the telomeric region of H7 was identical to tA01:011 (46). Group B haplotypes were represented by haplotypes H3 and H6 (**Figure 3B**), containing up to ten *KIR* genes, including multiple activating receptors. The telomeric region of haplotype H6 was identical to the previously reported tB01:001 region, and was combined with a cA01 region (cA01|tB01:001). *KIR* haplotype H3 confirmed the previously reported cB01|cA01 haplotype configuration (46). Haplotype H4 contained

only three *KIR* genes- *KIR3DL3*, *KIR2DL3*, and a *KIR2DL1/KIR3DL2* fusion gene- the latter suggesting a contracted haplotype. The absence of the framework gene *KIR2DL4* on haplotype H4 might suggest a deletion in the telomeric region of this haplotype. Recently, Roe and colleagues published a contracted haplotype on gDNA that appears to be identical to our haplotype H4, including the above-described recombinant gene (cA04) (42).

In most human disease association studies, *KIR* genes are typed by determination of their presence or absence on gDNA. However, multiple studies demonstrated that health and disease could be linked to certain *KIR* alleles instead of to *KIR* genes (48-50). Therefore, high-resolution characterization of the *KIR* genes might be clinically beneficial, and could improve future KIR disease association studies. To confirm that our transcriptome characterization approach did amplify all *KIR* genes, their presence or absence was also assayed at the gDNA level using the Olerup SSP® KIR typing kit. All seventeen *KIR* genes that were identified as present on gDNA were also detected at the transcription level, except for *KIR2DL5*, *KIR3DL3*, and the pseudo-genes (**Supp. Table III**). The pseudo-genes *KIR3DP1* and *KIR2DP1* were present in all genotyped individuals. *KIR3DP1* is a framework gene and was suggested to be present on all haplotypes, except for haplotype H4, which is assumed to lack the *KIR3DP1* gene due to a deletion in the telomeric region as described by Roe and colleagues (**Figure 3B**) (42). *KIR2DP1* has been described to be present on cA01 and cB01 regions (46), suggesting the presence of this pseudo-gene on all haplotypes of the studied human family. The framework gene *KIR3DL3* and the group B haplotype-specific *KIR2DL5* were found present at gDNA level in all individuals and on both group B haplotypes H3 and H6, respectively (**Figure 3B**; grey boxes). These findings supported the assumption that our transcriptome protocol amplified all *KIR* genes, except for *KIR2DL5, KIR3DL3*, and the pseudo-genes, and might be beneficial for future KIR disease association studies. In addition, the identification of multiple human *KIR* haplotype regions that had been reported by others further validated our protocol.

Discussion

More recently, sequencing and characterization studies provided insights into the complexity of the *KIR* gene system. However, due to large gene copy number variation and to the high similarity of the *KIR* genes, conventional sequencing methods hampered the accurate characterization of the complete *KIR* system at the transcription level. Here, we describe a comprehensive and relatively fast SMRT sequencing protocol using a PacBio Sequel platform to completely characterize the KIR transcriptomes in human and rhesus macaque families. The power of this approach is demonstrated by the fact that, in a relatively short time, novel *KIR* genes, alleles, and complex *KIR* haplotypes were defined by segregation studies in a family set-up. A relatively high number of human and macaque recombinant *KIR* genes were discovered, and seemed to be the result of several independent fusion events. This study allowed comparison of the

human and rhesus macaque KIR transcriptomes. Eventually, this may result in a better understanding and interpretation of KIR disease association studies.

A comparison of ten human and twelve rhesus macaque *KIR* haplotypes illustrates that both species share highly similar gene systems. Although there are subtle differences, such as different receptor lineage expansion and haplotype organisation, both species show extensive allelic polymorphism and gene copy number variation in their *KIR* repertoire. Hence, rhesus macaques may provide relevant models to study the impact of *KIR* genes on health and disease. Human *KIR* allelic polymorphism seems to have already been broadly mapped, as all the alleles we recovered are documented in a database containing over 900 alleles extracted from numerous population studies (14). In contrast, only 218 *Mamu-KIR* alleles were reported, but considering that almost half of the total *Mamu-KIR* alleles discovered in this study were unreported, it is reasonable to suggest that this number is only the tip of the iceberg. Therefore, the extent of allelic *KIR* polymorphism seems to be at least comparable in humans and rhesus macaques.

KIR gene copy number variation is observed on the human and rhesus macaque haplotypes, which contained from four to twelve or four to fourteen *KIR* (pseudo-) genes, respectively. In humans, the non-variable group A haplotypes rarely show copy number variation, whereas the group B haplotypes can have a variable number of - mainly activating - *KIR* genes, in part caused by duplications or chromosomal rearrangements (**Figure 3B**). In macaques, each haplotype can contain a different number of inhibitory and activating *KIR* genes, which can be magnified by deletions or insertions as a result of duplications and chromosomal rearrangements (**Figure 3A**). In comparison to the human situation, the overall CNV seems to be more extensive in the macaque *KIR* gene system, and might be explained by co-evolution with the expanded MHC class I repertoire in macaques, and by the absence of a haplotypic organisation as is observed in humans.

One human and four macaque haplotypes showed signs of contraction. In both species, the contraction of haplotypes was marked by the presence of recombinant genes and the apparent deletion of a haplotype segment. The generation of these contracted haplotypes is most likely mediated by repetitive elements present in the *KIR* introns (43). For some human recombinant haplotypes, these sequence repeats are identified and characterized as breakpoints that may facilitate chromosomal rearrangements (**Figure 2A, B**). In macaques, however, these repetitive elements are not yet characterized, but considering the observation of short *Mamu-KIR* haplotypes, in addition to the presence of recombinant genes, it is likely that the same mechanisms are responsible for *KIR* gene expansion and haplotype contraction as is observed in humans. The rapid loss and gain of *KIR* genes, driven by the repetitive sequence elements in the introns, might be an advantageous evolutionary strategy to expand gene variability using an existing gene repertoire, and thereby enhancing pathogen evasion. Hence, recombinant genes are composed of different heads (ligand interaction) and tails (signaling function), which may facilitate the exchange of functionalities and ligand interactions between receptors (**Figure 2C, D**).

In macaques, in addition to contraction, expansion of *KIR* haplotypes was also observed. Macaque haplotypes H7 and H11 seem to have expanded by gene duplication, whereas haplotype H9 showed evidence of unequal crossing-over events. However, it is arguable whether the duplication on macaque haplotypes H7 and H11 should be considered as copies of a *Mamu-KIR3DL01* gene, since the 'allelic' copies vary at 22 and 12 base pair positions, respectively. A sensible nomenclature system for *Mamu-KIR3DL01* sequences should be considered. A more extreme expansion is observed on macaque haplotype H9, which contains two 'allelic' copies of *Mamu-KIR3DL05* and *Mamu-KIR3DS02* and three *Mamu-KIR3DL07* copies. Although it is possible that this expansion is explained by multiple gene duplications, it seems more likely that these three duplicated genes are introduced as one entity or tandem, on which *Mamu-KIR3DL07* might already have been duplicated.

Although no expanded human *KIR* haplotypes were found in the present human family studied, other researchers reported *KIR* haplotypes with gene insertions, similar to the extended macaque haplotype H9 (42, 51, 52).

The expansion and contraction of *KIR* haplotypes might be a balancing selection for fighting infections on the one hand and for reproductive success on the other. A similar reproductive/immunological trade-off is illustrated by the haplotypic organization in human *KIR* (12). In humans, expansion and contraction of the *KIR* region is only observed on group B haplotypes, which indicates that structurally diversifying the non-variable group A haplotypes is not beneficial. The group B haplotypes, which can contain a variable number of *KIR* genes, show chromosomal rearrangements, whereby *KIR* genes are introduced and/or deleted. This diversifying selection might be associated with increasing immune response variability, successful reproduction, loss of unfavourable genes, or the generation of novel fusion genes. However, chromosomal recombination might also be driven by the specific content of group B haplotypes. In this case, sequence elements that are present on group B haplotype-specific genes, such as activating *KIR* and *KIR2DL5*, might drive recombination events, without necessarily requiring selective pressure. In macaques, as well as in all other non-human primates, there is no haplotypic organization that divides variable and non-variable haplotype content. All macaque haplotypes seem to be prone to diversifying selection, as great variability in gene content is observed, but association with reproductive success or pathogen evasion has not yet been demonstrated. On an individual level, contracted haplotypes should provide the essential functions of the *KIR* gene system. *Mamu-KIR3DL01* and *Mamu-KIR3DL20* are two highly polymorphic and frequently expressed genes, and all four contracted *Mamu-KIR* haplotypes identified in this study have expression of these genes in common. Three of the studied macaques (R12035, R13084, R14072) are homozygous for contracted haplotypes (Figs. 1, 4A), and do not show signs of an impaired immune system. This might indicate that *Mamu-KIR3DL01* and *Mamu-KIR3DL20* expression, in combination with a recombinant gene, is sufficient to provide functional NK cell activity.

Previously, two population studies (51, 52) and two smaller studies (42, 53) reported contracted and expanded *KIR* haplotypes at the gDNA level in humans. In three of

these studies, almost twice as many contracted haplotypes were observed in comparison to expanded haplotypes. However, the prevalence of contracted and expanded haplotypes in these populations with European ancestry was only 5-10%. In the selected rhesus macaque family, four of the twelve *Mamu-KIR* haplotypes were contracted, whereas only one showed extensive expansion. In 46% of the animals of the studied rhesus macaque family, a contracted or expanded haplotype was observed. Although a haplotype analysis of a larger rhesus macaque population is required to compare the prevalence of contraction and expansion to that observed in humans, this study suggests that chromosomal rearrangements in the *KIR* cluster is more common in rhesus macaques than it is in humans. Furthermore, the results indicate that contraction, accompanied by the generation of novel recombinant genes, seems to be more beneficial, or at least has a higher occurrence, than expansion of *KIR* haplotypes.

In the past, different typing strategies were reported to characterize the *KIR* cluster in humans and rhesus macaques, including cloning and Sanger sequencing, Roche/454 pyro-sequencing, and microsatellite analysis (22-24, 27, 33, 38). However, these methods were often insufficient to assemble full-length allele sequences, as well as being time consuming and mainly focused on the presence or absence of genes. More recent reports described high-resolution *KIR* characterization on gDNA by Next-Generation Sequencing (NGS) on a PacBio RS II platform, a MiSeq platform, or by exome capturing (42, 54, 55). These studies were able to define 'allelic'-level genotypes and to identify novel alleles, and less frequently, defined complete or partial haplotypes, and discovered recombinant genes at the gDNA level. Another recent study used SMRT sequencing on a PacBio RS II platform to characterize KIR transcription in an inbred Mauritian cynomolgus macaque population, which is restricted in its genetic diversity, and was able to identify nine novel alleles and to define *KIR* haplotypes (56). Our studies show that a similar approach is also applicable to outbred human and macaque populations.

Most current clinical methods to characterize the *KIR* gene system are primarily based on determining the presence or absence of the known human *KIR* gene segments, and they might miss substantial information such as allele-level typing, copy number variation, expanded haplotypes, and recombinant genes. Multiple disease association studies, however, illustrated the importance of distinguishing between alleles of *KIR* genes (48-50). Therefore, a high-resolution *KIR* characterization approach might be beneficial for future health and disease studies.

In this communication, a method is described to thoroughly characterize the KIR transcriptomes in humans and rhesus macaques, using a relatively fast high-resolution SMRT sequencing protocol on a PacBio Sequel platform. Novel alleles and recombinant genes were discovered, and transcribed haplotypes were defined based on transcription profiles in concert with segregation studies in families. Although sequencing at the transcription level might have minor drawbacks, such as the lack of intron and pseudo-gene information, it does have serious advantages over sequencing on gDNA; for example, transcriptional modifications can be observed, including splicing of transcripts or intron insertions. These transcriptional modifications might have an effect on the

function of the receptors; for instance, when KIR3D transcripts can be spliced to generate KIR2D transcripts, or when transmembrane regions are spliced out, which may result in soluble receptors. In addition, although the PacBio platform does not yet provide quantitative analysis, the number of identified transcripts might be indicative of the expression level of a certain allele. However, this indication might be affected by other factors, such as variegated KIR expression, the MHC class I gene repertoire, and previous pathogen exposure.

In conclusion, this comprehensive sequencing approach can eventually contribute to better understanding and characterizing the *KIR* gene cluster in different species, thereby improving not only the interpretation of disease association studies but also transplantation and reproduction biology.

Acknowledgements

We thank D. Devine for editing the manuscript and F. van Hassel for preparing the figures.

References

1. Mingari, M. C., A. Moretta, and L. Moretta. 1998. Regulation of KIR expression in human T cells: a safety mechanism that may impair protective T-cell responses. *Immunol Today.* 19: 153-157.

2. Trowsdale, J. 2001. Genetic and functional relationships between MHC and NK receptor genes. *Immunity.* 15: 363-374.

3. Parham, P., P. J. Norman, L. Abi-Rached, and L. A. Guethlein. 2012. Human-specific evolution of killer cell immunoglobulin-like receptor recognition of major histocompatibility complex class I molecules. *Philos Trans R Soc Lond B Biol Sci.* 367: 800-811.

4. Parham, P., and T. Ohta. 1996. Population biology of antigen presentation by MHC class I molecules. *Science.* 272: 67-74.

5. Khakoo, S. I., and M. Carrington. 2006. KIR and disease: a model system or system of models? *Immunological reviews.* 214: 186-201.

6. Purdy, A. K., and K. S. Campbell. 2009. Natural killer cells and cancer: regulation by the killer cell Ig-like receptors (KIR). *Cancer biology & therapy.* 8: 2211-2220.

7. Hershberger, K. L., R. Shyam, A. Miura, and N. L. Letvin. 2001. Diversity of the killer cell Ig-like receptors of rhesus monkeys. *Journal of immunology (Baltimore, Md : 1950).* 166: 4380-4390.

8. Bimber, B. N., and D. T. Evans. 2015. The killer-cell immunoglobulin-like receptors of macaques. *Immunological reviews.* 267: 246-258.

9. Marsh, S. G., P. Parham, B. Dupont, D. E. Geraghty, J. Trowsdale, D. Middleton, C. Vilches, M. Carrington, C. Witt, L. A. Guethlein, H. Shilling, C. A. Garcia, K. C. Hsu, and H. Wain. 2003. Killer-cell immunoglobulin-like receptor (KIR) nomenclature report, 2002. *Human immunology.* 64: 648-654.

10. Trowsdale, J., R. Barten, A. Haude, C. A. Stewart, S. Beck, and M. J. Wilson. 2001. The genomic context of natural killer receptor extended gene families. *Immunological reviews.* 181: 20-38.

11. Hsu, K. C., S. Chida, D. E. Geraghty, and B. Dupont. 2002. The killer cell immunoglobulin-like receptor (KIR) genomic region: gene-order, haplotypes and allelic polymorphism. *Immunological reviews.* 190: 40-52.

12. Wilson, M. J., M. Torkar, A. Haude, S. Milne, T. Jones, D. Sheer, S. Beck, and J. Trowsdale. 2000. Plasticity in the organization and sequences of human KIR/ILT gene families. *Proceedings of the National Academy of Sciences of the United States of America.* 97: 4778-4783.

13. Uhrberg, M., N. M. Valiante, B. P. Shum, H. G. Shilling, K. Lienert-Weidenbach, B. Corliss, D. Tyan, L. L. Lanier, and P. Parham. 1997. Human diversity in killer cell inhibitory receptor genes. *Immunity.* 7: 753-763.

14. Robinson, J., K. Mistry, H. McWilliam, R. Lopez, and S. G. E. Marsh. 2010. IPD—the Immuno Polymorphism Database. *Nucleic Acids Research.* 38: D863-869.

15. Giebel, S., F. Locatelli, T. Lamparelli, A. Velardi, S. Davies, G. Frumento, R. Maccario, F. Bonetti, J. Wojnar, M. Martinetti, F. Frassoni, G. Giorgiani, A. Bacigalupo, and J. Holowiecki. 2003. Survival advantage with KIR ligand incompatibility in hematopoietic stem cell transplantation from unrelated donors. *Blood.* 102: 814-819.

16. Hiby, S. E., J. J. Walker, M. O'Shaughnessy K, C. W. Redman, M. Carrington, J. Trowsdale, and A. Moffett. 2004. Combinations of maternal KIR and fetal HLA-C genes influence the risk of preeclampsia and reproductive success. *The Journal of experimental medicine.* 200: 957-965.

17. Abi-Rached, L., and P. Parham. 2005. Natural selection drives recurrent formation of activating killer cell immunoglobulin-like receptor and Ly49 from inhibitory homologues. *The Journal of experimental medicine.* 201: 1319-1332.

18. Kelley, J., L. Walter, and J. Trowsdale. 2005. Comparative Genomics of Natural Killer Cell Receptor Gene Clusters. *PLoS Genetics.* 1:

19. Kaushal, D., S. Mehra, P. Didier, and A. Lackner. 2012. The non-human primate model of tuberculosis. *Journal of medical primatology.* 41: 191-201.

20. Vierboom, M. P., M. Jonker, R. E. Bontrop, and B. t Hart. 2005. Modeling human arthritic diseases in nonhuman primates. *Arthritis research & therapy.* 7: 145-154.

21. Bontrop, R. E. 2001. Non-human primates: essential partners in biomedical research. *Immunological reviews.* 183: 5-9.

22. Blokhuis, J. H., M. K. van der Wiel, G. G. Doxiadis, and R. E. Bontrop. 2010. The mosaic of KIR haplotypes in rhesus macaques. *Immunogenetics.* 62: 295-306.

23. Blokhuis, J. H., M. K. van der Wiel, G. G. Doxiadis, and R. E. Bontrop. 2011. The extreme plasticity of killer cell Ig-like receptor (KIR) haplotypes differentiates rhesus macaques from humans. *European journal of immunology.* 41: 2719-2728.

24. Moreland, A. J., L. A. Guethlein, R. K. Reeves, K. W. Broman, R. P. Johnson, P. Parham, D. H. O'Connor, and B. N. Bimber. 2011. Characterization of killer immunoglobulin-like receptor genetics and comprehensive genotyping by pyrosequencing in rhesus macaques. *BMC genomics.* 12: 295.

25. Otting, N., C. M. Heijmans, R. C. Noort, N. G. de Groot, G. G. Doxiadis, J. J. van Rood, D. I. Watkins, and R. E. Bontrop. 2005. Unparalleled complexity of the MHC class I region in rhesus macaques. *Proceedings of the National Academy of Sciences of the United States of America.* 102: 1626-1631.

26. de Groot, N. G., J. H. Blokhuis, N. Otting, G. G. Doxiadis, and R. E. Bontrop. 2015. Co-evolution of the MHC class I and KIR gene families in rhesus macaques: ancestry and plasticity. *Immunological reviews.* 267: 228-245.

27. Sambrook, J. G., A. Bashirova, S. Palmer, S. Sims, J. Trowsdale, L. Abi-Rached, P. Parham, M. Carrington, and S. Beck. 2005. Single haplotype analysis demonstrates rapid evolution of the killer immunoglobulin-like receptor (KIR) loci in primates. *Genome research.* 15: 25-35.

28. Blokhuis, J. H., G. G. Doxiadis, and R. E. Bontrop. 2009. A splice site mutation converts an inhibitory killer cell Ig-like receptor into an activating one. *Molecular immunology.* 46: 640-648.

29. Colantonio, A. D., B. N. Bimber, W. J. Neidermyer, Jr., R. K. Reeves, G. Alter, M. Altfeld, R. P. Johnson, M. Carrington, D. H. O'Connor, and D. T. Evans. 2011. KIR polymorphisms modulate peptide-dependent binding to an MHC class I ligand with a Bw6 motif. *PLoS pathogens.* 7: e1001316.

30. Andersen, H., J. L. Rossio, V. Coalter, B. Poore, M. P. Martin, M. Carrington, and J. D. Lifson. 2004. Characterization of rhesus macaque natural killer activity against a rhesus-derived target cell line at the single-cell level. *Cellular immunology.* 231: 85-95.

31. Grendell, R. L., A. L. Hughes, and T. G. Golos. 2001. Cloning of rhesus monkey killer-cell Ig-like receptors (KIRs) from early pregnancy decidua. *Tissue antigens.* 58: 329-334.

32. Bostik, P., J. Kobkitjaroen, W. Tang, F. Villinger, L. E. Pereira, D. M. Little, S. T. Stephenson, M. Bouzyk, and A. A. Ansari. 2009. Decreased NK cell frequency and function is associated with increased risk of KIR3DL allele polymorphism in simian immunodeficiency virus-infected rhesus macaques with high viral loads. *Journal of immunology (Baltimore, Md : 1950).* 182: 3638-3649.

33. Kruse, P. H., C. Rosner, and L. Walter. 2010. Characterization of rhesus macaque KIR genotypes and haplotypes. *Immunogenetics.* 62: 281-293.

34. Chaichompoo, P., P. Bostik, S. Stephenson, S. Udompunturuk, J. Kobkitjaroen, K. Pattanapanyasat, and A. A. Ansari. 2010. Multiple KIR gene polymorphisms are associated with plasma viral loads in SIV-infected rhesus macaques. *Cellular immunology.* 263: 176-187.

35. Hellmann, I., S. Y. Lim, R. S. Gelman, and N. L. Letvin. 2011. Association of activating KIR copy number variation of NK cells with containment of SIV replication in rhesus monkeys. *PLoS pathogens.* 7: e1002436.

36. PacBio website [http://www.pacb.com]

37. Kearse, M., R. Moir, A. Wilson, S. Stones-Havas, M. Cheung, S. Sturrock, S. Buxton, A. Cooper, S. Markowitz, C. Duran, T. Thierer, B. Ashton, P. Meintjes, and A. Drummond. 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics (Oxford, England).* 28: 1647-1649.

38. Bimber, B. N., A. J. Moreland, R. W. Wiseman, A. L. Hughes, and D. H. O'Connor. 2008. Complete Characterization of Killer Ig-Like Receptor (KIR) Haplotypes in Mauritian Cynomolgus Macaques: Novel Insights into Nonhuman Primate KIR Gene Content and Organization. *Journal of immunology (Baltimore, Md : 1950).* 181: 6301-6308.

39. Rhoads, A., and K. F. Au. 2015. PacBio Sequencing and Its Applications. *Genomics, proteomics & bioinformatics.* 13: 278-289.

40. Westbrook, C. J., J. A. Karl, R. W. Wiseman, S. Mate, G. Koroleva, K. Garcia, M. Sanchez-Lockhart, D. H. O'Connor, and G. Palacios. 2015. No assembly required: Full-length MHC class I allele discovery by PacBio circular consensus sequencing. *Human immunology.* 76: 891-896.

41. Martin, M. P., A. Bashirova, J. Traherne, J. Trowsdale, and M. Carrington. 2003. Cutting Edge: Expansion of the KIR Locus by Unequal Crossing Over. *The Journal of Immunology.* 171: 2192-2195.

42. Roe, D., C. Vierra-Green, C. W. Pyo, K. Eng, R. Hall, R. Kuang, S. Spellman, S. Ranade, D. E. Geraghty, and M. Maiers. 2017. Revealing complete complex KIR haplotypes phased by long-read sequencing technology. *Genes and immunity.*

43. Traherne, J. A., M. Martin, R. Ward, M. Ohashi, F. Pellett, D. Gladman, D. Middleton, M. Carrington, and J. Trowsdale. 2010. Mechanisms of copy number variation and hybrid gene formation in the KIR immune gene complex. *Human molecular genetics.* 19: 737-751.

44. Ordonez, D., N. Gomez-Lozano, L. Rosales, and C. Vilches. 2011. Molecular characterisation of KIR2DS2*005, a fusion gene associated with a shortened KIR haplotype. *Genes and immunity.* 12: 544-551.

45. Cisneros, E., M. Moraru, N. Gomez-Lozano, M. Lopez-Botet, and C. Vilches. 2012. KIR2DL5: An Orphan Inhibitory Receptor Displaying Complex Patterns of Polymorphism and Expression. *Frontiers in immunology.* 3: 289.

46. Pyo, C. W., L. A. Guethlein, Q. Vu, R. Wang, L. Abi-Rached, P. J. Norman, S. G. Marsh, J. S. Miller, P. Parham, and D. E. Geraghty. 2010. Different patterns of evolution in the centromeric and telomeric regions of group A and B haplotypes of the human killer cell Ig-like receptor locus. *PloS one.* 5: e15115.

47. Vierra-Green, C., D. Roe, L. Hou, C. K. Hurley, R. Rajalingam, E. Reed, T. Lebedeva, N. Yu, M. Stewart, H. Noreen, J. A. Hollenbach, L. A. Guethlein, T. Wang, S. Spellman, and M. Maiers. 2012. Allele-Level Haplotype Frequencies and Pairwise Linkage Disequilibrium for 14 KIR Loci in 506 European-American Individuals. *PloS one.* 7:

48. Bari, R., P. Rujkijyanont, E. Sullivan, G. Kang, V. Turner, K. Gan, and W. Leung. 2013. Effect of Donor KIR2DL1 Allelic Polymorphism on the Outcome of Pediatric Allogeneic Hematopoietic Stem-Cell Transplantation. *Journal of Clinical Oncology.* 31: 3782-3790.

49. Martin, M. P., Y. Qi, X. Gao, E. Yamada, J. N. Martin, F. Pereyra, S. Colombo, E. E. Brown, W. L. Shupert, J. Phair, J. J. Goedert, S. Buchbinder, G. D. Kirk, A. Telenti, M. Connors, S. J. O'Brien, B. D. Walker, P. Parham, S. G. Deeks, D. W. McVicar, and M. Carrington. 2007. Innate partnership of HLA-B and KIR3DL1 subtypes against HIV-1. *Nature genetics.* 39: 733-740.

50. Berinstein, J., R. Pollock, F. Pellett, A. Thavaneswaran, V. Chandran, and D. D. Gladman. 2017. Association of variably expressed KIR3dl1 alleles with psoriatic disease. *Clinical rheumatology.*

51. Jiang, W., C. Johnson, J. Jayaraman, N. Simecek, J. Noble, M. F. Moffatt, W. O. Cookson, J. Trowsdale, and J. A. Traherne. 2012. Copy number variation leads to considerable diversity for B but not A haplotypes of the human KIR genes encoding NK cell receptors. *Genome research.* 22: 1845-1854.

52. Pyo, C. W., R. Wang, Q. Vu, N. Cereb, S. Y. Yang, F. M. Duh, S. Wolinsky, M. P. Martin, M. Carrington, and D. E. Geraghty. 2013. Recombinant structures expand and contract inter and intragenic diversification at the KIR locus. *BMC genomics.* 14: 89.

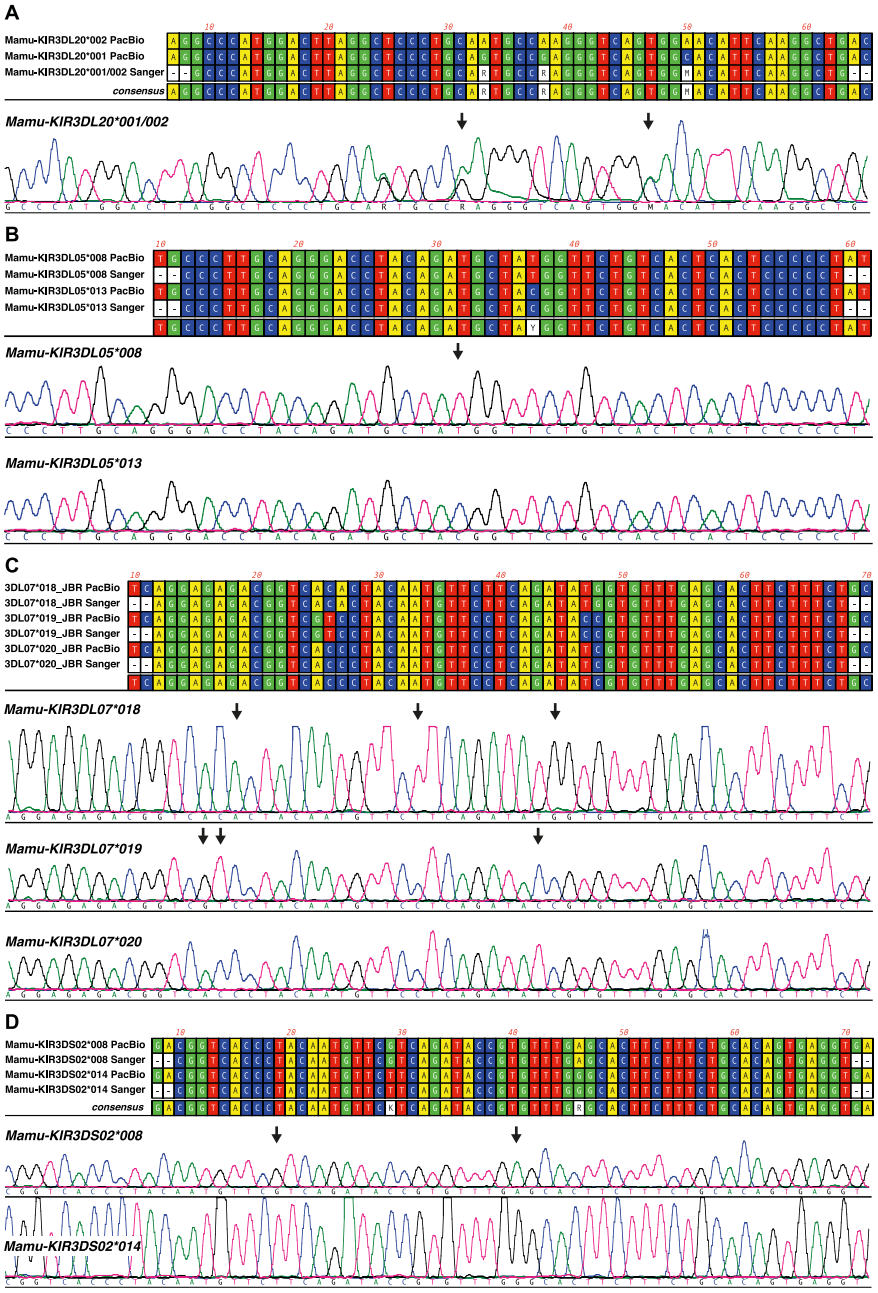
53. Hou, L., M. Chen, J. Ng, and C. K. Hurley. 2012. Conserved KIR allele-level haplotypes are altered by microvariation in individuals with European ancestry. *Genes and immunity.* 13: 47-58.

54. Maniangou, B., N. Legrand, M. Alizadeh, U. Guyet, C. Willem, G. David, E. Charpentier, A. Walencik, C. Retiere, and K. Gagne. 2017. Killer Immunoglobulin-Like Receptor Allele Determination Using Next-Generation Sequencing Technology. *Frontiers in immunology.* 8: 547.

55. Norman, P. J., J. A. Hollenbach, N. Nemat-Gorgani, W. M. Marin, S. J. Norberg, E. Ashouri, J. Jayaraman, E. E. Wroblewski, J. Trowsdale, R. Rajalingam, J. R. Oksenberg, J. Chiaroni, L. A. Guethlein, J. A. Traherne, M. Ronaghi, and P. Parham. 2016. Defining KIR and HLA Class I Genotypes at Highest Resolution via High-Throughput Sequencing. *American journal of human genetics.* 99: 375-391.

56. Prall, T. M., M. E. Graham, J. A. Karl, R. W. Wiseman, A. J. Ericsen, M. Raveendran, R. Alan Harris, D. M. Muzny, R. A. Gibbs, J. Rogers, and D. H. O'Connor. 2017. Improved full-length killer cell immunoglobulin-like receptor transcript discovery in Mauritian cynomolgus macaques. *Immunogenetics.* 69: 325-339.

Supplementary materials



◀ **Supplementary Figure 1. Sequence confirmation by Sanger sequencing.** Chromatograms and sequences obtained by Sanger sequencing to confirm results found by PacBio sequencing. Arrows in the chromatograms indicate sequence variation. Primers used for Sanger sequencing are listed in Supp. Table I. (A) Macaque haplotype H12 lacked informative offspring, and was therefore only identified in a single sire (95055). The haplotype was confirmed by multiple independent PCRs. Nonetheless, low PacBio read counts were obtained from the Mamu-KIR3DL20 alleles, and therefore these alleles were confirmed by Sanger sequencing. The arrows indicate double peaks, illustrating the presence of Mamu-KIR3DL20*001 and Mamu-KIR3DL20*002 in sire 95055, which allowed us to assign the alleles to the specific haplotypes. (B) Two copies of the Mamu-KIR3DL05 gene were found on macaque haplotype H9. One of these alleles was novel (Mamu-KIR3DL05*013), whereas the other was reported previously (Mamu-KIR3DL05*008).(C) Three copies of the Mamu-KIR3DL07 gene were found on macaque haplotype H9. These alleles were all novels, and confirmed by Sanger sequencing. (D) Two copies of the Mamu-KIR3DS02 gene were found on macaque haplotype H9. One of these alleles was novel (Mamu-KIR3DS02*014), whereas the other was reported previously (Mamu-KIR3DS02*008).

Supplementary table I. Overview of forward (Fw) and reverse (Rv) primers used to confirm alleles on macaque and human KIR haplotypes. To amplify different *Mamu-KIR3DL07* alleles, a wobble had to be introduced in the forward primer sequence: adenine or thymine (W). The PCR product sizes were sufficient to distinguish alleles and confirm their presence.

Target	Fw / Rv	Exon	Primer sequence 5' - 3'	Product size
Mamu-KIR3DL05	Fw	4	TCTGCAAAGTGAGGTGACCTT	337 bp
	Rv	5	AAGCCTAAGTTCATGGGTCTCC	
Mamu-KIR3DL07	Fw	3	GACCTTCTTGWTTGCCCGGC	556 bp
	Rv	4	ACTGGGAGCTGACAACACATAGTC	
Mamu-KIR3DS02	Fw	3	GATCATACCCGCACTCCCCAA	319 bp
	Rv	4	TGGGAGTGAGTGACAGAACCA	
Mamu-KIR3DL20	Fw	5	CCAGAGCTCGTTTGACATTTACCG	183 bp
	Rv	5	GGGTGTGACCACACATAGGGCAG	
Mamu-KIR3DS05	Fw	5	GTGTCAACGGAACATTCCAGGA	109 bp
	Rv	5	CTCGGGTGTGACCACTTGTAGGA	
Human KIR2DL1	Fw	3	ACTTTGCGCCTCATTGGAGA	359 bp
	Rv	4	CCCAGAGGAAAGTCAGCCTG	
Human KIR2DL4	Fw	9	GCCCTTCTCAGAGGAGCAAG	116 bp
	Rv	9	CATCAAGGCCTGACTGTGGT	
Human KIR2DS4	Fw	4	GGTTCAGGCAGGAGAGAAT	111 / 133 bp
	Rv	4	CTGGAATGTTCCGTKGATG	

Supplementary table II & III. These tables are too large to be readable in this thesis. We therefore chose to make them available online as a PDF file. We recommend to display them on a computer screen. Please use the QR-codes or type the link address in your browser window.

Supplementary table II

Supplementary table III

Supplementary table II
https://www.bprc.nl/sites/default/files/downloads/thesis/JB-figS02.pdf

Supplementary table III
https://www.bprc.nl/sites/default/files/downloads/thesis/JB-figS03.pdf

Chapter 3

Unparalleled Rapid Evolution of *KIR* Genes in Rhesus and Cynomolgus Macaque Populations

Jesse Bruijnesteijn[1], Nanine de Groot[1], Marit K. H. van der Wiel[1], Nel Otting[1], Annemiek J. M. de Vos-Rouweler[1], Natasja G. de Groot[1] and Ronald E. Bontrop[1,2,*]

[1]Department of Comparative Genetics and Refinement, Biomedical Primate Research Centre, 2288 GJ Rijswijk, the Netherlands

[2]Theoretical Biology and Bioinformatics Group, Utrecht University, 3527 Utrecht, the Netherlands

*Corresponding author:
Ronald Bontrop
Bontrop@bprc.nl
Tel. nr. +31-15-2842699
Fax nr. +31-15-2842600

Abstract

The killer cell immunoglobulin-like receptors (KIR) modulate immune responses through interactions with MHC class I molecules. The *KIR* region in large cohorts of rhesus and cynomolgus macaque populations were characterized, and the experimental design enabled the definition of a considerable number of alleles (n = 576) and haplotypes, which are highly variable with regard to architecture. Although high levels of polymorphism were recorded, only a few alleles are shared between species and populations. The rapid evolution of allelic polymorphism, accumulated by point mutations, was further confirmed by the emergence of a novel *KIR* allele in a rhesus macaque family. In addition to allelic variation, abundant orthologous and species-specific *KIR* genes were identified, the latter of which are frequently generated by fusion events. The concerted action of both genetic mechanisms, in combination with differential selective pressures at the population level, resulted in the unparalleled rapid evolution of the *KIR* gene region in two closely related macaque species. The variation of the *KIR* gene repertoire at the species and population level might have an impact on the outcome of preclinical studies with macaque models.

Introduction

Natural killer (NK) cells provide an early defense mechanism against infectious diseases and tumor formation by their ability to recognize and kill cells with aberrant MHC class I expression (1-3). This immune surveillance is modulated by killer cell immunoglobulin-like receptors (KIR), which are expressed on NK cells and subsets of T cells (4-6). These gene products are transmembrane receptors consisting of two or three extracellular domains, which can facilitate ligand interaction, and a long or short cytoplasmic tail that can utilize an intracellular immunoreceptor tyrosine-based inhibitory or activation motif (ITIM or ITAM), respectively (3, 7). In humans, the gene family encoding the *KIR* genes is located on chromosome 19 q13.4, and its complexity is reflected by allelic polymorphism, gene copy number variation (CNV), chromosomal recombination, and alternative splicing (8-11).

Comparison of the *KIR* gene cluster in humans and other primate species suggests a first round of expansion between 30 and 45 million years ago (12), which involved two progenitor genes. The *KIR3DX1* lineage is nowadays represented by a single copy in primates but expanded in cattle, whereas the *KIR3D* progenitor gene was subjected to diversification by tandem duplications, deletions, and recombinations (13, 14). This expansion resulted in a head-to-tail gene cluster encoding a broad repertoire of *KIR* genes, the overall architecture of which is conserved in primates. Species-specific diversification, however, may have resulted in differential lineage expansions and sequence variation, which is reflected by few *KIR* orthologs that are shared between distantly related primate species. Primate *KIR* genes are phylogenetically classified into

lineages based on receptor structure and ligand specificity. In humans, lineage I includes *KIR2DL4* and *KIR2DL5*, lineage II *KIR3DL1/L2/S1*, the expanded lineage III *KIR2DL1-3*, *KIR2DS1-5*, and the pseudogenes, and lineage V *KIR3DL3*, respectively. The initial expansion of lineage III members can be traced back to orangutans, and its emergence seems to have co-evolved with the presence of *HLA-C*-like genes, which are present on approximately 50% of the contemporary orangutan *MHC* haplotypes (15). In chimpanzees and humans, the lineage III *KIR* genes expanded further, and their genomic clusters comprise 17 and 13 *KIR* genes, respectively, but only four genes are considered orthologs (14). Old World Monkeys (OWM), like macaques (genus *macaca*), expanded mainly lineage II *KIR* genes (*KIR3D*), which may be associated with their expanded MHC class I repertoire (16, 17).

Macaques are geographically the most widespread non-human primates (NHP) that diversified from the human and great ape lineage about 25 million years ago, and include approximately 20 species that share a habitat spanning from northeast Africa to Asia. Rhesus and cynomolgus macaques (*Macaca mulatta*, *Macaca fascicularis*) are closely related species that diverged from each other approximately 1-3 million years ago. Rhesus macaques are distributed across south, east, and Southeast Asia, whereas cynomolgus macaques mainly inhabit the mainland and islands of Southeast Asia. Geographically distinct populations, such as the Indian, Burmese and Chinese rhesus macaques, and the insular cynomolgus macaques, emerged by means of natural barriers, and resulted in intraspecific variation. The Isthmus of Kra, which is the narrowest part of the Malaysian peninsula, separates the cynomolgus macaques that inhabit the mainland of Southeast Asia in a northern (Cambodia, Thailand, and Vietnam) and southern (Malaysian peninsula) population, and it is suggested that this geographical barrier restricts gene flow (18, 19). In Indochina, rhesus and cynomolgus macaques may come across each other, and bidirectional introgression is substantiated by shared genetic features (20-22). For example, ancestral haplotypes of the highly polymorphic *MHC* class I region are encountered in rhesus macaques and cynomolgus macaques (16, 23, 24), whereas extensive allele sharing had been documented for the *MHC* class II genes (25).

Several sequencing platforms have been used to characterize the macaque *KIR* gene region, particularly in Indian rhesus macaques (10, 26-30). Data on the *KIR* gene cluster and repertoire in other rhesus macaque populations is limited. For cynomolgus macaques, only the Mauritian animals were characterized thoroughly (31-33). This population was founded by a few animals that were introduced to the island by human interference, approximately 500 years ago, and therefore have a restricted *KIR* gene content.

Rhesus and cynomolgus macaques are used as preclinical models for many infectious and autoimmune diseases, as the immune response and pathologies reflect the human situation (34, 35). The origin of macaques, however, vary between different research facilities and might impact the disease phenotype, which is, for example, reported for SIV/AIDS studies in Indian and Chinese rhesus macaques (36, 37). The

presence or absence of certain *KIR* genes, in combination with the MHC class I ligands, have been associated with disease susceptibility in both humans (6) and macaques (38, 39). A comprehensive overview of the *KIR* gene content and repertoire of different natural macaque populations is, however, lacking, despite the potential refinement for macaque models. Therefore, we set out to analyze the *KIR* transcriptomes of cohorts of rhesus and cynomolgus macaques of different geographical origins, which probably experienced varying selective pressures. Our observations illustrate in both highly related macaque species and populations an unparalleled form of rapid evolution of *KIR* genes that is propelled by point mutations and complex chromosomal recombinations, which generate novel gene entities and result in highly variable haplotype architectures.

Methods

Samples and origin

Forty-six rhesus macaques, comprising 7 families, and 70 cynomolgus macaques, comprising 11 families, were selected from the self-sustaining colony housed at the Biomedical Primate Research Centre. During the annual health checks, EDTA or heparin whole blood samples were obtained, and PBMCs were isolated from the latter. PBMC samples from 16 Chinese rhesus macaques, comprising 7 families, were obtained from the BPRC Bio-bank.

The geographical origin of most rhesus macaques was known based on importation records, such as the families from the Indian, Chinese, and Burmese populations. Additional transcriptome data of Indian rhesus macaques was incorporated from a previous *KIR* study conducted by our lab (10). The geographical origin of the cynomolgus macaques was mainly deduced by phylogenetic comparison of mitochondrial 12S rRNA gene segments (40). With regard to this data, we defined three cynomolgus macaque populations, which originated from the mainland of Malaysia, from the Indonesian and Malaysian islands, and from Mauritius. The mainland population was further divided into populations north and south of the Isthmus of Kra. The origin of three cynomolgus macaques (Ji0603077, J15028, J16019) could not be determined unambiguously. In addition, previously reported KIR haplotypes of Mauritian cynomolgus macaques were added to the analysis (33).

RNA isolation and KIR transcriptome amplification

Total RNA was extracted directly from EDTA whole blood samples or from ± 15 x 10^6 PBMCs with RNeasy Mini Kit (Qiagen, Valencia, CA) in accordance with the manufacturer's instructions. First-strand cDNA was synthesized with the RevertAid First Strand cDNA Synthesis Kit (Invitrogen, Carlsbad, CA) using oligo(dT)18 primers. Genomic DNA was extracted from EDTA whole blood samples by a standard salting-out procedure, or from ± 15 x 10^6 PBMCs with an AllPrep RNA/DNA Mini Kit (Qiagen) according to the manufacturer's instructions.

Full-length KIR transcripts were obtained by amplification of total cDNA with a *KIR2DL04*-specific and two *KIR1D/KIR3D*-generic primer sets, in accordance with a previously reported protocol (10). These primer sets were cross-reactive for the different rhesus and cynomolgus macaque populations. PCR products were size-selected by gel-electrophoresis (± 1250 bp) and purified using a GeneJet Gel extraction kit (Invitrogen). The samples were pooled and purified twice using AMPure XP beads (Beckman Coulter, Woerden, the Netherlands) at a 1:1 bead to DNA volume ratio. The DNA concentration of purified pooled samples (>1 µg total DNA) was measured using the Qubit dsDNA HS assay kit and Qubit 2.0 Fluorometer (Thermo Fisher Scientific).

PacBio SMRTbell libraries were generated according to Pacific Biosystems "Procedure & Checklist - Amplicon Template Preparation and Sequencing", and sequencing was performed using a PacBio Sequel platform with a 10h movie time using sequencing kit versions 2.0, 2.1, and 3.0, which was performed by the Leiden Genome Technology Center.

PacBio data analysis

Circular consensus sequences were selected for high read quality (value of 0.99 or higher), and demultiplexed based on unique barcoding.

Geneious Prime 2019 software was used to map the CCS reads to a database, consisting of reported rhesus macaque and novel cynomolgus macaque KIR sequences, to identify 100% matching reads (100% overlap, 0% mismatch, maximum ambiguity = 1). The unused reads of related animals were grouped, and were *de novo* assembled. The consensus of each *de novo* contig was trimmed for the primer sequence, and phylogenetically aligned with the rhesus and cynomolgus macaque database. *De novo* sequences were confirmed when shown to segregate or when identified in two separate PCRs, and were subsequently submitted to the European Nucleotide Archive and assigned an accession number (https://www.ebi.ac.uk/ena/).

Macaque KIR nomenclature

The nomenclature of the KIR transcripts in rhesus and cynomolgus macaques follows the general guidelines of the KIR nomenclature report for non-human primates (NHP) (41). In brief, the name of the gene indicates the number of domains (1D, 2D, or 3D) and the signaling function (S or L). The inclusion of a "W" implies a workshop gene, which indicates a gene that is divergent on the basis of phylogenetic analysis but lacks sufficient reliability due to low frequency or to the absence of genomic sequencing or family studies. The inclusion of "Q" indicates that it is questionable whether the transcripts are feasible. Two digits distinguish the different genes, and an asterisk followed by three digits distinguishes alleles. Two additional digits indicate synonymous variation.

Novel cynomolgus macaque *KIR* sequences were compared with a database of 342 reported rhesus macaque sequences (10, 26, 29, 30) and newly identified transcripts by phylogenetic analysis, using the Neighbor-Joining Tree building method (best tree mode) in MacVector software (MacVector, Cambridge, U.K.). Phylogenetic clusters were confirmed by the Maximum Likelihood Comparison and Neighbor-Joining Tree

building methods in MEGA7 software, and all methods provided similar trees. Rhesus and cynomolgus macaque *KIR* sequences that clustered together with a close phylogenetic distance were considered orthologs, and received matching *KIR* gene names. Clusters of *Mafa-KIR* sequences that diverged from the other sequences according to sequence comparison and phylogenetic analysis received a workshop number. In addition, workshop numbers were assigned to cynomolgus macaque *KIR* genes that were thought to be the result of recombination events, as these are considered novel entities. In contrast, recombinant *KIR* genes in rhesus macaques are named after an allele of the gene that contributes the largest gene segment, as is described for this species in the NHP nomenclature report (41). The previously reported 46 *Mafa-KIR* sequences (33), all of which originated from the Mauritian cynomolgus macaque population, were also named.

Macaque KIR haplotype origin and gene frequencies

The origin of each *KIR* haplotype was categorized per macaque population. The populational origin of the rhesus macaques determined the haplotype origin, as none of the rhesus macaques had parents from different populations. The *KIR* haplotypes defined in cynomolgus macaques that had their roots in the mainland of Malaysia (north or south), or in the Malaysian/Indonesian islands, or in Mauritius were categorized based on the defined origin. In cynomolgus macaques with mixed roots (parents from the mainland and from islands), the origin of the *KIR* haplotypes was determined by the sequencing of parental genomic DNA, the origin of which was known, using a *Mafa-KIR3DL20* exon 4-specific primer set (forward: 5'-GAAGAGACGGTCATCCTGCAGT-3'; reverse: 5'-ACTCCCCCTATGTGTTGTCAGC-3') and a *Mafa-KIR1D* exon 4-specific primer set (forward: 5'-GAAGAGACGGTCATCCTGCAGT-3'; reverse: 5'-ACTCCCCCTATGTGTTGTCAGC-3'). Thermal cycling conditions were denaturation at 98°C for 2 min, followed by 32 cycles of 98°C for 20 s, 63°C for 25 s, and 72 °C for 1 min. Amplicons of approximately 180 bp were size-selected by gel electrophoresis and purified using a GeneJet Gel extraction kit (Invitrogen). Sanger sequencing was used, and the populational haplotype origin could be determined on the basis of three single nucleotide polymorphisms.

The frequency of a *KIR* gene in rhesus and cynomolgus macaques, or in one of the populations, was determined based on the presence of at least a single copy on a haplotype, the origin of which was determined, rather than on the presence of the gene in an individual.

Results

Definition of rhesus and cynomolgus macaque populations and their KIR transcriptomes

The *KIR* transcriptomes of 62 rhesus and 70 cynomolgus macaques covering different populations were subjected to analysis (**Figure 1**) (10). All macaque samples belong to families that comprised two or more individuals, which allowed us to confirm the segregation of alleles but also to define haplotypes (**Figure 2 and 3**). The origin of the rhesus macaques was documented thoroughly, and included Burmese (n = 14), Chinese (n = 16), and Indian (n = 32) origins (**Figure 1**). Based on the phylogeny of

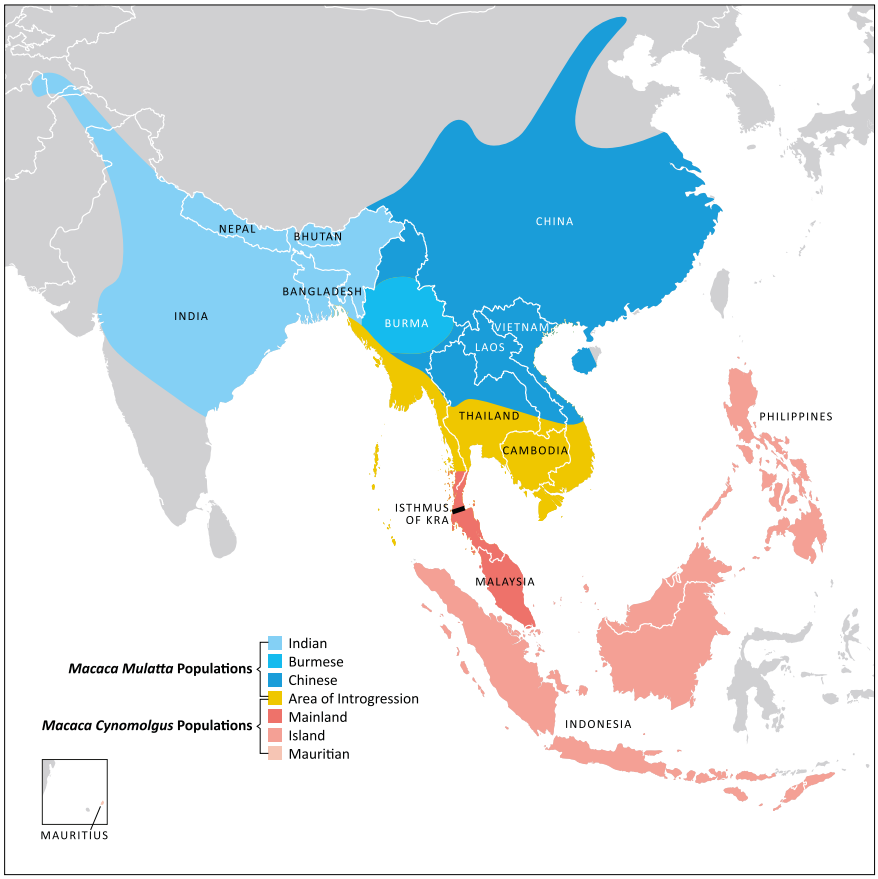


Figure 1. A schematic overview of the habitats of different rhesus and cynomolgus macaque populations. Three rhesus macaque populations (Indian, Burmese, and Chinese) are indicated by different blue colors, whereas four cynomolgus macaque populations (mainland: north and south of the Isthmus of Kra, Malaysian/ Indonesian islands, and Mauritius) are highlighted by red colors. The rhesus and cynomolgus macaque habitats include a hybrid zone (illustrated in yellow), in which introgression between the two species occurs. Mauritius is located approximately 4030 miles out of the South-east African coast, and is illustrated in a separate box.

mtDNA sequences (40), origins of the cynomolgus macaques were mapped to the mainland of Southeast Asia (n = 26), the Malaysian/Indonesian islands (n = 4), or Mauritius. The mainland population could be further divided into populations north (n = 23) and south (n = 3) of the Isthmus of Kra (**Figure 1**). For 19 cynomolgus macaques, a mixed origin was documented, whereas for 21 animals only the origin of a single parent could be determined. To expand our population panel, we included previously reported *KIR* transcriptome data from 30 Indian rhesus macaques (10) and 30 Mauritian cynomolgus macaques (33). Altogether, three rhesus and four cynomolgus macaque populations were subjected to comparison for their *KIR* repertoire.

Allele discovery: abundant levels of species-specific allelic variation in macaques

Up to now, 342 distinct rhesus macaque *KIR* alleles that were mainly isolated from Indian animals have been identified (10, 26, 29, 30). In the current cohort that comprises 32 Indian rhesus macaques, again another 48 unreported *KIR* alleles were discovered, indicating extensive allelic variation within this population. All Indian rhesus macaque *KIR* alleles could be clustered into 22 different *KIR* genes (**Table I**). From the Burmese and Chinese cohorts, 73 and 117 novel *KIR* alleles were isolated, respectively, which clustered to previously reported but also newly discovered *KIR* gene entities (**Table I**). During the course of this study, 34 rhesus macaque *KIR* genes were defined, which comprised 238 novel alleles, and 64 reported *Mamu-KIR* alleles were confirmed (Supp. **Figure 1**). The emergence of one of the novel alleles was observed in rhesus macaque R04104, which is expected to be homozygous for the *KIR* region, as it ought to receive two copies of *Mamu-KIR3DL05*006:01* via the H21-A haplotype (**Figure 2** and **3**). However, one copy of the *Mamu-KIR3DL05*006:01* allele shows nonsynonymous mutations at two positions in the D1 domain (T > C and G > T), thereby generating a novel allele, designated *Mamu-KIR3DL05*032*. This *de novo* allele segregated with its corresponding haplotype (H21-B) into two offspring of R04014, and its existence was further substantiated by independent Sanger sequencing (Supp. **Figure 2**).

Most allelic variation is controlled by *Mamu-KIR3DL07*, *-KIR3DL20*, and *-KIR3DL01*, and to a lesser extent by *-KIR2DL04*. The Indian and Burmese populations share four *KIR* alleles, whereas only a single allele was shared between the Indian and Chinese (*Mamu-KIR3DS06*016*), the Burmese and Chinese populations (*Mamu-KIR3DL05*007:01*), and all three populations (*Mamu-KIR3DL01*019:03*) (**Figure 4**).

Knowledge of the *KIR* cluster in cynomolgus macaques is mainly confined to the artificially introduced Mauritian population, and 49 alleles are documented (31, 33). In the current cohort from different populations, we identified 267 novel alleles that clustered into 55 distinct *KIR* genes (**Table II, Supp. Figure 1**). In addition, 10 of the 46 previously reported *Mafa-KIR* sequences identified in Mauritian cynomolgus macaques were confirmed (31, 33). The highest level of allelic variation was observed for *Mafa-KIR3DL20* and *-KIR1D*, followed by *-KIR2DL04*, *-KIR3DL01*, and *-KIR3DL07*. The different populations seem to have highly unique allelic *KIR* repertoires. A single allele was shared between the northern mainland and the Indonesian/Malaysian populations

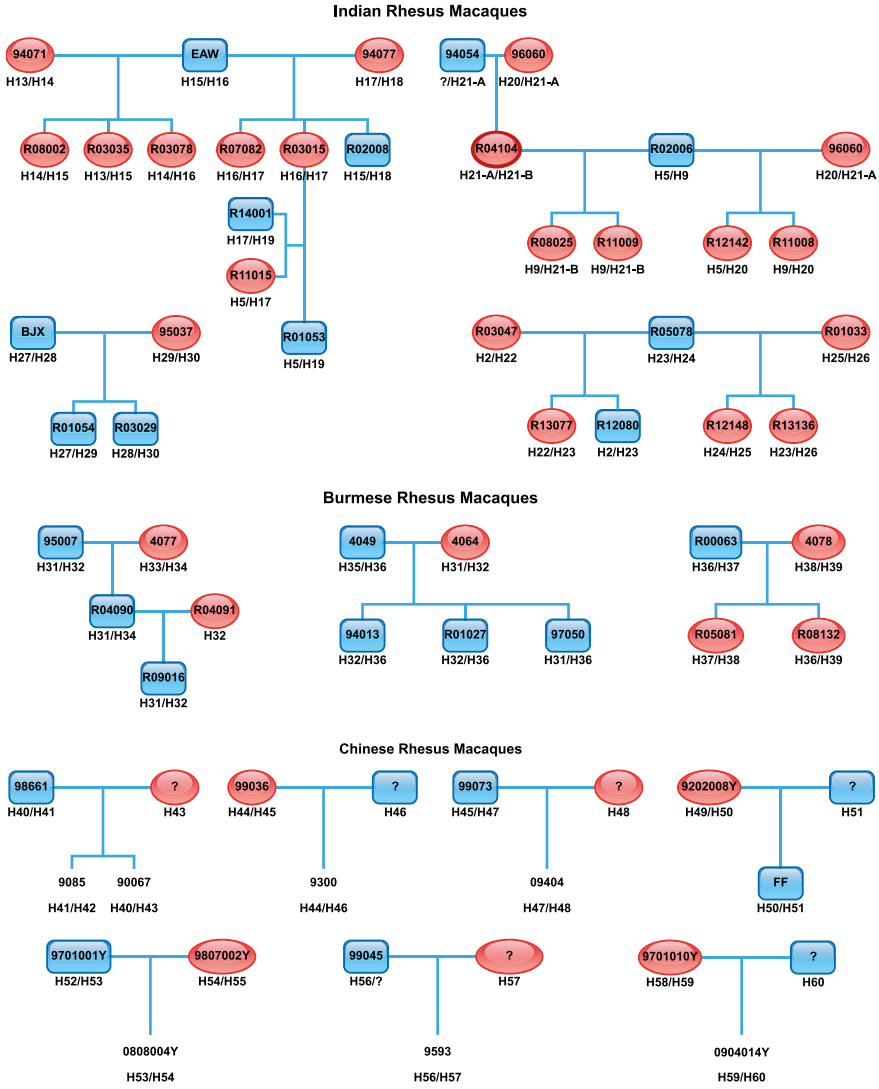


Figure 2. Rhesus macaque pedigrees. Fourteen rhesus macaque families are depicted and categorized by origin. Sires are indicated by blue squares, and dames by red ovals. For some offspring, the sex could not be determined. In six Chinese families, PBMC samples could be obtained from only a single parent, whereas the other parent is indicated with a question mark. Haplotype numbers are given for each animal and correspond to Figure 5.

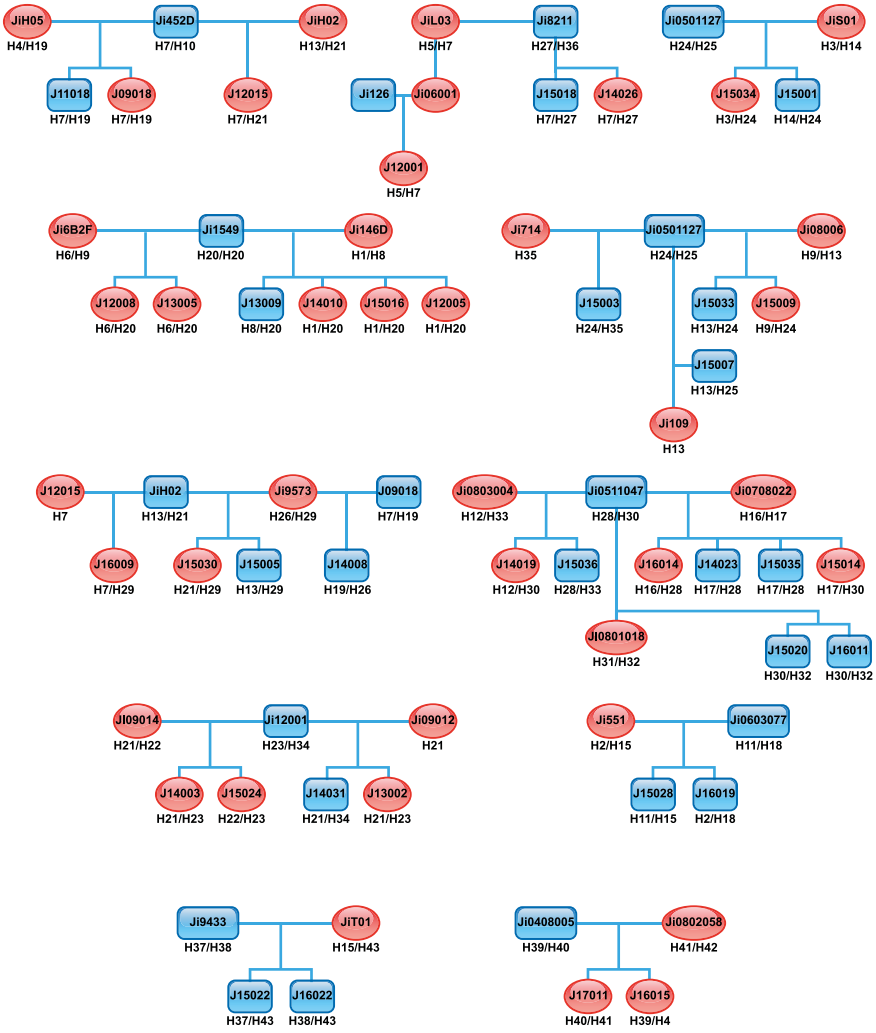


Figure 3. Cynomolgus macaque pedigrees. Eleven cynomolgus macaque families are depicted. Sires are indicated by blue squares, and dames by red ovals. Cynomolgus macaques from different populations are mixed in the families. Haplotype numbers are given for each animal and correspond to Figure 6.

(*Mafa-KIR3DLW23*001*), and the southern mainland and Mauritian populations (*Mafa-KIR2DL04*002*), whereas two alleles were in common between the Indonesian/Malaysian islands and Mauritian populations (*Mafa-KIR3DLW13*003*, *Mafa-KIR3DLW26*001*), and the southern mainland, the Indonesian/Malaysian islands, and Mauritian populations (*Mafa-KIR1D*030Q, Mafa-KIR3DL20*002*) (**Figure 4**).

Table I. Overview of known and novel *KIR* alleles in rhesus macaques, indicated per *KIR* gene and population. Known alleles were mainly reported in the Indian population [10, 25, 28, 29].

Gene	Known alleles	Novel alleles				Total
		Indian	Burmese	Chinese	Total novel	
KIR1D	5	2	3	3	8	13
KIR2DL04	36	0	3	10	13	49
KIR3DL20	30	10	8	15	33	63
KIR3DL01	42	3	7	13	23	65
KIR3DL02	15	0	2	4	6	21
KIR3DLW03	9	1	3	13	17	26
KIR3DL04	4	0	0	0	0	4
KIR3DL05	28	3	3	5	11	39
KIR3DL06	2	0	4	2	6	8
KIR3DL07	34	11	11	17	39	73
KIR3DL08	19	5	3	4	12	31
KIR3DL10	10	1	1	1	3	13
KIR3DL11	13	1	0	5	6	19
KIR3DLW12	0	0	0	2	2	2
KIR3DLW14	0	1	1	2	4	4
KIR3DLW17	0	0	1	2	3	3
KIR3DLW18	0	0	1	0	1	1
KIR3DLW25	0	0	1	0	1	1
KIR3DS01	8	1	0	1	2	10
KIR3DS02	21	2	8	4	14	35
KIR3DS03	5	1	0	0	1	6
KIR3DS04	11	1	2	2	5	16
KIR3DS05	10	0	1	0	1	11
KIR3DS06	14	2	1	4	7	21
KIR3DSW07	4	0	0	1	1	5
KIR3DSW08	13	2	2	3	7	20
KIR3DSW09	9	1	0	1	2	11
KIR3DSW10	0	0	1	0	1	1
KIR3DSW16	0	0	1	0	1	1
KIR3DSW18	0	0	1	0	1	1
KIR3DSW20	0	0	1	0	1	1
KIR3DSW21	0	0	2	2	4	4
KIR3DSW32	0	0	0	1	1	1
KIR3DSW34	0	0	1	0	1	1
Total	**342**	**48**	**73**	**117**	**238**	**580**

To sum up, 579 *KIR* alleles were identified in the rhesus and cynomolgus macaque populations studied. Only two alleles were shared between both highly related species: namely, *Mamu-KIR3DLW12*002*/*Mafa-KIR3DLW12*006* and *Mamu-* and *Mafa-KIR3DLW18*001* (**Figure 4**). The low number of allele sharing between the macaque species as well as the different populations suggests fast evolution. This is within

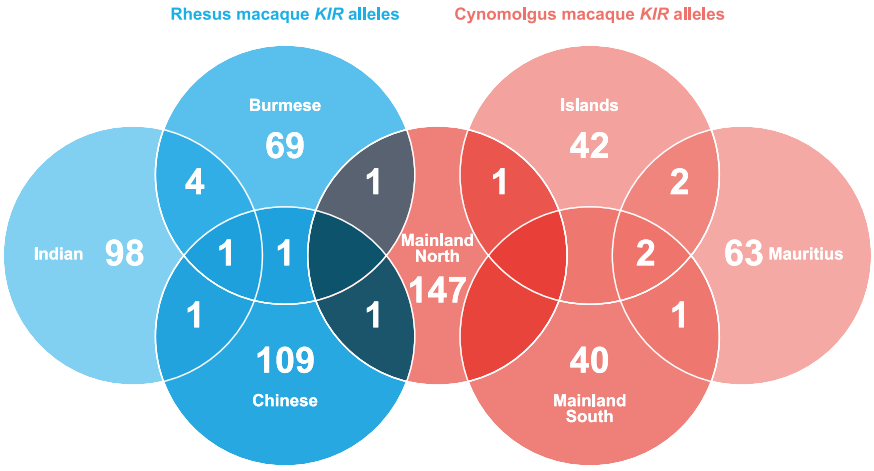


Figure 4. KIR allele distribution in rhesus and cynomolgus macaque populations. The three rhesus macaque populations are indicated in blue circles, and the four cynomolgus macaque populations are depicted in red circles. The total number of uniquely identified *KIR* alleles is provided per population and the number of shared *KIR* alleles is indicated for the involved populations.

lineages mainly mediated by point mutations, and contrasts the extensive sharing documented for *MHC* class II and, to a much lesser extent, for *MHC* class I alleles (42).

New KIR genes in macaques are generated by recombination

Considering a shared ancestor living 1-3 million years ago, one might expect highly similar repertoires of orthologous *KIR* genes in rhesus and cynomolgus macaques, as is observed for the closely related Bornean and Sumatran orangutan species (15). Apparently, however, this is not the case in both macaque species, as their *KIR* gene repertoires possess species-specific and a differential number of *KIR* gene moieties. Moreover, the 34 rhesus and 55 cynomolgus macaque *KIR* genes that are defined by sequence comparison and phylogenetic analysis indicate a greater expansion of the *KIR* gene repertoire in macaques as compared to humans and other primate species, for which 17 or fewer *KIR* genes were identified (41, 43, 44). The question to be answered therefore is how are new *KIR* genes generated. One mechanism that might explain the expanded macaque *KIR* gene repertoire is the occurrence of abundant recombination events, which result in the formation of hybrid genes composed of segments from two different *KIR* genes (**Tables III** and **IV**). Along with others, we found evidence of similar events in humans (10, 45), although this mechanism seems to happen more frequently in macaques. In rhesus macaques, hybrid *KIR* genes are named after the allele that contribute the largest segment (41). For example, multiple entities have a large *Mamu-KIR3DL07* segment, which is found in conjunction with a smaller segment of *-KIR3DL05*, *-KIR3DL08*, or *-KIR3DSW08*, but all are named and listed as alleles of *Mamu-KIR3DL07* (**Table III**). Another peculiar recombination event

Table II. Overview of known and novel *KIR* alleles in cynomolgus macaques, indicated per *KIR* gene.

Gene	Known alleles	Novel alleles	Total
KIR1D	4	30	34
KIR2DL04	3	38	41
3DL20	1	33	34
3DL01	2	14	16
3DLW03	0	4	4
3DL05	0	5	5
3DL06	0	1	1
3DL07	3	15	18
3DL11	2	16	18
3DLW12	3	11	14
3DLW13	3	6	9
3DLW14	2	6	8
3DLW15	1	3	4
3DLW16	2	2	4
3DLW17	0	3	3
3DLW18	0	1	1
3DLW19	1	1	2
3DLW21	0	1	1
3DLW22	0	1	1
3DLW23	0	1	1
3DLW24	1	0	1
3DLW25	2	2	4
3DLW26	1	0	1
3DLW27	0	1	1
3DLW28	1	0	1
3DLW29	0	1	1
3DLW30	0	1	1
3DS02	0	2	2
3DS04	0	1	1
3DS06	0	3	3
3DSW07	0	3	3
3DSW08	0	3	3
3DSW10	1	3	4
3DSW11	0	4	4
3DSW12	1	9	10
3DSW13	2	0	2
3DSW14	0	2	2
3DSW15	0	9	9
3DSW16	0	3	3
3DSW17	3	1	4
3DSW18	0	2	2
3DSW19	0	2	2
3DSW20	1	3	4
3DSW21	2	5	7
3DSW22	1	2	3
3DSW23	0	1	1
3DSW24	2	3	5
3DSW25	1	0	1
3DSW26	0	1	1
3DSW27	1	1	2
3DSW28	1	0	1
3DSW29	1	1	2
3DSW30	0	3	3
3DSW31	0	2	2
3DSW33	0	1	1
Total	**49**	**267**	**316**

resulted in *Mamu-KIR3DS04*011,* the extracellular domains (exons 1-5) of which originate from *-KIR3DS04*, whereas the cytoplasmic tail is similar to exons 6-9 of *-KIR3DL07*. The name is therefore somewhat confusing as this gene is listed as an allele of *Mamu-KIR3DS04*, although it encodes an inhibitory cytoplasmic tail. In rhesus macaques, at least 19 hybrid *KIR* genes were generated by recombination events (**Table III**). It would seem that for some of these hybrids the nomenclature is in need of attention (46). From a more general and functional perspective, hybrid gene entities could encode novel genes with potentially distinct functional features, due to differential combinations of ligand binding domains and signal transduction elements.

In cynomolgus macaques, at least seven hybrid *KIR* genes were detected (**Tables II** and **IV**). For example, the first 6 exons of *Mafa-KIR3DSW21* are highly similar (98-99%) to those of *-KIR3DL07*, whereas the transmembrane region and cytoplasmic tail of

Table III. Novel gene entities that are generated by chromosomal recombination events in rhesus macaques. The gene donors and corresponding donated segments are indicated. For some novel entities, only a single donor could be identified. The novel entities are named as an allele of the gene that contributed the largest segment.

	Gene segment 1		Gene segment 2	
Novel entity (allele name)	**Gene donor 1**	**Segment**	**Gene donor 2**	**Segment**
*Mamu-KIR3DL01*054*	*Mamu-KIR3DL05*	Exons 1- 3	*Mamu-KIR3DL01*	Exons 4- 9
*Mamu-KIR3DL02*005, *011*	*Mamu-KIR3DL02*	Exons 1- 6	*Mamu-KIR3DL01*	Exons 7- 9
*Mamu-KIR3DL02*006, *010*	*Mamu-KIR3DL02*	Exons 1- 7	*Mamu-KIR3DL01*	Exons 8- 9
*Mamu-KIR3DL02*016*	*Mamu-KIR3DL10*	Exons 1- 4	*Mamu-KIR3DL02*	Exons 5- 9
*Mamu-KIR3DLW03*023*	*Unknown donor*	Exons 1- 3	*Mamu-KIR3DLW03*	Exons 4- 9
*Mamu-KIR3DL05*029, *030, *033*	*Mamu-KIR3DL05*	Exons 1- 7	Unknown donor	Exons 8- 9
*Mamu-KIR3DL07*042*	*Mamu-KIR3DL07*	Exons 1- 5	*Mamu-KIR3DL08*	Exons 6- 9
*Mamu-KIR3DL07*045*	*Mamu-KIR3DL05*	Exons 1- 3	*Mamu-KIR3DL07*	Exons 4- 9
*Mamu-KIR3DL07*056*	*Mamu-KIR3DSW08*	Exons 1-3	*Mamu-KIR3DL07*	Exons 4- 9
*Mamu-KIR3DL07*064*	*Mamu-KIR3DL05*	Exons 1- 4	*Mamu-KIR3DL07*	Exons 5- 9
*Mamu-KIR3DL07*065*	Unknown donor	Exons 1- 3	*Mamu-KIR3DL07*	Exons 4- 9
*Mamu-KIR3DL08*018, *019, *020*	*Mamu-KIR3DL02*	Exons 1- 3	*Mamu-KIR3DL08*	Exons 4- 9
*Mamu-KIR3DL08*021*	*Mamu-KIR3DL01*	Exons 1- 4	*Mamu-KIR3DL08*	Exons 5- 9
*Mamu-KIR3DL20*030*	*Mamu-KIR3DL20*	Exons 1- 7	*Mamu-KIR2DL04*	Exons 8- 9
*Mamu-KIR3DL20*044*	*Mamu-KIR3DL20*	Exons 1- 7	*Mamu-KIR1D*	Exons 8- 9
*Mamu-KIR3DS02*012, *029*	*Mamu-KIR3DS02*	Exons 1- 4	*Mamu-KIR3DSW09*	Exons 5- 9
*Mamu-KIR3DS04*011*	*Mamu-KIR3DS04*	Exons 1- 5	*Mamu-KIR3DL07*	Exons 6- 9
*Mamu-KIR3DS06*019*	*Mamu-KIR3DSW07*	Exons 1- 3	*Mamu-KIR3DS06*	Exons 4- 9
*Mamu-KIR3DSW09*005, *011*	*Mamu-KIR3DSW08*	Exons 1- 3	Unknown donor	Exons 4- 9

-*KIR3DSW21* is identical to -*KIR3DSW12*. This suggests that *Mafa-KIR3DSW21* may interact with similar ligands as -*KIR3DL07*, but that it transduces activating instead of inhibitory signals. Seven *Mafa-KIR3DSW21* alleles are identified (**Table II**), suggesting positive selection for variation on the gene products generated by this recombination event.

The origin of both segments could not be identified for all hybrid *KIR* genes. *Mamu-KIR3DL05*029/*030/*033*, *Mafa-KIR3DLW24*, *Mafa-KIR3DLW29*, *Mafa-KIR3DSW18*, and *Mafa-KIR3DSW20* seem to have segments of *Mamu-KIR3DL05*, *Mafa-KIR3DLW12*,

Table IV. Novel gene entities that are generated by chromosomal recombination events in cynomolgus macaques. The gene donors and corresponding donated segments are indicated. For some novel entities, only a single donor could be identified. The novel entities received a gene workshop number.

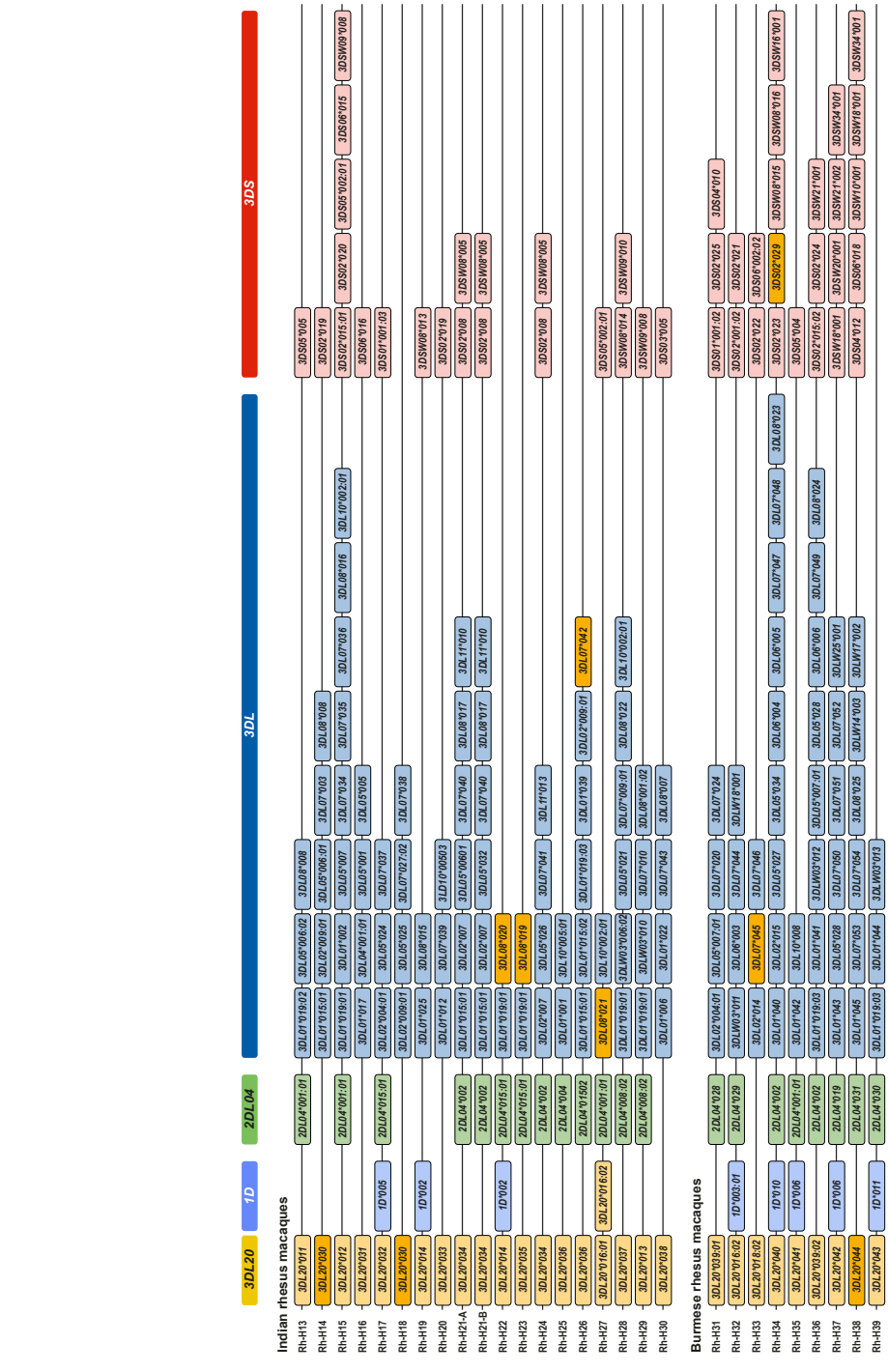
	Gene segment 1		Gene segment 2	
Novel entity (gene name)	**Gene donor 1**	**Segment**	**Gene donor 2**	**Segment**
Mafa-KIR3DLW24	Unknown donor	Exons 1- 3	*Mafa-KIR3DLW12*	Exons 4- 9
Mafa-KIR3DLW26	*Mafa-KIR3DSW15*	Exons 1- 3	*Mafa-KIR3DLW27*001*	Exons 4- 9
Mafa-KIR3DLW27	*Mafa-KIR3DSW22*001*	Exons 1- 4	*Mafa-KIR3DLW26*001*	Exons 5- 9
Mafa-KIR3DLW29	Unknown donor	Exons 1- 4	*Mafa-KIR3DLW13*	Exons 5- 9
Mafa-KIR3DSW18	*Mafa-KIR3DSW17*	Exons 1- 4	Unknown donor	Exons 5- 9
Mafa-KIR3DSW20	*Mafa-KIR3DSW19*	Exons 1- 4	Unknown donor	Exons 5- 9
Mafa-KIR3DSW21	*Mafa-KIR3DL07*	Exons 1- 6	*Mafa-KIR3DSW12*	Exons 7- 9

Mafa-KIR3DLW13, *Mafa-KIR3DSW17*, and *Mafa-KIR3DSW19*, respectively, but it was not possible to trace the donor of the other segment (**Tables III** and **IV**). This indicates that when more sequences become available, additional hybrid *KIR* gene entities and segments are likely to be defined in macaques.

Within the macaque *KIR* repertoire studied, 24 macaque *KIR* genes are highly similar, and are considered to be orthologs. These genes most likely represent a single locus in both species, although it is too early to elucidate their exact physical location, as the relevant genomic studies are in progress. One would expect that the number of 24 orthologs shared between two closely related macaque species reflects common ancestry, whereas the relatively high number of species-specific *KIR genes* indicates the rapid generation of novel gene entities, which can, in part, be explained by abundant recombination.

Definition of macaque KIR haplotypes

The family-based study design resulted in the thorough characterization of 49 rhesus and 43 cynomolgus macaque *KIR* haplotypes (**Figures 5** and **6**), which are categorized on the basis of geographical origin of the analysed animals (**Figure 1**). The rhesus macaque *KIR* haplotypes are referred to as Rh-H13 to Rh-H60, consecutively to the 12 previously reported haplotypes (**Figure 5**) (10). Cynomolgus macaque *KIR* haplotypes are referred to as Cy-H1 to Cy-H43, whereas the previously reported Mauritian chromosomal KIR configurations are listed as K1 to K8 (**Figure 6**) (33). All these haplotypes display extensive copy number variation. The rhesus macaque haplotypes encoded 4 to 17 *KIR* transcripts, whereas the cynomolgus macaque equivalents encoded 3 to 13 *KIR* transcripts. *KIR3DL20* was identified on most macaque haplotypes, except for haplotype Cy-H39, and seems to be absent on the haplotypes of the Mauritian animals. However, we assume that *KIR3DL20* should be considered a framework gene in macaques, and that a few transcripts were missed due to primer


3DL20
1D
2DL04
3DL
3DS
Indian rhesus macaques
Burmese rhesus macaques

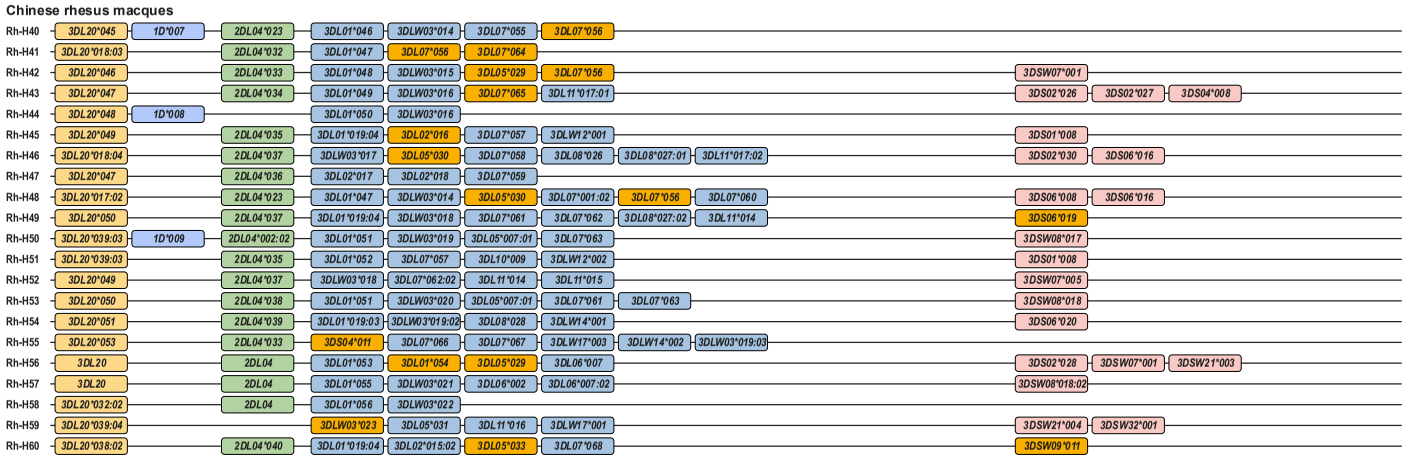
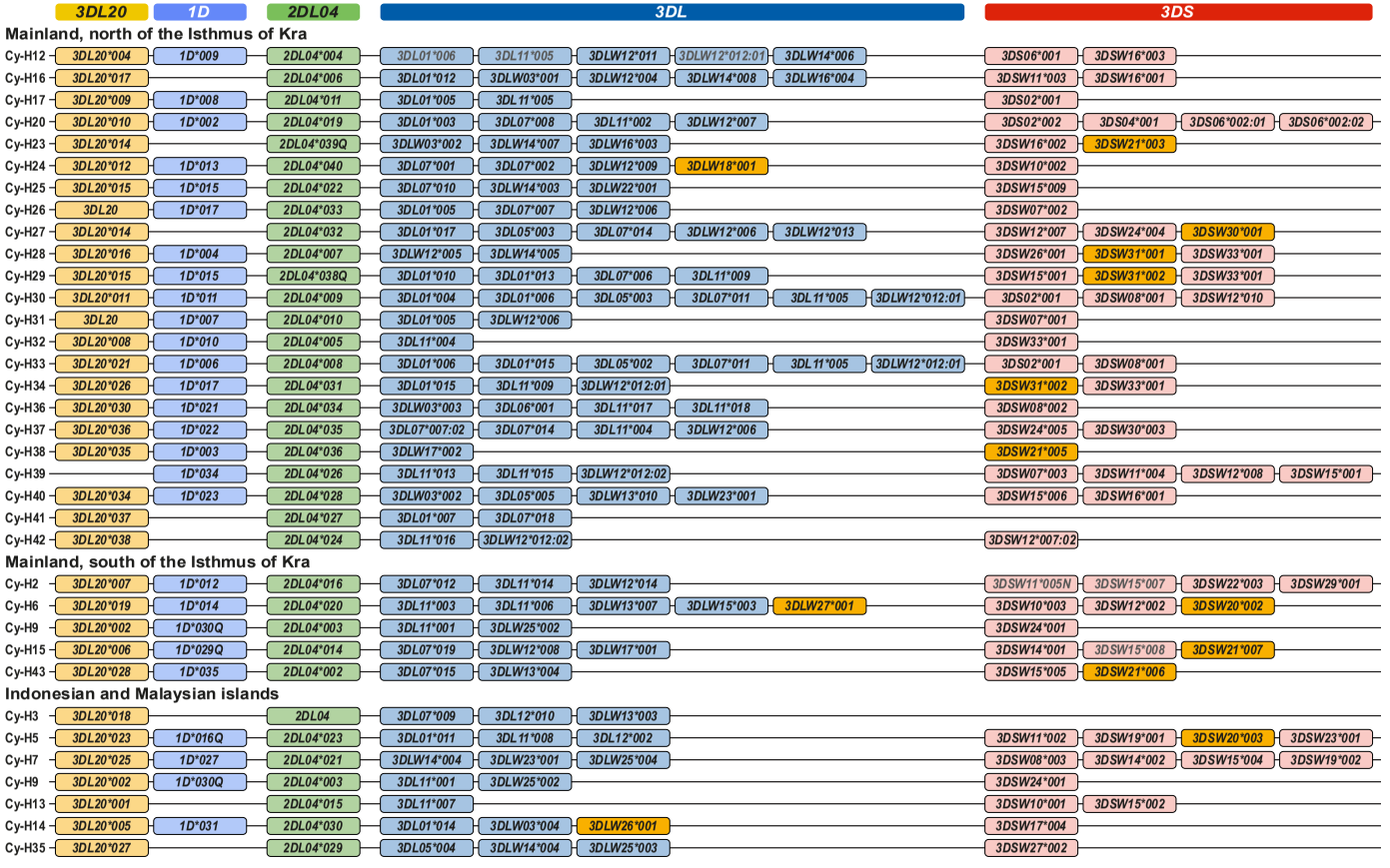


Figure 5. Rhesus macaque *KIR* haplotypes at the transcription level. A schematic overview of 49 defined rhesus macaque *KIR* haplotypes, categorized by population. Most of the haplotypes expressed *Mamu-KIR3DL20* (yellow boxes). Expression of *Mamu-KIR1D* and *Mamu-KIR2DL04* is indicated in light blue and green boxes, respectively. Inhibitory lineage II *KIR* genes are illustrated by blue boxes, whereas activating *KIR* genes of the same lineage are depicted by red boxes. Recombinant genes, for which there is evidence that they are hybrids consisting of segments of two different genes, are indicated by orange boxes. For several haplotypes, the presence of a gene was determined, but could not be distinguished at the allele level. The lineage II *KIR* genes are depicted in random order, whereas the physical locations of *Mamu-KIR3DL20*, *Mamu-KIR1D*, and *Mamu-KIR2DL04* are deduced from a genomically sequenced macaque haplotype (28). Haplotypes H21-A and H21-B are similar, except for the *de novo Mamu-KIR3DL05*032* allele.


3DL20
1D
2DL04
3DL
3DS
Mainland, north of the Isthmus of Kra
Cy-H12 3DL20*004 1D*009 2DL04*004 3DL01*006 3DL11*005 3DLW12*011 3DLW12*012:01 3DLW14*006 3DS06*001 3DSW16*003
Cy-H16 3DL20*017 2DL04*006 3DL01*012 3DLW03*001 3DLW12*004 3DLW14*008 3DLW16*004 3DSW11*003 3DSW16*001
Cy-H17 3DL20*009 1D*008 2DL04*011 3DL01*005 3DL11*005 3DS02*001
Cy-H20 3DL20*010 1D*002 2DL04*019 3DL01*003 3DL07*008 3DL11*002 3DLW12*007 3DS02*002 3DS04*001 3DS06*002:01 3DS06*002:02
Cy-H23 3DL20*014 2DL04*039Q 3DLW03*002 3DLW14*007 3DLW16*003 3DSW16*002 3DSW21*003
Cy-H24 3DL20*012 1D*013 2DL04*040 3DL07*001 3DL07*002 3DLW12*009 3DLW18*001 3DSW10*002
Cy-H25 3DL20*015 1D*015 2DL04*022 3DL07*010 3DLW14*003 3DLW22*001 3DSW15*009
Cy-H26 3DL20 1D*017 2DL04*033 3DL01*005 3DL07*007 3DLW12*006 3DSW07*002
Cy-H27 3DL20*014 2DL04*032 3DL01*017 3DL05*003 3DL07*014 3DLW12*006 3DLW12*013 3DSW12*007 3DSW24*004 3DSW30*001
Cy-H28 3DL20*016 1D*004 2DL04*007 3DLW12*005 3DLW14*005 3DSW26*001 3DSW31*001 3DSW33*001
Cy-H29 3DL20*015 1D*015 2DL04*038Q 3DL01*010 3DL01*013 3DL07*006 3DL11*009 3DSW15*001 3DSW31*002 3DSW33*001
Cy-H30 3DL20*011 1D*011 2DL04*009 3DL01*004 3DL01*006 3DL05*003 3DL07*011 3DL11*005 3DLW12*012:01 3DS02*001 3DSW08*001 3DSW12*010
Cy-H31 3DL20 1D*007 2DL04*010 3DL01*005 3DLW12*006 3DSW07*001
Cy-H32 3DL20*008 1D*010 2DL04*005 3DL11*004 3DSW33*001
Cy-H33 3DL20*021 1D*006 2DL04*008 3DL01*006 3DL01*015 3DL05*002 3DL07*011 3DL11*005 3DLW12*012:01 3DS02*001 3DSW08*001
Cy-H34 3DL20*026 1D*017 2DL04*031 3DL01*015 3DL11*009 3DLW12*012:01 3DSW31*002 3DSW33*001
Cy-H36 3DL20*030 1D*021 2DL04*034 3DLW03*003 3DL06*001 3DL11*017 3DL11*018 3DSW08*002
Cy-H37 3DL20*036 1D*022 2DL04*035 3DL07*007:02 3DL07*014 3DL11*004 3DLW12*006 3DSW24*005 3DSW30*003
Cy-H38 3DL20*035 1D*003 2DL04*036 3DLW17*002 3DSW21*005
Cy-H39 1D*034 2DL04*026 3DL11*013 3DL11*015 3DLW12*012:02 3DSW07*003 3DSW11*004 3DSW12*008 3DSW15*001
Cy-H40 3DL20*034 1D*023 2DL04*028 3DLW03*002 3DL05*005 3DLW13*010 3DLW23*001 3DSW15*006 3DSW16*001
Cy-H41 3DL20*037 2DL04*027 3DL01*007 3DL07*018
Cy-H42 3DL20*038 2DL04*024 3DL11*016 3DLW12*012:02 3DSW12*007:02
Mainland, south of the Isthmus of Kra
Cy-H2 3DL20*007 1D*012 2DL04*016 3DL07*012 3DL11*014 3DLW12*014 3DSW11*005N 3DSW15*007 3DSW22*003 3DSW29*001
Cy-H6 3DL20*019 1D*014 2DL04*020 3DL11*003 3DL11*006 3DLW13*007 3DLW15*003 3DLW27*001 3DSW10*003 3DSW12*002 3DSW20*002
Cy-H9 3DL20*002 1D*030Q 2DL04*003 3DL11*001 3DLW25*002 3DSW24*001
Cy-H15 3DL20*006 1D*029Q 2DL04*014 3DL07*019 3DLW12*008 3DLW17*001 3DSW14*001 3DSW15*008 3DSW21*007
Cy-H43 3DL20*028 1D*035 2DL04*002 3DL07*015 3DLW13*004 3DSW15*005 3DSW21*006
Indonesian and Malaysian islands
Cy-H3 3DL20*018 2DL04 3DL07*009 3DL12*010 3DLW13*003
Cy-H5 3DL20*023 1D*016Q 2DL04*023 3DL01*011 3DL11*008 3DL12*002 3DSW11*002 3DSW19*001 3DSW20*003 3DSW23*001
Cy-H7 3DL20*025 1D*027 2DL04*021 3DLW14*004 3DLW23*001 3DLW25*004 3DSW08*003 3DSW14*002 3DSW15*004 3DSW19*002
Cy-H9 3DL20*002 1D*030Q 2DL04*003 3DL11*001 3DLW25*002 3DSW24*001
Cy-H13 3DL20*001 2DL04*015 3DL11*007 3DSW10*001 3DSW15*002
Cy-H14 3DL20*005 1D*031 2DL04*030 3DL01*014 3DLW03*004 3DLW26*001 3DSW17*004
Cy-H35 3DL20*027 2DL04*029 3DL05*004 3DLW14*004 3DLW25*003 3DSW27*002

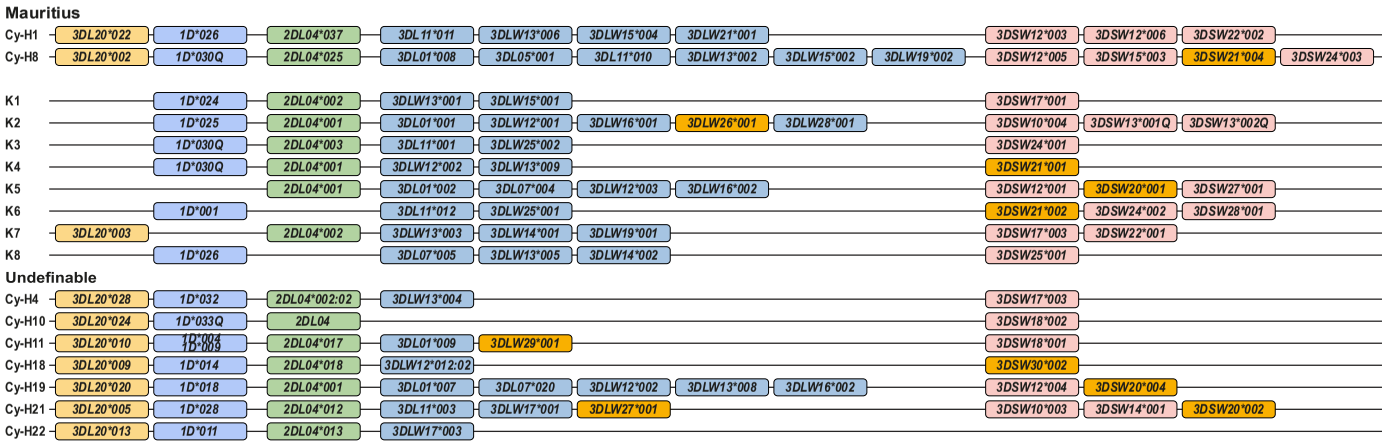


Figure 6. Cynomolgus macaque *KIR* haplotypes at the trans-cription level. A schematic over-view of 43 defined cynomolgus macaque *KIR* haplotypes, cate-gorized by population. Eight haplotypes previously reported in Mauritian animals are also illustrated (K1-K8). Expression of *Mafa-KIR3DL20*, *Mafa-KIR1D*, and *Mafa-KIR2DL04* is indicated by yellow, light blue, and green boxes, respectively. Inhibitory lineage II *KIR* genes are illust-rated by blue boxes, whereas activating *KIR* genes of the same linage are depicted by red boxes. Recombinant genes, which consist of segments of two different genes, are indicated by orange boxes. Some alleles were confirmed in two animals but were observed in a low number of reads (< 3 reads), which are indicated by grey text. The lineage II *KIR* genes are depicted in random order, whereas the physical location of the remaining genes is deduced from a previously reported complete macaque haplotype (28).

inconsistencies. For the Mauritian cynomolgus macaque, this assumption is confirmed by haplotype Cy-H9, which is identical to K3, defined by another research team, except for the presence of *Mafa-KIR3DL20*002*. Haplotype Cy-H9/K3 is identified in three populations, and may indicate an ancestral origin. *KIR2DL04* is observed on 70% and 94% of the rhesus and cynomolgus macaque *KIR* haplotypes, respectively, and represent the only reported macaque *KIR* gene that shares an apparent ortholog with humans. In humans, this gene is considered a framework gene, and there is support that this might also be the case for its cynomolgus macaque ortholog (47).

Recombination influences haplotype architecture and drives copy number variation

Chromosomal recombinations such as unequal crossing over, gene fusion and gene duplications can expand or contract a *KIR* haplotype, thereby affecting the genetic content. Two or more apparent allelic copies of a given *KIR* gene were identified on 23 of the 49 rhesus and 11 of the 43 cynomolgus macaque *KIR* haplotypes. It is likely that such genes were once orthologs, but owing to complex recombination events they might end up as paralogs. These duplications involved mainly lineage II inhibitory *KIR* genes, such as *KIR3DL01*, *KIR3DL07*, and *KIR3DL11*, but also *Mamu-KIR3DL20* (Rh-H27), *Mafa-KIR1D* (Cy-H11), and *Mafa-KIR3DSW12* (Cy-H1). In total, 15 different *KIR* genes are duplicated on the listed haplotypes, 11 of which are considered orthologs. This suggests that ancestral genes are more often subject to duplication as compared to more recently generated species-specific *KIR* genes. The most extensive copy number variation is witnessed for *Mamu-KIR3DL01* on haplotype Rh-H26 (**Figure 5**), on which four allelic copies exist.

Hybrid *KIR* genes (**Tables III** and **IV**) are associated with chromosomal recombination events, and might mark expanded and contracted *KIR* haplotypes. For example, two hybrid *Mamu-KIR3DL20* genes, composed of exons encoding the extracellular domains of *Mamu-KIR3DL20* (exons 1-7), and the cytoplasmic tail of *Mamu-KIR1D* (Rh-H38) or *Mamu-KIR2DL04* (Rh-H14, -H18), seem to coincide with a centromeric haplotype contraction. Also, haplotype Rh-H27 carries an example of a gene that consists of the first four exons of *Mamu-KIR3DL01* and the last five exons of *Mamu-KIR3DL08* (**Table III, Figure 5**). The formation of this gene probably resulted in another contracted haplotype, as only four *KIR* genes are present at the telomeric end. In the cynomolgus macaque, haplotypes Cy-H6 and-H21 seem to be expanded, marked by the presence of the recombinant genes *Mafa-KIR3DLW27* and *Mafa-KIR3DSW20*, whereas the relatively short haplotype Cy-H38 contains another hybrid gene *Mafa-KIR3DSW21*, the emergence of which might have resulted in a contraction (**Table IV, Figure 6**).

This in-frame fusion mechanism occurs rather frequently, as 21 rhesus macaque and 21 cynomolgus macaque haplotypes contain a recombinant gene (**Figure 5** and **6**), although not each hybrid gene seems to mark contraction or expansion. Thus, the *KIR* gene cluster in both macaque species seems to be subjected to frequent gene duplications and chromosomal recombination events, which not only generate novel gene entities but also result in a differential *KIR* haplotype architecture.

KIR gene frequencies differ between species

The occurrence of at least 24 orthologs in both macaque species is most likely due to the sharing of a common ancestor, but introgression between the two species may also have an impact. The frequency of these orthologs, however, differs considerably between both macaque species (**Figure 7**). The orthologous genes that are encountered more frequently on rhesus than on cynomolgus macaque haplotypes are *KIR3DL01*, *KIR3DLW03*, *KIR3DL05*, *KIR3DL06, KIR3DL07*, *KIR3DS02*, *KIR3DS06*, and *KIR3DSW08*. It is noted that for these genes the allelic variation is higher in rhesus macaques than in cynomolgus macaques (**Tables I** and **II**).

Other orthologs are more often present in cynomolgus macaques, such as *KIR3DL11*, *KIR3DLW12*, *KIR3DLW14*, *KIR3DLW25*, *KIR3DSW10*, *KIR3DSW20*, and *KIR3DSW21*, that, with the exception of *KIR3DL11*, display greater allelic variation compared to rhesus macaques (**Tables I** and **II**). An exceptional example is formed by *Mafa-KIR1D*, which is present on 82% of the haplotypes in cynomolgus monkeys but only on 22% of the haplotypes in rhesus macaques. Moreover, the allelic variation of *Mafa-KIR1D* exceeds that of *Mamu-KIR1D*, despite the difference in the number of animals studied per species (**Tables I** and **II**). These differences may indicate that *KIR1D* in cynomolgus macaques executes a more essential role.

On average, one more inhibitory *KIR* gene was present on haplotypes of rhesus macaques, whereas an additional activating KIR was encoded on cynomolgus macaque haplotypes. The differential gene and allele frequencies are indicators for species-specific selection, and might involve different infectious pathogen encounters due to varying habitats.

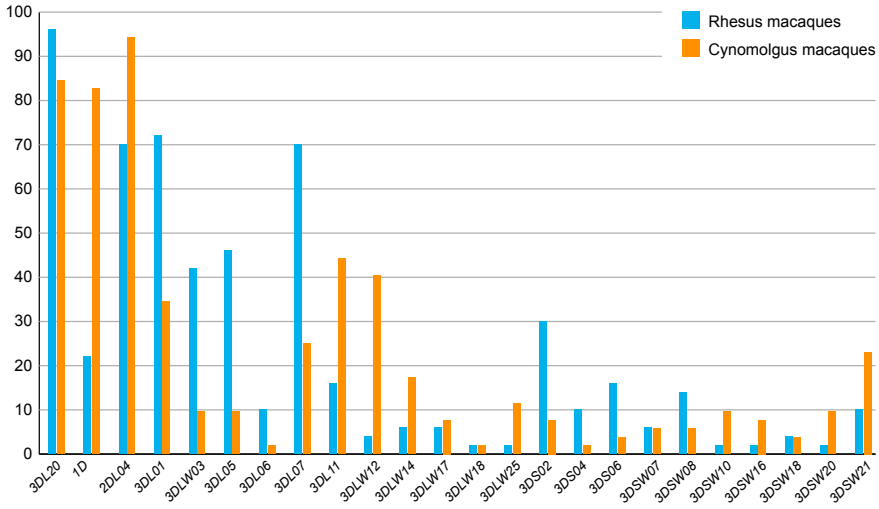


Figure 7. Gene frequencies of orthologous *KIR* genes in rhesus and cynomolgus macaques. The gene frequencies are given for orthologous *KIR* genes in rhesus (blue bars) and cynomolgus (orange bars) macaques.

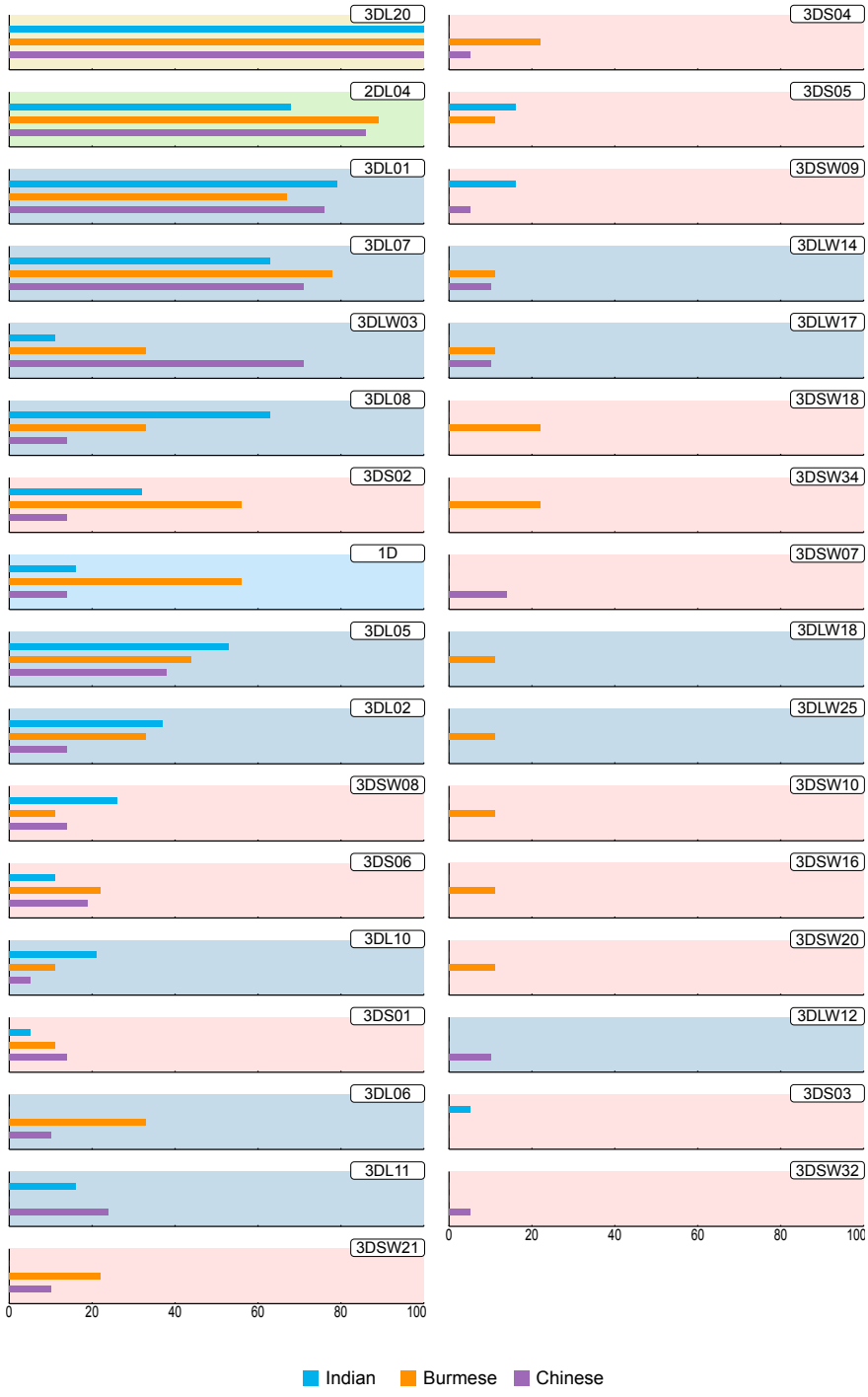
Differential KIR gene content and frequency in populations

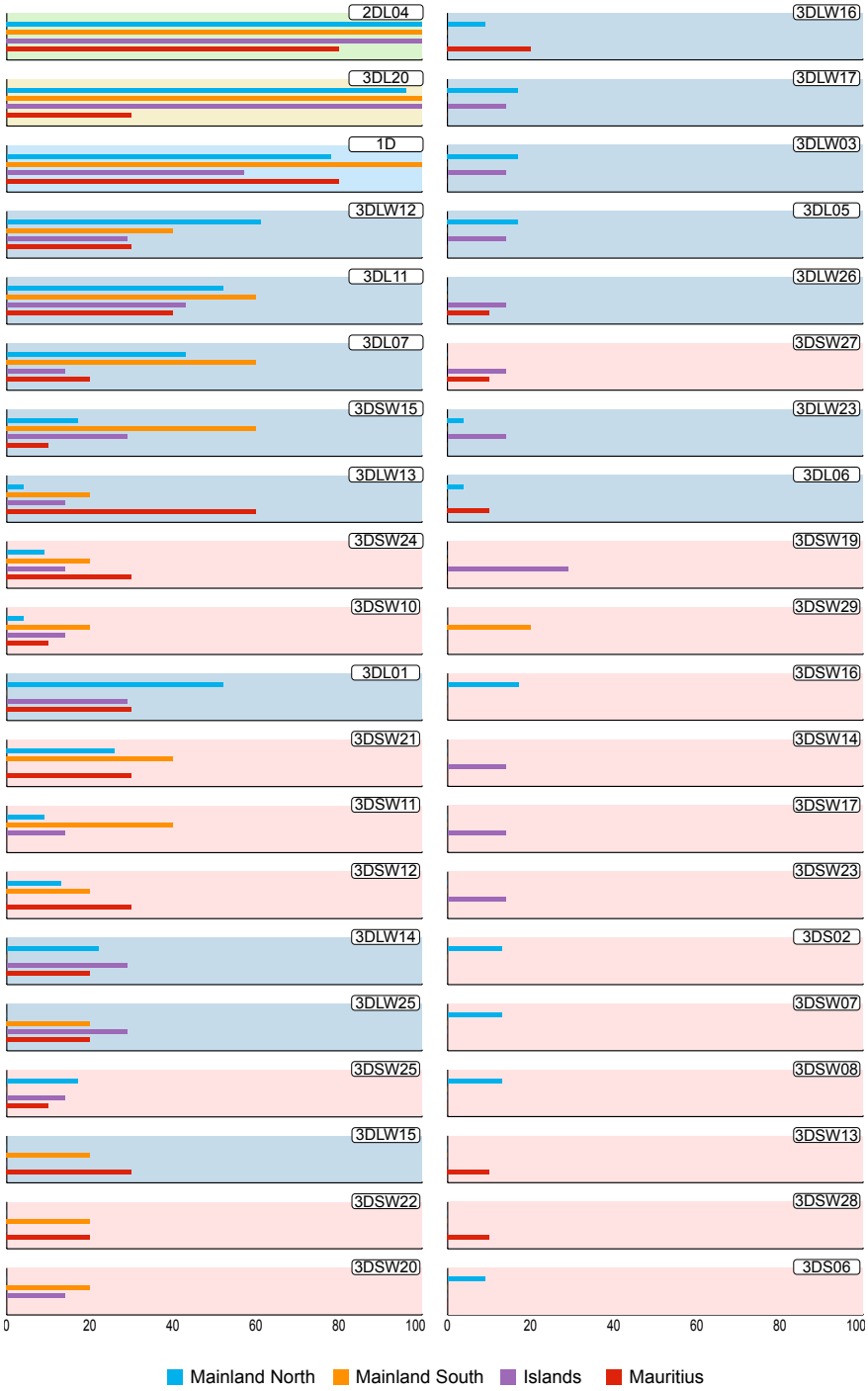
The populations of rhesus and cynomolgus macaques (**Figure 1**) parade differences in *KIR* gene content and gene frequency. Rhesus macaques from the Burmese population encoded, on average, one and two additional KIR3DL and KIR3DS receptors, respectively, as compared to the haplotypes that stem from the Indian and Chinese populations. Approximately 70% of the haplotypes contained at least one *Mamu-KIR3DL01* and/or *Mamu-KIR3DL07* copy, regardless of the origin, whereas multiple other *KIR* genes were differently distributed over the rhesus macaque populations (**Figure 8**). For example, *Mamu-KIR1D* is located on 56% of the haplotypes from Burmese animals, whereas it is present on only 16% and 14% of the Indian and Chinese rhesus macaque *KIR* haplotypes, respectively. Eleven *KIR* genes were identified in a single rhesus macaque population, including newly defined activating *KIR3DS* genes that were only encountered in the Burmese cohort studied.

In cynomolgus macaques, animals that originate from the mainland populations seem to have, on average, one additional inhibitory KIR receptor (KIR3DL) per haplotype as compared to the subjects that inhabit the Indonesian/Malaysian islands. Differential gene distribution trends are observed for several *KIR* genes (**Figure 9**). For example, *Mafa-KIR3DL01* and *-KIR3DLW12*, which were more frequently identified in the northern mainland population, and *-KIR3DLW19, -KIR3DLW21, -KIR3DLW28,* and *-KIR3DSW13*, which were found present only in the Mauritian population. Activating *KIR* genes with orthologs in Indian rhesus macaques were mainly identified in the northern mainland population, including *KIR3DS02* and *KIR3DSW07*, whereas the other cynomolgus macaque populations have species-specific activating *KIR* genes.

Genes that were identified in either three rhesus or four cynomolgus macaque populations mainly encode inhibitory receptors (**Figures 8** and **9**). The activating receptor are more often observed in two or one populations (**Figures 8** and **9**). Overall, the observed variable gene content and gene frequency in the different macaque populations support evidence pointing to rapid evolution of the *KIR* genes at the population level.

Figure 8. Gene distributions between the different rhesus macaque populations. Gene frequencies are listed for rhesus macaque *KIR* genes that were identified in the Indian (blue bars), Burmese (orange bars), and Chinese (purple bars) populations. The genes are listed from the most frequent genes in all populations to the least frequent species-specific genes. The background color indicates *Mamu-KIR3DL20* (yellow), *-KIR2DL04* (green), *-KIR1D* (light blue), and the inhibitory (blue) and activating (red) lineage II *KIR* genes. The frequencies are based on the presence on a haplotype of known origin rather than the presence in an individual. ▶


3DL20
2DL04
3DL01
3DL07
3DLW03
3DL08
3DS02
1D
3DL05
3DL02
3DSW08
3DS06
3DL10
3DS01
3DL06
3DL11
3DSW21
3DS04
3DS05
3DSW09
3DLW14
3DLW17
3DSW18
3DSW34
3DSW07
3DLW18
3DLW25
3DSW10
3DSW16
3DSW20
3DLW12
3DS03
3DSW32
0
20
40
60
80
100
Indian
Burmese
Chinese


2DL04
3DL20
1D
3DLW12
3DL11
3DL07
3DSW15
3DLW13
3DSW24
3DSW10
3DL01
3DSW21
3DSW11
3DSW12
3DLW14
3DLW25
3DSW25
3DLW15
3DSW22
3DSW20
3DLW16
3DLW17
3DLW03
3DL05
3DLW26
3DSW27
3DLW23
3DL06
3DSW19
3DSW29
3DSW16
3DSW14
3DSW17
3DSW23
3DS02
3DSW07
3DSW08
3DSW13
3DSW28
3DS06
0
20
40
60
80
100
Mainland North
Mainland South
Islands
Mauritius

Discussion

An essential step in the evolution of the primate *KIR* cluster started with the initial expansion of a lineage II *KIR* gene progenitor. Subsequently, other *KIR* lineages seem to have emerged through deletion and recombination events. In macaques, lineage II *KIR* genes (KIR3D) were subjected to substantial expansion (10, 17, 32), which coincides with an extended *MHC* class I gene repertoire (16, 24). The present study involves the comparative analysis of rhesus and cynomolgus macaque populations from distinct geographic areas. The *KIR* gene repertoires were found to reflect rapid evolution. Our data illustrate that not only within these closely related species, but even within their populations, new *KIR* gene entities are generated by complex recombination processes resulting in the formation of hybrid genes. In addition, a high level of allelic polymorphism was encountered in both macaque species, but the sharing of alleles was virtually absent. Moreover, recombination resulted in marked differences in the *KIR* haplotype architecture of both species, again testifying the rapid evolution of the macaque *KIR* genes, which has not been described in other non-human primate species.

In humans, the *KIR* gene cluster mainly diversifies at the allelic level, whereas gene expansion is modest and mainly confined to lineage III genes (43). Two major haplotype configurations are recognized, for which a trade-off has been suggested based on differential haplotype frequencies in human populations (47). The A haplotype configurations standardly express seven receptors and have an inhibitory profile, whereas the B haplotypes show moderate gene content variability including multiple activating *KIR* genes (7 to 13 *KIR* genes) (**Figure 10**) (44). Chimpanzees (*Pan troglodytes*) and humans diverged from a common ancestor about 5 million years ago, and, although the complexity of the *KIR* clusters is to some extent comparable, species-specific diversification is observed in receptor structure, haplotype architecture, *MHC* class I recognition potential, and gene content (48). The chimpanzee *KIR* region mainly comprises inhibitory genes, and resemble human A haplotypes. Several chimpanzee *KIR* genes are actually recombinant genes (48). The repertoire, however, is limited to 13 *KIR* genes, four of which are orthologous framework genes that are shared with humans (**Figure 10**) (41, 48). Although little is known about the allelic variation in chimpanzees, the limited *KIR* gene repertoire might suggest that the ancient selective sweep, which targeted the ancestral chimpanzee *MHC* class I region and was likely caused by a retroviral infection (49, 50), may also have had an indirect impact on its ligands within the *KIR* gene region. Bonobos (*Pan paniscus*) and chimpanzees shared a common ancestor approximately 2.3 million years ago. In this species, only 7 *KIR* genes

◀ **Figure 9. Gene distributions between the different cynomolgus macaque populations.** Frequencies are provided for *KIR* genes identified in cynomolgus macaque populations: mainland, north (blue bars) or south (orange bars) of the Isthmus of Kra, the Malaysian/Indonesian islands (purple bars), or Mauritius (red bars). The background color indicates *Mamu-KIR3DL20* (yellow), *-KIR2DL04* (green), *-KIR1D* (light blue), and the inhibitory (blue) and activating (red) lineage II *KIR* genes. The frequencies are based on the presence of a gene on haplotype of known origin rather than on the presence in an individual.

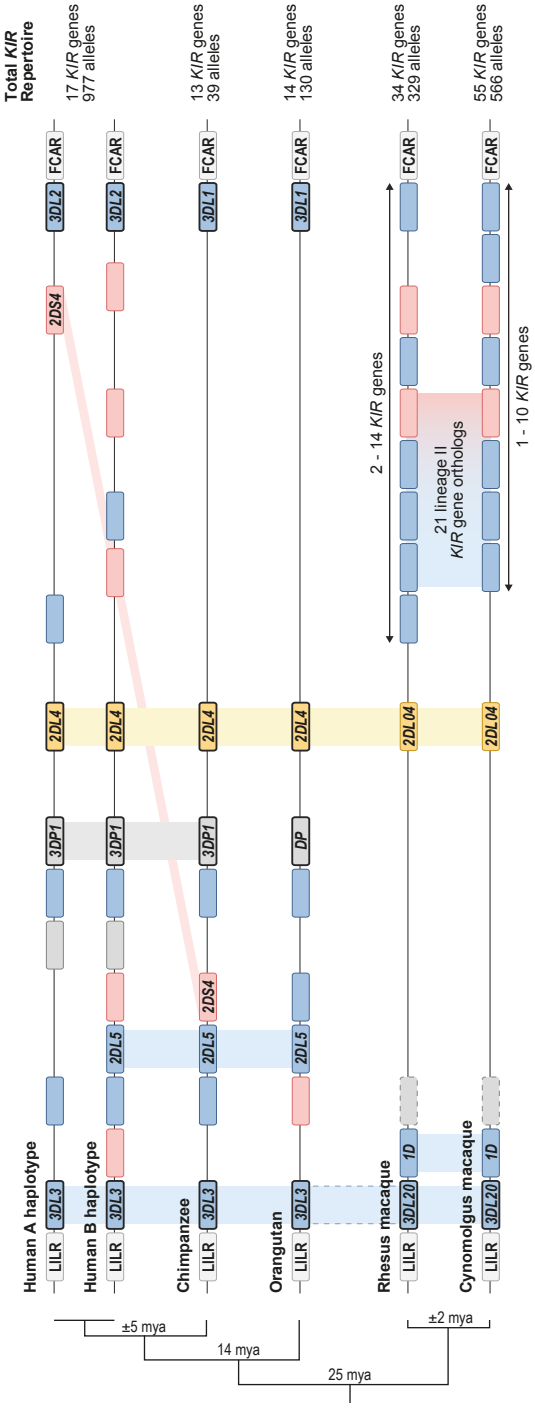


Figure 10. KIR haplotypes in different primate species. The KIR haplotype configurations of humans (A and B haplotypes), chimpanzees, orangutans, and rhesus and cynomolgus macaques are schematically illustrated. The evolutionary distance between the different species is depicted on the left, whereas the numbers of the documented *KIR* repertoires are provided on the right. The inhibitory and activating *KIR* genes are illustrated with blue and red boxes, whereas pseudogenes are in grey. Framework genes are indicated with black outlining, and orthologous and homologous genes shared in the depicted primate species are connected with each other. In total, 21 lineage II *KIR* genes are identified as orthologs in rhesus and cynomolgus macaques. *KIR3DL20*, identified as framework gene in both macaque species, is a potential ortholog of the other primate *KIR3DL3* genes. A putative pseudogene on the centromeric macaque KIR haplotype is indicated by a grey box with a dashed outlining. The *KIR* region is flanked by the *LILR* and *FCAR* genes in all primate species depicted.

are reported, 5 and 3 genes of which are orthologs, shared with common chimpanzees and humans, respectively (51). The short bonobo *KIR* haplotypes, the limited *KIR* gene repertoire, and the reduced bonobo *MHC* class I content (52-55) may result from subsequent selective sweeps (56). The expansion of lineage III *KIR* genes, which in macaques are represented by *KIR1D*, correlates with the emergence of *HLA*-C-like genes in orangutans (57-59). In orangutans, 13 *KIR* genes are identified, of which the framework genes share orthologs with humans (15, 41, 57, 58). Two sibling orangutan species inhabit Sumatra (*Pongo abelii*) and Borneo (*Pongo pygmaeus*) (60). Only one gene is species-specific (*KIR2DS15* in Bornean orangutans), whereas all other genes are orthologs. Ten of the 130 *KIR* alleles that were identified are shared between both sister species. For the human and great ape species discussed above, rapid evolution is mainly reflected by the gain in allelic variation, whereas the generation of novel gene entities and the formation of complex haplotype architectures seem to be relatively limited (**Figure 10**).

The present communication sheds light on the evolution of the *KIR* region in two highly related Old World monkey species, which share an introgression zone. In rhesus and cynomolgus macaques, the massive expansion of the lineage II *KIR* genes exceeds the modest lineage III expansion in great apes and humans. The rapid evolution of the macaque *KIR* region is reflected not only by allelic polymorphism but even more prominently by the large number of species-specific recombinant genes and haplotypes with a complex architecture in different macaque populations. In humans and other hominids, these events seem to be less abundant. Moreover, the allelic variation in macaques seems to exceed the numbers that are encountered in humans and higher primates. We recorded a total number of 579 *KIR* alleles, and it is important to realize that the number of samples that were analyzed is relatively small as compared to the human situation (43). The reason for the extensive expansion of the *KIR* gene cluster in macaques is of interest. Whereas in humans the less variable *KIR* gene content and haplotype configurations seem to be the result of a trade-off, such an indication appears to be absent in macaques. The most plausible driving forces of the rapid *KIR* cluster evolution in macaques might involve co-evolution with the extended *MHC* class I region, differential infectious pathogen encounters, a discontinuous habitat, and susceptibility to chromosomal recombination. Nonetheless, we cannot rule out that the extensive expansion of the macaque *KIR* gene system may have evolved due to the lack of evolutionary pressure on this system. The ligand of only a few receptors have been identified, and therefore the functional impact of the expanded macaque KIR repertoire remains largely unclear. However, diversification of ligand interactions is suggested by overlapping, but non-redundant, MHC class I specificity of multiple KIR (61-66). The extensive diversity of the macaque MHC and KIR clusters might facilitate interactions with allele-level specificity, differential affinity, and peptide dependency, and may contribute to rapid adaption driven by environmental conditions.

The general high levels of allelic polymorphism detected in the *KIR* region in primates might indicate that it is more prone to generate mutations than other regions of the genome. Mutation rates are elevated in CpG islands, which are genomic regions

that are enriched for CpG sites with an observed-to-expected ratio greater than 60 percent. All *KIR* genes indeed carry CpG islands (67). CpG site mutations, however, mainly involve cytosine to thymine transitions, whereas mutations in the generation of the novel *Mamu-KIR3DL05* allele involved T > C and G > T transitions, which are not commonly observed transition events. The two point mutations are separated by only two nucleotides, which suggests that one mutation initiated the other, and perhaps was caused by the recruitment of error-prone repair mechanisms (68). In addition to CpG islands and error-prone repair, other factors that might enhance the regional mutation rate may include recombination events, deletion and insertion events, chromatin configurations, distance to the telomere, and replication time (69, 70). Furthermore, relatively more single nucleotide polymorphisms were observed in regions that were homologous in humans, chimpanzees, and macaques, which substantiates the extensive variation of the *KIR* gene region (69). The birth of novel *KIR* alleles has been described previously in human families (71), and, together with the event recorded in macaques, this might suggest that point mutations substantially contribute to the extensive allelic *KIR* variation. Of course, it is clear that the generation of mutations is only one side of the coin, and that selection determines which polymorphisms will be enriched in the populations or are eventually rooted out.

In humans, genetic *KIR* variation is documented for over 250 populations and mainly records allelic variation and differential haplotype distribution in relation to gene frequencies (47, 72). Similar observations were made for the different macaque populations. Genes that are shared in all three rhesus or four cynomolgus macaque populations mainly involved inhibitory *KIR* genes (**Figures 8** and **9**). The conserved nature of these genes suggests an impact on essential functions, such as NK cell education, a process for which the involvement of inhibitory KIR is well established in humans (73-75). The role of activating KIR in humans is less understood, but associations with disease progression or protection are described, and *in vitro* studies demonstrate specific binding to certain peptide-MHC class I complexes (76-81). In macaques, activating *KIR* genes were mainly identified in one or two populations and may substantiate more specialized functions, like pathogen recognition (**Figures 8** and **9**). For instance, Mamu-*KIR3DSW18* is encountered only in Burmese rhesus macaques, and a similar observation was made for *Mafa-KIR3DSW08* in the northern-mainland cynomolgus macaques. In addition, the majority of the *KIR* alleles appear not to be shared at the population level (**Figure 4**). Again, this hints at a speedy generation of allelic polymorphism. For the Indian rhesus macaque population, a genetic bottleneck is evident (82), but it did not result in a reduced *KIR* gene content, which might indicate that the rapid evolution of the *KIR* repertoire erased the genetic footprint of a bottleneck.

Rhesus and cynomolgus macaques are widely used as preclinical models in translational biomedical research in order to further a better understanding of human diseases and the development of vaccines and therapies (34, 35). The genetic makeup of the different macaque species, however, can vary considerably and might potentially influence the outcome of studies. Even at the population level, a differential disease

susceptibility has been reported: for example, in SIV/AIDS-related experiments in rhesus macaques of Indian and Chinese origin (36, 37, 83). It is possible that the *KIR* repertoire may be one of the factors that have an impact on disease outcome, as correlations between *KIR* gene content and disease phenotypes in humans (6) and macaques are documented (38, 39).

This study design, including rhesus and cynomolgus macaque families from different geographical origin, allowed the transcriptomic characterization of the complex *KIR* cluster. The high level of allelic polymorphism, the number of novel gene entities, the plastic haplotype architecture, and the diversification at the species and population levels illustrate the unparalleled rapid evolution of the *KIR* gene region in macaques. This communication paves the way to study the impact of *KIR* genes in nonhuman primate models for human health and disease, but also may help in selecting animals with particular genetic markers for studies in the area of personalized medicine.

Acknowledgements

We thank D. Devine for editing the manuscript, and F. van Hassel for preparing the figures.

Materials & data

Requests for materials and data can be addressed to Prof. Dr. Ronald E. Bontrop (bontrop@bprc.nl). The sequences presented in this article have been submitted to the European Nucleotide Archive (https://www.ebi.ac.uk/ena/) under accession number PRJEB33481.

References

1. Benson, DM, Jr., Caligiuri MA. 2014. Killer immunoglobulin-like receptors and tumor immunity. *Cancer immunology research* 2(2):99-104.
2. Ljunggren, HG, Karre K. 1990. In search of the 'missing self': MHC molecules and NK cell recognition. *Immunology today* 11(7):237-244.
3. Lanier, LL. 2005. NK cell recognition. *Annual review of immunology* 23:225-274.
4. Battistini, L, Borsellino G, Sawicki G, Poccia F, Salvetti M, Ristori G, Brosnan CF. 1997. Phenotypic and cytokine analysis of human peripheral blood gamma delta T cells expressing NK cell receptors. *Journal of immunology (Baltimore, Md : 1950)* 159(8):3723-3730.
5. Trowsdale, J. 2001. Genetic and functional relationships between MHC and NK receptor genes. *Immunity* 15(3):363-374.
6. Kulkarni, S, Martin MP, Carrington M. 2008. The Yin and Yang of HLA and KIR in human disease. *Seminars in immunology* 20(6):343-352.
7. Parham, P. 2005. MHC class I molecules and KIRs in human history, health and survival. *Nature reviews Immunology* 5(3):201-214.
8. Trowsdale, J, Barten R, Haude A, Stewart CA, Beck S, Wilson MJ. 2001. The genomic context of natural killer receptor extended gene families. *Immunological reviews* 181:20-38.
9. Hsu, KC, Chida S, Geraghty DE, Dupont B. 2002. The killer cell immunoglobulin-like receptor (KIR) genomic region: gene-order, haplotypes and allelic polymorphism. *Immunological reviews* 190:40-52.
10. Bruijnesteijn, J, van der Wiel MKH, Swelsen WTN, Otting N, de Vos-Rouweler AJM, Elferink D, Doxiadis GG, Claas FHJ, Lardy NM, de Groot NG, Bontrop RE. 2018. Human and Rhesus Macaque KIR Haplotypes Defined by Their Transcriptomes. *Journal of immunology (Baltimore, Md : 1950)* 200(5):1692-1701.
11. Bruijnesteijn, J, van der Wiel MKH, de Groot N, Otting N, de Vos-Rouweler AJM, Lardy NM, de Groot NG, Bontrop RE. 2018. Extensive Alternative Splicing of KIR Transcripts. *Frontiers in Immunology* 9(2846).
12. Martin, AM, Freitas EM, Witt CS, Christiansen FT. 2000. The genomic organization and evolution of the natural killer immunoglobulin-like receptor (KIR) gene cluster. *Immunogenetics* 51(4-5):268-280.
13. Carrillo-Bustamante, P, Kesmir C, de Boer RJ. 2016. The evolution of natural killer cell receptors. *Immunogenetics* 68(1):3-18.
14. Parham, P, Norman PJ, Abi-Rached L, Guethlein LA. 2012. Human-specific evolution of killer cell immunoglobulin-like receptor recognition of major histocompatibility complex class I molecules. *Philosophical transactions of the Royal Society of London Series B, Biological sciences* 367(1590):800-811.
15. Guethlein, LA, Norman PJ, Heijmans CM, de Groot NG, Hilton HG, Babrzadeh F, Abi-Rached L, Bontrop RE, Parham P. 2017. Two Orangutan Species Have Evolved Different KIR Alleles and Haplotypes. *Journal of immunology (Baltimore, Md : 1950)* 198(8):3157-3169.
16. Daza-Vamenta, R, Glusman G, Rowen L, Guthrie B, Geraghty DE. 2004. Genetic divergence of the rhesus macaque major histocompatibility complex. *Genome research* 14(8):1501-1515.
17. de Groot, NG, Blokhuis JH, Otting N, Doxiadis GG, Bontrop RE. 2015. Co-evolution of the MHC class I and KIR gene families in rhesus macaques: ancestry and plasticity. *Immunological reviews* 267(1):228-245.
18. Kanthaswamy, S, Ng J, Satkoski Trask J, George DA, Kou AJ, Hoffman LN, Doherty TB, Houghton P, Smith DG. 2013. The genetic composition of populations of cynomolgus macaques (Macaca fascicularis) used in biomedical research. *Journal of medical primatology* 42(3):120-131.
19. Blancher, A, Bonhomme M, Crouau-Roy B, Terao K, Kitano T, Saitou N. 2008. Mitochondrial DNA sequence phylogeny of 4 populations of the widely distributed cynomolgus macaque (Macaca fascicularis fascicularis). *The Journal of heredity* 99(3):254-264.
20. Bunlungsup, S, Kanthaswamy S, Oldt RF, Smith DG, Houghton P, Hamada Y, Malaivijitnond S. 2017. Genetic analysis of samples from wild populations opens new perspectives on hybridization between long-tailed (Macaca fascicularis) and rhesus macaques (Macaca mulatta). *American journal of primatology* 79(12).
21. Stevison, LS, Kohn MH. 2009. Divergence population genetic analysis of hybridization between rhesus and cynomolgus macaques. *Molecular ecology* 18(11):2457-2475.

22. Yan, G, Zhang G, Fang X, Zhang Y, Li C, Ling F, Cooper DN, Li Q, Li Y, van Gool AJ, Du H, Chen J, Chen R, Zhang P, Huang Z, Thompson JR, Meng Y, Bai Y, Wang J, Zhuo M, Wang T, Huang Y, Wei L, Li J, Wang Z, Hu H, Yang P, Le L, Stenson PD, Li B, Liu X, Ball EV, An N, Huang Q, Zhang Y, Fan W, Zhang X, Li Y, Wang W, Katze MG, Su B, Nielsen R, Yang H, Wang J, Wang X, Wang J. 2011. Genome sequencing and comparison of two nonhuman primate animal models, the cynomolgus and Chinese rhesus macaques. *Nature Biotechnology* 29:1019.

23. Karl, JA, Bohn PS, Wiseman RW, Nimityongskul FA, Lank SM, Starrett GJ, O'Connor DH. 2013. Major histocompatibility complex class I haplotype diversity in Chinese rhesus macaques. *G3 (Bethesda, Md)* 3(7):1195-1201.

24. Otting, N, Heijmans CM, Noort RC, de Groot NG, Doxiadis GG, van Rood JJ, Watkins DI, Bontrop RE. 2005. Unparalleled complexity of the MHC class I region in rhesus macaques. *Proceedings of the National Academy of Sciences of the United States of America* 102(5):1626-1631.

25. Doxiadis, GG, Rouweler AJ, de Groot NG, Louwerse A, Otting N, Verschoor EJ, Bontrop RE. 2006. Extensive sharing of MHC class II alleles between rhesus and cynomolgus macaques. *Immunogenetics* 58(4):259-268.

26. Blokhuis, JH, van der Wiel MK, Doxiadis GG, Bontrop RE. 2010. The mosaic of KIR haplotypes in rhesus macaques. *Immunogenetics* 62(5):295-306.

27. Blokhuis, JH, van der Wiel MK, Doxiadis GG, Bontrop RE. 2011. The extreme plasticity of killer cell Ig-like receptor (KIR) haplotypes differentiates rhesus macaques from humans. *European journal of immunology* 41(9):2719-2728.

28. Sambrook, JG, Bashirova A, Palmer S, Sims S, Trowsdale J, Abi-Rached L, Parham P, Carrington M, Beck S. 2005. Single haplotype analysis demonstrates rapid evolution of the killer immunoglobulin-like receptor (KIR) loci in primates. *Genome research* 15(1):25-35.

29. Kruse, PH, Rosner C, Walter L. 2010. Characterization of rhesus macaque KIR genotypes and haplotypes. *Immunogenetics* 62(5):281-293.

30. Moreland, AJ, Guethlein LA, Reeves RK, Broman KW, Johnson RP, Parham P, O'Connor DH, Bimber BN. 2011. Characterization of killer immunoglobulin-like receptor genetics and comprehensive genotyping by pyrosequencing in rhesus macaques. *BMC genomics* 12:295.

31. Bimber, BN, Moreland AJ, Wiseman RW, Hughes AL, O'Connor DH. 2008. Complete characterization of killer Ig-like receptor (KIR) haplotypes in Mauritian cynomolgus macaques: novel insights into nonhuman primate KIR gene content and organization. *Journal of immunology (Baltimore, Md : 1950)* 181(9):6301-6308.

32. Bimber, BN, Evans DT. 2015. The killer-cell immunoglobulin-like receptors of macaques. *Immunological reviews* 267(1):246-258.

33. Prall, TM, Graham ME, Karl JA, Wiseman RW, Ericsen AJ, Raveendran M, Alan Harris R, Muzny DM, Gibbs RA, Rogers J, O'Connor DH. 2017. Improved full-length killer cell immunoglobulin-like receptor transcript discovery in Mauritian cynomolgus macaques. *Immunogenetics* 69(5):325-339.

34. Dijkman, K, Sombroek CC, Vervenne RAW, Hofman SO, Boot C, Remarque EJ, Kocken CHM, Ottenhoff THM, Kondova I, Khayum MA, Haanstra KG, Vierboom MPM, Verreck FAW. 2019. Prevention of tuberculosis infection and disease by local BCG in repeatedly exposed rhesus macaques. *Nature medicine* 25(2):255-262.

35. Harding, JD. 2017. Genomic Tools for the Use of Nonhuman Primates in Translational Research. *ILAR journal* 58(1):59-68.

36. Burdo, TH, Marcondes MC, Lanigan CM, Penedo MC, Fox HS. 2005. Susceptibility of Chinese rhesus monkeys to SIV infection. *AIDS (London, England)* 19(15):1704-1706.

37. Ling, B, Veazey RS, Luckay A, Penedo C, Xu K, Lifson JD, Marx PA. 2002. SIV(mac) pathogenesis in rhesus macaques of Chinese and Indian origin compared with primary HIV infections in humans. *AIDS (London, England)* 16(11):1489-1496.

38. Albrecht, C, Malzahn D, Brameier M, Hermes M, Ansari AA, Walter L. 2014. Progression to AIDS in SIV-Infected Rhesus Macaques is Associated with Distinct KIR and MHC class I Polymorphisms and NK Cell Dysfunction. *Front Immunol* 5:600.

39. Walter, L, Ansari AA. 2015. MHC and KIR Polymorphisms in Rhesus Macaque SIV Infection. *Front Immunol* 6:540.

40. Doxiadis, GG, de Groot N, de Groot NG, Rotmans G, de Vos-Rouweler AJ, Bontrop RE. 2010. Extensive DRB region diversity in cynomolgus macaques: recombination as a driving force. *Immunogenetics* 62(3):137-147.

41. Robinson, J, Guethlein LA, Maccari G, Blokhuis J, Bimber BN, de Groot NG, Sanderson ND, Abi-Rached L, Walter L, Bontrop RE, Hammond JA, Marsh SGE, Parham P. 2018. Nomenclature for the KIR of non-human species. *Immunogenetics* 70(9):571-583.

42. Doxiadis, GG, de Groot N, Otting N, Blokhuis JH, Bontrop RE. 2011. Genomic plasticity of the MHC class I A region in rhesus macaques: extensive haplotype diversity at the population level as revealed by microsatellites. *Immunogenetics* 63(2):73-83.

43. Robinson, J, Halliwell JA, McWilliam H, Lopez R, Marsh SG. 2013. IPD--the Immuno Polymorphism Database. *Nucleic acids research* 41(Database issue):D1234-1240.

44. Uhrberg, M, Valiante NM, Shum BP, Shilling HG, Lienert-Weidenbach K, Corliss B, Tyan D, Lanier LL, Parham P. 1997. Human diversity in killer cell inhibitory receptor genes. *Immunity* 7(6):753-763.

45. Roe, D, Vierra-Green C, Pyo CW, Eng K, Hall R, Kuang R, Spellman S, Ranade S, Geraghty DE, Maiers M. 2017. Revealing complete complex KIR haplotypes phased by long-read sequencing technology. *Genes and immunity* 18(3):127-134.

46. Bruijnesteijn, J, de Groot NG, Otting N, Maccari G, Guethlein LA, Robinson J, Marsh SGE, Walter L, O'Connor DH, Hammond JA, Parham P, Bontrop RE. 2019. Nomenclature report for killer-cell immunoglobulin-like receptors (KIR) in macaque species: new genes/alleles, renaming recombinant entities and IPD-NHKIR updates. *Immunogenetics*.

47. Middleton, D, Gonzelez F. 2010. The extensive polymorphism of KIR genes. *Immunology* 129(1):8-19.

48. Abi-Rached, L, Moesta AK, Rajalingam R, Guethlein LA, Parham P. 2010. Human-specific evolution and adaptation led to major qualitative differences in the variable receptors of human and chimpanzee natural killer cells. *PLoS genetics* 6(11):e1001192.

49. de Groot, NG, Otting N, Doxiadis GGM, Balla-Jhagjhoorsingh SS, Heeney JL, van Rood JJ, Gagneux P, Bontrop RE. 2002. Evidence for an ancient selective sweep in the MHC class I gene repertoire of chimpanzees. *Proceedings of the National Academy of Sciences* 99(18):11748-11753.

50. de Groot, NG, Heijmans CM, Zoet YM, de Ru AH, Verreck FA, van Veelen PA, Drijfhout JW, Doxiadis GG, Remarque EJ, Doxiadis, II, van Rood JJ, Koning F, Bontrop RE. 2010. AIDS-protective HLA-B*27/B*57 and chimpanzee MHC class I molecules target analogous conserved areas of HIV-1/SIVcpz. *Proceedings of the National Academy of Sciences of the United States of America* 107(34):15175-15180.

51. Rajalingam, R, Hong M, Adams EJ, Shum BP, Guethlein LA, Parham P. 2001. Short KIR haplotypes in pygmy chimpanzee (Bonobo) resemble the conserved framework of diverse human KIR haplotypes. *The Journal of experimental medicine* 193(1):135-146.

52. Wroblewski, EE, Guethlein LA, Norman PJ, Li Y, Shaw CM, Han AS, Ndjango JN, Ahuka-Mundeke S, Georgiev AV, Peeters M, Hahn BH, Parham P. 2017. Bonobos Maintain Immune System Diversity with Three Functional Types of MHC-B. *Journal of immunology (Baltimore, Md : 1950)* 198(9):3480-3493.

53. Maibach, V, Hans JB, Hvilsom C, Marques-Bonet T, Vigilant L. 2017. MHC class I diversity in chimpanzees and bonobos. *Immunogenetics* 69(10):661-676.

54. Maibach, V, Vigilant L. 2019. Reduced bonobo MHC class I diversity predicts a reduced viral peptide binding ability compared to chimpanzees. *BMC evolutionary biology* 19(1):14.

55. de Groot, NG, Heijmans CMC, Helsen P, Otting N, Pereboom Z, Stevens JMG, Bontrop RE. 2017. Limited MHC class I intron 2 repertoire variation in bonobos. *Immunogenetics* 69(10):677-688.

56. de Groot, NG, Stevens JMG, Bontrop RE. 2018. Does the MHC Confer Protection against Malaria in Bonobos? *Trends in immunology* 39(10):768-771.

57. Guethlein, LA, Flodin LR, Adams EJ, Parham P. 2002. NK cell receptors of the orangutan (Pongo pygmaeus): a pivotal species for tracking the coevolution of killer cell Ig-like receptors with MHC-C. *Journal of immunology (Baltimore, Md : 1950)* 169(1):220-229.

58. Guethlein, LA, Older Aguilar AM, Abi-Rached L, Parham P. 2007. Evolution of killer cell Ig-like receptor (KIR) genes: definition of an orangutan KIR haplotype reveals expansion of lineage III KIR associated with the emergence of MHC-C. *Journal of immunology (Baltimore, Md : 1950)* 179(1):491-504.

59. de Groot, NG, Heijmans CM, van der Wiel MK, Blokhuis JH, Mulder A, Guethlein LA, Doxiadis GG, Claas FH, Parham P, Bontrop RE. 2016. Complex MHC Class I Gene Transcription Profiles and Their Functional Impact in Orangutans. *Journal of immunology (Baltimore, Md : 1950)* 196(2):750-758.

60. Steiper, ME. 2006. Population history, biogeography, and taxonomy of orangutans (Genus: Pongo) based on a population genetic meta-analysis of multiple loci. *Journal of human evolution* 50(5):509-522.

61. Ries, M, Reynolds MR, Bashkueva K, Crosno K, Capuano S, 3rd, Prall TM, Wiseman R, O'Connor DH, Rakasz EG, Uno H, Lifson JD, Evans DT. 2017. KIR3DL01 upregulation on gut natural killer cells in response to SIV infection of KIR- and MHC class I-defined rhesus macaques. *PLoS pathogens* 13(7):e1006506.

62. Banerjee, P, Ries M, Janaka SK, Grandea AG, 3rd, Wiseman R, O'Connor DH, Golos TG, Evans DT. 2018. Diversification of Bw4 Specificity and Recognition of a Nonclassical MHC Class I Molecule Implicated in Maternal-Fetal Tolerance by Killer Cell Ig-like Receptors of the Rhesus Macaque. *Journal of immunology (Baltimore, Md : 1950)* 201(9):2776-2786.

63. Schafer, JL, Colantonio AD, Neidermyer WJ, Dudley DM, Connole M, O'Connor DH, Evans DT. 2014. KIR3DL01 recognition of Bw4 ligands in the rhesus macaque: maintenance of Bw4 specificity since the divergence of apes and Old World monkeys. *Journal of immunology (Baltimore, Md : 1950)* 192(4):1907-1917.

64. Colantonio, AD, Bimber BN, Neidermyer WJ, Jr., Reeves RK, Alter G, Altfeld M, Johnson RP, Carrington M, O'Connor DH, Evans DT. 2011. KIR polymorphisms modulate peptide-dependent binding to an MHC class I ligand with a Bw6 motif. *PLoS pathogens* 7(3):e1001316.

65. Maloveste, SM, Chen D, Gostick E, Vivian JP, Plishka RJ, Iyengar R, Kruthers RL, Buckler-White A, Brooks AG, Rossjohn J, Price DA, Lafont BA. 2012. Degenerate recognition of MHC class I molecules with Bw4 and Bw6 motifs by a killer cell Ig-like receptor 3DL expressed by macaque NK cells. *Journal of immunology (Baltimore, Md : 1950)* 189(9):4338-4348.

66. Rosner, C, Kruse PH, Hermes M, Otto N, Walter L. 2011. Rhesus macaque inhibitory and activating KIR3D interact with Mamu-A-encoded ligands. *Journal of immunology (Baltimore, Md : 1950)* 186(4):2156-2163.

67. Santourlidis, S, Trompeter HI, Weinhold S, Eisermann B, Meyer KL, Wernet P, Uhrberg M. 2002. Crucial role of DNA methylation in determination of clonally distributed killer cell Ig-like receptor expression patterns in NK cells. *Journal of immunology (Baltimore, Md : 1950)* 169(8):4253-4261.

68. Walser, JC, Ponger L, Furano AV. 2008. CpG dinucleotides and the mutation rate of non-CpG DNA. *Genome research* 18(9):1403-1414.

69. Hodgkinson, A, Eyre-Walker A. 2011. Variation in the mutation rate across mammalian genomes. *Nature reviews Genetics* 12(11):756-766.

70. Zhu, Y, Neeman T, Yap VB, Huttley GA. 2017. Statistical Methods for Identifying Sequence Motifs Affecting Point Mutations. *Genetics* 205(2):843-856.

71. Halfpenny, IA, Middleton D, Barnett YA, Williams F. 2004. Investigation of killer cell immunoglobulin-like receptor gene diversity: IV. KIR3DL1/S1. *Human immunology* 65(6):602-612.

72. Gonzalez-Galarza, FF, Takeshita LY, Santos EJ, Kempson F, Maia MH, da Silva AL, Teles e Silva AL, Ghattaoraya GS, Alfirevic A, Jones AR, Middleton D. 2015. Allele frequency net 2015 update: new features for HLA epitopes, KIR and disease and HLA adverse drug reaction associations. *Nucleic acids research* 43(Database issue):D784-788.

73. Anfossi, N, Andre P, Guia S, Falk CS, Roetynck S, Stewart CA, Breso V, Frassati C, Reviron D, Middleton D, Romagne F, Ugolini S, Vivier E. 2006. Human NK cell education by inhibitory receptors for MHC class I. *Immunity* 25(2):331-342.

74. Yawata, M, Yawata N, Draghi M, Partheniou F, Little AM, Parham P. 2008. MHC class I-specific inhibitory receptors and their ligands structure diverse human NK-cell repertoires toward a balance of missing self-response. *Blood* 112(6):2369-2380.

75. Kim, S, Poursine-Laurent J, Truscott SM, Lybarger L, Song YJ, Yang L, French AR, Sunwoo JB, Lemieux S, Hansen TH, Yokoyama WM. 2005. Licensing of natural killer cells by host major histocompatibility complex class I molecules. *Nature* 436(7051):709-713.

76. Sim, MJW, Rajagopalan S, Altmann DM, Boyton RJ, Sun PD, Long EO. 2019. Human NK cell receptor KIR2DS4 detects a conserved bacterial epitope presented by HLA-C. *Proceedings of the National Academy of Sciences of the United States of America* 116(26):12964-12973.

77. Venstrom, JM, Pittari G, Gooley TA, Chewning JH, Spellman S, Haagenson M, Gallagher MM, Malkki M, Petersdorf E, Dupont B, Hsu KC. 2012. HLA-C–Dependent Prevention of Leukemia Relapse by Donor Activating KIR2DS1. *New England Journal of Medicine* 367(9):805-816.

78. Chewning, JH, Gudme CN, Hsu KC, Selvakumar A, Dupont B. 2007. KIR2DS1-positive NK cells mediate alloresponse against the C2 HLA-KIR ligand group in vitro. *Journal of immunology (Baltimore, Md : 1950)* 179(2):854-868.

79. Stewart, CA, Laugier-Anfossi F, Vely F, Saulquin X, Riedmuller J, Tisserant A, Gauthier L, Romagne F, Ferracci G, Arosa FA, Moretta A, Sun PD, Ugolini S, Vivier E. 2005. Recognition of peptide-MHC class I complexes by activating killer immunoglobulin-like receptors. *Proceedings of the National Academy of Sciences of the United States of America* 102(37):13224-13229.

80. Ivarsson, MA, Michaëlsson J, Fauriat C. 2014. Activating Killer Cell Ig-Like Receptors in Health and Disease. *Frontiers in Immunology* 5(184).

81. 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. 2002. Epistatic interaction between KIR3DS1 and HLA-B delays the progression to AIDS. *Nature genetics* 31(4):429-434.

82. Xue, C, Raveendran M, Harris RA, Fawcett GL, Liu X, White S, Dahdouli M, Rio Deiros D, Below JE, Salerno W, Cox L, Fan G, Ferguson B, Horvath J, Johnson Z, Kanthaswamy S, Kubisch HM, Liu D, Platt M, Smith DG, Sun B, Vallender EJ, Wang F, Wiseman RW, Chen R, Muzny DM, Gibbs RA, Yu F, Rogers J. 2016. The population genomics of rhesus macaques (Macaca mulatta) based on whole-genome sequences. *Genome research* 26(12):1651-1662.

83. Sui, Y, Gordon S, Franchini G, Berzofsky JA. 2013. Nonhuman primate models for HIV/AIDS vaccine development. *Current protocols in immunology* 102:Unit 12.14.

Chapter 4

Extensive alternative splicing of KIR transcripts

Jesse Bruijnesteijn[1]*, Marit K. H. van der Wiel[1], Nanine de Groot[1], Nel Otting[1], Annemiek J. M. de Vos-Rouweler[1], Neubury M. Lardy[2], Natasja G. de Groot[1], and Ronald E. Bontrop[1,3]

[1]Comparative Genetics and Refinement, Biomedical Primate Research Centre, Rijswijk, The Netherlands
[2]Dept. of Immunogenetics, Sanquin, Amsterdam, The Netherlands
[3]Theoretical Biology and Bioinformatics, Utrecht University, Utrecht, The Netherlands

*Corresponding author:
Jesse Bruijnesteijn
Bruijnesteijn@bprc.nl
Tel. nr. +31-15-2842574
Fax nr. +31-15-2842600

Abstract

The killer-cell Ig-like receptors (KIR) form a multigene entity involved in modulating immune responses through interactions with MHC class I molecules. The complexity of the *KIR* cluster is reflected by, for instance, abundant levels of allelic polymorphism, gene copy number variation, and stochastic expression profiles. The current transcriptome study involving human and macaque families demonstrates that *KIR* family members are also subjected to differential levels of alternative splicing, and this seems to be gene dependent. Alternative splicing may result in the partial or complete skipping of exons, or the partial inclusion of introns, as documented at the transcription level. This post-transcriptional process can generate multiple isoforms from a single *KIR* gene, which diversifies the characteristics of the encoded proteins. For example, alternative splicing could modify ligand interactions, cellular localization, signaling properties, and the number of extracellular domains of the receptor. In humans, we observed abundant splicing for *KIR2DL4*, and to a lesser extent in the lineage III *KIR* genes. All experimentally documented splice events are substantiated by *in silico* splicing strength predictions. To a similar extent, alternative splicing is observed in rhesus macaques, a species that shares a close evolutionary relationship with humans. Splicing profiles of *Mamu-KIR1D* and *Mamu-KIR2DL04* displayed a great diversity, whereas *Mamu-KIR3DL20* (lineage V) is consistently spliced to generate a homolog of human *KIR2DL5* (lineage I). The latter case represents an example of convergent evolution. Although just a single KIR splice event is shared between humans and macaques, the splicing mechanisms are similar, and the predicted consequences are comparable. In conclusion, alternative splicing adds an additional layer of complexity to the *KIR* gene system in primates, and results in a wide structural and functional variety of KIR receptors and its isoforms, which may play a role in health and disease.

Introduction

Natural killer (NK) cells express killer-cell immunoglobulin-like receptors (KIR) that interact with major histocompatibility complex (MHC) class I molecules expressed on the cell surface of nucleated cells. Through these interactions, KIR may modulate the NK-cell activity, thereby providing regulation of the immune system in infectious diseases, pregnancy, and transplantation (1-4). KIR belong to a multigene family, which in humans comprises seventeen members that are categorized into four lineages based on structure and ligand interactions; lineage I includes *KIR2DL4/5*, lineage II includes *KIR3DL1/L2/S1*, lineage III includes *KIR2DL1-3/2DS1-5* and the pseudogenes, and lineage V includes *KIR3DL3*. The *KIR* gene cluster is a complex entity, as is reflected by allelic polymorphism (5), gene copy number variation resulting in different haplotypic configurations (6), variegated expression (7, 8), and complex chromosomal recombination events (9-11).

The *KIR* genes are tandemly arranged on chromosome 19q13.4, each spanning 10,000-15,000 base pairs (bp), and are separated by ~1,000 bp (12). The receptors are encoded by up to nine exons, of which the first two exons encode the leader peptide, followed by exons encoding two or three extracellular Ig-like domains (2D or 3D; exons 3-5), a stem structure (exon 6), a transmembrane region (exon 7), and a cytoplasmic tail (exons 8-9) (13, 14). A long cytoplasmic tail (L) contains two immunoreceptor tyrosine-based inhibitory motifs (ITIM) and characterizes inhibitory KIR. Activating KIR feature a short cytoplasmic tail (S) and a positively charged residue in the transmembrane region, which interacts with molecules that contain the immunoreceptor tyrosine-based activation motif (ITAM).

In the past, numerous *KIR* characterization studies were mainly performed at the genomic DNA (gDNA) level, thereby lacking information about transcription and post-transcriptional modifications (PTM) of the transcripts. Recently, next-generation sequencing (NGS) has improved and speeded up characterization of the *KIR* gene cluster, resulting in the identification of novel alleles, recombinant genes, and haplotypes (9-11, 15-17). In addition, NGS also enables the characterization of transcripts that are subjected to alternative splicing. The alternative splicing of transcripts is a prevalent form of PTM, which is observed for approximately 95% of the human multi-exon genes, and it plays a crucial role in the regulation of protein diversity and tissue-specific gene expression (18, 19). Normally, precursor messenger RNA (pre-mRNA) is converted to mRNA by constitutive splicing, which involves the removal of introns and the ligation of exons by the spliceosome, which is a complex of five small nuclear RNAs (U1, U2, U4, U5, and U6), and multiple associated core proteins (20-22). To correctly identify the splice sites, a precise interplay of conserved sequence elements present on pre-mRNA (*cis*-acting), along with spliceosome factors, (*trans*-acting) is required. Deviation from constitutive splicing, caused by variation and mutations in the pre-mRNA sequence and/or an imbalance in the *trans*-acting splice factors, can result in alternative splicing. This process can be either beneficial, as a wide variety of isoforms can originate from a single gene, or detrimental, as different isoforms can be involved in the development of various diseases, such as spinal muscular atrophy and different forms of cancer (23-25). Despite its clinical relevance, at present the splicing pattern of only a few multi-exon genes has been described thoroughly (26, 27).

Three groups reported alternatively spliced KIR transcripts that lacked complete exons (28-32). For example, *KIR2DL4* transcripts may lack exon 3 (D0 domain), suggesting the existence of a protein structure with only the D2 extracellular domain (30). Likewise, *KIR2DL5* splice variants with deletions of exon 5, exons 3 and 5, and exon 7 have been reported; these might encode protein structures lacking the extracellular D2 domain, both the D0 and D2 domains, and the transmembrane region, respectively (31). The latter suggests the existence of a soluble KIR2DL5 isoform. In addition, deletions of only fragments of the Ig-like domains have been described.

Over the past few years, the *KIR* gene family has also been characterized in non-human primate species, which provide insights into the evolution of this section of the immune system, and eventually may help to optimize and refine animal models (9, 33-37). Rhesus macaques (*Macaca mulatta*, *Mamu*), for example, share a close evolutionary relationship with humans, as is reflected by similar immune responses and pathologies in many models for infectious and autoimmune diseases (38-40). Although there are certain subtle differences such as different receptor lineage expansion (lineage II and III in macaques and humans, respectively), and the absence of a haplotype A and B organization in macaques, the *KIR* cluster in macaques is highly similar to that observed in humans (9, 41). Within the rhesus macaque *KIR* family, twenty-two genes are identified, including receptors with a single extracellular domain (*Mamu-KIR1D*), a homolog of human *KIR2DL4,* and multiple *KIR3D* gene structures (33, 42). Rhesus macaque *KIR* haplotypes can contain four to fourteen genes, illustrating the extensive copy number variation (9, 33). As observed in humans, the macaque *KIR* cluster is also characterized by allelic variation, variegated expression, and chromosomal recombination.

Alternatively spliced KIR transcripts with deletions of complete or partial domain-encoding exons are described for a few *Mamu-KIR3D* genes, identified by Sanger sequencing (36, 42, 43). For example, the *Mamu-KIR3DL20* gene, which shows sequence similarity with human lineages I and V *KIR*, is hypothesized to consistently generate *Mamu-KIR2DL05* transcripts by the skipping of exon 4 (33, 42, 44, 45).

Although a few KIR splice variants were already reported in different primate species, the current literature lacks a comprehensive overview of the modifications of *KIR* genes generated by alternative splicing, as well as an indication of its possible functional consequences. By using a Single-Molecule, Real-Time (SMRT) sequencing approach on the Pacific Bioscience's (PacBio) Sequel platform, we were able to thoroughly characterize the alternative splicing of *KIR* gene transcripts in both human and rhesus macaque families. The chosen high-resolution method provides insights into the segregation of alternatively spliced KIR transcripts and the potential splicing mechanisms. The data illustrated that alternative splicing adds another layer of complexity to the KIR family in both humans and rhesus macaques. Moreover, the alternatively spliced *KIR* gene isoforms might encode receptors having a modified structure, function, and/or expression profile, which consequently might play a custom role in health and disease.

Materials and Methods

Transcriptome datasets

The KIR transcriptomes of 15 related humans and 30 related rhesus macaques were reported previously (9). In addition, during the course of this study the KIR transcriptomes of another three rhesus macaque families, which in total comprised 25 macaques, and one human family, comprising six individuals, samples of whom were provided by Sanquin (Amsterdam, The Netherlands), were analyzed as previously described (9). In short, total RNA was isolated from human and rhesus macaque PBMCs, and cDNA was synthesized. Primer sets were designed for human *KIR2DL4* and *KIR2D/3D*, which amplified all human *KIR* genes except for *KIR3DL3*, *KIR2DL5*, and the pseudogenes. A *KIR2DL04*-specific primer set along with two *KIR2D/3D* primer sets amplified macaque *KIR*. Tagged KIR amplicons were pooled and purified, and SMRTbell libraries were generated. Sequencing was performed on a PacBio Sequel platform using P6-C4 sequencing chemistry. Informed consent was obtained from all participants.

Identification of alternative spliced KIR transcripts and splice elements

Subsequent to PacBio sequencing, the circular consensus sequences were selected for high read quality, and were demultiplexed based on unique barcoding. Geneious Pro R10 software (46) was used to map the PacBio reads to reference databases that included all reported full-length and partial human and rhesus macaque *KIR* allele sequences, which were derived from the IPD-KIR database and the literature (5, 33, 43, 45, 47-51), to identify 100% matched reads (0% mismatch, maximum ambiguity = 1, minimum mapping quality = 30, 80% minimum overlap identity, minimum overlap = 400). Next, the unused reads were subjected to deletion and structural variant discovery, which can align paired and unpaired reads that include structural rearrangements, deletions, and insertions, to reference sequences from the databases (0% mismatch, maximum ambiguity = 1, minimum mapping quality = 30, 10 gaps per read allowed, minimum overlap = 100). A splice variant was confirmed when observed in two or more individuals with at least three supporting reads. For each gene for which a specific splice event was confirmed, the sequence was submitted to the ENA database and received an accession number (**Suppl. Table 1** and **2**). In addition, Sanger sequencing was used to confirm alternative splicing in transcripts of human *KIR2DL5* (deletion of 294 bp), human *KIR2DL4* (deletion of 104/105 bp), *Mamu-KIR3DL01* (inclusion of 170 bp), *Mamu-KIR3DL20* (deletion of 294 bp and 300 bp), and *Mamu-KIR2DL04* (inclusion of 245 bp), using primers designed in the inserted region, or at the boundary of the deleted region. The skipping of exon 4 in *Mamu-KIR3DL20* transcripts was visualized by gel electrophoresis, using gene-specific primers situated at the boundary of exons 1/2 and at the end of exon 5.

***KIR* intron sequences**

Macaque *KIR* intron sequences are almost absent from the literature, except for introns in a completely sequenced rhesus macaque KIR haplotype (45). Therefore, we extracted genomic DNA (gDNA) from EDTA whole blood samples by a standard salting-out procedure, or from ~15 x 10^6 PBMCs with an AllPrep RNA/DNA Mini Kit (Qiagen), in accordance with the manufacturer's instructions. We designed generic primer sets, tagged with PacBio barcodes, which amplified one or multiple introns, and flanking exons, of *Mamu-KIR3DL/S*, *Mamu-KIR1D*, and *Mamu-KIR2DL04* (**Suppl. Table 3**). Thermal cycling conditions were denaturation at 98°C for 2 min, followed by "x" cycles of 98°C for 20 s, "x" °C for 30 s, and 72°C for 2 min (number of cycles and annealing temperature are indicated in **Suppl. Table 3** for each primer set). Sequencing was performed on a PacBio Sequel platform. The majority of the human *KIR* intron sequences were derived from the IPD-KIR database (52). Additional sequences of human *KIR* introns 6, 7, and 8, and flanking exons, were obtained by amplification with two primer sets, using the above-mentioned thermal cycling conditions **(Suppl. Table 3**). The obtained intron sequences could be assigned to the corresponding *KIR* genes or alleles based on the flanking exon sequences.

Splicing strength prediction

Multiple prediction tools have been developed and compared to score sequence elements that are involved in splicing, such as the 3' splice site (ss) region, including the polypyrimidine tract (PPT), and the 5' ss (53-57). In all studies, the Maximum Entropy Modeling Scan (MaxEntScan; MES) (58), the Position Weight Matrix (PWM) via SpliceView (59), and the Human Splice Finder (HSF) (60) outperformed the other tools, and were therefore selected to predict the splicing strength of the KIR splice elements. The different prediction tools use varying nucleotide ranges to score the splicing strength of the 3' ss, which is likely due to the degenerate nature of this motif (**Suppl. Table 4**). The 5' ss was mainly defined by three exonic (-3) and six intronic nucleotides (+6). The output value of the tools also has different ranges, but a higher score always implies a more precise prediction (**Suppl. Table 4**). It should be noted that the scores are not a measure of effect sizes, and there are no thresholds that can predict whether or not a splice event will occur. The scores should only be used to facilitate a comparison between related splice sites. In addition to MES, PWM, and HSF, the Weight Matrix Model (WMM) (58) and NNSplice tool (61) were evaluated, and used when other tools failed to provide a splicing strength score.

Results

Overview of alternative splicing events in human KIR genes

In a preceding family study, single Molecule, Real-Time (SMRT) sequencing on a Pacific Bioscience's (PacBio) Sequel platform was used to obtain *KIR2DL4* and *KIR2D/3D* transcript profiles of fifteen related individuals (9). These transcript profiles partly consisted of reads that matched 100% to known *KIR* alleles. The dataset, however, also comprised a considerable number of partial sequences, and sequences that contained single nucleotide gaps. In this communication, we performed an in-depth analysis of these latter datasets, and determined that approximately 53% and 4% of the 100%-matched reads (error-free reads) accounted for alternatively spliced *KIR2DL4* and *KIR2D/3D* transcripts, respectively.

In total, 29 distinct KIR splice events were identified (≥ 3 PacBio reads), of which 18 were observed in two or more related individuals (**Table 1**). These independently confirmed splice events involved both insertions (6 events) and deletions (12 events), and can be categorized into common types of splicing mechanisms, such as exon skipping, alternative 3'- and 5'-splice sites (ss), and cryptic exon inclusion (**Figure 1** and

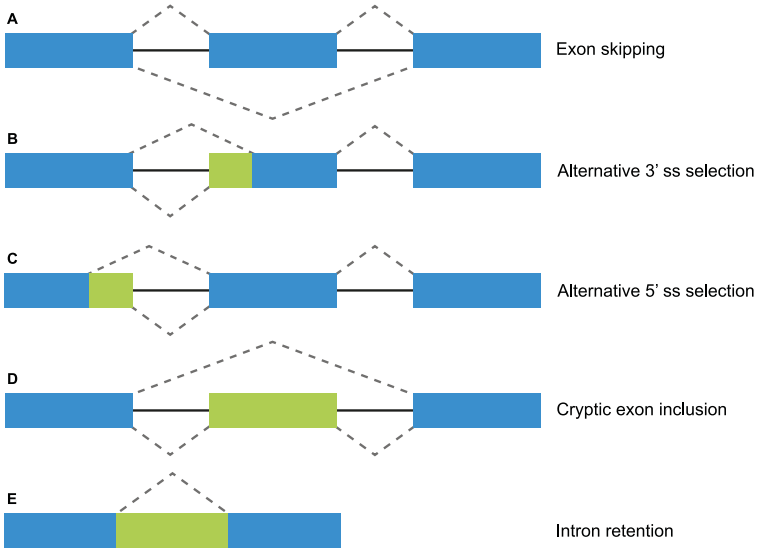


Figure 1. Different mechanisms of alternative splicing. Blue boxes indicate exonic regions. Green boxes indicate intronic sequences that are included in mature mRNA by alternative splicing. The upper dashed lines indicate constitutive splicing, whereas the lower dashed lines indicate alternative splicing. (A) Exon skipping (or 'cassette exon') is the most prevalent form of alternative splicing, which involves the complete deletion of one or multiple exons. (B, C) An alternative 3' or 5' splice site (ss) can result in a partial intron retention (observed as insertion in the mRNA) or deletion. (D) Intron retention is a less common form of alternative splicing, and involves the inclusion of a complete intron. (E) Introns can include exonic sequences that can contain intact 3' and 5' splice sites. The incorporation of these cryptic exons is prohibited during constitutive splicing, but they can be included into the mature mRNA by alternative splicing.

Table 1. Eighteen splice events identified in fifteen human individuals by PacBio sequencing at the transcription level. The events can be catagorized into exon skipping (or 'cassette' exons), alternative 3' and/or 5' splice sites, and cryptic exon inclusion. The size of deletions or inclusions are indicated in base pairs (bp). The result of the alternative splice event is predicted.

Splice event	Deletion / Inclusion	Size	Position	Observed in *KIR* genes	Result
Exon skipping ('cassette' exon)	Deletion	51 bp	Exon 6	*2DL1/3/4, 2DS2/4, 3DL2*	Missing stem region
	Deletion	294 bp	Exon 5	*2DL2/3/5, 3DL1*	Missing D2 domain
	Deletion	594 bp	Exons 4 and 5	*3DL1, 3DL3*	Missing D1 and D2 domains
	Deletion	104/105 bp	Exon 7	*2DL4, 3DL2*	Missing transmembrane region; possible soluble receptor
	Deletion	155/156 bp	Exons 6 and 7	*2DL4, 3DL2*	Missing stem and TM region; possible soluble receptor
	Deletion	158 bp	Exons 7 and 8	*2DL4*	Missing TM region, and part of cytoplasmic tail; possible soluble receptor
	Deletion	209 bp	Exons 6, 7, and 8	*2DL4*	Missing stem, TM region, and part cytoplasmic tail; possible soluble receptor
Alternative 3' ss	Inclusion	49 bp	Following exon 7	*2DL1*	Out-frame; stopcodon introduced
	Inclusion	170 bp	Following exon 5	*3DL1*	Stopcodon introduced
	Deletion	150 bp	Start exon 5	*2DL3*	In-frame deletion of the first 150 bp in the D2 domain
Alternative 5' ss	Inclusion	67 bp	Following exon 7	*2DL4*	"9A" KIR2DL4 alleles: frameshift restored ORF "10A" KIR2DL4 alleles: out-frame; stopcodon introduced
	Inclusion	129 bp	Following exon 4	*3DL2*	Stopcodon introduced
	Deletion	66 bp	End exon 3	*2DL4*	In-frame deletion of the first 66 bp in the D0 domain
	Deletion	73/74 bp	End exon 7	*2DL4*	Out-frame; stopcodon introduced
	Deletion	198 bp	End exon 3	*2DL4*	Missing 66 AA in end of D0 domain. Only in combination with a deletion in the TM region.
Alternative 3' and 5' ss	Deletion	294 bp	End exon 4, begin exon 5	*2DL3, 2DS4, 3DL1/2*	In-frame deletion of end D1 domain and begin D2 domain
Cryptic exon	Inclusion	78 bp	Following exon 6	*2DL1/2/3, 2DS1*	In-frame; positively and negatively charged residues introduced
	Inclusion	54/57 bp	Following exon 5	*2DL1, 2DS1/4/5*	Stopcodon introduced

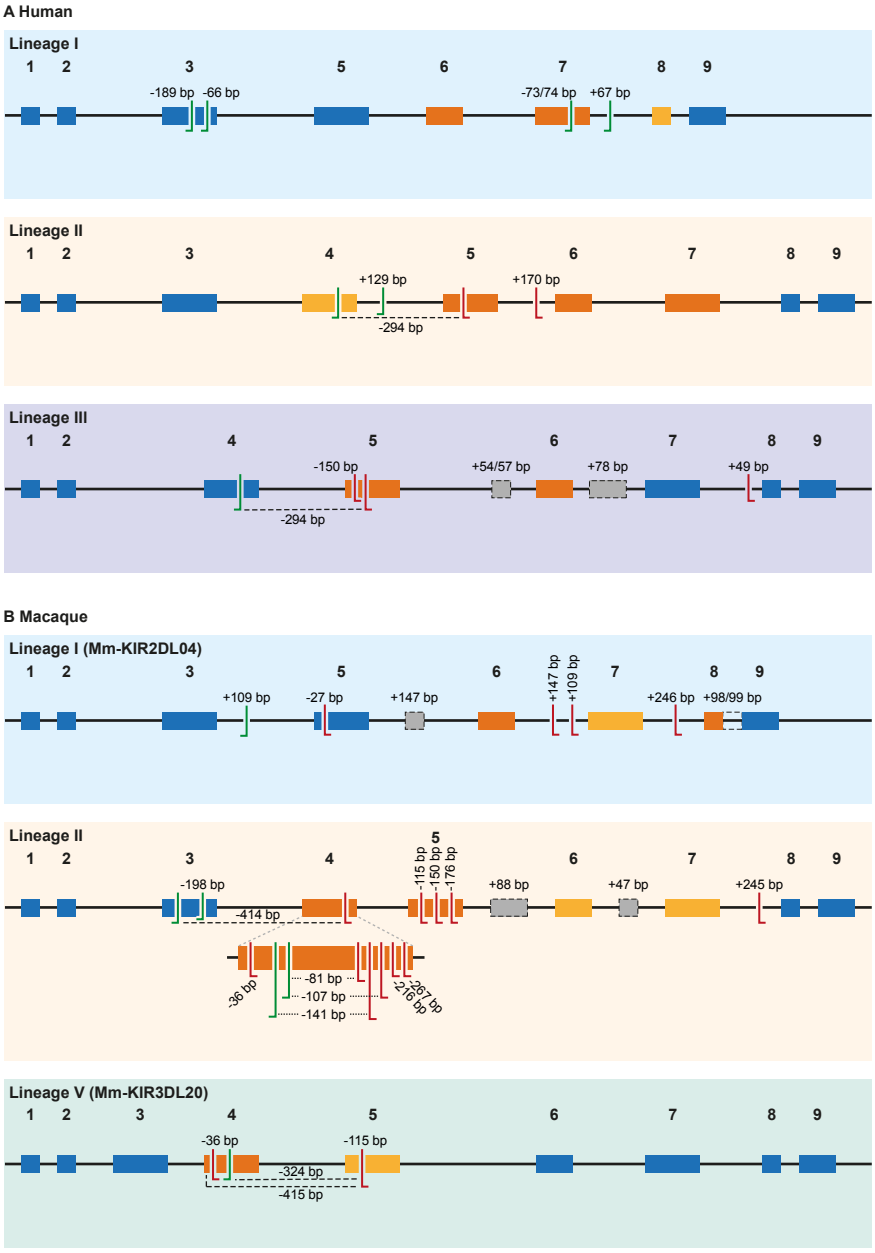
Table 1). In **figure 2A**, a schematic overview is provided of the confirmed splice events summarized in **Table 1**. The excision of exon 6, which encodes the stem region, represented the most frequently observed splice event, and was identified in alleles of six different *KIR* genes (**Table 1**). Other commonly observed splice events were the deletion of exon 5 (D2 domain), a deletion of 294 bp mediated by an alternative 5' ss at the end of exon 4 and an alternative 3' ss at the beginning of exon 5, an insertion of 54/57 bp following exon 5, and an insertion of 78 bp subsequent to exon 6. The remaining splice events were specific for one or two *KIR* genes. In transcripts of *KIR2DL1*, *KIR2DL3*, *KIR2DL4*, *KIR3DL1*, and *KIR3DL2,* at least four different splice events were observed, resulting in a diverse range of isoforms for these genes. The most diverse alternative splicing profile was observed for *KIR2DL4*, for which eight different splice events were identified. Most of these events were *KIR2DL4*-specific, including a frequently observed insertion of 67 bp subsequent to exon 7, and a deletion of 66 bp in exon 3. For transcripts encoded by activating *KIR* genes less alternative splicing events were observed, which might be explained by the lower frequency of these genes in the individuals studied. Similar splice events were observed in an additional human family comprising six individuals, confirming the obtained splice profiles, and suggesting that the data provides a comprehensive overview of alternative splicing in human *KIR*.

In silico prediction of cis-acting splicing elements

Constitutive and alternative splicing of pre-mRNA is regulated by *trans*-acting factors (small nuclear RNAs, spliceosome core proteins), and their cognate nucleotide sequence *cis*-elements near the intron-exon boundaries (54, 62). Essential splicing *cis*-elements are the 3' splice site (ss), the 5' ss, the branch point sequence (BPS), and the polypyrimidine tract (PPT) (**Figure 3**). Additional enhancer and silencer elements can be identified in the exons (Exon Splicing Enhancer, ESE; Exon Splicing Silencer, ESS) and introns (Intron Splicing Enhancer, ISE; Intron Splicing Silencer, ISS). These regulatory splicing sequences are degenerate, and the consensus sequences can only be loosely followed (63). Although software tools are available to predict and score the splicing strength of all different *cis*-elements (54), we mainly focused on the better modeled prediction of the splice site elements (3'- and 5' ss, BPS, and PPT). In the following sections, different observed events (**Table 1**, **Figure 2A**) are substantiated by the identified splice sites, and by their corresponding *in silico*-predicted splicing strength scores, per alternative splicing mechanism (**Figure 1**).

Exon skipping in human KIR transcripts

The skipping of one or multiple exons was the most frequently observed alternative splicing mechanism in the KIR transcriptomes of the human family studied (**Figure 1A**, **Table 1**, and suppl. **Table 1**). The skipping of exon 7, which encodes the transmembrane region, was observed in *KIR2DL4* and *KIR3DL2* transcripts, and might be explained by variation in the splicing *cis*-elements (**Figure 4**). In all *KIR* genes, identical BPS and 3' ss sequences were identified in intron 6 preceding exon 7, and were in agreement with


A Human
Lineage I
1 2 3 5 6 7 8 9
-189 bp -66 bp
-73/74 bp +67 bp
Lineage II
1 2 3 4 5 6 7 8 9
+129 bp
+170 bp
-294 bp
Lineage III
1 2 4 5 6 7 8 9
-150 bp
+54/57 bp
+78 bp
+49 bp
-294 bp
B Macaque
Lineage I (Mm-KIR2DL04)
1 2 3 5 6 7 8 9
+109 bp
-27 bp
+147 bp
+147 bp
+109 bp
+246 bp
+98/99 bp
Lineage II
1 2 3 4 5 6 7 8 9
-198 bp
-414 bp
-115 bp
-150 bp
-176 bp
+88 bp
+47 bp
+245 bp
-36 bp
-81 bp
-107 bp
-141 bp
-216 bp
-267 bp
Lineage V (Mm-KIR3DL20)
1 2 3 4 5 6 7 8 9
-36 bp
-115 bp
-324 bp
-415 bp

◀ **Figure 2. Overview of alternative splice sites in human and rhesus macaque KIR.** A schematic representation of the different splice events observed in human (A) and rhesus macaque (B) KIR transcripts categorized by gene lineage. The splice events illustrated correspond to the splice events summarized in tables 1 and 2, and are indicated with the size of the inclusion (+) or deletion (-) in base pairs (bp), or by color-coding. The black line indicates the introns, whereas colored boxes represent the exons. Exons that are subjected to exon skipping are illustrated with dark orange boxes, and the exons that are only skipped in combination with one or more exons are indicated in light orange boxes. Exons that are not subjected to exon skipping are colored blue. The actual splice sites, which map to the exon/intron boundaries, have not been indicated. Alternative 3' splice sites (ss) are indicated with green left-directed hooks, whereas alternative 5' ss are indicated with red right-directed hooks. An alternative splice site always pair with the adjunct complement actual 3' or 5' splice site, except splice events that are mediated by a set of 3' and 5' alternative ss, which are marked with a dashed line. Cryptic exons are illustrated as grey boxes with a dashed line, and the (alternative) splice sites of these cryptic exons are not explicitly indicated. The intron retention event observed in rhesus macaque KIR2DL04 (lineage I) is indicated with a dashed line between exons 8 and 9. Exon 4 in rhesus macaque lineage II KIR genes is enlarged to more precisely illustrate the high number of alternative splice events observed.

the consensus sequences YUNAY and NYAG/G (**Figures 3** and **4**; / marks actual splice site), respectively (64). Compared to the 5' ss sequence of lineage III *KIR* genes (MES: 9.72; HSF: 88.47; PWM: 87), a single nucleotide variation (T/C) was observed in the *KIR2DL4* 5' ss sequence of exon 7, and this resulted in a lower *in silico*-predicted splicing strength score (MES: 7.31; HSF: 86.29; PWM: 84) (**Figure 4**). Also for *KIR2DL5* and *KIR3DL2*, a decreased splicing strength score was predicted for the 5' ss sequence of exon 7 (MES: 9.35; HSF: 83.61; PWM: 84), and these genes are discriminated from other *KIR* genes by a single nucleotide as well (A/G). Furthermore, the PPT of exon 7 in *KIR2DL1* and *KIR2DL2* contained a single adenine substitution as compared to the remaining lineage III *KIR* genes, but despite the predicted lowered 3' ss splicing strength score, the skipping of exon 7 was not observed for the corresponding transcripts. The

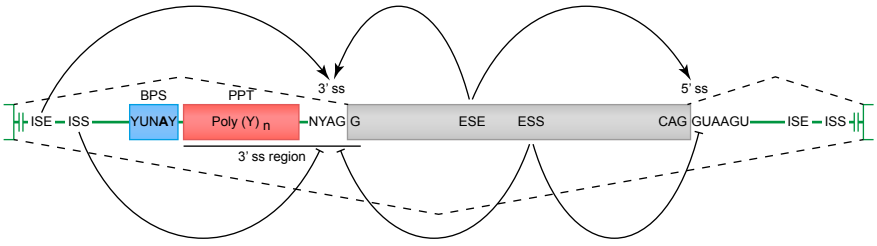


Figure 3. *Cis*-acting motifs that mediate constitutive and alternative splicing. The boundaries of exons (grey box) and introns (green line) are marked by splice sites (ss). At the 3' ss, the end of the intron is characterized by an adenine and guanine (AG), and forms the basis of the 3' ss motif. At the 5' ss, the start of the intron is marked by a guanine and thymine (GT), and forms the basis of the 5' ss motif (-3 bp in exon, and +6 bp in intron). Prior to the 3' ss, a branch point sequence (BPS) and polypyrimidine tract (PPT) can be identified, and these elements are involved in spliceosome binding and intron exclusion. The 3' splice site and PPT together are referred to as the 3' region (-20 bp in intron, and +3 bp in exon), and can be used to predict the splicing strength. In addition to the splice site motifs, enhancer and silencer motifs can be identified in the exons (Exon Splicing Enhancer, ESE; Exon Splicing Silencer, ESS) and introns (Intron Splicing Enhancer, ISE; Intron Splicing Silencer, ISS), and can stimulate or inhibit splicing of an exon.

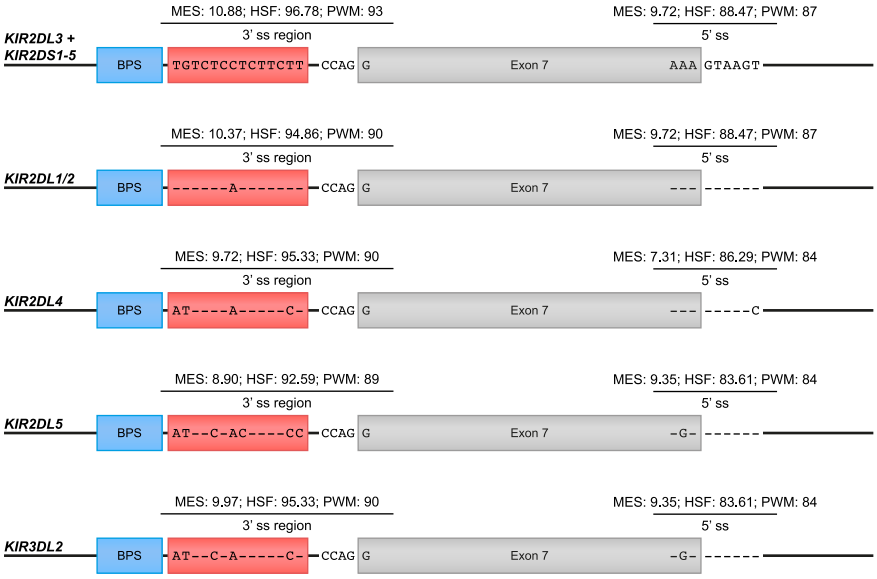


Figure 4. Suboptimal splice sites might mediate the skipping of exon 7 in KIR2DL4, KIR2DL5, and KIR3DL2 transcripts. Exon 7 is indicated in grey boxes, whereas the BPS and PPT elements are indicated in blue and red boxes, respectively. The 3' ss region (3' ss and PPT) and the 5' ss sequences of lineage III *KIR* genes, except for *KIR2DL1/2*, are used as consensus sequences. Dashes (-) indicate sequence identity with the consensus sequence. The splicing strength scores of the 3' and 5' splice sites are provided. The BPS and 3' splice sites of exon 7 were identical, but variation was observed in the PPT and 5' splice site.

PPT of exon 7 of *KIR2DL4*, *KIR2DL5*, and *KIR3DL2*, however, varied from the lineage III *KIR* genes at four to seven nucleotide positions. This variation included the presence of two adenines that interrupted the guanine- and thymine-rich tract, and although a long continuous PPT is not required for splicing, it does appear to increase the splicing efficiency (65, 66). Indeed, a decreased splicing strength score of the 3' ss region (3' ss + PPT) of exon 7 in *KIR2DL4*, *KIR2DL5*, and *KIR3DL2* was predicted (**Figure 4**). Thus, the observed skipping of exon 7 in *KIR2DL4* and *KIR3DL2* transcripts might be explained by deviations in the 5' ss and the PPT together, and suggest the existence of soluble isoforms of these receptors. In addition, the absence of exon 7 in KIR2DL4 molecules results in the loss of their activating signaling potential, which is facilitated by a positive residue in the transmembrane region. Based on the data derived from the prediction tools, the skipping of exon 7 could be expected in *KIR2DL5* transcripts as well, and was indeed reported previously (31). However, despite the presence of the *KIR2DL5* gene in some individuals, we did not identify the event in the human family studied.

Other exon skipping events involved the complex of exons 4 and 5, exon 5 only, and exon 6, respectively, encoding the extracellular domains and the stem region (**Figure 2A**). In particular, the skipping of exon 6 was frequently identified, and observed

in *KIR* genes of lineages I, II, and III, suggesting the presence of conserved suboptimal *cis*-elements. However, between the different *KIR* genes, extensive nucleotide variation was observed in the BPS, PPT, and 5' ss of exon 6, which resulted in a variety of predicted splicing strength scores, implying that conserved suboptimal splice sites did not mediate the splice event. In addition to less efficient splice sites, skipped exons are often characterized by longer flanking introns that can obstruct exon recognition, or that contain splice enhancer and silencer motifs (62). The two largest introns of the *KIR* gene are those flanking exon 6, and might mediate exon skipping. Phylogenetic analysis of introns 5 and 6 illustrated, however, that the introns do not have a close evolutionary relationship across the different *KIR* lineages. Despite lineage variation in the introns, ISE and ISS motifs, or elements that induce secondary intron structures, might be conserved between these introns, but these elements are hard to predict using the available *in silico* models.

Alternative splice sites in human KIR

Alternative 3' or 5' splice sites are thought to be an intermediate between constitutively spliced and skipped exons, and can introduce in- and out-frame deletions and insertions in transcripts (**Figures 1B, 1C,** and **2A**) (62). An example of an alternative splice event, caused by an alternative 3' ss, is the retention of 170 bp of intron 5, which was observed in *KIR3DL1* transcripts (**Table 1**). This partial intron retention introduced a premature stop codon subsequent to exon 5, resulting in a transcript that encodes only the extracellular domains, and could be explained by the presence of an additional 3' splice site upstream of the actual splice site (**Figure 5A**). However, according to the *in silico* models, the splicing strength of the alternative 3' ss region (MES: 4.72; HSF: 77.69; PWM: 81) is remarkably lower compared to the actual splice site (MES: 11.54; HSF: 80.64; PWM: 86), which might indicate that this splice event is not common. Additionally, a BPS prior to the alternative 3' ss that matches the consensus sequence was not observed. The low number of PacBio reads (≤ 6 reads) for this KIR3DL1 splice variant might already be indicative that this splice event, although observed in three individuals, is not favorable over constitutive splicing. Even more, the introduction of a premature stopcodon might be indicative for the degradation of the alternatively spliced transcript by the nonsense-mediated decay. Nonetheless, the presence of this splice variant was confirmed by Sanger sequencing, and might still have functional relevance in certain NK cell subsets that are resident in specific tissues.

A partial deletion at the end of exon 7 was observed in *KIR2DL4* transcripts, and was mediated by an alternative 5' splice site (**Figure 5B**). The end of exon 7 in *KIR2DL4* is marked by a poly-adenine sequence that can be nine (9A) or ten (10A) nucleotides long (30). The "9A" *KIR2DL4* alleles have a premature stopcodon subsequent to exon 7, suggesting the absence of a cytoplasmic tail, and thereby the loss of their inhibitory potential (**Figure 5B**). The "10A" *KIR2DL4* transcripts encode a complete receptor, including a cytoplasmic tail with a single ITIM. Deletions of 73 and 74 bp at the end of exon 7 were observed in transcripts of "9A" and "10A" *KIR2DL4* alleles, respectively (**Figure 5B**, **Table 1**). These deletions were mediated by an alternative 5' ss that is

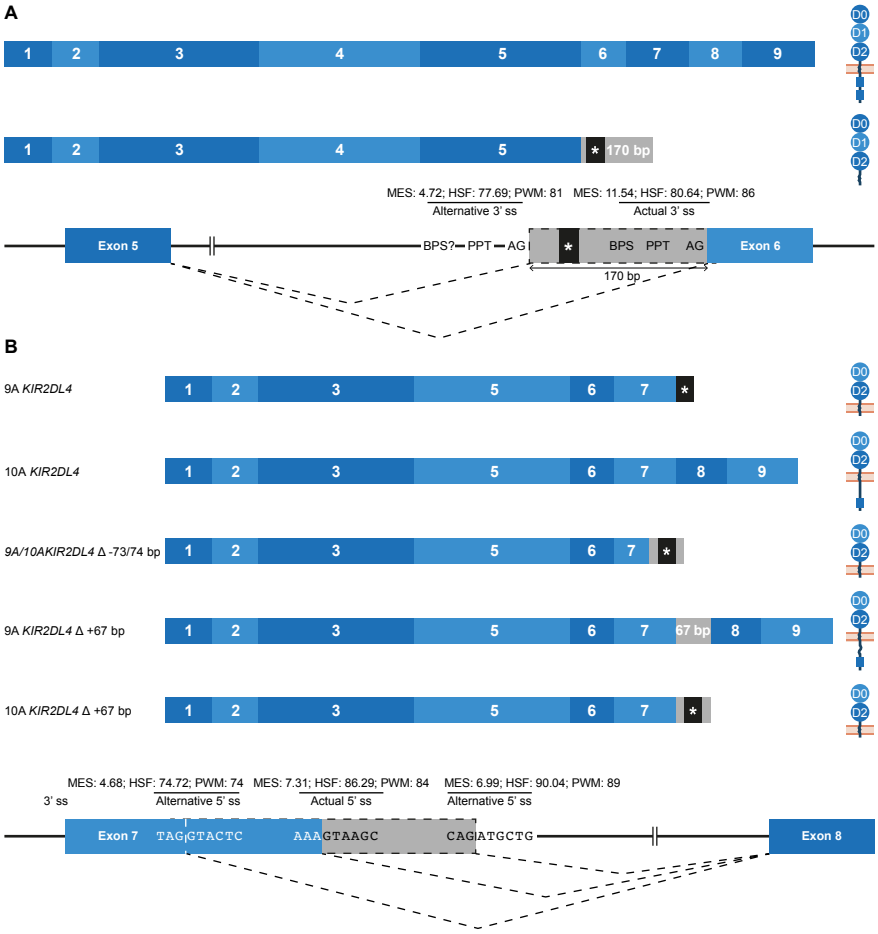


Figure 5. Alternative splice sites mediate deletions and insertions at the transcription level. The observed transcripts are illustrated, in which exons are indicated by blue boxes, and corresponding protein structures are schematically depicted adjacent to the transcript. Intron inclusions are indicated in grey boxes. Dashed lines indicate the potential splice events, and predicted splicing strength scores are provided for actual and alternative splice sites. Stopcodons are indicated by black boxes with an asterisk (*). (A) The constitutive splicing of human *KIR3DL1* results in a transcript including nine exons, and encodes a KIR3DL molecule. Alternative splicing, mediated by an alternative 3' splice site located 170 bp downstream of the actual splice site, results in a transcript encoding only three extracellular domains. "BPS?" refers to the potential lacking of a BPS for the alternative 3' ss. (B) The constitutive splicing of human "9A" and "10A" KIR2DL4 alleles results in membrane-bound molecules containing two extracellular domains, or molecules that contain two extracellular domains and a cytoplasmic tail including a single ITIM, respectively. An alternative 5' ss located in exon 7 results in a partial deletion of the transmembrane region in both "9A" and "10A" *KIR2DL4* alleles, and the introduction of a stopcodon. A second alternative 5' ss located in intron 7 results in a partial intron inclusion of 67 bp. In "9A" *KIR2DL4* alleles, this inclusion restores the open reading frame, and these isoforms probably express an inhibitory cytoplasmic tail. In contrast, the same splice event in "10A" *KIR2DL4* alleles results in a frame-shift that introduces a stopcodon subsequent to exon 7.

located within exon 7, and caused a frameshift that introduced a premature stop codon, suggesting a soluble KIR2D molecule. The *in silico* models predicted that the splicing strength score of the actual 5' ss is higher (MES: 7.31; HSF: 86.29; PWM: 84) than the splicing strength score of the alternative 5' ss located in exon 7 (MES: 4.68; HSF: 74.72; PWM: 74), suggesting that constitutive splicing would be more prevalent. In addition, another alternative 5' ss was observed in intron 7 of *KIR2DL4*, which resulted in a partial intron inclusion of 67 bp subsequent to exon 7. This alternative 5' ss scored a higher predicted splicing strength (MES: 6.99; HSF: 90.04; PWM: 89) then the alternative 5' ss located in exon 7, and even scored higher compared to the actual 5' ss according to the HSF and PWM models. This might indicate that the inclusion of 67 bp subsequent to exon 7 in *KIR2DL4* transcripts is a prevalent splicing event, which is also supported by high PacBio read counts observed for this splice variant (an average of 115 PacBio reads per individual). In the "10A" *KIR2DL4* transcripts, the partial intron inclusion mediated by the alternative 5' ss in intron 7 caused a frameshift that introduced a stopcodon subsequent to exon 7, and they thereby lack the cytoplasmic tail that includes an ITIM. In contrast, in the "9A" *KIR2DL4* alleles, which normally encode a truncated receptor, the open reading frame (ORF) was restored by the partial intron inclusion, resulting in transcripts that encode a KIR protein including a cytoplasmic tail. These examples suggest that alternative splicing might regulate whether the KIR2DL4 receptors contain a cytoplasmic tail, and thereby maintain their inhibitory function, or not.

"Cryptic" exons in human KIR

Some potential exons – referred to as cryptic exons – are located within intronic regions, and are normally not spliced into mature mRNA by constitutive splicing (**Figure 1E**); this could be due to intrinsic defects, the presence of splice silencer elements, or the formation of inhibiting RNA secondary structures (67). Nonetheless, alternative splicing can mediate the inclusion of cryptic exons in the transcript, as is previously described for *KIR2DL1*, and this might play a role in health and disease (67-70). In the family studied, multiple alternative splice events that introduced a cryptic exon were identified. For example, an inclusion of 78 bp that originated from intron 6 was observed in transcripts of *KIR2DL1*, *KIR2DL2*, *KIR2DL3*, and *KIR2DS1* (**Figure 6A**). The extended transcripts remained in-frame, and the 26 introduced amino acids (cryptic exon) between the stem and transmembrane region included positively and negatively charged residues. The stretch of amino acids was found to be highly conserved in the four *KIR* gene products mentioned above, with only one single nucleotide variation present in the alleles studied. In all other *KIR* genes, except for *KIR2DL5* and *KIR3DL3*, the cryptic exon could be identified at the same position in intron 6 (~1426 bp 3' of exon 6, ~2755 bp 5' of exon 7), but sequence variability was observed, varying from five to fifteen nucleotides, as compared to the above-mentioned conserved sequence. This variation might involve ESE and ESS motifs, which could inhibit cryptic exon inclusion. In *KIR* genes that contained the cryptic exon or the variant cryptic exon, a BPS, PPT, and 3' ss were identified, and similar 3' ss splicing

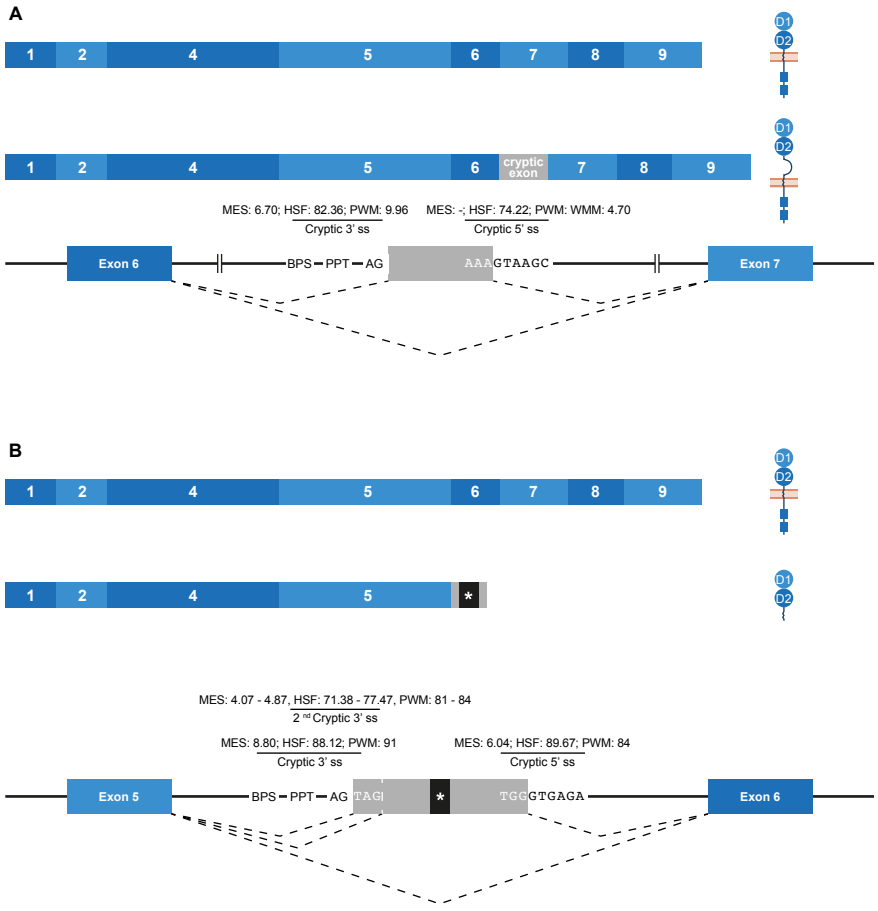


Figure 6. Inclusion of cryptic exons by alternative splicing. The observed transcripts are illustrated, in which exons are indicated by blue boxes, and corresponding protein structures are schematically depicted adjacent to the transcript. Cryptic exons are indicated in grey boxes. Dashed lines indicate the potential splice events, and predicted splicing strength scores are provided for cryptic splice sites. Stopcodons are indicated by black boxes with an asterisk (*). (A) In transcripts of KIR2DL1-3/2DS1, a cryptic exon of 78 bp was observed that originated from intron 6. This inclusion extends the region between the stem and transmembrane region by 26 amino acids, including positively and negatively charged residues. (B) In transcripts of *KIR2DL1* and *KIR2DS5*, a cryptic exon of 57 bp was observed, which originated from intron 5. Three bp upstream, a second cryptic 3' ss was identified, which explained the cryptic exon inclusion of 54 bp in transcripts of *KIR2DS1* and *KIR2DS4*. At the gDNA level, the presence of one or both of these cryptic exons was also identified in other lineage III *KIR* genes and *KIR2DL5*.

strength scores were predicted (data not shown). The 5' ss of *KIR2DL1-3* and *KIR2DS1* could be distinguished from the 5' ss of other *KIR* genes by a substitution of a cytosine with an adenine, which resulted in a higher predicted 5' ss splicing strength score in *KIR2DL1-3* and *KIR2DS1* (MES:-; HSF: 74.22; PWM:-; WMM: 4.70) compared to the *KIR* genes that had a cytosine in the 5' ss (MES:-; HSF: 65.41; PWM:-; WMM: 1.71). Thus, this mutation might contribute to the inclusion of the cryptic exon at the transcription level. Furthermore, phylogenetic analysis illustrated that intron 6 of each *KIR* gene clustered separately, but that the evolutionary distance of *KIR2DL1*, *KIR2DL2*, *KIR2DL3*, and *KIR2DS1* was small compared to the other *KIR* genes. Although hard to predict, the variation in intron 6 sequences might involve ISE and ISS motifs that, in combination with the cryptic exon variation and 5' ss mutation, contribute to the inclusion of the cryptic exon.

Another example of a cryptic exon inclusion is the insertion of 57 bp that was observed in transcripts of *KIR2DL1* and *KIR2DS5* (**Figure 6B** and **Table 1**). This cryptic exon originated from intron 5 (~837 bp 3' of exon 5, ~2259 bp 5' of exon 6), and introduced a stopcodon subsequent to exon 5, resulting in transcripts encoding only the D1 and D2 domains. In all *KIR* genes, this cryptic exon could be identified at the same position in intron 5, with variation up to nine nucleotides. However, only in four genes (*KIR2DL1*, *KIR2DS1*, *KIR2DS3*, and *KIR2DS5*) does the cryptic exon have an intact 3' ss region (MES: 8.80, HSF: 88.12, PWM: 91), whereas the other *KIR* genes are missing a 3' ss at this position. Three nucleotides upstream, however, another 3' ss could be identified in all lineage III *KIR* genes as well as in *KIR2DL5* (MES: 4.07-4.87, HSF: 71.38-77.47, PWM: 81-84), which could result in the inclusion of 54 bp subsequent to exon 5, as was observed in transcripts of *KIR2DS1* and *KIR2DS4* (**Table 1**). Since the predicted splicing strength score is lower in the second cryptic 3' ss, a cryptic exon of 57 bp might be more prevalent then a 54 bp inclusion for genes that have both cryptic 3' ss, but quantitative techniques are required to confirm this. The predicted splicing strength of the 5' ss is similar in all genes (MES: 6.04, HSF: 89.67, PWM: 84), except for *KIR3DL2*, in which a 5' ss was not identified. Although cryptic exon inclusion events (54 or 57 bp) were only observed in four *KIR* genes in the human family studied (**Table 1**), these observations suggest that the cryptic exon of 57 bp can be expected in transcripts of *KIR2DL1*, *KIR2DS1*, *KIR2DS3*, and *KIR2DS5*, whereas an identical cryptic exon of 54 bp might be observed in transcripts of all lineage III *KIR* genes and *KIR2DL5*.

Overview of alternative splicing in rhesus macaque KIR transcripts

In addition to the splicing profiles of human *KIR*, we also analyzed alternative splicing of *KIR* transcripts in rhesus macaques. From a preceding family-based study, macaque KIR transcriptome profiles were obtained, which consisted of 100% matched *Mamu-KIR* sequences (20-45%), partial sequences, and sequences that contained a single nucleotide gap (9). In-depth analysis demonstrated that approximately 24% and 13% of the 100%-matched *Mamu-KIR3DL/S* and -*KIR2DL04* reads (error-free reads) accounted for alternatively spliced transcripts, respectively. In total, 48 different alternative splice events were identified (≥ 3 PacBio reads), of which 29 were confirmed

in two or more rhesus macaques (**Table 2**, **Figure 2B**, and suppl. **Table 2**). To verify whether we had obtained a complete overview of the alternative KIR splicing profiles, the PacBio read coverage of some previously typed macaque transcriptomes was increased by pooling samples of three instead of twelve rhesus macaques in a single PacBio Sequel sequencing run. This resulted in an average of 40.000 PacBio reads per rhesus macaque, which is approximately four times the number of reads we obtained per macaque from the previous study. Three additional splice events were identified (≥ 10 PacBio reads, or confirmed in two macaques; **Table 3**), and although a few splice events may have been missed, this indicated that the coverage of the formerly obtained KIR transcriptomes is sufficient to provide a fairly complete overview of the alternative splicing profiles. Furthermore, three additional families, which in total comprised 25 rhesus macaques, were sequenced for their KIR transcriptome, in accordance with the previously described protocol (9). The alternative splicing profiles of the KIR transcripts in these families revealed only one novel splice event (deletion of 112 bp in exons 4 and 5), and confirmed twenty-four splice events that were already present in the alternative KIR splicing profiles of the formerly studied family. Moreover, three events that were previously identified in a single macaque (**Table 3**, events in italic) could be confirmed by analyses of the three additional families (≥ 3 PacBio reads, in two or more macaques). This illustrated that most, but not all, KIR splice events are shared between macaque families.

Common alternative splicing events in rhesus macaque KIR

As in human KIR, all independently confirmed splice events observed in rhesus macaque *KIR* could be categorized into common alternative splicing mechanisms (**Figure 1**), which are listed in **Table 2** and schematically illustrated in **Figure 2B**. Splice events that involved *Mamu-KIR1D* are not included in this table and schematic figure, and will be discussed separately in the following section. The deletion of the first 36 bp of exon 4 was the most frequently observed alternative splice event, and is mediated by a conserved alternative 3' ss that is present in most *Mamu-KIR1D/3D* alleles (**Table 2**). The predicted splicing strength score of the alternative 3' ss (MES: 8.24-10.27, HSF: 90.11-90.53, PWM: 88-92) was higher (HSF and PWM), or similar (MES), compared to the predicted score of the actual 3' ss (MES: 8.25-10.20, HSF: 86.61-88.52, PWM: 84-87). Transcripts with this in-frame deletion lacked 12 amino acids at the start of the D1 domain, including three positively charged residues that might be involved in protein folding or ligand binding. In exon 4 of most *Mamu-KIR3D* genes, two other alternative 3' ss could be identified, resulting in transcripts with in-frame deletions of 216 bp and 267 bp (**Figure 2B** and **Table 2**). Likewise, four alternative 3' ss

Table 2. Thirty splice events identified inthirty related rhesus macaques by PacBio sequencing at the transcription level. The events can be catagorized into exon skipping (or 'cassette' exons), alternative 3' and/or 5' splice sites, cryptic exon inclusion, and intron retention. The size of deletions or inclusions are indicated in base pairs (bp). The result of the alternative splice event is predicted. ▶

Splice event	Deletion / Inclusion	Size (bp)	Position	Observed in *KIR* genes	Result
Exon skipping ('cassette' exon)	Deletion	300	Exon 4	*3DL05, 3DL20, 3DS05*	Missing D1 domain
	Deletion	294	Exon 5	*3DL01/02/05/07/10, 3DL10A/3DL08*	Missing D2 domain
	Deletion	51	Exon 6	*2DL04*	Missing stem region
	Deletion	104/105	Exon 7	*2DL04*	Missing transmembrane region; possible soluble receptor
	Deletion	53	Exon 8	*2DL04*	Missing first region cytoplasmic tail
	Deletion	594	Exons 4 + 5	*3DL07, 3DL20*	Missing D1 and D2 domain; possible KIR1D receptor
	Deletion	446	Exons 5, 6 and beginning 7	*3DL01/02/05/07/10, 3DL10A/3DL08*	Missin D2 domain, stem region, and start transmembrane region
	Deletion	155/156	Exons 6 and 7	*2DL04*	Missing stem and transmembrane region; possible soluble receptor
Alternative 3' ss	Deletion	36	Start exon 4	*3DL01/02/05/07/08/10, 3DL20, 3DS02/05/w08*	In-frame; Missing 12 AA in the beginning of the D1 domain
	Deletion	216	Start exon 4	*3DL02/08, 3DS05/w08, 3DL02/3DL08A*	In-frame; Missing 72 AA in the beginning of the D1 domain
	Deletion	267	Start exon 4	*3DL05, 3DS05*	In-frame; Missing 89 AA in the beginning of the D1 domain
	Deletion	27 bp	Start exon 5	*2DL04*	In-frame; Missing 9 AA in the beginning of the D2 domain
	Deletion	115	Start exon 5	*3DL07/08/10, 3DL20, 3DS03, 3DL02/3DL08A, 3DL10A/3DL02*	Out-frame; Stopcodon introduced
	Deletion	150	Start exon 5	*3DL02/w03/10*	In-frame; Missing 50 AA in the beginning of the D2 domain
	Deletion	176	Start exon 5	*3DL07, 3DL10A/3DL02*	Out-frame; Stopcodon introduced
	Inclusion	109	Following exon 7	*2DL04*	Out-frame; Stopcodon introduced
	Inclusion	147	Following exon 7	*2DL04*	Stopcodon introduced
	Inclusion	245/246	Following exon 7	*2DL04, 3DL07*	Stopcodon introduced
Alternative 5' ss	Deletion	198	End exon 3	*3DL02, 3DS05, 3DL02/3DL08A*	In-frame; Missing 66 AA in the end of the D0 domain
	Inclusion	109	Following exon 3	*2DL04*	Stopcodon introduced
Alternative 3' and 5' ss	Deletion	81	Within exon 4	*3DS02*	In-frame; Missing 27 AA in center of the D1 domain
	Deletion	107	Within exon 4	*3DS02*	Out-frame; Stopcodon introduced
	Deletion	141	Within exon 4	*3DL08, 3DSw08, 3DL02/3DL08A*	In-frame; Missing 47 AA in center of the D1 domain
	Deletion	324	Parts exons 4 and 5	*3DL20*	In frame; Missing parts D1 and D2 domains
	Deletion	414	Parts exons 3 and 4	*3DS05*	In frame; Missing parts D0 and D1 domains
Exon skipping + Alt. 3' ss	Deletion	415	Exon 4 and start exon 5	*3DL20*	Out-frame; Missing the D1 domain and stopcodon introduced
Cryptic exon	Inclusion	88	Following exon 5	*3DL01*	Out-frame; Stopcodon introduced
	Inclusion	147	Following exon 5	*2DL04*	Stopcodon introduced
	Inclusion	47	Following exon 6	*3DL07*	Out-frame; Stopcodon introduced
Intron retention	Inclusion	98/99	Following exon 8	*2DL04*	"98 bp" alleles: out-frame; Stopcodon introduced "99 bp" alleles: in-frame; 33 additional AA's subsequent to exon 8

could be identified in exon 5 of most *Mamu-KIR3D* genes, of which the one that mediated an out-frame deletion of 115 bp was most frequently observed. Other common alternative splice events observed in rhesus macaque KIR involved the skipping of exon 5 and the deletion of 446 bp (exons 5, 6, and 7). *Mamu-KIR1D*, *-KIR3DL20*, and *-KIR2DL04* displayed a remarkable alternative splicing profile, and will be discussed in more detail in the next sections.

Extensive alternative splicing in Mamu-KIR1D

Mamu-KIR1D, which is the only lineage III *KIR* gene in rhesus macaques*,* was identified in approximately 25-30% of the defined *Mamu*-KIR haplotypes, and, so far, only three different alleles are reported, suggesting a high level of conservation at the exon level (9). Hershberger and colleagues described nine different splice variants using Sanger sequencing (43). In our KIR transcriptome profiles obtained by PacBio sequencing, we identified a complex array of nineteen different alternatively spliced *Mamu-KIR1D* transcripts that originated from a single allele (*Mamu-KIR1D*002*) (**Figure 7A**). Up to fifteen different *Mamu-KIR1D* isoforms could be identified in a single individual (≥ 3 PacBio reads), which illustrates extensive alternative splicing. These splice variants could be explained by exon skipping, alternative 3'- and 5' ss, and cryptic exons. At the genomic DNA (gDNA) level, three domain-encoding exons could be identified, but only exon 4 (D1 domain) was present in all transcribed *Mamu-KIR1D* isoforms. On the basis of gDNA analysis, it was revealed that exon 3 of *Mamu-KIR1D* contained a 5 bp deletion (71), and was constitutively skipped, similar to what is observed in human lineage III *KIR* genes. However, an intact BPS, PPT, 3' ss (MES: 6.62, HSF: 86.67, PWM: 85), and 5' ss (MES: 7.41, HSF: 92.64, PWM: 86) could be identified for exon 3 of *Mamu-KIR1D*, which may suggest that another mechanism plays a role in the constitutive skipping of this exon. An explanation might be the absence of 33 bp in intron 2, which characterizes all lineage III *KIR* genes in both humans and macaques (**Figure 7B**). This intron part is a purine-rich element that might be essential for spliceosome binding, and leads in its absence to the exclusion of exon 3 at the transcription level.

In most of the identified *Mamu-KIR1D* isoforms, exon 5 was present or partially included (**Figure 7A**, #1-13). Due to a 7 bp deletion in exon 5 at the gDNA level, complete inclusion of this exon at the mRNA level by constitutive splicing resulted in a frameshift that introduced a stopcodon in the beginning of exon 7 (**Figure 7A**; #1, 2). The remaining transcripts that included exon 5 either skipped exon 7, but had in-frame exons 8 and 9 (**Figure 7A**; #3), or had intronic/exonic inclusions subsequent to exon 6 that introduced a stopcodon (**Figure 7A**; #4-6). These transcripts probably encode soluble and truncated KIR1D receptors, respectively. In other *Mamu-KIR1D* isoforms, the first part of exon 5 was skipped, which resulted in in-frame transcripts that encoded the D1 domain, the second part of the D2 domain, and an intact cytoplasmic tail (**Figure 7A**; #7-9), or out-frame transcripts that had a stopcodon subsequent to exon 5 (**Figure 7A**; #10-13). Transcripts that completely lacked exon 5 encoded membrane-bound isoforms with the D1 domain and an ITIM-containing cytoplasmic tail (**Figure 7A**, #14, 15), or isoforms

Splice event		Deletion / Inclusion	Size (bp)	Position	Observed in *KIR* genes	Result
High coverage	Alt. 3' ss + Alt. 5' ss	Deletion	48	In exon 5	*3DL05, 3DL08, 3DS02*	In-frame; Missing 16 AA in the beginning of the D2 domain
	Alt. 5' ss	Deletion	209	End exon 4	*3DL20*	Out-frame; Stopcodon introduced
	Exon skipping	Deletion	345	Exons 5 + 6	*3DL20*	Missing the D2 domain and stem region
Additional families	Alt. 3' ss	Deletion	79	Start exon 5	*3DL10A/3DL08, 3DL11*	Out-frame; Stopcodon introduced. 3DL11*009: ORF restored in cytoplasmic tail
	Alt. 3' ss + Alt. 5' ss	Deletion	97	In exon 4	*3DL02, 3DSw08, 3DSw09*	Out-frame; Stopcodon introduced
	Alt. 3' ss + Alt. 5' ss	Deletion	112	Parts exons 4 + 5	*3DL07, 3DL11*	Out-frame; Stopcodon introduced
	Exon skipping	Deletion	645	Exons 4 + 5+ 6	*3DL05, 3DL20*	Missing D1 and D2 domains, and stem region

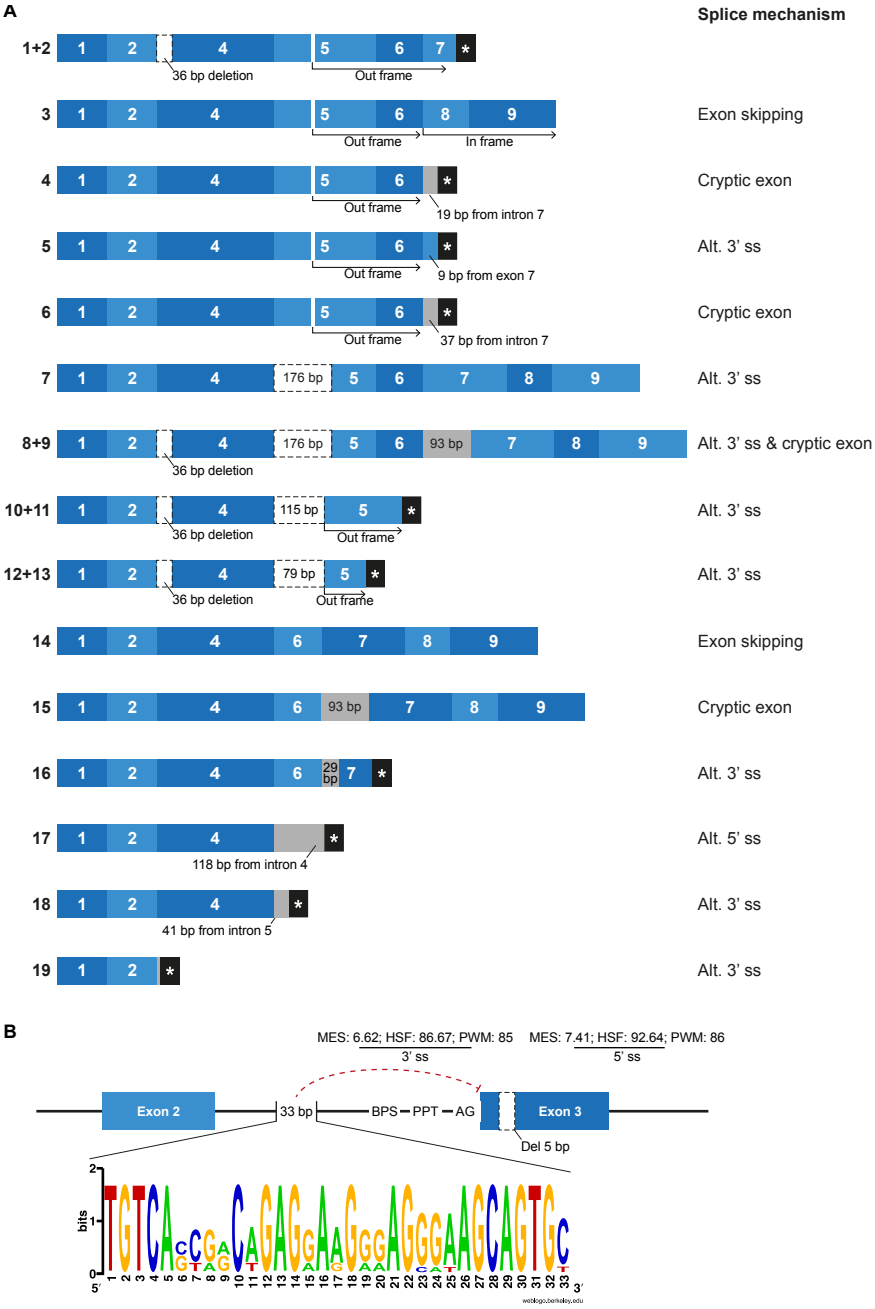
Table 3. To verify whether a complete overview of Mamu-KIR alternative splicing was obtained, four samples were pooled for a PacBio Sequel sequencing run, which provided a fourfold coverage compared to the obtained transcriptome profiles. Three additional splice events were confirmed (≥ 10 PacBio reads, or confirmed in two macaques), indicating that a rather complete overview of alternative splicing can be provided by data from the preceding study (9). Also, the alternative splicing profiles of three additional families, consisting of 25 rhesus macaques in total, were characterized. Only a single splice event was identified that was not observed in the main dataset, whereas 24 splice events were confirmed, indicating that most splice events are shared between rhesus macaque families.

that encoded only the D1 domain (**Figure 7A**, #16-18). The deletion of 36 bp at the beginning of exon 4, as is observed in most lineage II *Mamu-KIR* genes (**Table 2**), was also identified in multiple Mamu-KIR1D transcripts; these isoforms could appear with and without this splice event (**Figure 7A**, #1/2, 8/9, 10/11, 12/13).

In other Old World monkey species such as cynomolgus macaques (*Macaca fascicularis*), olive baboons (*Papio Anubis*), and vervet monkeys (*Chlorocebus aethiops*), orthologs of *Mamu-KIR1D* were observed that also displayed the 5 bp and 7 bp deletions in exons 3 and 5, respectively (36, 71). In addition, a comparison of intron 2 of these genes revealed that they also lack the purine-rich element of 33 bp, which might explain the constitutive skipping of exon 3. It is not known whether these orthologs are also subjected to extensive alternative splicing. In humans, no orthologs of *Mamu-KIR1D* were identified. However, multiple human *KIR2DS4* alleles that skip exon 3 – and that have a 22 bp deletion at the gDNA level in exon 5, which introduces a frameshift that resulted in an early stopcodon subsequent to exon 6 – have been described as *Mamu-KIR1D* analogs (72).

Consistent alternative splicing of Mamu-KIR3DL20

The *Mamu-KIR3DL20* gene is present on most, but not all, reported *Mamu*-KIR haplotypes, and has been considered a framework gene. Phylogenetic analysis has illustrated a relationship


A
Splice mechanism
1+2
36 bp deletion
Out frame
3
Out frame
In frame
Exon skipping
4
Out frame
19 bp from intron 7
Cryptic exon
5
Out frame
9 bp from exon 7
Alt. 3' ss
6
Out frame
37 bp from intron 7
Cryptic exon
7
176 bp
Alt. 3' ss
8+9
176 bp
93 bp
36 bp deletion
Alt. 3' ss & cryptic exon
10+11
115 bp
36 bp deletion
Out frame
Alt. 3' ss
12+13
79 bp
36 bp deletion
Out frame
Alt. 3' ss
14
Exon skipping
15
93 bp
Cryptic exon
16
29 bp
Alt. 3' ss
17
118 bp from intron 4
Alt. 5' ss
18
41 bp from intron 5
Alt. 3' ss
19
Alt. 3' ss
B
MES: 6.62; HSF: 86.67; PWM: 85
3' ss
MES: 7.41; HSF: 92.64; PWM: 86
5' ss
Exon 2
33 bp
BPS
PPT
AG
Exon 3
Del 5 bp
bits

between *Mamu-KIR3DL20* and human linage I (*KIR2DL4*, *KIR2DL5*) and V (*KIR3DL3*) *KIR* genes (45, 71). Indeed, exon 3 of *Mamu-KIR3DL20* showed similarity to human *KIR2DL5*, and exons 4 and 5 displayed similarity to human *KIR3DL3*. The exons encoding the cytoplasmic tail of *Mamu-KIR3DL20* are more related to macaque *KIR* genes. Multiple studies have suggested that the frequently identified *Mamu-KIR2DL05* transcript is a splice variant of the *Mamu-KIR3DL20* gene, in which exon 4 (D1 domain) is spliced out (33, 43, 44, 71, 73). Our results substantiates that *Mamu-KIR2DL05* is indeed a splice variant of *Mamu-KIR3DL20*, and that this splice event is consistent for every identified *Mamu-KIR3DL20* allele in the rhesus macaques studied (**Figure 8**). In addition, gel electrophoresis indicates that the amount of exon 4 skipping in Mamu-KIR3DL20 transcripts is considerable (additional file **figure S1** in Supplementary Material). The 3' ss region of exon 4 in *Mamu-KIR3DL20* is intact (MES: 8.25, HSF: 86.96, PWM: 85), although the predicted splicing strength is slightly lower compared to the 3' ss of most other macaque *KIR* genes (MES: 9.84, HSF: 88.52, PWM: 85). Due to a single substitution of a cytosine with a thymine, the 5' ss of exon 4 in *Mamu-KIR3DL20* alleles (MES: 6.95, HSF: 94.52, PWM: 88) also scored lower compared to the remaining *KIR* genes (MES: 9.22, HSF: 96.51, PWM: 92). Therefore, these suboptimal splice sites might contribute to the skipping of exon 4 in *Mamu-KIR3DL20*, resulting in *Mamu-KIR2DL05* transcripts. Moreover, the skipping of exon 4 might be mediated by intron 3 of *Mamu-KIR3DL20*, which is 450-650 bp shorter compared to intron 3 of other *Mamu-KIR* genes. This may result in modified or missing splicing elements in intron 3, thereby influencing the spliceosome efficiency. Of note is that in human *KIR3DL3*, which has an exon 4 similar to *Mamu-KIR3DL20*, intron 3 is also shorter, but the consistent skipping of exon 4 is not reported for this human gene. In contrast to exon 4 in all reported human and macaque *KIR* genes, exon 4 of *Mamu-KIR3DL20* is completely conserved in all 22 reported rhesus macaque alleles. This observation suggests selective pressure, and indicates an important function of exon 4 (D1 domain)

◀ **Figure 7. Overview of alternative splicing in *Mamu-KIR1D*.** (A) Nineteen different *Mamu-KIR1D* transcripts were observed; each transcript is illustrated by blue boxes per exon. White boxes with a dashed outline indicate partial exon deletions, and intronic inclusions are indicated in grey boxes. Multiple transcripts have an out-frame region, due to a deletion of 7 bp in exon 5 at the gDNA level, which can introduce a stopcodon (#1, 2, 4-6). However, the complete inclusion of exon 5 combined with the skipping of exon 7 can restore the open reading frame (#3). The partial deletion of exon 5 might result in Mamu-KIR1D molecules containing an intact inhibitory cytoplasmic tail (#7-9), or in molecules that only consist of a single D1 domain (#10-14). Transcripts that completely skipped exon 5 might encode membrane-bound molecules including an inhibitory cytoplasmic tail (#14, 15), or molecules that only encode the D1 domain, with or without the stem region (#16-18). A second variant was observed for some transcripts, which involved the deletion of 36 bp at the beginning of exon 4, which was also observed in lineage II KIR genes (#1/2, 8/9, 10/11, 12/13; Table II). (B) Exon 3 is present at the gDNA level in *Mamu-KIR1D*, but none of the KIR1D isoforms contain the D0 domain encoded by this exon. The BPS, PPT, and both 3' and 5' splice sites of exon 3 are intact, just as in human lineage III *KIR* genes, and predicted splicing strength scores are provided. Intron 2 of all lineage III *KIR* genes, including *Mamu-KIR1D*, lack 33 bp in intron 2 compared to all other *KIR* genes. The lack of this 33 bp stretch might inhibit constitutive splicing of exon 3, as indicated by the red dashed line. The weblogo plot shows the nucleotide sequence composition of this 33 bp stretch that is present in all *KIR* genes except the KIR lineage III genes, which might mediate the constitutive splicing of exon 3.

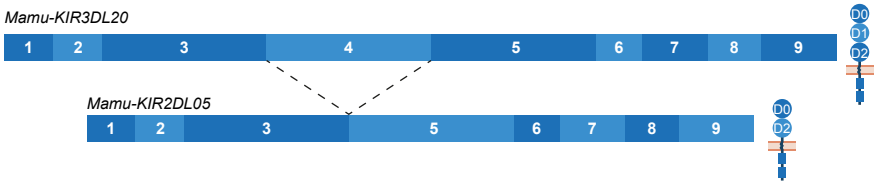


Figure 8. The skipping of exon 4 in *Mamu-KIR3DL20* transcripts to generate *Mamu-KIR2DL05* transcripts. Transcripts are illustrated with the exons indicated in blue boxes, and corresponding protein structures are schematically depicted adjacent to the transcript. The splice event is indicated with dashed lines. Exon 4 (D1 domain) is consistently skipped for all studied *Mamu-KIR3DL20* alleles to generate *Mamu-KIR2DL05* transcripts.

in the recognition of Mamu-KIR3DL20 receptors, or in the formation of Mamu-KIR2DL05 splice variants via splicing enhancer or silencer motifs.

The consensus sequence of the *Mamu-KIR2DL05* transcripts, generated by the alternative splicing of the *Mamu-KIR3DL20* gene, showed 89.5% similarity to the consensus sequence of constitutively spliced human *KIR2DL5* transcripts, and suggests a convergent evolution of this gene in humans and macaques. Although the exact mechanism and function of the consistent skipping of exon 4 in *Mamu-KIR3DL20* resulting in *Mamu-KIR2DL05* transcripts is not completely understood, it does illustrate that alternative splicing in macaques can introduce a second *KIR2DL* transcript additional to *Mamu-KIR2DL04*. As well as the skipping of exon 4, *Mamu-KIR3DL20* transcripts that lacked exons 4 and 5 (594 bp deletion; **Table 2**) were also frequently observed. These transcripts were not consistently observed in all macaques, however, and seem to encode inhibitory receptors with a single extracellular domain (D0).

Alternative splicing in Mamu-KIR2DL04 is mainly gene-specific

Whereas human *KIR2DL4* is a framework gene, the macaque ortholog *Mamu-KIR2DL04* is identified on approximately 65-75% of the reported *Mamu*-KIR haplotypes (9, 33). As with *KIR2DL4* in human, the most diverse splicing profile in macaques was observed for *Mamu-KIR2DL04*, including ten confirmed splice events, of which nine were gene-specific (**Table 2**). Exon skipping events were observed in exons 6 to 8. In *Mamu-KIR2DL04* transcripts, the skipping of exon 7 is only observed in combination with the skipping of exon 6 (155 bp in total) (**Figure 2B**). The skipping of exon 8, which encodes a part of the cytoplasmic tail, was only observed in *Mamu-KIR2DL04*015*. In two macaques that expressed this allele, no complete transcripts were identified, indicating allele-specific consistent exon skipping. Other events involved alternative splice sites, of which three were located in intron 7, which resulted in partial intron retentions subsequent to exon 7 (**Figure 9** and **Table 2**). These intronic insertions introduced a stopcodon, and the three corresponding transcripts probably encode a membrane-bound receptor that lacks a cytoplasmic tail. Similar alternatively spliced transcripts were observed for human "10A" *KIR2DL4* alleles with an insertion of 67 bp

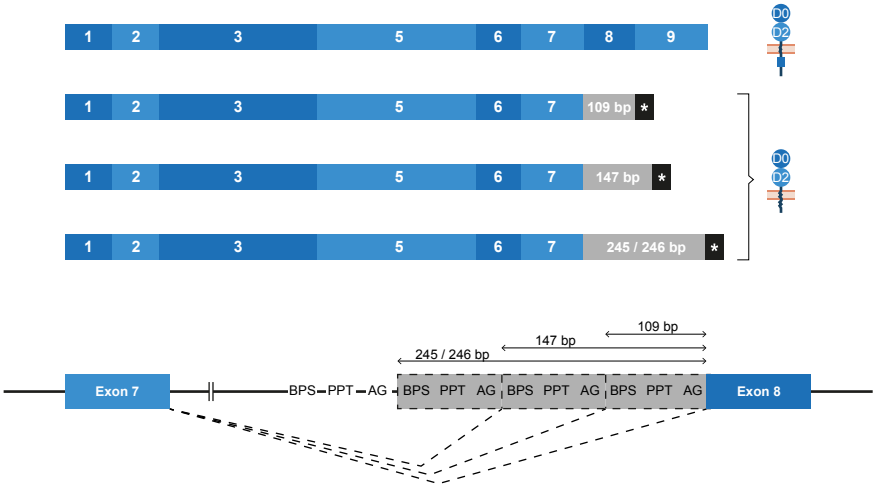


Figure 9. Alternative splice sites in intron 7 of *Mamu-KIR2DL04*. Transcripts are illustrated with the exons indicated in blue boxes, and corresponding protein structures are schematically depicted adjacent to the transcript. Intronic inclusions are indicated in grey boxes, and stopcodons are marked by a black box with asterisk (*). Splice events are indicated with dashed lines. Three alternative 3' splice sites were identified in intron 7 of *Mamu-KIR2DL04*, and result in intron inclusions of 109, 147, and 245/246 bp subsequent to exon 7. All these inclusions result in a frameshift that introduces a stopcodon, indicating membrane-bound KIR2D receptors without an inhibitory cytoplasmic tail.

subsequent to exon 7 (**Figure 5B**). However, whereas in human *KIR2DL4* the splice event is mediated by an alternative 5' ss, resulting in an inclusion of the first part of intron 7, in macaques the intron inclusions originate from the end of intron 7, and are mediated by alternative 3' splice sites. Notably, three out of the four alternative splice sites in intron 7 could be identified in the human and macaque *KIR2DL4* orthologs, but no similar alternatively spliced transcripts were shared between the species.

Discussion

The plasticity of the *KIR* genes is manifested by allelic polymorphism, copy number variation, the expansion and contraction of haplotypes, variegated expression, and the generation of hybrid genes by recombination. Here we demonstrate that alternative splicing adds an additional layer of complexity by the generation of isoforms originating from a single *KIR* gene. This phenomenon appears to be a structural aspect of the *KIR* gene cluster in different primate species. In total, 18 human and 55 macaque splice events were documented (**Figures 2** and **7A**, **Tables 1-3**), including gene-specific events, and events that were observed across different *KIR* lineages. The potential mechanisms that mediate these splice events were categorized into common types of

alternative splicing, and the responsible motifs were predicted and scored with different software tools. Overall, the current overview of the *KIR* splicing profiles in humans and rhesus macaques indicates that the generation of different KIR isoforms might be of functional relevance in health and disease.

Alternative splicing can diversify the characteristics of the encoded protein, including the domains it contains, its ligand interactions, cellular localization, and signaling properties. As a result, different isoforms encoded by the same gene can execute distinct functionalities. The receptor structure, and to a lesser extent the functional characteristics, of some KIR isoforms could be predicted based on the alternatively spliced KIR transcripts. The skipping of one or multiple exons that encode the extracellular domains (exons 3-5) likely results in the formation of KIR1D or KIR2D receptors, which can have distinct binding properties compared to the constitutively spliced isoform. A frequently observed splice event in human *KIR* is the skipping of exon 6, which encodes the stem region. The function of this region is not yet clear, but it might be involved in the flexibility and orientation of the receptor. Crystallography and *in vitro* lysis assays illustrated, for instance, that the stem region is not involved in ligand binding, but may contribute to the inhibitory signaling function (74, 75). Therefore, isoforms that lack the stem region might be less stringent in delivering inhibitory signals. Transcripts that completely or partially lack exon 7 probably encode soluble receptors, whereas the absence of exons 8 and/or 9 indicates a loss of inhibitory signaling function. The consequences of events facilitated by splice mechanisms other then exon skipping is harder to predict. For example, the in-frame deletion of 36 bp in exon 4, which is observed in thirteen different *Mamu-KIR* genes and mediated by an alternative 3' ss, might result in a different D1 domain orientation, in distinct ligand interactions, or in an aberrant folding of the complete receptor. The conservation of this alternative 3' splice site may indicate a selective pressure on a functional characteristic of the KIR isoform generated by this splice event. *In vitro* binding and inhibition assays with KIR isoform-transfected cells could elucidate the function of these less predictable splice variants. However, it should be noted that a large proportion of the splice events may not result in functional receptors and these product may be subjected to the nonsense-mediated decay pathway, as was previously reported in human proteomic studies (76-78). This nonproductive splicing is likely a redundancy relating to the rapid evolution of splice sites, from which beneficial isoforms are positively selected, although it has also been suggested that nonfunctional splicing is a mechanism to downregulate expression of the protein encoded by constitutive splicing (79, 80). Nonetheless, the number of splice events that we identified (**Figures 2** and **7A**, and **Tables 1-3**), together with the observed segregation of splice events and the sharing of splice mechanisms resulting in similar consequences on protein level in humans and rhesus macaques, suggests that at least a part of the alternative splicing profiles contributes to a structural and functional variety of KIR receptors.

All common alternative splicing mechanisms were observed in human and rhesus macaque *KIR* genes, except for intron retention, which was only observed in a single splice event in *Mamu-KIR2DL04*. In both species, similar exon skipping events were observed (**Figure 2** and **Tables 1-3**), although in humans more events involved the skipping of multiple exons, especially exon 7 and its flanking exons. In addition to most exon skipping events, only the deletion of 198 bp in the end of exon 3, which is mediated by an alternative 5' ss, was shared between humans and rhesus macaques. In macaques, this deletion was observed in three different genes, whereas in humans this splice event was specific for *KIR2DL4*, and was only observed in combination with a second deletion in the transmembrane region. These isoforms probably have an aberrant D0 domain, which might result in modified binding properties. All other splice events were only identified in one of the two species. In contrast to human KIR splice events, most splice events in macaque KIR involved the domain-encoding exons. This could be related to the expansion of lineage II *KIR* genes in macaques, which contain three extracellular domains, and therefore might have more flexibility to modify the domains without compromising ligand binding.

The skipping of exon 4 (D1 domain) was consistently observed in transcripts of *Mamu-KIR3DL20* (lineage V), which is considered a framework gene in rhesus macaques, and resulted in Mamu-KIR2D transcripts that only encoded the D0 and D2 domains. These alternatively spliced transcripts seem to be a functional analog of human *KIR2DL5* (lineage I) and they share a similarity of 89.5%, suggesting a convergent evolution of this structure. In macaques, the *Mamu-KIR2DL05* transcripts were identified in all individuals, whereas in humans, *KIR2DL5* is only present on specific haplotypes (group B haplotypes). This is an indication that the alternatively spliced *Mamu-KIR2DL05* transcripts, or the *Mamu-KIR3DL20* gene itself, are essential to rhesus macaques. The complete conservation of exon 4 in all *Mamu-KIR3DL20* alleles further supports this. Exon 4 in *Mamu-KIR3DL20* seems to be essential in facilitating its own consistent skipping, or in the interaction of Mamu-KIR3DL20 molecules, and might therefore be conserved by selective pressure. This conserved character is not observed for any other exon in human or macaque *KIR*. The skipping of exon 4 in *Mamu-KIR3DL20* transcripts illustrates how alternative splicing can expand the plasticity of the KIR repertoire by generating isoforms of two different *KIR* lineages from a single gene.

The skipping of exon 5, and exons 4 and 5 together, was observed in *KIR* genes of humans and macaques, and the events might have similar consequences. The skipping of exon 3 was not observed in human and macaque lineage I and II *KIR* genes, which implies essential properties for the D0 domain. It has been reported that the D0 domain is involved in the direct binding of MHC class I molecules (81), whereas others described only a modulatory role for this domain in KIR3D receptors (82, 83). Furthermore, the cell surface level of human KIR3DL1 could be modified by D0 polymorphisms, such as the substitution of a valine with a leucine at position 18 (V18L), which prevents the surface expression of KIR3DL1*053 (84). The characteristics of the D0 domain might be essential for KIR3D function in both species, and therefore the alternative splicing of this domain might be subjected to negative selection.

A large number of alternative splicing events were observed for the *Mamu-KIR1D* gene (**Figure 7A**), which is the only lineage III *KIR* gene in macaques and is highly conserved, as only three alleles have been documented in the apparent functional sections of the gene (exons 1, 2 and 4). Despite an intact BPS, PPT, and 3' and 5' splice sites, exon 3 is constitutively skipped in this gene. An explanation for this phenomenon can be found in the absence of a purine-rich stretch of 33 bp in intron 2 of *Mamu-KIR1D* (**Figure 7B**), which is also lacking in human lineage III *KIR* genes. Human and macaque *KIR* genes, which do include exon 3 in their transcripts, have an intron 2 that contains the purine-rich 33 bp. Within these genes, the 33 bp appear highly conserved, which might indicate its essential role for spliceosome recognition. Furthermore, the *Mamu-KIR1D* gene shows extensive alternative splicing subsequent to exon 4, and this might be due to the introns flanking exons 5 and 6. In rhesus macaques, introns 5 and 6 of lineage II *KIR* genes are approximately 2000 bp and 914 bp in length, respectively. Introns 5 and 6 of *Mamu-KIR1D* (lineage III) are approximately 3290 bp and 4330 bp, respectively, and similar intron lengths are observed in human lineage III *KIR* genes. In humans, lineage III *KIR* genes mainly interact with HLA-C molecules. In rhesus macaques, however, no homolog of HLA-C is identified, and most *Mamu*-KIR probably interacts with members of the expanded repertoire of MHC-A and -B molecules. Therefore, selective pressure to conserve the lineage III gene might be low, which makes the large introns observed for Mamu-KIR1D prone to mutations that might induce alternative splicing. As such, an initial macaque KIR2D gene is now translated into Mamu-KIR1D. At present it is unclear whether any of the different identified Mamu-KIR1D isoforms are functional in macaques.

The extent of the impact that alternative splicing has on the KIR repertoire is dependent not only on which splice variants are formed but also on the frequency of the splice events. In this study, the PacBio Sequel platform was used to determine the alternative splicing profiles, but this method does not provide quantification, and may only be used as a quantitative indication. Approximately 4% and 53% of the 100%-matched human KIR2D/3D and KIR2DL4 reads accounted for alternatively spliced transcripts. In rhesus macaques, 24% and 13% splice variants were obtained from the 100%-matched KIR2D/3D and KIR2DL04 PacBio reads, respectively. These percentages indicate abundant alternative splicing for human *KIR2DL4* and macaque *KIR2D/3D*, but one has to be cautious with the interpretation of these numbers, as the quantification value of the PacBio platform is low. Furthermore, preferential amplification of the used primer sets can not be ruled out, which may also have an effect on the calculated numbers. More reliable quantification methods, like droplet digital PCR (ddPCR) or RNA-seq, are hard to adapt on a multigene family such as KIR. Therefore, we are currently only able to provide quantitative indications of alternative KIR splicing.

In this study, and in other studies that reported KIR isoforms (28-30, 36, 42, 43), the splice variants were identified in whole blood samples, and might therefore give a representation of the complete splicing profile. However, tissue-specific alternative splicing has been reported, and suggests isoforms with a local specialized function

(85). Especially in the brain, testis, and liver, increased alternative splicing events can be identified, which mainly involved exon skipping and alternative splice sites. NK cells that reside in tissues are reported in multiple organs, such as the intestines, lungs, liver, spleen, lymph nodes, brain, eye retina, and uterus, and can be phenotypically and functionally distinct from the NK cells in peripheral blood (86-93). The diversity of NK-cell subsets in the different tissues includes selective expression of KIR, but might also involve distinct alternative splicing profiles of the *KIR* genes. The regulation of tissue-specific alternative splicing is complex, and involves the differential expression of splicing factors and epigenetic modifications, such as methylation and histone acetylation (94-96). In tissues that should dampen the immune response to avoid inflammation, like the eye retina, or tissues that require high immune surveillance, like the intestines and liver, alternative splicing might provide the required isoforms. Also in the uterine tissue, phenotypically and functionally distinct NK cells (uNK cells) have been identified that mainly express *KIR2DL4*, and are involved in pregnancy. This NK cell subset can interact with the highly expressed HLA-G molecules in the uterus, which are subjected to alternative splicing, to maintain the fetal-maternal interface and induce cytokine production (97-99). Alternative splicing might modify the activity and interactions of KIR2DL4 expressed on the uNK cells, for example, by skipping the transmembrane region to generate soluble KIR2DL4 receptors (**Figure 4**), or by an insertion of 67 bp that can regulate the presence of an inhibitory cytoplasmic tail (**Figure 5B**).

Overall, we characterized the alternative splicing profiles of *KIR* genes in human and macaque families, which provides an illustration of the potential formation of protein isoforms. These posttranscriptional modifications might contribute to the complexity of the *KIR* gene family of both species, human and macaque, and result in a wide structural and functional variety of receptors that might be involved in health and disease.

Acknowledgements

We thank D. Devine for editing the manuscript and F. van Hassel for preparing the figures. Human PBMC samples from a preceding study (9) were kindly provided by Prof. dr. Frans H. J. Claas (Immunohematology and Blood Transfusion, Leiden University Medical Center, Leiden, The Netherlands).

References

1. Giebel, S., Locatelli, F., Lamparelli, T., Velardi, A., Davies, S., Frumento, G., et al. (2003). Survival advantage with KIR ligand incompatibility in hematopoietic stem cell transplantation from unrelated donors. *Blood* 102(3)**,** 814-819. doi: 10.1182/blood-2003-01-0091.

2. Hiby, S.E., Walker, J.J., O'Shaughnessy K, M., Redman, C.W., Carrington, M., Trowsdale, J., et al. (2004). Combinations of maternal KIR and fetal HLA-C genes influence the risk of preeclampsia and reproductive success. *J Exp Med* 200(8)**,** 957-965. doi: 10.1084/jem.20041214.

3. Kulkarni, S., Martin, M.P., and Carrington, M. (2008). The Ying and Yang of HLA and KIR in Human Disease. *Semin Immunol* 20(6)**,** 343-352. doi: 10.1016/j.smim.2008.06.003.

4. Rajagopalan, S., and Long, E.O. (2005). Understanding how combinations of HLA and KIR genes influence disease. *J Exp Med* 201(7)**,** 1025-1029. doi: 10.1084/jem.20050499.

5. Robinson, J., Mistry, K., McWilliam, H., Lopez, R., and Marsh, S.G.E. (2010). IPD—the Immuno Polymorphism Database. *Nucleic Acids Res* 38(Database issue)**,** D863-869. doi: 10.1093/nar/gkp879.

6. Uhrberg, M., Valiante, N.M., Shum, B.P., Shilling, H.G., Lienert-Weidenbach, K., Corliss, B., et al. (1997). Human diversity in killer cell inhibitory receptor genes. *Immunity* 7(6)**,** 753-763.

7. Held, W., and Kunz, B. (1998). An allele-specific, stochastic gene expression process controls the expression of multiple Ly49 family genes and generates a diverse, MHC-specific NK cell receptor repertoire. *Eur J Immunol* 28(8)**,** 2407-2416. doi: 10.1002/(sici)1521-4141(199808)28:08<2407::aid-immu2407>3.0.co;2-d.

8. Valiante, N.M., Uhrberg, M., Shilling, H.G., Lienert-Weidenbach, K., Arnett, K.L., D'Andrea, A., et al. (1997). Functionally and structurally distinct NK cell receptor repertoires in the peripheral blood of two human donors. *Immunity* 7(6)**,** 739-751.

9. Bruijnesteijn, J., van der Wiel, M.K.H., Swelsen, W.T.N., Otting, N., de Vos-Rouweler, A.J.M., Elferink, D., et al. (2018). Human and Rhesus Macaque KIR Haplotypes Defined by Their Transcriptomes. *J Immunol.* doi: 10.4049/jimmunol.1701480.

10. Jiang, W., Johnson, C., Jayaraman, J., Simecek, N., Noble, J., Moffatt, M.F., et al. (2012). Copy number variation leads to considerable diversity for B but not A haplotypes of the human KIR genes encoding NK cell receptors. *Genome Res* 22(10)**,** 1845-1854. doi: 10.1101/gr.137976.112.

11. Traherne, J.A., Martin, M., Ward, R., Ohashi, M., Pellett, F., Gladman, D., et al. (2010). Mechanisms of copy number variation and hybrid gene formation in the KIR immune gene complex. *Hum Mol Genet* 19(5)**,** 737-751. doi: 10.1093/hmg/ddp538.

12. Martin, A.M., Freitas, E.M., Witt, C.S., and Christiansen, F.T. (2000). The genomic organization and evolution of the natural killer immunoglobulin-like receptor (KIR) gene cluster. *Immunogenetics* 51(4-5)**,** 268-280.

13. Vilches, C., and Parham, P. (2002). KIR: diverse, rapidly evolving receptors of innate and adaptive immunity. *Annu Rev Immunol* 20**,** 217-251. doi: 10.1146/annurev.immunol.20.092501.134942.

14. Mingari, M.C., Moretta, A., and Moretta, L. (1998). Regulation of KIR expression in human T cells: a safety mechanism that may impair protective T-cell responses. *Immunol Today* 19(4)**,** 153-157.

15. Roe, D., Vierra-Green, C., Pyo, C.W., Eng, K., Hall, R., Kuang, R., et al. (2017). Revealing complete complex KIR haplotypes phased by long-read sequencing technology. *Genes Immun* 18(3)**,** 127-134. doi: 10.1038/gene.2017.10.

16. Robinson, J., Halliwell, J.A., McWilliam, H., Lopez, R., and Marsh, S.G. (2013). IPD--the Immuno Polymorphism Database. *Nucleic Acids Res* 41(Database issue)**,** D1234-1240. doi: 10.1093/nar/gks1140.

17. Pyo, C.W., Wang, R., Vu, Q., Cereb, N., Yang, S.Y., Duh, F.M., et al. (2013). Recombinant structures expand and contract inter and intragenic diversification at the KIR locus. *BMC Genomics* 14**,** 89. doi: 10.1186/1471-2164-14-89.

18. Nilsen, T.W., and Graveley, B.R. (2010). Expansion of the eukaryotic proteome by alternative splicing. *Nature* 463(7280)**,** 457-463. doi: 10.1038/nature08909.

19. Blencowe, B.J. (2006). Alternative splicing: new insights from global analyses. *Cell* 126(1)**,** 37-47. doi: 10.1016/j.cell.2006.06.023.

20. Will, C.L., and Luhrmann, R. (2011). Spliceosome structure and function. *Cold Spring Harb Perspect Biol* 3(7). doi: 10.1101/cshperspect.a003707.

21. Pagani, F., and Baralle, F.E. (2004). Genomic variants in exons and introns: identifying the splicing spoilers. *Nat Rev Genet* 5(5)**,** 389-396. doi: 10.1038/nrg1327.

22. Roy, B., Haupt, L.M., and Griffiths, L.R. (2013). Review: Alternative Splicing (AS) of Genes As An Approach for Generating Protein Complexity. *Curr Genomics* 14(3)**,** 182-194. doi: 10.2174/1389202911314030004.

23. Tazi, J., Bakkour, N., and Stamm, S. (2009). Alternative splicing and disease. *Biochim Biophys Acta* 1792(1)**,** 14-26. doi: 10.1016/j.bbadis.2008.09.017.

24. Venables, J.P. (2006). Unbalanced alternative splicing and its significance in cancer. *Bioessays* 28(4)**,** 378-386. doi: 10.1002/bies.20390.

25. Skotheim, R.I., and Nees, M. (2007). Alternative splicing in cancer: noise, functional, or systematic? *Int J Biochem Cell Biol* 39(7-8)**,** 1432-1449. doi: 10.1016/j.biocel.2007.02.016.

26. Voorter, C.E., Gerritsen, K.E., Groeneweg, M., Wieten, L., and Tilanus, M.G. (2016). The role of gene polymorphism in HLA class I splicing. *Int J Immunogenet* 43(2)**,** 65-78. doi: 10.1111/iji.12256.

27. Romero, A., García-García, F., López-Perolio, I., Ruiz de Garibay, G., García-Sáenz, J.A., Garre, P., et al. (2015). BRCA1 Alternative splicing landscape in breast tissue samples. *BMC Cancer* 15(1)**,** 219. doi: 10.1186/s12885-015-1145-9.

28. Chwae, Y.J., Cho, S.E., Kim, S.J., and Kim, J. (1999). Diversity of the repertoire of p58 killer cell inhibitory receptors in a single individual. *Immunol Lett* 68(2-3)**,** 267-274.

29. Dohring, C., Samaridis, J., and Colonna, M. (1996). Alternatively spliced forms of human killer inhibitory receptors. *Immunogenetics* 44(3)**,** 227-230.

30. Goodridge, J.P., Lathbury, L.J., Steiner, N.K., Shulse, C.N., Pullikotil, P., Seidah, N.G., et al. (2007). Three common alleles of KIR2DL4 (CD158d) encode constitutively expressed, inducible and secreted receptors in NK cells. *Eur J Immunol* 37(1)**,** 199-211. doi: 10.1002/eji.200636316.

31. Vilches, C., Rajalingam, R., Uhrberg, M., Gardiner, C.M., Young, N.T., and Parham, P. (2000). KIR2DL5, a novel killer-cell receptor with a D0-D2 configuration of Ig-like domains. *J Immunol* 164(11)**,** 5797-5804.

32. Goodridge, J.P., Witt, C.S., Christiansen, F.T., and Warren, H.S. (2003). KIR2DL4 (CD158d) genotype influences expression and function in NK cells. *J Immunol* 171(4)**,** 1768-1774.

33. Blokhuis, J.H., van der Wiel, M.K., Doxiadis, G.G., and Bontrop, R.E. (2010). The mosaic of KIR haplotypes in rhesus macaques. *Immunogenetics* 62(5)**,** 295-306. doi: 10.1007/s00251-010-0434-3.

34. Guethlein, L.A., Norman, P.J., Heijmans, C.M., de Groot, N.G., Hilton, H.G., Babrzadeh, F., et al. (2017). Two Orangutan Species Have Evolved Different KIR Alleles and Haplotypes. *J Immunol* 198(8)**,** 3157-3169. doi: 10.4049/jimmunol.1602163.

35. Guethlein, L.A., Older Aguilar, A.M., Abi-Rached, L., and Parham, P. (2007). Evolution of killer cell Ig-like receptor (KIR) genes: definition of an orangutan KIR haplotype reveals expansion of lineage III KIR associated with the emergence of MHC-C. *J Immunol* 179(1)**,** 491-504.

36. Prall, T.M., Graham, M.E., Karl, J.A., Wiseman, R.W., Ericsen, A.J., Raveendran, M., et al. (2017). Improved full-length killer cell immunoglobulin-like receptor transcript discovery in Mauritian cynomolgus macaques. *Immunogenetics* 69(5)**,** 325-339. doi: 10.1007/s00251-017-0977-7.

37. Rajalingam, R., Hong, M., Adams, E.J., Shum, B.P., Guethlein, L.A., and Parham, P. (2001). Short KIR haplotypes in pygmy chimpanzee (Bonobo) resemble the conserved framework of diverse human KIR haplotypes. *J Exp Med* 193(1)**,** 135-146.

38. Bontrop, R.E. (2001). Non-human primates: essential partners in biomedical research. *Immunol Rev* 183**,** 5-9.

39. Kaushal, D., Mehra, S., Didier, P.J., and Lackner, A.A. (2012). The non-human primate model of tuberculosis. *J Med Primatol* 41(3)**,** 191-201. doi: 10.1111/j.1600-0684.2012.00536.x.

40. Vierboom, M.P.M., Jonker, M., Bontrop, R.E., and t Hart, B. (2005). Modeling human arthritic diseases in nonhuman primates. *Arthritis Res Ther* 7(4)**,** 145-154. doi: 10.1186/ar1773.

41. de Groot, N.G., Blokhuis, J.H., Otting, N., Doxiadis, G.G., and Bontrop, R.E. (2015). Co-evolution of the MHC class I and KIR gene families in rhesus macaques: ancestry and plasticity. *Immunol Rev* 267(1)**,** 228-245. doi: 10.1111/imr.12313.

42. Moreland, A.J., Guethlein, L.A., Reeves, R.K., Broman, K.W., Johnson, R.P., Parham, P., et al. (2011). Characterization of killer immunoglobulin-like receptor genetics and comprehensive genotyping by pyrosequencing in rhesus macaques. *BMC Genomics* 12**,** 295. doi: 10.1186/1471-2164-12-295.

43. Hershberger, K.L., Shyam, R., Miura, A., and Letvin, N.L. (2001). Diversity of the killer cell Ig-like receptors of rhesus monkeys. *J Immunol* 166(7)**,** 4380-4390.

44. Bimber, B.N., Moreland, A.J., Wiseman, R.W., Hughes, A.L., and O'Connor, D.H. (2008). Complete Characterization of Killer Ig-Like Receptor (KIR) Haplotypes in Mauritian Cynomolgus Macaques: Novel Insights into Nonhuman Primate KIR Gene Content and Organization. *J Immunol* 181(9)**,** 6301-6308.

45. Sambrook, J.G., Bashirova, A., Palmer, S., Sims, S., Trowsdale, J., Abi-Rached, L., et al. (2005). Single haplotype analysis demonstrates rapid evolution of the killer immunoglobulin-like receptor (KIR) loci in primates. *Genome Res* 15(1)**,** 25-35. doi: 10.1101/gr.2381205.

46. Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., et al. (2012). Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28(12)**,** 1647-1649. doi: 10.1093/bioinformatics/bts199.

47. Blokhuis, J.H., Doxiadis, G.G., and Bontrop, R.E. (2009). A splice site mutation converts an inhibitory killer cell Ig-like receptor into an activating one. *Mol Immunol* 46(4)**,** 640-648. doi: 10.1016/j.molimm.2008.08.270.

48. Colantonio, A.D., Bimber, B.N., Neidermyer, W.J., Jr., Reeves, R.K., Alter, G., Altfeld, M., et al. (2011). KIR polymorphisms modulate peptide-dependent binding to an MHC class I ligand with a Bw6 motif. *PLoS Pathog* 7(3)**,** e1001316. doi: 10.1371/journal.ppat.1001316.

49. Bostik, P., Kobkitjaroen, J., Tang, W., Villinger, F., Pereira, L.E., Little, D.M., et al. (2009). Decreased NK cell frequency and function is associated with increased risk of KIR3DL allele polymorphism in simian immunodeficiency virus-infected rhesus macaques with high viral loads. *J Immunol* 182(6)**,** 3638-3649. doi: 10.4049/jimmunol.0803580.

50. Grendell, R.L., Hughes, A.L., and Golos, T.G. (2001). Cloning of rhesus monkey killer-cell Ig-like receptors (KIRs) from early pregnancy decidua. *Tissue Antigens* 58(5)**,** 329-334.

51. Kruse, P.H., Rosner, C., and Walter, L. (2010). Characterization of rhesus macaque KIR genotypes and haplotypes. *Immunogenetics* 62(5)**,** 281-293. doi: 10.1007/s00251-010-0433-4.

52. Robinson, J., Waller, M.J., Stoehr, P., and Marsh, S.G. (2005). IPD--the Immuno Polymorphism Database. *Nucleic Acids Res* 33(Database issue)**,** D523-526. doi: 10.1093/nar/gki032.

53. Jian, X., Boerwinkle, E., and Liu, X. (2014). In silico tools for splicing defect prediction: a survey from the viewpoint of end users. *Genet Med* 16(7)**,** 497-503. doi: 10.1038/gim.2013.176.

54. Ohno, K., Takeda, J.I., and Masuda, A. (2018). Rules and tools to predict the splicing effects of exonic and intronic mutations. *Wiley Interdiscip Rev RNA* 9(1). doi: 10.1002/wrna.1451.

55. Houdayer, C., Caux-Moncoutier, V., Krieger, S., Barrois, M., Bonnet, F., Bourdon, V., et al. (2012). Guidelines for splicing analysis in molecular diagnosis derived from a set of 327 combined in silico/in vitro studies on BRCA1 and BRCA2 variants. *Hum Mutat* 33(8)**,** 1228-1238. doi: 10.1002/humu.22101.

56. Jian, X., Boerwinkle, E., and Liu, X. (2014). In silico prediction of splice-altering single nucleotide variants in the human genome. *Nucleic Acids Res* 42(22)**,** 13534-13544. doi: 10.1093/nar/gku1206.

57. Colombo, M., De Vecchi, G., Caleca, L., Foglia, C., Ripamonti, C.B., Ficarazzi, F., et al. (2013). Comparative in vitro and in silico analyses of variants in splicing regions of BRCA1 and BRCA2 genes and characterization of novel pathogenic mutations. *PLoS One* 8(2)**,** e57173. doi: 10.1371/journal.pone.0057173.

58. Yeo, G., and Burge, C.B. (2004). Maximum Entropy Modeling of Short Sequence Motifs with Applications to RNA Splicing Signals. *Journal of Computational Biology* 11(2-3)**,** 377-394. doi: 10.1089/1066527041410418.

59. Rogozin, I.B., and Milanesi, L. (1997). Analysis of donor splice sites in different eukaryotic organisms. *J Mol Evol* 45(1)**,** 50-59.

60. Desmet, F.O., Hamroun, D., Lalande, M., Collod-Beroud, G., Claustres, M., and Beroud, C. (2009). Human Splicing Finder: an online bioinformatics tool to predict splicing signals. *Nucleic Acids Res* 37(9)**,** e67. doi: 10.1093/nar/gkp215.

61. Reese, M.G., Eeckman, F.H., Kulp, D., and Haussler, D. (1997). Improved splice site detection in Genie. *J Comput Biol* 4(3)**,** 311-323. doi: 10.1089/cmb.1997.4.311.

62. Keren, H., Lev-Maor, G., and Ast, G. (2010). Alternative splicing and evolution: diversification, exon definition and function. *Nat Rev Genet* 11(5)**,** 345-355. doi: 10.1038/nrg2776.

63. Schwartz, S.H., Silva, J., Burstein, D., Pupko, T., Eyras, E., and Ast, G. (2008). Large-scale comparative analysis of splicing signals and their corresponding splicing factors in eukaryotes. *Genome Res* 18(1)**,** 88-103. doi: 10.1101/gr.6818908.

64. Cornish-Bowden, A. (1985). Nomenclature for incompletely specified bases in nucleic acid sequences: recommendations 1984. *Nucleic Acids Res* 13(9)**,** 3021-3030.

65. Coolidge, C.J., Seely, R.J., and Patton, J.G. (1997). Functional analysis of the polypyrimidine tract in pre-mRNA splicing. *Nucleic Acids Res* 25(4)**,** 888-896.

66. Sickmier, E.A., Frato, K.E., Shen, H., Paranawithana, S.R., Green, M.R., and Kielkopf, C.L. (2006). Structural basis for polypyrimidine tract recognition by the essential pre-mRNA splicing factor U2AF65. *Mol Cell* 23(1)**,** 49-59. doi: 10.1016/j.molcel.2006.05.025.

67. Dhir, A., and Buratti, E. (2010). Alternative splicing: role of pseudoexons in human disease and potential therapeutic strategies. *Febs j* 277(4)**,** 841-855. doi: 10.1111/j.1742-4658.2009.07520.x.

68. Rump, A., Rosen-Wolff, A., Gahr, M., Seidenberg, J., Roos, C., Walter, L., et al. (2006). A splice-supporting intronic mutation in the last bp position of a cryptic exon within intron 6 of the CYBB gene induces its incorporation into the mRNA causing chronic granulomatous disease (CGD). *Gene* 371(2)**,** 174-181. doi: 10.1016/j.gene.2005.11.036.

69. Fackenthal, J.D., Lee, Y., and Olopade, O.I. (2012). Hidden dangers: a cryptic exon disrupts BRCA2 mRNA. *Clin Cancer Res* 18(18)**,** 4865-4867. doi: 10.1158/1078-0432.ccr-12-2090.

70. Wright, P.W., Li, H., Huehn, A., O'Connor, G.M., Cooley, S., Miller, J.S., et al. (2014). Characterization of a weakly expressed KIR2DL1 variant reveals a novel upstream promoter that controls KIR expression. *Genes Immun* 15(7)**,** 440-448. doi: 10.1038/gene.2014.34.

71. Palacios, C., Cuervo, L.C., and Cadavid, L.F. (2011). Evolutionary patterns of killer cell Ig-like receptor genes in Old World monkeys. *Gene* 474(1-2)**,** 39-51. doi: 10.1016/j.gene.2010.12.006.

72. Maxwell, L.D., Wallace, A., Middleton, D., and Curran, M.D. (2002). 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* 60(3)**,** 254-258.

73. Bimber, B.N., and Evans, D.T. (2015). The killer-cell immunoglobulin-like receptors of macaques. *Immunol Rev* 267(1)**,** 246-258. doi: 10.1111/imr.12329.

74. Fan, Q.R., Long, E.O., and Wiley, D.C. (2000). A disulfide-linked natural killer cell receptor dimer has higher affinity for HLA-C than wild-type monomer. *Eur J Immunol* 30(9)**,** 2692-2697. doi: 10.1002/1521-4141(200009)30:9<2692::aid-immu2692>3.0.co;2-0.

75. Kumar, S., Rajagopalan, S., Sarkar, P., Dorward, D.W., Peterson, M.E., Liao, H.S., et al. (2016). Zinc-Induced Polymerization of Killer-Cell Ig-like Receptor into Filaments Promotes Its Inhibitory Function at Cytotoxic Immunological Synapses. *Mol Cell* 62(1)**,** 21-33. doi: 10.1016/j.molcel.2016.03.009.

76. Pickrell, J.K., Pai, A.A., Gilad, Y., and Pritchard, J.K. (2010). Noisy splicing drives mRNA isoform diversity in human cells. *PLoS Genet* 6(12)**,** e1001236. doi: 10.1371/journal.pgen.1001236.

77. Leoni, G., Le Pera, L., Ferre, F., Raimondo, D., and Tramontano, A. (2011). Coding potential of the products of alternative splicing in human. *Genome Biol* 12(1)**,** R9. doi: 10.1186/gb-2011-12-1-r9.

78. Sorek, R., Shamir, R., and Ast, G. (2004). How prevalent is functional alternative splicing in the human genome? *Trends Genet* 20(2)**,** 68-71. doi: 10.1016/j.tig.2003.12.004.

79. Lewis, B.P., Green, R.E., and Brenner, S.E. (2003). Evidence for the widespread coupling of alternative splicing and nonsense-mediated mRNA decay in humans. *Proc Natl Acad Sci U S A* 100(1)**,** 189-192. doi: 10.1073/pnas.0136770100.

80. Lareau, L.F., Brooks, A.N., Soergel, D.A., Meng, Q., and Brenner, S.E. (2007). The coupling of alternative splicing and nonsense-mediated mRNA decay. *Adv Exp Med Biol* 623**,** 190-211.

81. Sharma, D., Bastard, K., Guethlein, L.A., Norman, P.J., Yawata, N., Yawata, M., et al. (2009). Dimorphic motifs in D0 and D1+D2 domains of killer cell Ig-like receptor 3DL1 combine to form receptors with high, moderate, and no avidity for the complex of a peptide derived from HIV and HLA-A*2402. *J Immunol* 183(7)**,** 4569-4582. doi: 10.4049/jimmunol.0901734.

82. Khakoo, S.I., Geller, R., Shin, S., Jenkins, J.A., and Parham, P. (2002). The D0 domain of KIR3D acts as a major histocompatibility complex class I binding enhancer. *J Exp Med* 196(7)**,** 911-921.

83. O'Connor, G.M., Vivian, J.P., Widjaja, J.M., Bridgeman, J.S., Gostick, E., Lafont, B.A., et al. (2014). Mutational and structural analysis of KIR3DL1 reveals a lineage-defining allotypic dimorphism that impacts both HLA and peptide sensitivity. *J Immunol* 192(6)**,** 2875-2884. doi: 10.4049/jimmunol.1303142.

84. Thomas, R., Yamada, E., Alter, G., Martin, M.P., Bashirova, A.A., Norman, P.J., et al. (2008). Novel KIR3DL1 alleles and their expression levels on NK cells: convergent evolution of KIR3DL1 phenotype variation? *J Immunol* 180(10)**,** 6743-6750.

85. Yeo, G., Holste, D., Kreiman, G., and Burge, C.B. (2004). Variation in alternative splicing across human tissues. *Genome Biol* 5(10)**,** R74. doi: 10.1186/gb-2004-5-10-r74.

86. Peng, H., Wisse, E., and Tian, Z. (2016). Liver natural killer cells: subsets and roles in liver immunity. *Cell Mol Immunol* 13(3)**,** 328-336. doi: 10.1038/cmi.2015.96.

87. Burt, B.M., Plitas, G., Zhao, Z., Bamboat, Z.M., Nguyen, H.M., Dupont, B., et al. (2009). The lytic potential of human liver NK cells is restricted by their limited expression of inhibitory killer Ig-like receptors. *J Immunol* 183(3)**,** 1789-1796. doi: 10.4049/jimmunol.0900541.

88. Hesker, P.R., and Krupnick, A.S. (2013). The role of natural killer cells in pulmonary immunosurveillance. *Front Biosci (Schol Ed)* 5**,** 575-587.

89. Sanos, S.L., and Diefenbach, A. (2010). Isolation of NK cells and NK-like cells from the intestinal lamina propria. *Methods Mol Biol* 612**,** 505-517. doi: 10.1007/978-1-60761-362-6_32.

90. Garrod, K.R., Wei, S.H., Parker, I., and Cahalan, M.D. (2007). Natural killer cells actively patrol peripheral lymph nodes forming stable conjugates to eliminate MHC-mismatched targets. *Proc Natl Acad Sci U S A* 104(29)**,** 12081-12086. doi: 10.1073/pnas.0702867104.

91. Niederkorn, J.Y. (2010). "NK cells in the eye," in *Natural Killer Cells*. Elsevier Ltd.), pp. 385-401

92. Gaynor, L.M., and Colucci, F. (2017). Uterine Natural Killer Cells: Functional Distinctions and Influence on Pregnancy in Humans and Mice. *Front Immunol* 8**,** 467. doi: 10.3389/fimmu.2017.00467.

93. Carrega, P., and Ferlazzo, G. (2012). Natural killer cell distribution and trafficking in human tissues. *Front Immunol* 3**,** 347. doi: 10.3389/fimmu.2012.00347.

94. Chen, M., and Manley, J.L. (2009). Mechanisms of alternative splicing regulation: insights from molecular and genomics approaches. *Nat Rev Mol Cell Biol* 10(11)**,** 741-754. doi: 10.1038/nrm2777.

95. Grosso, A.R., Gomes, A.Q., Barbosa-Morais, N.L., Caldeira, S., Thorne, N.P., Grech, G., et al. (2008). Tissue-specific splicing factor gene expression signatures. *Nucleic Acids Res* 36(15)**,** 4823-4832. doi: 10.1093/nar/gkn463.

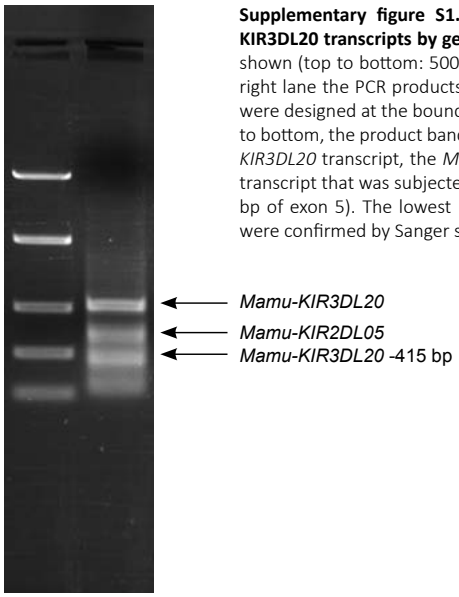
96. Zhou, H.L., Luo, G., Wise, J.A., and Lou, H. (2014). Regulation of alternative splicing by local histone modifications: potential roles for RNA-guided mechanisms. *Nucleic Acids Res* 42(2)**,** 701-713. doi: 10.1093/nar/gkt875.

97. Moreau, P., Carosella, E., Teyssier, M., Prost, S., Gluckman, E., Dausset, J., et al. (1995). Soluble HLA-G molecule. An alternatively spliced HLA-G mRNA form candidate to encode it in peripheral blood mononuclear cells and human trophoblasts. *Hum Immunol* 43(3)**,** 231-236.

98. Kovats, S., Main, E.K., Librach, C., Stubblebine, M., Fisher, S.J., and DeMars, R. (1990). A class I antigen, HLA-G, expressed in human trophoblasts. *Science* 248(4952)**,** 220-223.

99. Dahl, M., Djurisic, S., and Hviid, T.V. (2014). The many faces of human leukocyte antigen-G: relevance to the fate of pregnancy. *J Immunol Res* 2014**,** 591489. doi: 10.1155/2014/591489.

Supplementary material



Supplementary figure S1. Visualization of alternative splicing in Mamu-KIR3DL20 transcripts by gel electrophoresis. In the left lane marker bands are shown (top to bottom: 5000 bp, 2000 bp, 850 bp, 400 bp, 100 bp), and in the right lane the PCR products of *Mamu-KIR3DL20* are shown, using primers that were designed at the boundary of exons 1/2 and at the end of exon 5. From top to bottom, the product bands correspond with the constitutively spliced *Mamu-KIR3DL20* transcript, the *Mamu-KIR2DL05* transcript (exon 4 skipped), and the transcript that was subjected to the excision of 415 bp (exon 4 and the first 115 bp of exon 5). The lowest band shows aspecific amplification. The sequences were confirmed by Sanger sequencing.

Supplementary table 1. Eighteen different splice events were observed for human KIR. For each *KIR* gene for which a splice event was observed, an accession number is listed.

Splice mechanism	Insertion/deletion size	Gene	Accession number
Exon skipping	Deletion 294 bp (exon 5)	*2DL2*	LS974084
		2DL3	LS974085
		2DL5	LS974086
		3DL1	LS974087
	Deletion 51 bp (exon 6)	*2DL1*	LS974078
		2DL3	LS974079
		2DL4	LS974080
		2DS2	LS974081
		2DS4	LS974082
		3DL2	LS974083
	Deletion 104/105 bp (exon 7)	*2DL4*	LS974091, LS974092
		3DL2	LS974090
	Deletion 594 bp (exons 4 + 5)	*3DL1*	LS974088
		3DL3	LS974089
	Deletion 155/156 bp (exons 6 + 7)	*2DL4*	LS974093 ,LS974095
		3DL2	LS974094
	Deletion 158 bp (exons 7 + 8)	*2DL4*	LS974097
	Deletion 209 bp (exons 6 + 7 + 8)	*2DL4*	LS974096
Alternative 3’ ss	Deletion 150 bp	*2DL3*	LS974098
	Inclusion 49 bp	*2DL1*	LS974099
	Inclusion 170 bp	*3DL1*	LS974100
Alternative 5’ ss	Deletion 66 bp	*2DL4*	LS974103
	Deletion 73/74 bp	*2DL4*	LS974105, LS974104
	Deletion 198 bp	*2DL4*	LS974106
	Inclusion 67 bp	*2DL4*	LS974101
	Inclusion 129 bp	*3DL2*	LS974102
Alternative 3’ + 5’ ss	Deletion 294 bp	*2DL3*	LS974108
		2DS4	LS974107
		3DL1	LS974110
		3DL2	LS974109
Cryptic exon	Inclusion 78 bp	*2DL1*	LS974111
		2DL2	LS974112
		2DL3	LS974113
		2DS1	LS974114
	Inclusion 54/57 bp	*2DL1*	LS974115
		2DS1	LS974117
		2DS4	LS974077
		2DS5	LS974116

Supplementary table 2. Twenty-nine different splice events were observed for rhesus macaque KIR. For each *KIR* gene for which a splice event was observed, an accession number is listed.

Splice mechanism	Insertion/deletion size	Gene	Accession number
Exon skipping	Deletion 300 bp (exon 4)	*3DL05*	LR030267
		3DS05	LR030254
	Deletion 294 bp (exon 5)	*3DL01*	LR030249, LR030461
		3DL02	LR030261
		3DL05	LR030491
		3DL07	LR030266
		3DL10	LR030478
		3DL10A/3DL02	LR030248
	Deletion 51 bp (exon 6)	*2DL04*	LR030486
	Deletion 53 bp (exon 8)	*2DL04*	LR030481
	Deletion 594 bp (exon 4 + 5)	*3DL07*	LR030255
		3DL20	LR030240
	Deletion 446 bp (exon 5 + 6 + 7)	*3DL01*	LR030258
		3DL05	LR030463
		3DL07	LR030492
		3DL10	LR030259
		3DL10A/3DL02	LR030253
	Deletion 115 bp (exon 6 + 7)	*2DL04*	LR030490
Alternative 3' ss	Deletion 36 bp	*3DL01*	LR030242
		3DL02	LR030495
		3DL05	LR030477, LR030494
		3DL07	LR030264
		3DL08	LR030260
		3DL10	LR030466, LR030469
		3DS01	LR030505
		3DS02	LR030268
		3DS03	LR030503
		3DS05	LR030251
		3DSw08	LR030493
		3DL02/3DL08A	LR030504
		3DL20	LR030472
	Deletion 216 bp	*3DL02*	LR030496
		3DL08	LR030245
		3DS05	LR030252
		3DSw08	LR030474
		3DL02/3DL08A	LR030243
	Deletion 267 bp	*3DL05*	LR030497
		3DS05	LR030462
	Deletion 27 bp	*2DL04*	LR030487

Supplementary table 2. Continued.

Splice mechanism	Insertion/deletion size	Gene	Accession number
Alternative 3' ss	Deletion 115 bp	*3DL07*	LR030265
		3DL08	LR030471
		3DL10	LR030479
		3DS03	LR030262
		3DL02/3DL08A	LR030247
		3DL10A/3DL02	LR030241
		3DL20	LR030464
	Deletion 150 bp	*3DL02*	LR030498
		3DLw03	LR030476
		3DL10	LR030480
	Deletion 176 bp	*3DL07*	LR030256
		3DL10A/3DL02	LR030250
	Inclusion 109 bp (exon. 7)	*2DL04*	LR030483
	Inclusion 147 bp (exon 7)	*2DL04*	LR030488
	Inclusion 245/246 bp	*3DL07*	LR030263
		2DL04	LR030489
Alternative 5' ss	Deletion 198 bp	*3DL02*	LR030499
		3DS05	LR030467
		3DL02/3DL08A	LR030244
	Inclsuion 109 bp (exon 3)	*2DL04*	LR030484
Alternative 3' and 5' ss	Deletion 81 bp	*3DS02*	LR030502
	Deletion 107 bp	*3DS02*	LR030501
	Deletion 141 bp	*3DL08*	LR030473
		3DSw08	LR030475
		3DL02/3DL08A	LR030246
	Deletion 414 bp	*3DS05*	LR030500
	Deletion 324 bp	*3DL20*	LR030468
Exon skipping and alt. 3' ss	Deletion 415 bp	*3DL20*	LR030465
Cryptic exon	Inclusion 88 bp	*3DL01*	LR030470
	Inclusion 147 bp (exon 5)	*2DL04*	LR030485
	Inclusion 47 bp	*3DL07*	LR030257
Intron retention	Inclusion 98/99 bp	*2DL04*	LR030482

Supplementary table 3. Primer sets to amplify human and macaque intron sequences. In human, two generic primer sets amplified most *KIR2D* genes. In macaques, different primer sets were used to amplify certain regions of different *KIR* genes.

Macaque					
Gene	**Region**	**Forward**	**Reverse**	**Annealing temp. (C°)**	**# of cycli**
KIR3DL	Exon 6 to exon 9	TCACCCACTGAACCAAGCTC	ACAYGCTGGTATCTGTTGGG	62	30
KIR3DL	Intron 4 to exon 9	TGGGGAAGTGAGGTCAGAGA	ACACGCTGGTATCTGTTGGG	62	32
KIR2DL04	Exon 5 to exon 9	ACATTCCAGGCTGACTTCCC	GCTGTGAAGCCGGTTTTGAG	62	32
KIR3DL01	Exon 5 to exon 9	TGACCTTGTCCTGCAGTTCC	GCTGTGAAGCCGGTTTTGAG	62	32
KIR3DL20	Exon 1 to intron 4	GCRTGTGTTGGGTTCTTCTTG	TCCTGTTTCTCTACCTCTGTTGGT	62	32

Human					
Gene	**Region**	**Forward**	**Reverse**	**Annealing temp. (C°)**	**# of cycli**
KIR2DL	Exon 6 to exon 9	TCACCCACTGAACCAAGCTC	GGGCAGGAGACAACTTTGGA	60	32
KIR2DS	Exon 6 to exon 9	TCACCCACTGAACCAAGCTC	TGCGTATGACACCTCCTGATG	60	32

Supplementary table 4. The Maximum Entropy Modeling Scan (MaxEntScan; MES) (68), the Position Weight Matrix (PWM) via SpliceView (69), and the Human Splice Finder (HSF) were used to predict most splicing strenght scores. If these tools failed to provide a splicing strenght score, the Weight Matrix Model (WMM) (68) and NNSplice tool (71) were used. The definition of the 3' and 5' splice site could differ between the different models and are listed: + determine the number of nucleotides within the exon, and- display the number of nucleotides within the intron. Also, the models define different score ranges, in which a higher value always implies a better predicted splice site. Only the HSF model provided scores for the branch point sequence (BPS), exonic splice enhancers (ESE), and exonic splice silencers (ESS).

Software tool	**3' ss definition**	**5' ss definition**	**Splice strenght score range**	**Additional elements**
MES	-20 to +3	+3 to-6	-20 to +20	-
PWM	-9 to +4	+3 to-6	0 to 100	-
HSF	-12 to +2	+3 to-6	0 to 100	BP, ESE, ESS
WMM	-20 to +3	+3 to-6	-20 to +20	-
NNSplice tool	-21 to +20	+7 to-8	0 to 1	-

Chapter 5

Rapid characterization of complex killer cell immunoglobulin-like receptor (KIR) regions using Cas9 enrichment and Nanopore sequencing

Jesse Bruijnesteijn[1]*, Marit van der Wiel[1], Natasja G. de Groot[1] and Ronald E. Bontrop[1,2]

[1]Comparative Genetics and Refinement, Biomedical Primate Research Centre, Rijswijk, the Netherlands
[2]Theoretical Biology and Bioinformatics, Utrecht University, Utrecht, the Netherlands

*Corresponding author:
Jesse Bruijnesteijn
bruijnesteijn@bprc.nl
Tel. nr. +31-15-2842574
Fax nr. +31-15-2842600

Abstract

Long-read sequencing approaches have considerably improved the quality and contiguity of genome assemblies. Such platforms bear the potential to resolve even extremely complex regions, such as multigenic immune families and repetitive stretches of DNA. Deep sequencing coverage, however, is required to overcome low nucleotide accuracy, especially in regions with high homopolymer density, copy number variation, and sequence similarity, such as the *MHC* and *KIR* gene clusters of the immune system. Therefore, we have adapted a targeted enrichment protocol in combination with long-read sequencing to efficiently annotate complex *KIR* gene regions. Using Cas9 endonuclease activity, segments of the *KIR* gene cluster were enriched and sequenced on an Oxford Nanopore Technologies platform. This provided sufficient coverage to accurately resolve and phase highly complex *KIR* haplotypes. Our strategy eliminates PCR-induced amplification errors, facilitates rapid characterization of large and complex multigenic regions, including its epigenetic footprint, and is applicable in multiple species, even in the absence of a reference genome.

Introduction

Repetitive regions are difficult to resolve using short-read sequencing approaches, and often remain registered for years as incomplete gaps in draft genomes (1-7). These complex stretches often involve transposable elements, microsatellites, and multi-copy gene clusters, the latter of which is represented by multiple gene families that encode essential components of the immune system (8, 9). For example, the major histocompatibility complex (MHC) genes, known in humans as human leukocyte antigens (HLA) genes, are considered the most polymorphic gene cluster. The *MHC* genes co-evolved with their receptors of the killer cell immunoglobulin-like receptor (KIR) gene family, which also features striking levels of complexity (10, 11).

The fundamental limitations to characterizing these complex regions by short-read sequencing strategies are potentially overcome by third-generation techniques that generate high yields of long reads (12, 13). Oxford Nanopore sequencing may produce reads far above 100 kb by recording changes in the electrical current as nucleotides pass through synthetic nanopores. The data quality and throughput of nanopore sequencing is improving rapidly, and has allowed the *de novo* assembly of multiple human genomes (14-16). These genome assemblies contiguously span multigenic clusters, such as the *MHC* and *KIR* gene regions, but correct annotation is hampered by the relatively low coverage, which precludes at this stage an accurate allele level resolution. Considering the important role of different multi-copy gene families in health and disease, a cost-efficient and high-resolution characterization approach regarding these types of regions is urgent.

Instead of whole genome sequencing, specific genes and regions might be enriched during library preparation. For instance, the *MHC class II DRB* gene region was enriched

by long-ranged PCR, and characterized using a hybrid sequencing approach that combined Illumina and Oxford Nanopore platforms (17). Amplification steps, however, might introduce nucleotide errors during synthesis and, in addition, erase all epigenetic footprints. An amplification-free enrichment technique involves Cas9-mediated targeting of chromosome segments and nanopore sequencing (18-22). The Cas9 endonuclease activity may specifically excise genomic regions of interest that are subsequently ligated to nanopore adapters. This allows the direct sequencing of genomic segments while avoiding error prone DNA synthesis and maintaining epigenetic modifications. Efficient and specific enrichment using this approach has been demonstrated for single genes, including several cancer-related fusion genes (21), but an application for multigenic regions is absent in the literature.

In this study, we adapted the Cas9-mediated enrichment potential to resolve complex immune regions and validated this approach by the targeted characterization of *KIR* gene clusters in two different primate species. We focused on the *KIR* region in humans, which has been thoroughly characterized at the genomic level, and is important, for instance, in AIDS susceptibility and transplantation biology (23, 24). Rhesus macaques, however, represent a frequently used species in preclinical health research concerning, for example, COVID-19 and AIDS (25-27), but the physical location of the *KIR* genes is poorly understood. The *KIR* receptor family is involved in the regulation of NK cell activity and comprises activating and inhibitory members that may recognize particular epitopes on MHC class I molecules. A comprehensive nomenclature system distinguishes the variety of KIR receptors, and reflects the number of extracellular domains (KIR1D, KIR2D, KIR3D) and the length of the cytoplasmic tail (long, L; short, S) (28, 29). Subsequent numbering defines structurally similar but phylogenetically distinct genes (e.g., *KIR2DL1*), whereas three additional digits distinguish allotypes (e.g., *KIR2DL1*001*). In humans, a total of 17 *KIR* genes are defined, 1110 alleles of which are documented (IPD-KIR, release 2.9.0) (30).

The *KIR* gene repertoire is shaped by abundant tandem duplications, deletions, and chromosomal recombination events, and exceeds the plasticity of the *MHC* gene cluster (31-33). The *KIR* genes are 10 to 15 kb long and are arranged in a head-to-tail manner, separated by intergenic regions of approximately 2 kb. Sequence similarity characterizes the genetic cluster, with any two *KIR* genes sharing 80%–90% homology, and allelic variants of a certain gene tend to be over 98% similar. The KIR haplotypes, defined as a segregating unit of genes located on a single chromosome, distinguish a centromeric and telomeric segment, which display diverse configurations and extensive copy number variation. The *KIR* genes display a variegated expression pattern, which is modulated by methylation of the promotor regions (34). For this study, we resolved six human *KIR* haplotypes, derived from three randomly selected human donors, at an allele-level resolution and additionally determined their methylation profiles.

The continuous evolution of the *KIR* gene system is reflected by the genomic diversification at a species, population, and individual level. To validate our concept, we enriched and assembled KIR haplotypes in rhesus macaques. So far, only two completely sequenced rhesus macaque *KIR* haplotypes have been documented, but

previous transcriptome and segregation studies indicate extensive variation (35-39). Annotation of this complex immune region is a difficult enterprise, which is reflected by a poorly annotated *KIR* region in the rhesus macaque reference genome (Mmul_10) (40). Our genomic characterization of rhesus macaque *KIR* haplotypes demonstrated the rapid construction of complex multigenic haplotypes, even in the absence of reference sequences. Hence, adaption of this technique allows a speedy and cost-efficient characterization of other immune regions, from which whole genome assemblies and clinical implications might benefit.

Materials and methods

Cells and genomic DNA extraction

Human buffy coat samples from healthy donors were obtained from the Dutch blood bank (Sanquin, the Netherlands). Informed consent was obtained from all participants. Rhesus macaques with a characterized KIR transcriptome were selected from the self-sustaining colony housed at the Biomedical Primate Research Centre (BPRC) (28, 36). Heparin whole blood samples from these animals were obtained during annual health checks. PBMCs were isolated from human buffy coats and rhesus macaque heparin samples.

High-molecular-weight (HMW) gDNA was isolated from human and rhesus macaque PBMC samples (± 7 x 10^6 cells), using the Circulomics Nanobind CBB Big DNA Kit (Circulomics, NB-900-001-01) and following the manufacturer's instructions. The concentration and purity of the gDNA samples were determined using a Nanodrop and a Qubit platform. The (HMW) gDNA fragment length was determined by pulsed field gel electrophoresis (PFGE) in reference to a lambda PFGE ladder.

Designing guiding crRNAs and constructing RNPs

For both humans and macaques, sets of generic and specific CRISPR RNAs (crRNA) were designed within the *KIR* gene cluster and the flanking *LILR* and *FCaR* genes by using Benchling, a freely available online software tool (41). The crRNAs have a guiding length of 20 bp, and are designed in front of a protospacer adjacent motif (PAM) sequence "NGG", in which "N" represents any nucleotide base. The software tool provides on- and off-target scores for specific crRNAs. The on-target scores are based on optimized calculations from Doench and colleagues, with the higher score indicating the better crRNA target binding (42). The off-target scores reflect the specificity of the crRNAs, and were based on different builds of the human (NCBI36, GRCh37, GRCh38) and macaque (MMUL_1, Mmul_8.0.1) reference genomes (43). CRISPR RNAs with high off-target scores were considered specific for one *KIR* gene. Relatively low off-target scores (ranging from 3 to 60) were recorded for crRNAs that potentially target multiple *KIR* genes and were included in the panel as putative generic crRNAs. In total, 54 and 45 custom crRNAs were selected to enrich the human and macaque *KIR* gene cluster, respectively (**Sup. Table 1** and **2**) (IDT, custom Alt-R® CRISPR-Cas9 crRNA).

Sets of different crRNAs were pooled based on the different strand-directed orientations (sense versus anti-sense) to avoid cleavage and the sequencing of unintended short on-target fragments. The pools were defined to generate DNA fragments that comprise *KIR* genes from exon 1 to exon 9, or fragments that connect neighboring *KIR* genes (**Sup. Table 1** and **2**). The pooled crRNAs were mixed with trans-activating crRNAs (tracrRNA) (IDT, #1072534) in a 1:1 ratio and further diluted in Duplex Buffer to a final concentration of 10 µM. The crRNAs and tracrRNA were annealed by heating the duplex solution for 5 min at 95˚C, followed by cooling to room temperature (RT) on a benchtop. To subsequently construct the Cas9 ribonucleoprotein particles (RNPs), the crRNA-tracrRNA duplexes were assembled with HiFi Cas9 endonuclease (IDT, #1081060) in 1x NEB CutSmart Buffer (NEB, #B7204S) at a total volume of 30 µl by incubating the solution for 30 min at RT. The Cas9 RNPs were stored until use at 4˚C for up to a week.

Cas9-mediated target enrichment and Oxford Nanopore sequencing

Throughout the protocol, unintended fragmentation of gDNA will decrease the capturing efficiency and enrich off-target fragments. Therefore, samples should be handled with care and processed with wide-bore pipette tips. Input gDNA (5-10 µg) was resuspended in 10x NEB CutSmart Buffer (8:1) and dephosphorylated by incubation with Quick calf intestinal phosphatase (CIP) (NEB, # M0525S) at 37˚C for 20 min, followed by heating at 80˚C for 2 min to deactivate the enzyme (**Figure 1**). After the sample returned to room temperature, the dephosphorylated gDNA (30 µl) was gently mixed with a Cas9 RNP pool (10 µl), 10 mM dATP (1 µl), and Taq polymerase (1 µl), followed by incubation at 37˚C for 60 min, then 72˚C for 5 min, and hold at 4˚C. Ligation buffer (20 µl) and sequencing adaptors (5 µl) from the Ligation Sequencing Kit (ONT, #LSK109) and Quick T4 DNA Ligase (10 µl) (NEB, M2200S) were added to the cleaved and dA-tailed gDNA sample, followed by an incubation of 60 min at RT. Adapter-ligated samples from the same individual, which were treated with different crRNA sets, were pooled, and diluted (1:1) in TE buffer. The excess of adaptors and short DNA fragments were removed using 0.3x AMPure XP Beads (Beckman Coulter, #A63881), which were washed twice on a magnetic rack with Long Fragment Buffer (ONT, #LSK109). The beads were eluted in 15 µl Elution Buffer (ONT, #LSK109). A sequencing library was prepared by adding 37.5 µl Sequencing Buffer and 25.5 µl Loading Beads (ONT, #LSK109) to the processed DNA sample. Eluted samples were sequenced on a R9.4.1 flowcell using an Oxford Nanopore MinION device. Prior to sequencing, the flowcells were primed according to the manufacturer's instructions using the Flow Cell Priming Expansion Pack (ONT, #LSK109). After 24 hours of sequencing, flowcells with over 500 active pores remaining were washed and reloaded with a second enriched library of the same gDNA sample according to the manufacturer's instructions using the Flow Cell Wash Kit (ONT, #EXP-WSH003).

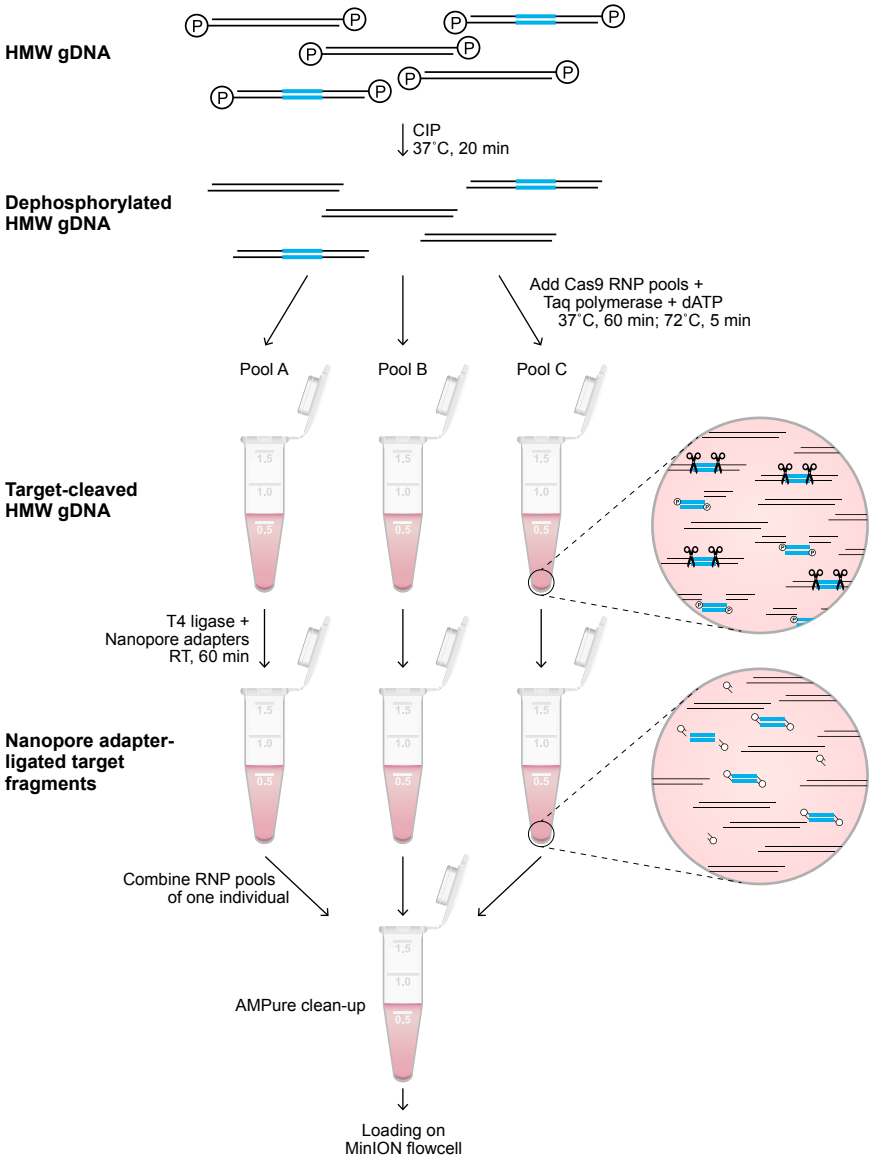


Figure 1. Library preparation of the targeted enrichment protocol. A schematic overview of the library preparation to enrich a target region (blue). Freshly isolated HMW genomic DNA was dephosphorylated by incubation with CIP for 20 min at 37˚C. Subsequently, selected pools of RNPs (Sup. Table 1 and 2) were added to aliquots of the dephosphorylated gDNA, and incubated for 60 min at 37˚C in the presence of Taq polymerase and dATP, followed by incubation for 5 min at 72˚C. The available phosphate groups at the terminus of the targeted region were ligated to Nanopore adaptors during incubation with T4 ligase for 60 min at room temperature (RT). The adapter-ligated aliquots were pooled, followed by clean-up with 0.3X AMPure beads. The eluted sample was loaded on a flowcell for Nanopore sequencing.

Sequence data analysis

Base calling and read quality assessment (min qscore 7) were performed using Guppy V3.4.1 software on a Linux platform that utilized a GeForce RTX 2080 Ti graphics processing unit (GPU). Base called reads were imported into Geneious Prime software (v.2020.1.2) for further analysis. Exon libraries, including sequences of exons 3, 4, and 7 from the human and rhesus macaque KIR databases (IPD-KIR, release 2.9.0; IPD-NHKIR, release 1.3.0.0), were used as references to map the Nanopore reads into contigs based on similarity. Each on-target contig resembled a specific *KIR* gene or a pair of neighboring *KIR* genes.

For each human individual, the haplotype configurations (e.g., cA01-tA01) could be largely deduced from the read contigs generated by exon library mapping. On the basis of this knowledge, the reads of a single flowcell were re-mapped to the complete human reference genome (HG38), complemented with the particular *KIR* haplotype configuration references, using minimap2 (version 2.17). Consensus sequences that covered *KIR* genes from start to end, or that comprised segments of neighboring *KIR* genes, were generated from the alignments based on 65% nucleotide similarity. The consensus accuracy was determined by comparison to the reference haplotype. The on-target coverage was resolved for each individual by mapping all reads with a read length above 7,000 bp to the human reference genome (HG38) using minimap2 in Geneious software. The enrichment factor was determined as the ratio of on- and off-target mean coverage, and reflects the enrichment efficiency.

For each rhesus macaque sample, reads from the on-target exon contigs were re-mapped to the contig consensus sequence using minimap2. Additional re-mapping steps were performed using the generated consensus sequences to optimize accuracy and diminish homopolymer errors. To further optimize these consensus sequences, shorter on-target reads (< 7,000 bp) were included for consensus calculations. The eventual consensus sequences were generated based on 65% nucleotide similarity. The consensus accuracy was estimated by the alignment with exon references available from previous KIR transcriptome studies (IPD-NHKIR, release 1.3.0.0). On-target coverage was defined by mapping all reads with a length of 7,000 bp or more to the rhesus macaque reference genome (Mmul_10), which was complemented with the appropriate assembled *KIR* haplotypes, using minimap2 in Geneious software.

DNA modification profiles

Raw signal data including information on DNA modifications, such as methylation, were processed using Guppy V4.0.12 software with the dna_r9.4.1_450bps_modbases_dam-dcm-cpg_hac.cfg configuration and fast5_out flag. Multi-read fast5 files were converted to single-read files using multi_to_single_fast5 provided by ONT. The base called reads were mapped to a reference genome using minimap2, and subsequently called for modifications using Nanopolish (44). The methylation likelihood and frequencies were visualized by Methplotlib (45)d, using an annotated reference genome. When the reference genome did not contain the appropriate *KIR* genes, like

Mmul_10 for all macaque individuals, methylation calls were annotated using a modified reference genome, which was complemented with the newly assembled haplotypes as artificial chromosomes.

Results

A 'tiling' approach to enrich complex immune clusters without amplification

The characterization of large and repetitive immune regions requires the generation and sequencing of genomic DNA (gDNA) fragments that share overlapping segments. Allelic variation in these overlaps allows the phasing of haplotypes. To achieve this goal, dephosphorylated high molecular weight (HMW) gDNA needs to be cleaved, using sets of CRISPR RNAs (crRNA) in complex with Cas9 endonuclease (**Figure 1**). These crRNAs are designed to target conserved stretches that are shared by members of a multigenic family. This approach will allow generic enrichment. Only at the terminus of the cleaved target sites is a phosphate group available, which is utilized for dA-tailing and subsequent ligation to Nanopore sequencing adaptors. This 'tiling' approach facilitates the selective enrichment of large overlapping DNA-segments, and allows the subsequent sequencing of polymorphic and multigenic immune regions without the need for amplification.

Enrichment of complex KIR regions

To validate our approach, the *KIR* gene regions in humans and rhesus macaques were enriched and characterized (**Figure 2**). The nature of *KIR* gene complexity required the design of species-specific sets of crRNAs to enrich complete genomic clusters (**Figure 3**; **Sup. Table 1** & **2**). In humans, the presence of four framework genes with distinctive physical locations marks the centromeric (*KIR3DL3* and *KIR3DP1*) and telomeric (*KIR2DL4* and *KIR3DL2*) haplotype segments (**Figure 2**). Expansion and contraction of both regions resulted in haplotypes that contain 9 to 14 *KIR* genes, including two pseudogenes (*KIR2DP1* and *KIR3DP1*). Important to note, however, is that some genes that are present on a given *KIR* haplotype may be absent from another. Human *KIR* haplotypes can be roughly categorized based on their gene content, into those with more an inhibitory (group A) or an activating (group B) gene profile (46). Recombination events, possibly owing to the high transposon density, might rearrange haplotype organizations (**Figure 2**). To determine the high content variability of this genetic cluster, 35 generic crRNAs were designed to target the differential presence of human *KIR* genes that may be encountered on a haplotype, whereas 12 crRNA were specific for one particular framework gene (**Figure 3A; Sup. Table 1**). In addition, seven crRNA were included to target the genes that flank the *KIR* gene cluster (*LILR* and *FcAR*), in order to define both ends of the *KIR* haplotype.

The *KIR* haplotypes in rhesus macaques display even more content diversity, with 4 to 17 KIR transcripts encoded as defined by segregation studies (35, 36). The only framework gene present on all haplotypes is *KIR3DL20*, which marks the centromeric

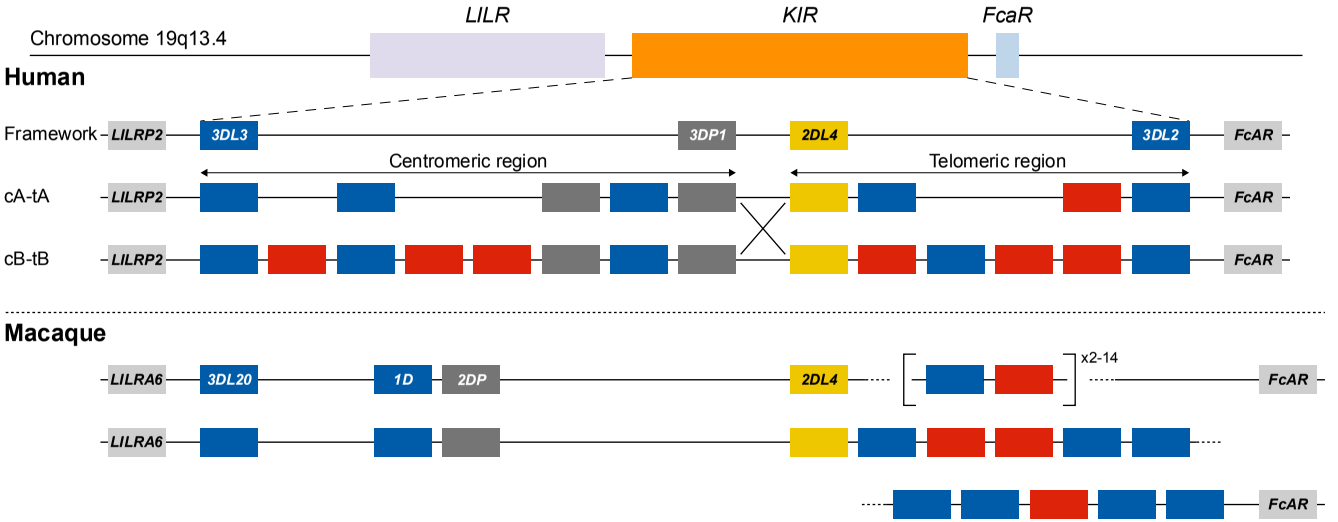


Figure 2. KIR haplotype configurations in humans and rhesus macaques. The *KIR* gene family maps between the *LILR* cluster and the *FcAR* gene. In humans, four genes are considered to reflect a framework reference haplotype, with *KIR3DL3* and *KIR3DP1* defining the centromeric region (c), and *KIR2DL4* and *KIR3DL2* bounding the telomeric region (t). In the human population, both the c and t regions may feature different gene contents. The red and blue boxes depict activating or inhibitory receptors, respectively. Pseudogenes have been indicated as well (grey). The only homologous gene shared in humans and macaques is *KIR2DL4* (yellow). Human *KIR* haplotypes are categorized based on a more inhibitory (group A) or activating (group B) gene profile, and reflect different segment configurations (e.g., cA-tA and cB-tB) [31]. Chromosomal recombination events might shuffle haplotype segments (e.g., cA-tB defined by the large X). In rhesus macaques, the framework is represented by *KIR3DL20* at the centromeric region. The telomeric region is characterized by highly variable gene expansions and contractions. In contrast to the situation for humans, physical maps of rhesus macaque KIR haplotypes are virtually absent from the literature. For that reason, the lower panel represents a hypothetical configuration as the proposed situation at the start of this study.

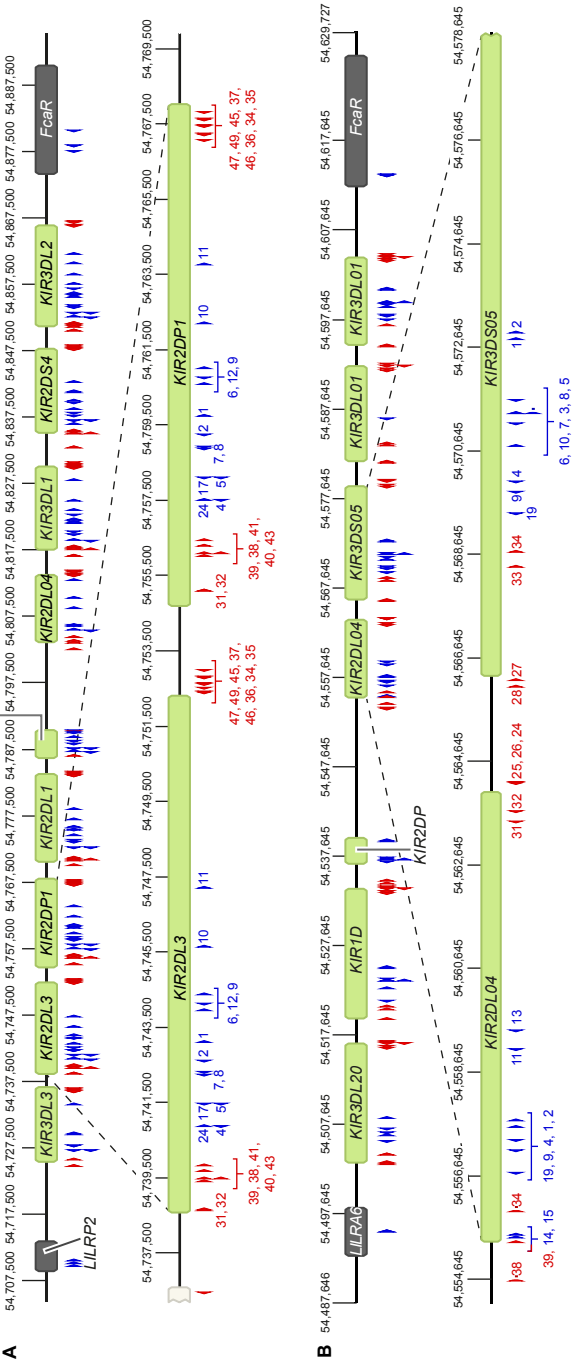


Figure 3. Overview of crRNA sets to target human and rhesus macaque KIR haplotypes. Sets of crRNAs were designed to guide the Cas9 nuclease enzyme to the *KIR* gene cluster in humans (A) and rhesus macaques (B). The cluster location at the reference genomes, HG38 and Mmul_10, is provided. Representative regions are enlarged for a more specific indication of the cut sites. Blue and red arrows reflect crRNAs that excise gene to gene fragments and complete gene fragments, respectively, and are indicated with numbers corresponding to Sup. Tables 1 and 2.

region (**Figure 2**). Only two other *KIR* genes are differentially located within this haplotype segment (*KIR1D* and *KIR2DP*). The only macaque KIR ortholog that is shared with humans, *KIR2DL04*, is present on approximately 80% of the telomeric haplotype segments. This haplotype region is further characterized by a differential number of *KIR* genes in highly diverse frequencies. To enrich the macaque *KIR* cluster, we designed 24 generic crRNAs that target the variable gene tandem, and complemented those with crRNAs specific for *KIR1D*, *KIR2DL04*, and *KIR3DL20*, and the flanking genes, *LILRA6* and *FcAR* (**Figure 3B; Sup. Table 2**).

Assembly of human KIR haplotypes

For this communication, the human target region was defined at the start of exon 5 of *LILRP2* to the end of exon 1 of *FcAR*, thereby comprising the complete *KIR* haplotype (**Figure 2**). On the human reference genome (HG38), this target region spans approximately 161 kb and contains nine *KIR* genes, one of which encodes an activating KIR (group A haplotype). To account for the variability of *KIR* haplotype configurations, four additional reference sequences, which were assembled by Fosmid sequencing and reflected different *KIR* genotypes, were included in our panel (47). In total, five reference *KIR* regions were used to assemble captured reads and to determine sequence accuracy.

For each of the three randomly selected human individuals, two Nanopore flowcells were loaded with HMW gDNA samples that were enriched either for fragments that comprise a complete *KIR* gene or fragments that connect and distinguish neighboring *KIR* genes. An average of 1.26 million Nanopore reads were yielded per flowcell (**Table 1**). Of these reads, 4.2% to 19.3% had a length of 7,000 bp or longer, which is required to connect and distinguish neighboring *KIR* genes. The percentage of size-selected reads that mapped to the target region ranged from 1.6% to 2.5% *(±* 4,015 reads), which provided a median coverage ranging from 269 to 323X. The enrichment factor, which reflects the efficiency of the targeted enrichment, ranged from 215 to 394X. The length of the consensus sequences that were generated from the on-target reads ranged from 6.5 to 29.7 kb, and their assembly covered complete reference *KIR* haplotype configurations (**Figure 4A**, **4B**, and **4C**). The consensus accuracy compared to the reference sequences ranged from 96.7% to 99.9%.

The centromeric and telomeric regions were completely assembled and grouped into four different segment configurations (cA01, cB02, tA01, tB01). Phasing of complete *KIR* haplotypes was achieved in two individuals, including a homozygous haplotype configuration using allele level resolution (**Figure 4A**, **4B** and **4D**). This result demonstrates the resolution power of our approach. For another homozygous setting, *KIR* haplotypes were phased for all genes, except for *KIR2DL1* (**Figure 4C**). This gene was identical on both haplotypes, and defined a SNP desert (25 kb) that hampered complete phasing (**Figure S1**). The largest completely assembled and phased haplotype comprised 11 *KIR* genes and covered a total of approximately 176 kb (**Figure 4B**).

Table 1. Overview of the total reads, read length, on-target hits, and the coverage of the target region in human samples.

	Gene fragments (flowcell 1)		
	#1	**#2**	**#3**
Total # reads	789705	1175600	1413919
Total # reads > 7.000 bp	57830	49535	59575
Percentage reads > 7.000 bp	7.3	4.2	4.2
Average read length (bp)	2603	2114	2092
Average read length (bp) of reads > 7.000 bp	13410	10895	13317
	Gene to gene fragments (flowcell 2)		
	#1	**#2**	**#3**
Total # reads	1060000	1078502	2012903
Total # reads > 7.000 bp	204797	152372	108688
Percentage reads > 7.000 bp (%)	19.3	14.1	5.4
Average read length (bp)	5062	3728	2021
Average read length (bp) of reads > 7.000 bp	17206	13327	13455
	Total reads (flowcells 1 and 2)		
	#1	**#2**	**#3**
On-target reads (> 7.000 bp reads)	4470	3308	4268
Percentage on-target reads	1.7	1.6	2.5
Mean coverage target region	323	269	315
Enrichment factor	215	299	394

Assembly of rhesus macaque KIR haplotypes without reference genome

The successful deciphering of highly variable rhesus macaque *KIR* haplotypes validated our enrichment and characterization approach further, even in the absence of genomic reference sequences. The haplotype content was initially determined by previous transcriptome and segregation studies (35, 36), but the physical location of *KIR* genes remained elusive. The target region was defined at the start of exon 7 in *LILRA6* to the end of exon 2 in *FcAR*, thereby comprising the complete rhesus macaque *KIR* cluster. This region covered 330 kb on the rhesus macaque reference genome (Mmul_10). Ironically, the *KIR* gene cluster is poorly assembled and annotated on this reference genome and is at present not suitable as a reference to assemble *KIR* haplotypes.

The rhesus macaque *KIR* cluster was enriched from samples of three animals. One or two flowcells were used to enrich fragments that comprised *KIR* genes from start to end, or connected neighboring *KIR* genes. A range of 2.5% to 36.3% of the total reads had a read-length of 7,000 bp or longer (**Table 2**). The percentage captured on-target reads ranged from 0.5% to 4.2% (**Table 2**). The mean coverage of the target region was determined by mapping all reads to the reference genome (Mmul_10), which was complemented with the assembled *KIR* haplotypes as artificial chromosomes and reached 26X to 91X. The enrichment factor ranged from 128X to 637X. Consensus

Table 2. Overview of the total reads, read length, on-target hits, and the coverage of the target region in rhesus macaque samples. *For #3, a single flowcell was used for all different crRNA pools, generating fragments containing a *KIR* gene from start to end, and fragments that span from one *KIR* gene to another.

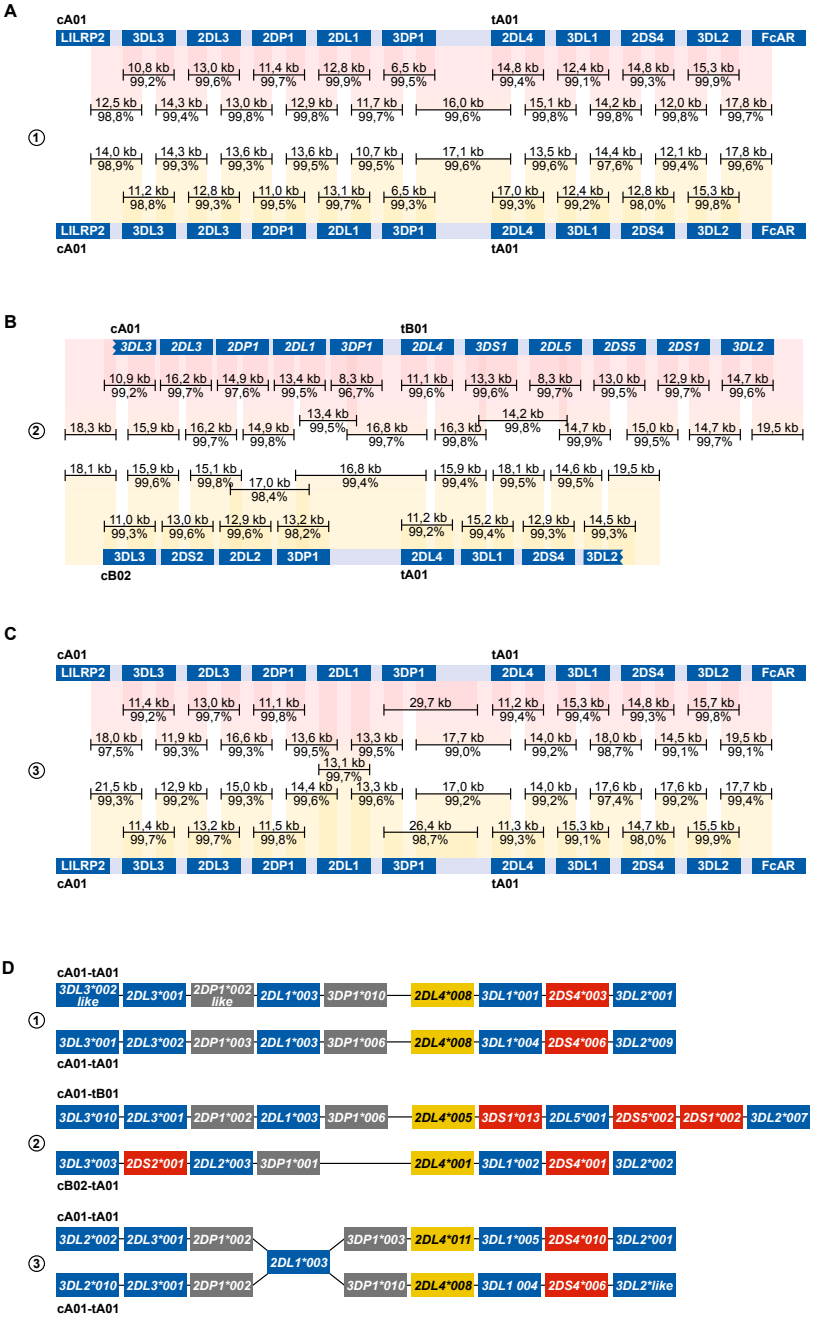
	Gene fragments (flowcell 1)		
	#1	**#2**	**#3**
Total reads	252981	453649	2208000
Total reads > 7.000 bp	25062	54772	55490
Percentage reads > 7.000 bp	9.9	12.1	2.5
Average read length (bp)	3098	3375	1978
Average read length (bp) of reads > 7.000 bp	11258	11875	9319
	Gene to gene fragments (flowcell 2)		
	#1	**#2**	**-**
Total reads	48586	20956	-
Total reads > 7.000 bp	5179	7607	-
Percentage reads > 7.000 bp (%)	10.7	36.3	-
Average read length (bp)	2884	7110	-
Average read length (bp) of reads > 7.000 bp	12226	14789	-
	Total reads (flowcells 1 and 2)		
	#1	**#2**	**#3**
On-target reads (> 7.000 bp reads)	1070	1005	292
Percentage on-target reads	4.3	1.8	0.5
Mean coverage target region	60.8	91.8	26.8
Enrichment factor	637	404	128

sequences were generated, which displayed lengths ranging from 7.2 to 20.0 kb. The accuracy of the consensus sequences is estimated by a comparison with rhesus macaque exon sequences extracted from the non-human KIR Database (IPD-NHKIR, release 1.3.0.0), and reached 96.7% to 100% similarity at these coding regions.

An allele level resolution allowed the phasing of six rhesus macaque *KIR* haplotypes (**Figure 5**). Even with a single flowcell, sufficient coverage was reached to define haplotypes at an allele level resolution (**Figure 5C**). The largest haplotype contained 16 *KIR* genes and spanned 280 kb (H15). The different allelic copies of *KIR3DL01*, *KIR3DL07* and *KIR3DS01* that could be distinguished on this extended haplotype suggests its generation by multiple chromosomal recombination events. The shortest *KIR* haplotype encoded five *KIR* gene members (H10). A fusion gene, which consists of segments from two distinct *KIR* genes, and that are occasionally generated by chromosomal recombination events, was identified on H14 (*KIR3DL20*030R*) (**Figure 5A**). These recombined entities often remain undetected by current genotyping approaches.

Methylation profile of the multigenic KIR region

Amplification-free enrichment and nanopore sequencing allow the characterization of DNA modification profiles. *KIR* genes display a variegated expression on NK cells and subsets of T cells, which is tightly regulated by methylation of the promotor region


A
cA01
tA01
LILRP2 3DL3 2DL3 2DP1 2DL1 3DP1 2DL4 3DL1 2DS4 3DL2 FcAR
10,8 kb 99,2% 13,0 kb 99,6% 11,4 kb 99,7% 12,8 kb 99,9% 6,5 kb 99,5% 14,8 kb 99,4% 12,4 kb 99,1% 14,8 kb 99,3% 15,3 kb 99,9%
12,5 kb 98,8% 14,3 kb 99,4% 13,0 kb 99,8% 12,9 kb 99,8% 11,7 kb 99,7% 16,0 kb 99,6% 15,1 kb 99,8% 14,2 kb 99,8% 12,0 kb 99,8% 17,8 kb 99,7%
①
14,0 kb 98,9% 14,3 kb 99,3% 13,6 kb 99,3% 13,6 kb 99,5% 10,7 kb 99,5% 17,1 kb 99,6% 13,5 kb 99,6% 14,4 kb 97,6% 12,1 kb 99,4% 17,8 kb 99,6%
11,2 kb 98,8% 12,8 kb 99,3% 11,0 kb 99,5% 13,1 kb 99,7% 6,5 kb 99,3% 17,0 kb 99,3% 12,4 kb 99,2% 12,8 kb 98,0% 15,3 kb 99,8%
LILRP2 3DL3 2DL3 2DP1 2DL1 3DP1 2DL4 3DL1 2DS4 3DL2 FcAR
cA01
tA01
B
cA01
tB01
3DL3 2DL3 2DP1 2DL1 3DP1 2DL4 3DS1 2DL5 2DS5 2DS1 3DL2
10,9 kb 99,2% 16,2 kb 99,7% 14,9 kb 97,6% 13,4 kb 99,5% 8,3 kb 96,7% 11,1 kb 99,6% 13,3 kb 99,6% 8,3 kb 99,7% 13,0 kb 99,5% 12,9 kb 99,7% 14,7 kb 99,6%
②
18,3 kb 15,9 kb 16,2 kb 99,7% 14,9 kb 99,8% 13,4 kb 99,5% 16,8 kb 99,7% 16,3 kb 99,8% 14,2 kb 99,8% 14,7 kb 99,9% 15,0 kb 99,5% 14,7 kb 99,7% 19,5 kb
18,1 kb 15,9 kb 99,6% 15,1 kb 99,8% 17,0 kb 98,4% 16,8 kb 99,4% 15,9 kb 99,4% 18,1 kb 99,5% 14,6 kb 99,5% 19,5 kb
11,0 kb 99,3% 13,0 kb 99,6% 12,9 kb 99,6% 13,2 kb 98,2% 11,2 kb 99,2% 15,2 kb 99,4% 12,9 kb 99,3% 14,5 kb 99,3%
3DL3 2DS2 2DL2 3DP1 2DL4 3DL1 2DS4 3DL2
cB02
tA01
C
cA01
tA01
LILRP2 3DL3 2DL3 2DP1 2DL1 3DP1 2DL4 3DL1 2DS4 3DL2 FcAR
11,4 kb 99,2% 13,0 kb 99,7% 11,1 kb 99,8% 29,7 kb 11,2 kb 99,4% 15,3 kb 99,4% 14,8 kb 99,3% 15,7 kb 99,8%
18,0 kb 97,5% 11,9 kb 99,3% 16,6 kb 99,3% 13,6 kb 99,5% 13,3 kb 99,5% 17,7 kb 99,0% 14,0 kb 99,2% 18,0 kb 98,7% 14,5 kb 99,1% 19,5 kb 99,1%
③
13,1 kb 99,7%
21,5 kb 99,3% 12,9 kb 99,2% 15,0 kb 99,3% 14,4 kb 99,6% 13,3 kb 99,6% 17,0 kb 99,2% 14,0 kb 99,2% 17,6 kb 97,4% 17,6 kb 99,2% 17,7 kb 99,4%
11,4 kb 99,7% 13,2 kb 99,7% 11,5 kb 99,8% 26,4 kb 98,7% 11,3 kb 99,3% 15,3 kb 99,1% 14,7 kb 98,0% 15,5 kb 99,9%
LILRP2 3DL3 2DL3 2DP1 2DL1 3DP1 2DL4 3DL1 2DS4 3DL2 FcAR
cA01
tA01
D
cA01-tA01
3DL3*002 like 2DL3*001 2DP1*002 like 2DL1*003 3DP1*010 2DL4*008 3DL1*001 2DS4*003 3DL2*001
①
3DL3*001 2DL3*002 2DP1*003 2DL1*003 3DP1*006 2DL4*008 3DL1*004 2DS4*006 3DL2*009
cA01-tA01
cA01-tB01
3DL3*010 2DL3*001 2DP1*002 2DL1*003 3DP1*006 2DL4*005 3DS1*013 2DL5*001 2DS5*002 2DS1*002 3DL2*007
②
3DL3*003 2DS2*001 2DL2*003 3DP1*001 2DL4*001 3DL1*002 2DS4*001 3DL2*002
cB02-tA01
cA01-tA01
3DL2*002 2DL3*001 2DP1*002
2DL1*003
3DP1*003 2DL4*011 3DL1*005 2DS4*010 3DL2*001
③
3DL2*010 2DL3*001 2DP1*002
3DP1*010 2DL4*008 3DL1 004 2DS4*006 3DL2*like
cA01-tA01

(34). Therefore, we have investigated whether we could determine methylation signatures for the *KIR* gene region. The modification likelihood (low, blue; high, red) was defined for all positions on individual reads that mapped to the *KIR* cluster (**Figure A** and **B**, top part). Based on up to 100 reads, the frequency of modification was determined for different *KIR* gene promotor regions (**Figure 6A** and **B**, black line at the bottom part). A high modification likelihood and frequency was determined for the promotor region of all *KIR* genes in the human samples (**Figure 6A**). The modification frequency of the complete intergenic region ranged from 62% to 93% and seems to slightly increase in the proximal promotor adjacent to exon 1. The intergenic regions in the rhesus macaque *KIR* cluster also displayed high likelihood and frequency of modification (**Figure 6B**). However, the slight increase of modification frequency at the proximal promotor is not observed. Because all DNA samples were isolated from peripheral blood lymphocytes, the high methylation frequency is in line with scare expression of KIR on most prominent lymphocytes. The true resolution power of this approach would emerge using isolated cell populations or single-cell clones.

Discussion

Most multigenic regions are subjected to complex evolutionary processes, in which species-specific gene duplications, deletions, and recombination events shape a genetic cluster. Immune gene families, such as KIR and MHC, diversified under selective pressure by a continuous arms race with pathogens, and individuals might benefit from extensive variation at the population level (48)**.** The high gene content diversity and sequence similarity challenges rapid genomic characterization of these multigenic families.

We make available a rapid enrichment and characterization approach for complex immune regions, which overcomes many of the limitations of short-read or whole-genome sequencing. The Oxford Nanopore MinION sequencing device and flow cells are relatively inexpensive. This makes the descirbed method widely accessible to

◀ **Figure 4. Enrichment and phasing of human KIR haplotypes.** Human donors 1, 2, and 3 were randomly selected, and displayed one heterozygous (#2) and two homozygous (#1, #3) *KIR* haplotype configurations. Consensus sequences (black bars) that covered KIR genes from start to end, or that comprised segments from neighbouring *KIR* genes, were mapped to corresponding haplotype references derived from the human reference genome (HG38) or from previous Fosmid sequencing studies (Roe et al. 2017). The consensus sequence length and accuracy are indicated. Colored columns indicate the mapped regions, in which a darker color illustrates overlapping consensus sequences. KIR haplotypes were completely phased in two individuals (A, B). The third individual shared an identical *KIR2DL1* gene on both haplotypes, which located an extended SNP desert (25 kb; Figure S2) (C). *KIR* haplotypes of this individual were phased at all genes at an allele level resolution, except for the *KIR2DL1* gene. Four differential *KIR* region configurations from three randomly selected donors were resolved. Based on allelic polymorphism, each of these haplotypes appears to be unique (D). In addition, three new variants of documented alleles were defined, which are indicated with "like".

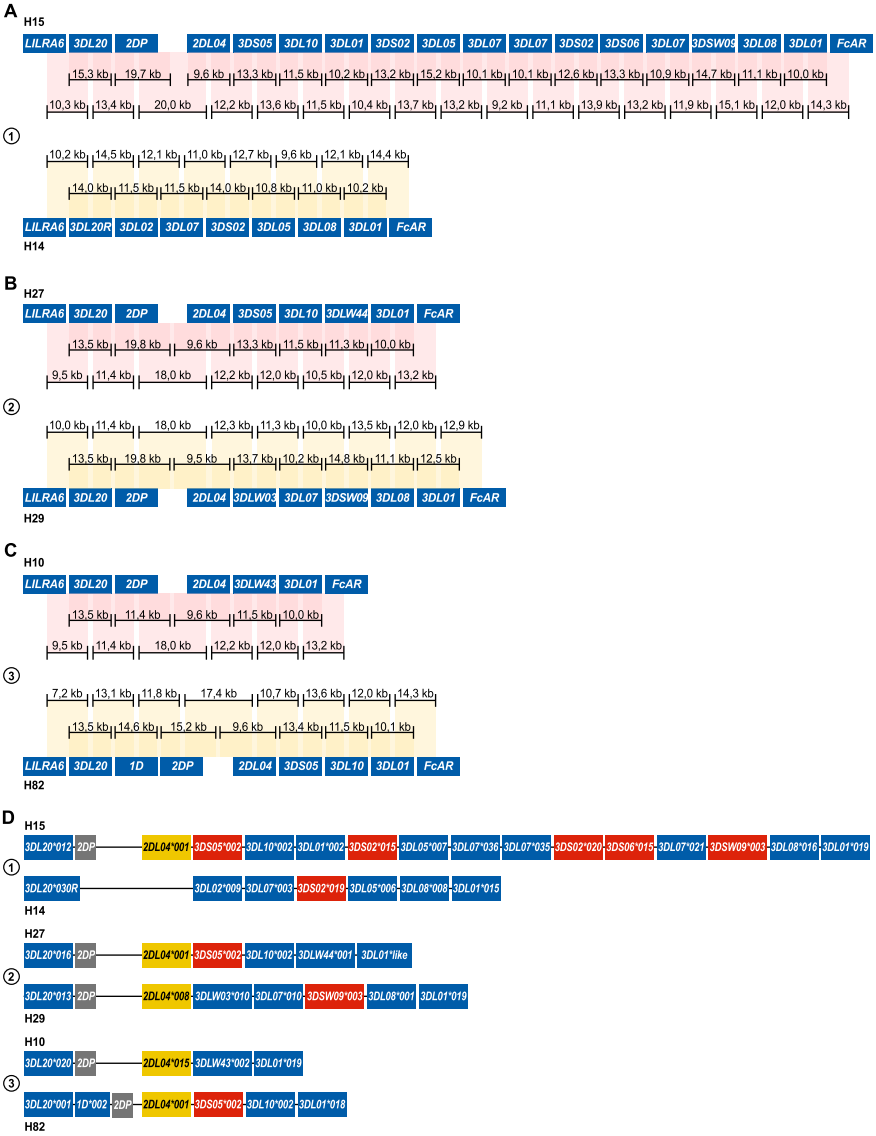


Figure 5. Enrichment and phasing of rhesus macaque *KIR* haplotypes. For rhesus macaque samples (#1, #2, and #3), the *KIR* gene content was determined by a combination of transcriptome and segregation studies [32-34]. Our approach resulted in an accurate annotation of the complex *KIR* regions present in these animals. Consensus sequences (black bars) that covered *KIR* genes from exon 1 to exon 9, or that comprised segments from neighboring *KIR* genes, were assembled. The physical order of *KIR* genes at the haplotype was determined by the extensive overlaps and the alignment with exon sequences from the IPD-NHKIR. The consensus sequence lengths are indicated. Phasing was achieved for all haplotypes, and reflected six different configurations (A, B, C). For each sample, haplotypes were sorted out at an allele level resolution (D). One new allele was defined for *KIR3DL01* (H27), which is indicated with "like".

laboratories without the need for dedicated sequencing facilities, which contrasts other long-read sequencing platforms, such as PacBio sequencing. Extensive diversity in gene content of multigene families was resolved by targeting long overlapping genomic fragments that covered genes from start to end, or that contained segments of neighboring genes. These enriched fragments display a median length of approximately 10 kb with outliers up to 25 kb, which is in line with the expected target fragments. Isolation of longer fragments might be achieved using Cas9-enrichment but was not required for this study design. Efficient targeting was based on using generic crRNAs, which yielded a relatively high median coverage at the region of interest and allowed the generation of accurate consensus sequences. Most inaccuracies involved homopolymers, here defined as a stretch of three or more identical bases, which simultaneously impact the current measured by the nanopore. The consensus sequence accuracy was, however, sufficient to define alleles and to phase complex *KIR* haplotypes. Only at large identical stretches in which overlapping fragments lack allelic variation might this technique hamper the phasing of haplotypes (**Figure S1**). These SNP deserts might be resolved by additional crRNAs that flank the identical stretch. In addition, the continuous improvement of library preparation chemistry, base-calling algorithms, and post-sequencing correction software might eventually increase the length of enriched fragments and the accuracy of nanopore sequencing.

The target coverage was sufficient for the characterization of *KIR* haplotypes at the allele level resolution, but the enrichment performance displayed some deviations in the different samples (**Tables 1** & **2**). The variance in total read numbers and read length is likely to be affected by flowcell variations, gDNA sample quality and short read clean-up efficiency during library preparation. Despite the deviations in read counts, the on-target coverage was sufficient, even when only a single flowcell was used (**Table 2**, #3; **Figure 5C**). Further optimalization of the enrichment protocol, by designing additional crRNAs or by enhancing the Cas9 efficiency with improved chemistry, might provide higher on-target coverage. Samples might then be multiplexed on a single flow cell to further reduce the costs of KIR cluster characterization. Nonetheless, the current approach already ensured highly accurate and overlapping consensus sequences that cost-efficiently resolved complete human and macaque *KIR* haplotypes at an allele level resolution (**Figure 4D** & **5D**).

The KIR receptors display a selective and variegated expression on the cell surface of NK cells and subsets of T cells (49, 50). The majority of these cells only express a single *KIR* gene, whereas smaller fractions might have two to four KIR receptors on their cell surface. This selective expression pattern is most likely regulated by methylation of the promotor region, which is initiated during NK cell differentiation (51). The high modification frequency of the *KIR* promotor regions (**Figure 6**) indicates low levels of expression in the characterized samples (34). Our HMW DNA samples were isolated from whole blood, which includes only a small fraction of KIR-positive cells that might display hypomethylated promotor regions. Nevertheless, the epigenetic modification of members from highly complex multigenic families could be determined using our approach.

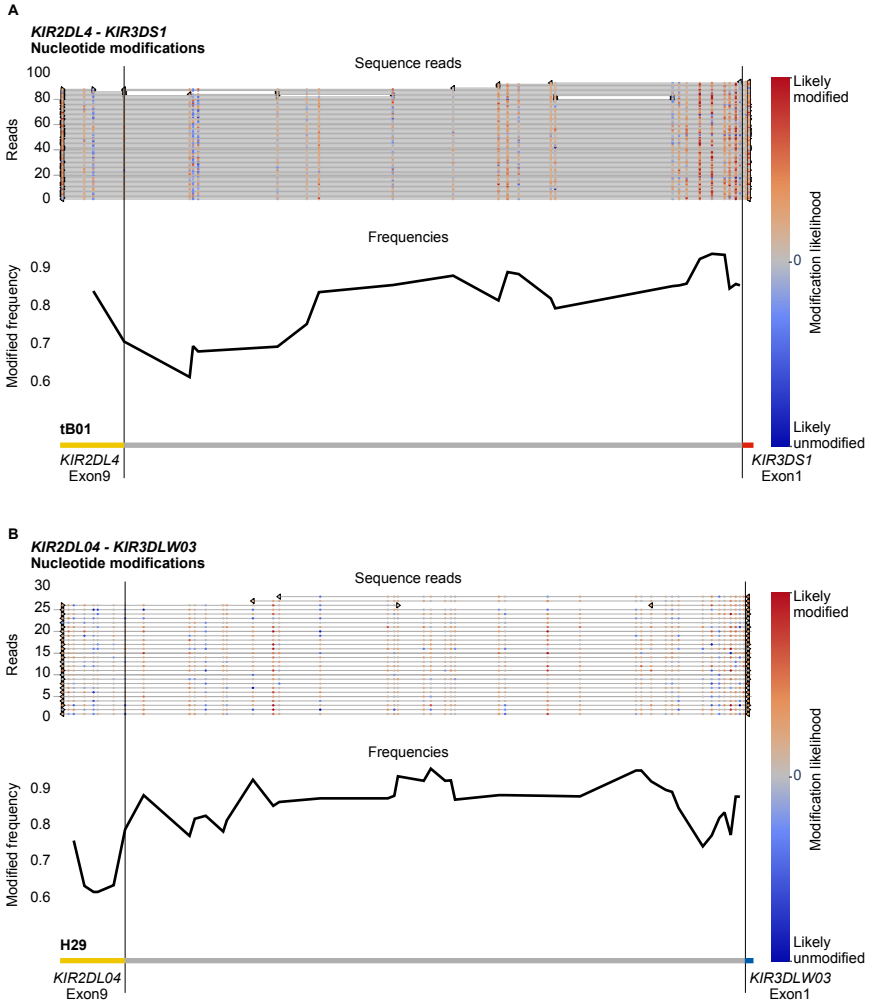


Figure 6. DNA modification profiles. DNA modification profiles are displayed for one human (sample #2) (A) and one rhesus macaque (sample #2) (B) *KIR* gene. The epigenetic predictions were calculated with up to 100 randomly selected reads per enriched *KIR* segment. Indicated are the modification likelihoods (red, high; blue, low), the modification frequency, which ranges from 0 to 1, and the annotation that is based on the (complemented) reference genomes (HG38 and Mmul_10). Plots were generated by Methplotlib (De Coster et al. 2020). The *KIR* promotor region, which is approximately 300 bp in front of exon 1 (Chan et al. 2003), is highly modified for human *KIR3DS1* (A) and macaque *KIR3DLW03* (B), with modification frequencies ranging from 85 to 95%. These two regions are representative for all other *KIR* gene regions studied in our human and rhesus macaque cohorts.

Resolving complete multigenic haplotypes and their epigenetic profiles at the allele level resolution might provide biological and diagnostic insights into the role of these complex systems in health and disease. This is thoroughly demonstrated for different combinations of KIR and HLA that might have an impact on human health, susceptibility or resistance to disease, or may affect graft survival success in transplantation biology (23, 24, 52). There is, however, a lack of consensus from these association studies. One of the causes that might explain these contradicting outcomes is the predominant strategy to characterize common structural motifs, such as centromeric and telomeric KIR haplotype segments, thereby simplifying the plasticity of these immune regions. Using our rapid enrichment and sequencing approach, more comprehensive associations might be defined based on completely defined haplotypes at the allele level resolution.

Members of multigenic families often encode components of essential immune responses and exhibit extensive copy number variation and allelic polymorphism. The continuous diversification and selection of these genes enables adaption to pathogens but might also generate candidates that enhance susceptibility to immune-related disorders. These adaptions might involve SNPs that are mapping in coding regions, whereas others are not, and may impact, for instance, gene (expression) regulation and alternative splicing potential. This enrichment technique fosters a cost-efficient and rapid strategy to characterize complex immune-receptor families at an allele level. Insights regarding complex immune clusters might provide a comprehensive perspective on biological interpretations, and lift SNP disease association studies from the allele to the haplotype/region level, in which all polymorphisms are considered.

Acknowledgements

We thank D. Devine for editing the manuscript and F. van Hassel for preparing the figures.

Data availability

All raw Nanopore files and processed sequencing data generated in this study have been submitted to the European Nucleotide Archive (ENA) (https://www.ebi.ac.uk/ena/browser/home) under accession number PRJEB43311.

References

1 Alkan, C, Cardone, MF, Catacchio, CR, Antonacci, F, O'Brien, SJ, Ryder, OA, et al., Genome-wide characterization of centromeric satellites from multiple mammalian genomes. *Genome research* 21 (2011) 137-45. doi: 10.1101/gr.111278.110

2 Chinwalla, AT, Cook, LL, Delehaunty, KD, Fewell, GA, Fulton, LA, Fulton, RS, et al., Initial sequencing and comparative analysis of the mouse genome. *Nature* 420 (2002) 520-562. doi: 10.1038/nature01262

3 Gibbs, RA, Rogers, J, Katze, MG, Bumgarner, R, Weinstock, GM, Mardis, ER, et al., Evolutionary and biomedical insights from the rhesus macaque genome. *Science* (New York, N.Y.) 316 (2007) 222-34. doi: 10.1126/science.1139247

4 International Human Genome Sequencing, C, Finishing the euchromatic sequence of the human genome. *Nature* 431 (2004) 931-945. doi: 10.1038/nature03001

5 Lindblad-Toh, K, Wade, CM, Mikkelsen, TS, Karlsson, EK, Jaffe, DB, Kamal, M, et al., Genome sequence, comparative analysis and haplotype structure of the domestic dog. *Nature* 438 (2005) 803-819. doi: 10.1038/nature04338

6 Marques-Bonet, T, Ryder, OA, and Eichler, EE, Sequencing primate genomes: what have we learned? *Annual review of genomics and human genetics* 10 (2009) 355-86. doi: 10.1146/annurev.genom.9.081307.164420

7 Waterson, RH, Lander, ES, Wilson, RK, The Chimpanzee, S, and Analysis, C, Initial sequence of the chimpanzee genome and comparison with the human genome. *Nature* 437 (2005) 69-87. doi: 10.1038/nature04072

8 Peona, V, Blom, MPK, Xu, L, Burri, R, Sullivan, S, Bunikis, I, et al., Identifying the causes and consequences of assembly gaps using a multiplatform genome assembly of a bird-of-paradise. *bioRxiv* (2019) 2019.12.19.882399. doi: 10.1101/2019.12.19.882399

9 Thomma, BPHJ, Seidl, MF, Shi-Kunne, X, Cook, DE, Bolton, MD, van Kan, JAL, et al., Mind the gap; seven reasons to close fragmented genome assemblies. *Fungal Genetics and Biology* 90 (2016) 24-30. doi: https://doi.org/10.1016/j.fgb.2015.08.010

10 Guethlein, LA, Norman, PJ, Hilton, HG, and Parham, P, Co-evolution of MHC class I and variable NK cell receptors in placental mammals. *Immunol Rev* 267 (2015) 259-82. doi: 10.1111/imr.12326

11 de Groot, NG, Blokhuis, JH, Otting, N, Doxiadis, GGM, and Bontrop, RE, Co-evolution of the MHC class I and KIR gene families in rhesus macaques: ancestry and plasticity. *Immunological Reviews* 267 (2015) 228-245. doi: https://doi.org/10.1111/imr.12313

12 Eid, J, Fehr, A, Gray, J, Luong, K, Lyle, J, Otto, G, et al., Real-Time DNA Sequencing from Single Polymerase Molecules. *Science* (New York, N.Y.) 323 (2009) 133. doi: 10.1126/science.1162986

13 Jain, M, Fiddes, IT, Miga, KH, Olsen, HE, Paten, B, and Akeson, M, Improved data analysis for the MinION nanopore sequencer. *Nature Methods* 12 (2015) 351-356. doi: 10.1038/nmeth.3290

14 Cai, R, Dong, Y, Fang, M, Guo, C, and Ma, X, De novo genome assembly of a Han Chinese male and genome-wide detection of structural variants using Oxford Nanopore sequencing. *Molecular Genetics and Genomics* 295 (2020) 871-876. doi: 10.1007/s00438-020-01672-y

15 Jain, M, Koren, S, Miga, KH, Quick, J, Rand, AC, Sasani, TA, et al., Nanopore sequencing and assembly of a human genome with ultra-long reads. *Nature Biotechnology* 36 (2018) 338-345. doi: 10.1038/nbt.4060

16 Shafin, K, Pesout, T, Lorig-Roach, R, Haukness, M, Olsen, HE, Bosworth, C, et al., Nanopore sequencing and the Shasta toolkit enable efficient de novo assembly of eleven human genomes. *Nature Biotechnology* (2020). doi: 10.1038/s41587-020-0503-6

17 Fuselli, S, Baptista, RP, Panziera, A, Magi, A, Guglielmi, S, Tonin, R, et al., A new hybrid approach for MHC genotyping: high-throughput NGS and long read MinION nanopore sequencing, with application to the non-model vertebrate Alpine chamois (Rupicapra rupicapra). *Heredity* 121 (2018) 293-303. doi: 10.1038/s41437-018-0070-5

18 Gabrieli, T, Sharim, H, Fridman, D, Arbib, N, Michaeli, Y, and Ebenstein, Y, Selective nanopore sequencing of human BRCA1 by Cas9-assisted targeting of chromosome segments (CATCH). *Nucleic acids research* 46 (2018) e87. doi: 10.1093/nar/gky411

19 Giesselmann, P, Brändl, B, Raimondeau, E, Bowen, R, Rohrandt, C, Tandon, R, et al., Analysis of short tandem repeat expansions and their methylation state with nanopore sequencing. *Nature Biotechnology* 37 (2019) 1478-1481. doi: 10.1038/s41587-019-0293-x

20 Gilpatrick, T, Lee, I, Graham, JE, Raimondeau, E, Bowen, R, Heron, A, et al., Targeted nanopore sequencing with Cas9-guided adapter ligation. *Nature Biotechnology* 38 (2020) 433-438. doi: 10.1038/s41587-020-0407-5

21 Stangl, C, de Blank, S, Renkens, I, Westera, L, Verbeek, T, Valle-Inclan, JE, et al., Partner independent fusion gene detection by multiplexed CRISPR-Cas9 enrichment and long read nanopore sequencing. *Nature Communications* 11 (2020) 2861. doi: 10.1038/s41467-020-16641-7

22 Watson, CM, Crinnion, LA, Hewitt, S, Bates, J, Robinson, R, Carr, IM, et al., Cas9-based enrichment and single-molecule sequencing for precise characterization of genomic duplications. *Laboratory Investigation* 100 (2020) 135-146. doi: 10.1038/s41374-019-0283-0

23 Bashirova, AA, Thomas, R, and Carrington, M, HLA/KIR restraint of HIV: surviving the fittest. *Annu Rev Immunol* 29 (2011) 295-317. doi: 10.1146/annurev-immunol-031210-101332

24 Farag, SS, Bacigalupo, A, Eapen, M, Hurley, C, Dupont, B, Caligiuri, MA, et al., The Effect of KIR Ligand Incompatibility on the Outcome of Unrelated Donor Transplantation: *A Report from the Center for International Blood and Marrow Transplant Research, the European Blood and Marrow Transplant Registry, and the Dutch Registry. Biology of Blood and Marrow Transplantation* 12 (2006) 876-884. doi: https://doi.org/10.1016/j.bbmt.2006.05.007

25 Yu, P, Qi, F, Xu, Y, Li, F, Liu, P, Liu, J, et al., Age-related rhesus macaque models of COVID-19. *Animal Models and Experimental Medicine* 3 (2020) 93-97. doi: https://doi.org/10.1002/ame2.12108

26 Hatziioannou, T, Ambrose, Z, Chung, NPY, Piatak, M, Yuan, F, Trubey, CM, et al., A macaque model of HIV-1 infection. *Proceedings of the National Academy of Sciences* 106 (2009) 4425. doi: 10.1073/pnas.0812587106

27 Estes, JD, Wong, SW, and Brenchley, JM, Nonhuman primate models of human viral infections. *Nature Reviews Immunology* 18 (2018) 390-404. doi: 10.1038/s41577-018-0005-7

28 Bruijnesteijn, J, de Groot, NG, Otting, N, Maccari, G, Guethlein, LA, Robinson, J, et al., Nomenclature report for killer-cell immunoglobulin-like receptors (KIR) in macaque species: new genes/alleles, renaming recombinant entities and IPD-NHKIR updates. *Immunogenetics* 72 (2020) 37-47. doi: 10.1007/s00251-019-01135-8

29 Marsh, SG, Parham, P, Dupont, B, Geraghty, DE, Trowsdale, J, Middleton, D, et al., Killer-cell immunoglobulin-like receptor (KIR) nomenclature report, 2002. *Immunogenetics* 55 (2003) 220-6. doi: 10.1007/s00251-003-0571-z

30 Maccari, G, Robinson, J, Hammond, JA, and Marsh, SGE, The IPD Project: a centralised resource for the study of polymorphism in genes of the immune system. *Immunogenetics* 72 (2020) 49-55. doi: 10.1007/s00251-019-01133-w

31 Guethlein, LA, Older Aguilar, AM, Abi-Rached, L, and Parham, P, Evolution of Killer Cell Ig-Like Receptor (KIR) Genes: Definition of an Orangutan KIR Haplotype Reveals Expansion of Lineage III KIR Associated with the Emergence of MHC-C. *The Journal of Immunology* 179 (2007) 491. doi: 10.4049/jimmunol.179.1.491

32 de Groot, NG, Blokhuis, JH, Otting, N, Doxiadis, GG, and Bontrop, RE, Co-evolution of the MHC class I and KIR gene families in rhesus macaques: ancestry and plasticity. *Immunol Rev* 267 (2015) 228-45. doi: 10.1111/imr.12313

33 Pyo, C-W, Wang, R, Vu, Q, Cereb, N, Yang, SY, Duh, F-M, et al., Recombinant structures expand and contract inter and intragenic diversification at the KIR locus. *BMC Genomics* 14 (2013) 89. doi: 10.1186/1471-2164-14-89

34 Chan, HW, Kurago, ZB, Stewart, CA, Wilson, MJ, Martin, MP, Mace, BE, et al., DNA methylation maintains allele-specific KIR gene expression in human natural killer cells. *J Exp Med* 197 (2003) 245-55. doi: 10.1084/jem.20021127

35 Bruijnesteijn, J, de Groot, N, van der Wiel, MKH, Otting, N, de Vos-Rouweler, AJM, de Groot, NG, et al., Unparalleled Rapid Evolution of KIR Genes in Rhesus and Cynomolgus Macaque Populations. *The Journal of Immunology* (2020) ji1901140. doi: 10.4049/jimmunol.1901140

36 Bruijnesteijn, J, van der Wiel, MKH, Swelsen, WTN, Otting, N, de Vos-Rouweler, AJM, Elferink, D, et al., Human and Rhesus Macaque KIR Haplotypes Defined by Their Transcriptomes. *Journal of immunology* (Baltimore, Md. : 1950) 200 (2018) 1692-1701. doi: 10.4049/jimmunol.1701480

37 Bimber, BN, Moreland, AJ, Wiseman, RW, Hughes, AL, and O'Connor, DH, Complete characterization of killer Ig-like receptor (KIR) haplotypes in Mauritian cynomolgus macaques: novel insights into nonhuman primate KIR gene content and organization. *Journal of immunology* (Baltimore, Md. : 1950) 181 (2008) 6301-8. doi: 10.4049/jimmunol.181.9.6301

38 Dutcher, S, Fulton,R. and Lindsay,T., Macaca mulatta isolate AG07107 chromosome 19, whole genome shotgun sequence, 2018.

39 Sambrook, JG, Bashirova, A, Palmer, S, Sims, S, Trowsdale, J, Abi-Rached, L, et al., Single haplotype analysis demonstrates rapid evolution of the killer immunoglobulin-like receptor (KIR) loci in primates. *Genome research* 15 (2005) 25-35. doi: 10.1101/gr.2381205

40 Warren, WC, Harris, RA, Haukness, M, Fiddes, IT, Murali, SC, Fernandes, J, et al., Sequence diversity analyses of an improved rhesus macaque genome enhance its biomedical utility. *Science* 370 (2020) eabc6617. doi: 10.1126/science.abc6617

41 Benchling, [Biology Software]. Retrieved from https://benchling.com. (2020).

42 Doench, JG, Fusi, N, Sullender, M, Hegde, M, Vaimberg, EW, Donovan, KF, et al., Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nature Biotechnology* 34 (2016) 184-191. doi: 10.1038/nbt.3437

43 Hsu, PD, Scott, DA, Weinstein, JA, Ran, FA, Konermann, S, Agarwala, V, et al., DNA targeting specificity of RNA-guided Cas9 nucleases. *Nature Biotechnology* 31 (2013) 827-832. doi: 10.1038/nbt.2647

44 impson, JT, Workman, RE, Zuzarte, PC, David, M, Dursi, LJ, and Timp, W, Detecting DNA cytosine methylation using nanopore sequencing. *Nature Methods* 14 (2017) 407-410. doi: 10.1038/nmeth.4184

45 De Coster, W, Stovner, EB, and Strazisar, M, Methplotlib: analysis of modified nucleotides from nanopore sequencing. *Bioinformatics* 36 (2020) 3236-3238. doi: 10.1093/bioinformatics/btaa093

46 Uhrberg, M, Valiante, NM, Shum, BP, Shilling, HG, Lienert-Weidenbach, K, Corliss, B, et al., Human diversity in killer cell inhibitory receptor genes. *Immunity* 7 (1997) 753-63. doi: 10.1016/s1074-7613(00)80394-5

47 Roe, D, Vierra-Green, C, Pyo, CW, Eng, K, Hall, R, Kuang, R, et al., Revealing complete complex KIR haplotypes phased by long-read sequencing technology. *Genes & Immunity* 18 (2017) 127-134. doi: 10.1038/gene.2017.10

48 Nei, M, and Rooney, AP, Concerted and birth-and-death evolution of multigene families. *Annu Rev Genet* 39 (2005) 121-52. doi: 10.1146/annurev.genet.39.073003.112240

49 Li, H, Pascal, V, Martin, MP, Carrington, M, and Anderson, SK, Genetic control of variegated KIR gene expression: polymorphisms of the bi-directional KIR3DL1 promoter are associated with distinct frequencies of gene expression. *PLoS Genet* 4 (2008) e1000254. doi: 10.1371/journal.pgen.1000254

50 Björkström, NK, Béziat, V, Cichocki, F, Liu, LL, Levine, J, Larsson, S, et al., CD8 T cells express randomly selected KIRs with distinct specificities compared with NK cells. *Blood* 120 (2012) 3455-65. doi: 10.1182/blood-2012-03-416867

51 Wu, CY, Zhang, B, Kim, H, Anderson, SK, Miller, JS, and Cichocki, F, Ascorbic Acid Promotes KIR Demethylation during Early NK Cell Differentiation. *J Immunol* 205 (2020) 1513-1523. doi: 10.4049/jimmunol.2000212

52 Hiby, SE, Walker, JJ, O'Shaughnessy, KM, Redman, CWG, Carrington, M, Trowsdale, J, et al., Combinations of Maternal KIR and Fetal HLA-C Genes Influence the Risk of Preeclampsia and Reproductive Success. *Journal of Experimental Medicine* 200 (2004) 957-965. doi: 10.1084/jem.20041214

Supplementary materials

Supplementary Table I. Overview of crRNAs to enrich for the human KIR gene region. Benchling software was used to predict potential crRNA sequences for the KIR gene region [43]. The different crRNAs are combined into four different pools. All target sequences were followed by an NGG PAM. The on-target score is determined by the optimized metric of Doench, et al. [44].

#	Pool	Target gene(s)	Cut site	Direction	crRNA sequence	On-target score
Gene to gene fragments: *Generic crRNAs*						
1	1	*KIR2DL/S, KIR3DL/S, KIR2DP1, KIR3DP1*	Intron 3	Reverse	CATAGAAAGATGGAGATGTG	71.1
2	2	*KIR2DL/S, KIR3DL/S, KIR2DP1, KIR3DP1*	Intron 3	Forward	TGTGAGCAAAGTCAGCATGG	73.9
3	2	*KIR2DL/S, KIR3DL/S, KIR2DL4, KIR3DP1*	Intron 4	Reverse	GACAAGGAAGAACCTCCCTG	76.0
4	1, 2	*KIR2DL/S, KIR3DL/S, KIR2DL4, KIR2DP1, KIR3DP1*	Intron 2	Forward	TGTGATTACACTGAGCCCAG	74.7
5	1, 2	*KIR2DL/S, KIR3DL/S, KIR2DP1, KIR3DP1*	Intron 2	Forward	CCTATGTGGATGGAGCCTGG	71.7
6	1, 2	*KIR2DL/S, KIR3DL/S, KIR2DL4, KIR2DP1*	Exon 4	Reverse	CATGTACCATCTATCCAGGG	71.8
7	1, 2	*KIR2DL/S, KIR3DL/S, KIR2DP1, KIR3DP1*	Exon 3	Reverse	GCATCTGTAGGTCCCTGCAA	70.1
8	1, 2	*KIR2DL/S, KIR2DP1*	Intron 3	Reverse	CCTGTGATGACGATGTCCAG	73.6
9	1, 2	*KIR2DL/S, KIR2DP1*	Intron 4	Forward	AGCGGTCAGGAGAGACCCAG	65.8
10	1, 2	*KIR2DL/S, KIR2DP1*	Intron 4	Forward	CAAACTCATGACCTCAACTG	72.8
11	1, 2	*KIR2DL/S, KIR2DP1*	Intron 4	Forward	ACCTTGAAGTCTCAAGACAG	66.4
12	3	*KIR2DL/S, KIR3DL/S, KIR2DL4, KIR2DP1, KIR3DP1*	Exon/Intron 4	Forward	CCTGTGACAGAAACAAGCAG	75.4
13	3	*KIR3DL/S*	Intron 5	Reverse	GGTGAATGGATAGAGAACTG	79.2
14	3	*KIR3DL/S*	Intron 5	Forward	TTCCTACCAACAGGGTACCA	61.0
15	3	*KIR3DL/S*	Intorn 5	Forward	GGAGATACAGATAGATCATG	71.4
Gene to gene fragments: *Specific crRNAs*						
16	1	*KIR3DL3*	Exon 4	Reverse	TGACCTGGGAACCCGCATCG	71.4
17	1	*KIR3DL3*	Intron 3	Reverse	CCTATGTGGATCGAGCCTGG	72.9
18	1	*LILRP2*	Exon 2	Forward	GCATTACACGAACTTCCCCT	69.8
19	1	*LILRP2*	Intron 3	Forward	GGAGGTGTCAGCTCAGAACG	73.6
20	1	*LILRP2*	Exon 3	Forward	CAAATTCACTCTGTACAAGG	71.5
21	1	*FcAR*	Intron 2	Reverse	CCGACAGTATTAGATCATTG	67.1
22	1	*FcAR*	Intron 2	Reverse	GAGGAACTAAGAGAACCAAG	69.3
23	1, 2	*KIR2DL4*	Intron 6	Forward	TAGGCACAACTCCCACACTG	76.0
24	1, 2	*KIR2DL4*	Intron 3	Forward	TGGGCTCAGTGTAATCACAA	74.3
25	1, 2	*KIR3DP1*	Exon 5	Reverse	CCTGCCTTAACCATGGGGCG	58.3
26	1, 2	*KIR3DP1*	Exon 5	Reverse	CTAAGGTCTGACCACTCGTA	59.8
27	1, 3	*KIR2DL4*	Intron 6	Reverse	TCTACACTCAGCTTCCCGAG	74.8
28	1, 3	*KIR3DL2*	Intron 6	Forward	GATAAGAGGCATGAGCCACG	72.8
29	1, 3	*FcAR*	Intron 6	Reverse	ACTGTGCTCAAATTACATCA	72.2
30	1, 3	*FcAR*	Intron 2	Reverse	GACAACCACACAATAATAGG	69.2

Supplementary Table I. Continued.

#	Pool	Target gene(s)	Cut site	Direction	crRNA sequence	On-target score
Gene fragments: *Generic crRNAs*						
31	4	*KIR2DL/S, KIR3DL/S, KIR2DP1, KIR3DP1*	Exon/Intron 1	Forward	TGTGTTGGTGAGTCCTGGAA	67.7
32	4	*KIR2DS, KIR3DL/S*	Intron 1	Forward	TGAGTCCTGGAAGGGAATCG	66.3
33	4	*KIR2DL/S, KIR3DL/S, KIR2DP1, KIR3DP1*	Intron 1	Forward	GTGGAGATATAGGCCTGGAG	63.3
34	4	*KIR2DL/S, KIR3DL/S, KIR2DP1, KIR3DP1*	Exon 9	Reverse	TCAAGTGAAATGGAGAATTG	60.1
35	4	*KIR2DL/S, KIR3DL/S, KIR2DP1*	3' UTR	Reverse	AGGTGGAACAGCATGAGGGA	63.2
36	4	*KIR2DL/S, KIR3DL/S, KIR2DP1, KIR3DP1*	Exon 9	Reverse	TTTCTCTCCAGCAGGCAGTG	66.4
37	4	*KIR3DL3, KIR2DL4*	Exon 9	Reverse	CTCCCTAGAAGATCCCATCA	64.7
38	4	*KIR2DL, KIR2DP1*	Intron 1	Forward	CCAAGACTCACAGCCCAGTG	67.0
39	4	*KIR2DL, KIR2DP1*	Intron 1	Forward	TTGGGTGCAGGTAGGCACTG	63.9
40	4	*KIR2DS, KIR2DP1*	Intron 2	Forward	ACTAGGAAGAGGGGACCCTG	70.7
41	4	*KIR2DL/S*	Intron 1	Forward	CAAGACTCACAGCCCAGTGG	63.4
42	4	*KIR3DL*	Intron 2	Forward	GTCAAGCTCTGTGAAGACTG	72.8
43	4	*KIR3DL, KIR2DL4*	Intron 2	Forward	AGCAGGTCCTCTGAGGACAA	67.9
44	4	*KIR3DL3, KIR2DL4*	Intron 2	Forward	ACCCTCCAGCGTTTCCATGA	62.4
45	4	*KIR2DL*	Exon 9	Reverse	TACACGATGATATCTGTTGG	72.7
46	4	*KIR2DL/S, KIR3DL3, KIR2DP1*	Exon 9	Reverse	GTACATGGGAGCTGGCAACC	63.7
47	4	*KIR2DL/S, KIR3DL/S*	Intron 8	Reverse	GGGAAATGCTGAGTGAGGGA	64.6
48	4	*KIR3DL, KIR2DL4*	Intron 7/Exon 8	Reverse	ACAGCAGCATCTGTAGGAGA	63.7
49	4	*KIR2DL, KIR3DL/S, KIR2DL4, KIR2DP1*	Intron 8/Exon 9	Reverse	TTCATCAGAGTCCTGGAGAG	62.3
50	4	*KIR3DP1*	Intron 8	Reverse	CATTGACCTTGGGCACTGCA	62.9
Gene fragments: *Specific crRNAs*						
51	4	*KIR2DL4*	5' UTR	Forward	AATTCAATCGTCCCGTGCAG	62
52	4	*KIR2DL4*	5' UTR	Forward	TGTGTGTTGAGAAGATCATG	73.2
53	4	*KIR2DL4*	3' UTR	Reverse	GGAAAGGTGGAACAGCACGA	78.7
54	4	*KIR3DL2*	Intron 1	Forward	CTTGGCAGCAGGTAGCAGGG	73.3

Supplementary Table II. Overview of crRNAs to enrich for the rhesus macaque KIR gene region. Benchling software was used to predict potential crRNA sequences for the KIR gene region [43]. The different crRNAs are combined into six different pools. All target sequences were followed by an NGG PAM. The on-target score is determined by the optimized metric of Doench, et al. [44].

#	Pool	Target gene(s)	Cut site	Direction	crRNA sequence	On-target score
Gene to gene fragments: *Generic crRNAs*						
1	1	*KIR2DL04, KIR2DP, KIR3DL20, KIR3DL/S*	Intron 4	Forward	GACAAGGAAGAACCTCCCTG	76
2	1	*KIR1D, KIR2DL04, KIR1D, KIR3DL/S*	Exon 5	Forward	CATGTACCATCTATCCAGGG	72
3	1	*KIR1D, KIR2DP, KIR3DL/S*	Intron 4	Forward	CAGACCAGGTGTCATAACAG	72
4	1	*KIR2DL04, KIR3DL20, KIR3DL/S*	Intron 3	Reverse	GAAAATTATCACAAACATGG	70
5	1	*KIR1D, KIR2DP, KIR3DL/S*	Intron 4	Reverse	CGGCATCTGTAGGTGCCACA	66.5
6	2	*KIR3DL20, KIR3DL/S*	Intron 3	Forward	AGATGGAGGGACCTGCACCA	76
7	2	*KIR1D, KIR2DP, KIR3DL/S*	Intron 5	Forward	ATAACAGAGGACAGACACAG	81.2
8	2	*KIR1D, KIR2DP, KIR3DL/S*	Intron 5	Forward	CAGACCAGGTGTCATAACAG	77
9	1, 2	*KIR2DL04, KIR3DL/S*	Intron 3	Reverse	TCCACCTTCAAAGCCCACAG	72.6
Gene to gene fragments: *Specific crRNAs*						
10	1	*KIR3DL20*	Exon 4	Reverse	CCTATGATCACGATGTCCAG	71
11	2	*KIR2DL04*	Intron 5	Reverse	CCTAAAAAGAGAACTTCCAG	71
12	3	*LILRA6*	Intron 6	Forward	GGGAGGACTGAAATGAGATA	61.5
13	3	*KIR2DL04*	Intron 5	Reverse	AATTAGTACAACTTTCACAG	74
14	4	*KIR2DL04*	Intron 1	Forward	AGATTCGGGGTCTCCCAACA	70.4
15	4	*KIR2DL04*	Intron 1	Forward	AGTGAACGTTCAGCCCAGCG	70
16	4	*FcaR*	Intron 2	Reverse	CCTTTAAATGAAGTGACAGG	66.6
17	1, 2, 4	*FcaR*	Intron 2	Reverse	TGGGCAGTCTAACACAGGCA	67.5
18	1, 2, 4	*FcaR*	Intron 2	Reverse	CTTCATGGGCAGTCTAACAC	62
19	1, 2, 4	*KIR1D*	Exon 3	Reverse	CTGTGACCATGATCGCCAGG	78.3
20	1, 3	*LILRA6*	Intron 6	Forward	AGACTGAGGATAGAAGATGG	72
21	1, 3	*LILRA6*	Intron 6	Forward	AGGACTGAAATGAGATACGG	72

Supplementary Table II. Continued.

#	Pool	Target gene(s)	Cut site	Direction	crRNA sequence	On-target score
Gene fragments: *Generic crRNAs*						
22	5	*KIR3DL20, KIR1D, KIR3DL/S*	3' UTR	Reverse	GGACTAGGAAGTAAGCCAGT	63
23	5	*KIR3DL20, KIR1D, KIR3DL*	3' UTR	Reverse	ACGTTCAGATTCGTGGTGTG	63
24	5	*KIR3DL20, KIR2DL04, KIR3DL/S*	3' UTR	Reverse	CTGAAAAATCGTCTGAGGGG	64
25	5	*KIR3DL20, KIR1D, KIR2DL04, KIR3DL/S*	3' UTR	Reverse	AGGTGGAACAGCATGAGGGA	63.7
26	5	*KIR3DL20, KIR2DL04, KIR3DS*	3' UTR	Reverse	AAGACTGAAAAATCGTCTGA	63.3
27	5	*KIR1D, KIR3DL/S*	5' UTR	Forward	GGTCAGCACGTCAACTGCAT	61.4
28	5	*KIR3DL/S*	5' UTR	Forward	TGGTCATAGTGAAGGACACA	66
29	5	*KIR3DL*	3' UTR	Reverse	ATGTTCAGATTCGTGGTGTG	63
30	6	*KIR3DL20, KIR1D*	3' UTR	Forward	CATCCAAAAATGCAGCACAT	67
31	6	*KIR3DL20, KIR3DL/S*	Intron 7	Reverse	AAGTCATCAAAACAGCTGGG	76.4
32	6	*KIR3DL20 , KIR1D, KIR2DL04, KIR3DL/S*	Exon 9	Reverse	GTGCGTATGTCACCTCCTGA	61.8
33	6	*KIR1D, KIR3DL/S*	Exon 2	Forward	CAGAGGGCCTGTCCACACAC	68.7
34	6	*KIR1D, KIR2DL04, KIR3DL/S*	Intron 2	Forward	TAGGAACAGCAGGTCCTCTG	67.3
35	6	*KIR3DL*	Exon 9	Reverse	TGGAAGTTCCGTGTACACGC	61.6
36	6	*KIR3DL*	Exon 9	Reverse	TTTTTCCCTGTGTGAGAACG	66.7
Gene fragments: *Specific crRNAs*						
37	5	*KIR2DL04*	5' UTR/Exon 1	Forward	CACCAAGTCCATGTCGCCCA	61.1
38	5	*KIR2DL04*	5' UTR	Forward	GCAGTGTTCCGGTATCAAGT	62.8
39	5	*KIR3DL20*	5' UTR	Forward	TGGCAATAGTGAAGGATGCA	62.4
40	5	*KIR3DL20*	5' UTR	Forward	TACGTCATCCTCTCATGATG	61.2
41	6	*KIR2DL04*	5' UTR	Reverse	TACGAACAAACTGAATCCCG	70
42	6	*KIR2DL04*	5' UTR	Reverse	ATGATCTTCTCAATACATGT	62.5
43	6	*KIR3DL20*	Intron 2	Forward	ACATGAGGCAGGCAGCACAA	70.3
44	6	*KIR3DL20*	Intron 2	Forward	GATGATTTTCTTGAAACATG	63.3
45	6	*KIR3DL20*	3' UTR	Forward	TTGAATTGACACAAATGGAG	64.8

Chapter 6

The genetic mechanisms driving diversification of the *KIR* gene cluster in primates

Jesse Bruijnesteijn[1], Natasja G. de Groot[1] and Ronald E. Bontrop[1,2]

[1]Comparative Genetics and Refinement,
Biomedical Primate Research Centre, Rijswijk, The Netherlands
[2]Theoretical Biology and Bioinformatics,
Utrecht University, Utrecht, The Netherlands

Abstract

The activity and function of natural killer (NK) cells are modulated through the interactions of multiple receptor families, of which some recognize MHC class I molecules. The high level of *MHC class I* polymorphism requires their ligands either to interact with conserved epitopes, as is utilized by the NKG2A receptor family, or to co-evolve with the MHC class I allelic variation, which task is taken up by the killer cell immunoglobulin-like receptor (KIR) family. Multiple molecular mechanisms are responsible for the diversification of the *KIR* gene system, and include abundant chromosomal recombination, high mutation rates, alternative splicing, and variegated expression. The combination of these genetic mechanisms generates a compound array of diversity as is reflected by the contraction and expansion of *KIR* haplotypes, frequent birth of fusion genes, allelic polymorphism, structurally distinct isoforms, and variegated expression, which is in contrast to the mainly allelic nature of MHC class I polymorphism in humans. A comparison of the thoroughly studied human and macaque *KIR* gene repertoires demonstrates a similar evolutionarily conserved toolbox, through which selective forces drove and maintained the diversified nature of the *KIR* gene cluster. This hypothesis is further supported by the comparative genetics of *KIR* haplotypes and genes in other primate species. The complex nature of the *KIR* gene system has an impact upon the education, activity, and function of NK cells in coherence with an individual's MHC class I repertoire and pathogenic encounters. Although selection operates on an individual, the continuous diversification of the *KIR* gene system in primates might protect populations against evolving pathogens.

Introduction

The innate and adaptive arms of the immune system are interconnected, and feature several effector functions that provide efficient and specific protection against infection and tumor formation. Major components of the adaptive arm comprise T and B lymphocytes characterized by rearranging antigen receptors, which exert cytotoxic and humoral immunity, respectively. The cytotoxicity mediated by T lymphocytes highly depends on the presentation of intracellular antigen segments derived from pathogens by MHC class I molecules and subsequent clonal expansion of cells with specific receptors. A third type of lymphocytes bridge the innate and adaptive immune response, and comprises natural killer (NK) cells, which participate, for instance, in the recognition and elimination of aberrant cells that downregulate their MHC class I expression to evade detection by T lymphocytes [1]. Without prior priming or clonal expansion, inhibitory and activating receptors on the NK cell surface interact with MHC class I molecules on nucleated cells to modulate NK cell effector functions, which include the killing of target cells by the release of cytolytic proteins and the regulation of other immune cells by the secretion of cytokines [2]. The genes encoding the MHC class I molecules are considered the most polymorphic genes known in vertebrates, a

phenomenon that resulted from selective pressure to adapt to the rapid diversification of pathogens. This extended repertoire of *MHC class I* genes and alleles requires the NK cell receptors to co-evolve to maintain a functional relation with their ligands. The recognition of MHC class I molecules by NK cells involves two receptor families: the conserved CD94:NKG2A receptors and the highly polymorphic and diverse killer cell immunoglobulin-like receptors (KIR). Both receptor families consist of inhibitory and activating members. Their engagement with MHC class I molecules calibrates the responsiveness of NK cells through a continuous educational process, which largely controls subsequent NK cell activity [3, 4]. The KIR receptors are encoded within the Leukocyte Receptor Complex (LRC) on chromosome 19q13.4, and share this genomic region with other structurally similar immune-regulators, such as the leukocyte Ig-like receptors (LILRs) and the leukocyte-associated Ig-like receptors (LAIRs) (**Figure 1**) [5]. Based on different *Alu* elements that can be regarded as a molecular clock, the initial expansion of the primate *KIR* gene cluster is estimated to date back to approximately 31 to 44 million years ago. This process continued, and is currently reflected by extensive gene duplications and point mutations [6]. Different diversifying mechanisms

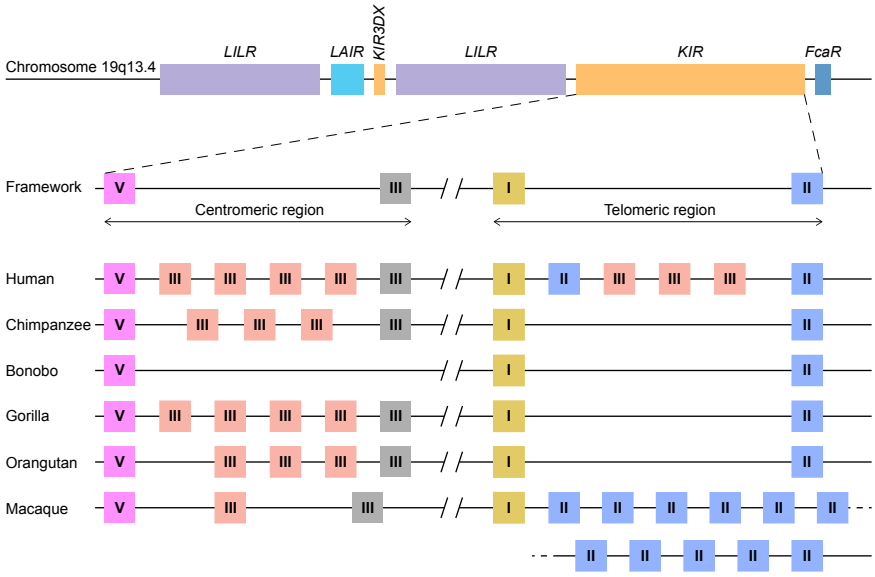


Figure 1. KIR haplotype organizations in different primate species. A schematic overview of the Leukocyte Receptor Complex (LRC) on chromosome 19q13.4 and KIR haplotype organizations in different primate species. A fixed copy of the *KIR3DX* gene is located within the primate *LILR* gene cluster, whereas the expanded *KIR* gene cluster is flanked by the *LILR* and *FcaR* genes. The expansion involved four different *KIR* gene lineages – I, II, III, and V – the members of which are indicated as yellow, blue, red/grey, and pink boxes, respectively. The grey lineage III boxes represent pseudogenes. In most hominoids, KIR haplotype organizations follow a standard framework, in which the centromeric and telomeric regions are bordered by genes from lineages V and III, and lineages I and II, respectively. A relatively large non-coding segment separates the centromeric and telomeric haplotype sections

in combination with evolutionary selective factors propel the complex *KIR* gene content at the individual level but also at the population and species-specific level, which all together contribute to the heterogeneity of NK cell subsets and their activity. The *KIR* gene diversification is not limited to humans. Comparative analyses that include other primate species might help in gaining a thorough understanding of the evolutionary processes that resulted in the diversification of this gene system. In the following sections, we will discuss the different genetic mechanisms that drove the evolution of the highly plastic *KIR* gene system in hominoids (humans and great apes) and Old World monkeys, and how this might influence their NK cell response.

Co-evolution of *MHC* and *KIR* genes

The complex *KIR* gene system requires a comprehensive nomenclature guideline for the different genes and allotypes in order to distinguish the corresponding receptors by their structure and signaling potential [7-9]. Receptors may contain one to three Ig-like domains, which are encoded by exon 3 (D0 domain), exon 4 (D1 domain), and exon 5 (D2 domain), and are referred to as KIR1D, KIR2D, and KIR3D in the official nomenclature. Further classification defines the inhibitory or activating signaling function of the KIR receptors, which is characterized by either a long or short cytoplasmic tail, respectively, and specified with an "L" or an "S" following the domain number denotation. The long cytoplasmic tail contains one or two immune tyrosine-based inhibitory motifs (ITIMs), whereas the signal transduction of activating KIR depends on the interaction with an adaptor molecule that includes an immune tyrosine-based activating motif (ITAM) such as DAP12. Pseudogenes are indicated with a "P" (e.g., *KIR3DP*). In addition, a four-character species designation is included in front of the KIR acronym (e.g., *Mamu-KIR3DL20* in rhesus macaques; *macaca mulatta*).

The mammalian *KIR* genes originate from two progenitor gene lineages: KIR3DX and KIR3DL. The KIR3DX lineage is represented by a single gene copy located in the center of the *LILR* gene cluster (**Figure 1**). The gene is fixed in most primate species, and its function is currently unknown [10]. This lineage is, however, expanded in cattle, and encodes multiple inhibitory and a single activating functional KIR3DX receptor, which interact with an expanded repertoire of classical MHC molecules [11, 12]. In contrast, the KIR3DL lineage expanded in primates and was diversified by duplications, deletions, and recombinations, which resulted in an elaborated *KIR* gene family. Based on their structure, ligand specificity, and/or phylogenetic analysis, the primate KIR receptors are divided into four lineages. Lineage I genes encode receptors with a D0-D2 domain configuration; lineage II is defined by the specificity for subtypes of HLA-A and -B in humans; lineage III includes receptors with D1-D2 and D0-D1-D2 domain configurations; and lineage V is represented by human KIR3DL3 and its orthologs. In the primate species studied, at least one *KIR* gene was discovered for each lineage, which indicates that gene duplication and diversification predates primate speciation. The subsequent lineage expansions are, however, species specific (**Table I**).

Table I. The number of KIR genes defined per primate species indicated per lineage.

		Lineage I			Lineage II			Lineage III			Lineage V			Total			
		I	**A***	**P**	**I**	**A**	**P**	**I**	**A**	**P**	**I**	**A**	**P**	**I**	**A**	**P**	**Total**
Human	*Hosa*	2	1	0	2	1	0	3	5	2	1	0	0	8	7	2	17
Chimpanzee	*Patr*	1	1	0	1	0	0	6	3	0	1	0	0	9	4	0	13
Bonobo	*Papa*	1	1	0	3	1	0	2	0	0	1	0	0	7	2	0	9
Gorilla	*Gogo*	1	1	0	1	0	0	5	1	0	1	0	0	8	2	0	10
Bornean organgutan	*Popy*	1	1	0	1	1	0	2	3	0	1	0	0	5	5	0	10
Sumatran organgutan	*Poab*	1	1	0	1	1	0	3	3	0	1	0	0	6	5	0	11
Rhesus macaque	*Mamu*	0	1	0	31	23	0	1	0	1	1	0	0	33	24	1	58
Cynomolgus macaque	*Mafa*	0	1	0	26	30	0	1	0	1	1	0	0	28	31	1	60

*KIR2DL4 is considered an activating *KIR* gene.
I=Inhibitory, **A**=Activating, **P**=Pseudogene

Lineage I and V *KIR* genes have a conserved nature in all primate species examined, and comprise, respectively, *KIR2DL4* and *KIR2DL5*, and *KIR3DL3*, or a similar structure, such as *Mamu-KIR3DL20* in rhesus macaques. More extensive and species-specific expansions are reported for *KIR* genes that cluster into lineages II and III (**Table I**), and the data suggest that this coincides with the evolution of their MHC class I ligands. Therefore, diversification of the lineage II and III *KIR* genes might be indirectly propelled by the adaption of the MHC class I molecules to pathogenic encounters. For hominoids, this section of co-evolution of KIR and MHC has been comprehensively reviewed by Wroblewski and colleagues [13]. In short, the *MHC* gene content in great apes displays to a limited extent a variable number of *MHC-A*, *-B*, and *-C* genes per haplotype (**Table II**). *MHC-C*, which originated from a duplication of an *MHC-B* gene, is fixed in all hominoids except for orangutans, where it is present on about half of the haplotypes [14]. In addition, the epitopes recognized by the relevant KIR are differentially distributed across the different *MHC class I* genes (**Table II**). The C1 and C2 epitopes, for example, are absent in bonobos and orangutans, respectively, whereas the A3/A11 epitope is only defined on HLA-A molecules. The hominoid MHC class I evolution is accompanied by the reduction and refinement of KIR specific for MHC-A and-B, which is reflected in their limited number of lineage II KIR receptors, whereas the emergence and fixation of MHC-C in humans, chimpanzees, and gorillas drove the expansion and specialization of lineage III KIR (**Table I**) [13].

Old World monkeys, including macaques, lack an MHC-C ortholog, but instead display extensive copy number variation regarding polymorphic *MHC-A* and *-B* genes, as opposed to the fixed number of *MHC class I* genes in hominoids (**Table II**) [15-18]. The expression level of the different MHC-A and -B molecules, however, varies

Table 2. The expansion of *MHC class I* genes in different primate species. Indicated are the number of genes present on a single chromosome and the KIR-recognizing epitopes that may be encoded by allotypes. The frequencies of the different epitopes vary per gene and species. In macaques, on average 1- 3 *MHC-B* genes are highly transcribed (majors), whereas the total number of genes on a single MHC haplotype can reach up to 19 copies, including low transcribed genes (minors) as well as pseudogenes [15-20].

	MHC-A		MHC-B		MHC-C	
	# genes	KIR-epitopes	# genes	KIR-epitopes	# genes	KIR-epitopes
Human	1	A3/A11, Bw4	1	Bw4, C1	1	C1, C2
Chimpanzee	1	-	1	Bw4, C1	1	C1, C2
Bonobo	1	-	1	Bw4, C1	1	C2
Gorilla	1 (2*)	Bw4	1- 2	Bw4, C1	1	C1, C2
Orangutan	1	-	2- 4	Bw4, C1	0-1	C1
Macaque	1- 3	Bw4, Bw6	1- 3 (< 19)	Bw4, Bw6	-	-

*Gorilla's have an additional MHC-A related gene, named Gogo-Oko

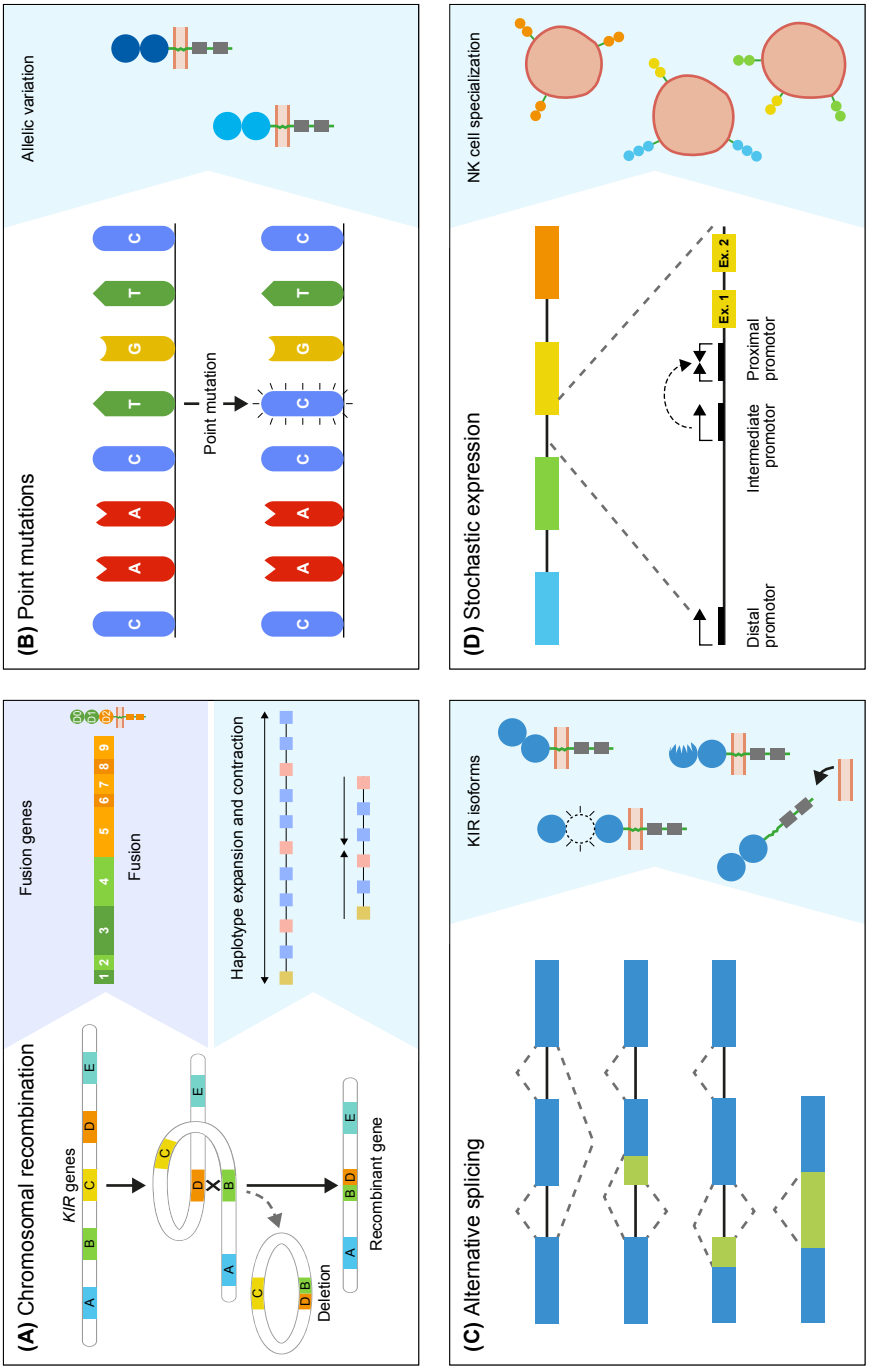
considerably in macaques. It is generally accepted that per haplotype at least a single *MHC-A* and 1 to 3 *MHC-B* genes are characterized by high transcription, and are referred to as "majors", whereas the other *MHC class I* genes have lower transcription levels ("minors"), or may be pseudogenes. The differential transcription suggests a more classical function for the major MHC molecules, such as antigen presentation, whereas the minors might exert more specialized functions [19, 20]. Only a few interactions of macaque MHC and KIR are documented, and, so far, all interactions involved lineage II KIR that recognize Bw4 and Bw6 epitopes on MHC-A and-B allotypes (**Table II**) [21-25]. This putative lineage II specificity for the copious macaque MHC class I repertoire coincides with an extensive ligand expansion, and, thus far, 54 and 56 different lineage II *KIR* genes have been documented for rhesus and cynomolgus macaques, respectively (**Table I**) [7]. Like the majors and minors for the MHC system, the *KIR* genes may display differential expression levels, which are modulated by sequence polymorphisms and by an individual's *MHC class I* repertoire [26-28]. Lineage III *KIR* genes, which encode ligands for MHC-C in hominoids and were subject to expansion, are represented in macaques by a single gene and encodes a receptor with only the D1 extracellular domain (KIR1D). Its presence on 22% and 82% of the rhesus and cynomolgus macaque KIR haplotypes, respectively, suggests a balancing selection for this structurally modified receptor, which might execute a function other than conventional MHC recognition [29].

The maximal expression of six distinct *MHC class I* genes in most hominoids and the specialization of MHC-C as ligand for lineage III KIR is in line with their modest *KIR* gene expansion (Tables I and II). Macaques may harbor over 20 distinct *MHC class I* genes in one individual, of which only a few are dominantly expressed and considered to be majors. The expanded MHC repertoire in macaques probably propelled the extensive expansion and differential expression of their lineage II KIR. The balanced expansion of the *MHC* and *KIR* gene systems in primates indicates co-evolution in order to maintain a functional relation.

Transposable elements facilitate chromosomal recombination

One of the mechanisms responsible for the extensive *KIR* gene diversification in macaques, and to a lesser extent in hominoids, involves chromosomal rearrangements that are accompanied by deletions and recombination, which may result in the generation of fusion genes (**Figure 2A**). This type of gene formation may shuffle the binding and signaling domains of different KIR receptors, thereby functionally altering the response potential of KIR family members. The dense head-to-tail arrangement of the *KIR* genes is likely to facilitate at least in part the chromosomal instability of this gene cluster. A KIR haplotype spans approximately 150 to 350 kb, depending on the number of genes present. Most *KIR* genes are separated by only 2.5 kb, as opposed to the wider haplotype configurations of more stable and less expanded gene families, such as the *LILR* gene cluster [6, 30-33]. In addition, the presence of transposable elements, including Alu and LINE elements, in the intergenic and intragenic KIR sequences is another factor that further promotes genetic instability [6, 34-36]. These repetitive elements are present in all primate *KIR* genes, although with species-specific variation, and drive recombination and genetic deletions [35, 37-39]. For the few completely sequenced fusion *KIR* genes in humans, the chromosomal breakpoints indeed map in the intragenic transposable elements. This supports the idea that the abundant presence of transposons in the *KIR* cluster facilitates chromosome fragility, which is reflected by genetic expansion and contraction, and the formation of fusion genes [34, 40, 41]. A considerable number of human fusion *KIR* genes were generated by reshuffling that involved segments of pseudogenes [34]. The conservation of two pseudogenes in the human KIR repertoire, *KIR2DP1* and *KIR3DP1*, might be explained by their role in promoting recombination events. The human KIR haplotypes that include an apparent fusion gene are represented by relatively low frequencies [42-45]. Positive selection of fusion entities might, however, increase their frequencies in certain populations [45]. Ancient recombination events and subsequent selection might have contributed substantially to the current human KIR repertoire, but the modest expansion of the human *KIR* genes nowadays indicates limited recent recombination events. In contrast, an excessive number of recombination events are recorded in rhesus and cynomolgus macaques, with the presence of at least one fusion *KIR* gene on 42% and 49% of their haplotypes, respectively [29, 43]. The abundant presence of fusion genes indicates that in these species the reshuffling of *KIR* gene segments is an ongoing process that expands the macaque KIR repertoire. Although information on the non-coding regions in the macaque *KIR* cluster is limited at present, the chromosome instability and consequential recombinations in concert with selection are likely to have driven the extensive expansion of lineage II *KIR* genes. This fast mode of evolution is further reflected in the relatively low number of orthologs that are shared between the closely related rhesus and cynomolgus macaques and their populations [29].

In all hominoids and Old World monkeys, the 5' section of the *KIR* gene cluster is occupied by *KIR3DL3* or similar structures, which are considered framework genes and


(A) Chromosomal recombination
KIR genes
A
B
C
D
E
Fusion genes
1 2 3 4 5 6 7 8 9
Fusion
Deletion
Recombinant gene
Haplotype expansion and contraction
(B) Point mutations
Allelic variation
Point mutation
(C) Alternative splicing
KIR isoforms
(D) Stochastic expression
NK cell specialization
Ex. 2
Ex. 1
Proximal promotor
Intermediate promotor
Distal promotor

might carry out essential functions. The structure and evolutionary pathway of these lineage V *KIR* genes is a complex outcome of multiple recombination events [46]. Additional chromosomal rearrangements in rhesus macaques involved the exchange of the cytoplasmic tail of *KIR3DL20* with the tails of *KIR2DL04* (lineage I) and *KIR1D* (lineage III). These recombination events are not conserved in macaque populations, which implies the relatively recent formation of novel gene entities propelled by ancient recombination hotspots [29].

Chromosomal recombination events generate genetic variability in the *KIR* gene cluster by the formation of fusion genes. Subsequent selection of these novel genes might supply an adaptive and protective strategy in the arms race with rapidly evolving pathogens.

KIR haplotype diversity in primate species

Chromosomal rearrangements not only generate novel *KIR* gene entities by recombination but also diversify the haplotype gene content by insertions and deletions of genes (**Figure 2A**). In general, hominoid KIR haplotypes consist of two genomic regions that are bordered by four framework genes (**Figure 1**). The proximal half of the haplotype is termed the centromeric region and is defined by *KIR3DL3* to *KIR3DP1/KIRDP*, whereas the distal part, or telomeric region, ranges from *KIR2DL4* to *KIR3DL2/KIR3DL1*. Within these sections, *KIR* genes of different lineages expanded and contracted during hominoid speciation. In humans, the expansion involved lineage III *KIR* genes in their centromeric and telomeric regions, whereas expansion in chimpanzees, gorillas, and orangutans expansion took place in the same lineage in the centromeric region only (**Figure 1**). The human haplotype content ranges from 7 to 12 *KIR* genes, whereas the number in chimpanzee and orangutan haplotypes stretches from 5 to 11 and 5 to 10 functional *KIR* genes, respectively. In contrast to other hominoids, bonobos are characterized by a contraction of their KIR region, with only 3-7 *KIR* genes expressed on a haplotype. The shortest bonobo KIR haplotype consists

◀ **Figure 2. The genetic mechanisms propelling diversification.** The primate KIR cluster diverged as a result of multiple molecular processes, which together modulate the *KIR* gene content and expression status. (A) The expansion and contraction of KIR haplotypes is mediated by chromosomal recombinations, which can introduce or remove one or multiple *KIR* genes. Occasionally, a recombination event is accompanied by the generation of a fusion gene, which functionally and structurally expands the gene repertoire. (B) The *KIR* genes are further diversified by point mutations in coding and non-coding regions, which generate alleles that encode receptors with different structures, localization, function, and expression. (C) Alternative splicing is another mechanism that has a similar impact on the function and structure of receptors. The blue and green boxes indicate exons and introns. The isoforms are generated by different splice events, which involve alternative splice sites and exon skipping. (D) The differential expression of subsets of KIR receptors on different NK cell clones forms another level of variation that is mediated and maintained by sequence variability in the complex promoter regions and epigenetic modifications. A conjunction of the proximal, intermediate, and distal promoters is required to induce KIR expression.

of only the framework genes [47]. The contracted bonobo KIR cluster coincides with a reduced nucleotide variation in their MHC class I repertoire, which might be caused by a bottleneck or pathogen-driven selective sweep after divergence from the chimpanzee's lineage [48-51]. In contrast, a highly variable *KIR* haplotype content is encountered in the macaque, with 4 to 17 functional *KIR* genes that mainly map to the telomeric region (**Figure 1**). The haplotype framework in macaques is less fixed than in hominoids, with only *KIR3DL20* expressed on all haplotypes, whereas *KIR2DL04* is present on 70% of the rhesus macaque haplotypes. A gene orthologous to hominoid *KIR3DL2*/*KIR3DL1* that usually marks the telomeric region is absent.

The diversified *KIR* haplotypes in hominoids and Old World monkeys stem from a primordial configuration, for which a model has been proposed by Guethlein and colleagues [35]. This model describes abundant duplications and recombination events that eventually formed a conserved haplotype framework in hominoids. The previously mentioned transposable elements are likely propagating these chromosomal rearrangements that continue to mediate the diversification of haplotype configurations. One major hotspot for recombination is mapping in between the centromeric and telomeric regions, which facilitates the reorganization of the different haplotype regions. In addition, KIR haplotypes also display the insertion and deletion of one or multiple *KIR* genes propelled by unequal crossing-over, which is occasionally accompanied by the formation of a fusion gene [29, 40, 42-44]. In humans, these contractions and expansions, which are mediated by double-stranded breaks at intragenic and intergenic repetitive elements, resulted in haplotypes that expressed 3 to 15 *KIR* genes [40]. The short haplotypes do not express all framework genes. For instance, the deletion of *KIR2DL4* is commonly observed on genotypes defined across different populations [52, 53]. Approximately 7% of the human *KIR* haplotypes are showing indications for contraction and expansion [42]. Although the number of completely defined *KIR* haplotypes in other hominoids is low, several rare *KIR* configurations in chimpanzees and orangutans illustrate genetic footprints for insertion and deletion events, which is also occasionally accompanied by the formation of a fusion gene [54, 55]. In macaques, only two completely sequenced haplotypes are available at present, whereas an abundant number of haplotypes are deduced at the transcription level by segregation studies [26, 29, 31, 43, 56, 57]. The presence of multiple highly similar allotypes, encoded by highly similar *KIR* genes, on a single haplotype indicates an expansion by the insertion of one or more genes. Such events were recorded for 47% and 26% of the rhesus and cynomolgus macaque haplotypes, respectively [29]. In contrast, a minimal *KIR* gene content and the presence of a fusion gene often are indicative of a haplotype contraction. An example of a prominent haplotype reduction in rhesus macaques involved the deletion of the complete centromeric region by an intragenic recombination of *KIR3DL20* and *KIR2DL04* [29]. The variable haplotype content and the relatively high number of fusion genes indicate extensive recombination as a mechanism to diversify the macaque *KIR* gene system in a still ongoing process. This phenomenon is observed to a lesser extent for the *KIR* haplotypes in hominoids, where the process seems to have relaxed.

The *KIR* gene allelic repertoire is expanded by point mutations

Another level of variation is displayed by allelic polymorphisms, which is explained to a large extent by the occurrence of single nucleotide polymorphisms (SNP) (**Figure 2B**). These nucleotide variations have a wide-ranging impact, and may modulate the expression level at the cell surface, ligand specificity, interaction strength, and localization of the KIR receptor. Single nucleotide variations in the extracellular D0 and D1 domains of human *KIR2DL2*004* and *KIR3DL1*004*, for example, retain the receptors within the cell, which might be caused by misfolding [58, 59]. Polymorphisms in *KIR2DL3* alleles affect the avidity of the receptor to bind their HLA-C ligands. The low-avidity *KIR2DL3*001* and the high-avidity *KIR2DL3*005* only differ at three nucleotides in their D1 domain, which alters the orientation of their extracellular domains and thereby their binding strength [60]. Although most KIR disease association studies determine the gene content by the presence and absence of *KIR* gene sections, and thereby lack allele-level resolution, several studies demonstrated that the functional differences of *KIR* alleles might also impact health and disease. For example, two *KIR2DL1* alleles in the African KhoeSan population evolved by single nucleotide mutations and are associated with a reduced risk for pregnancy disorders [61]. Other associations demonstrated that the highly expressed *KIR3DL1* alleles are more protective against disease progression in HIV-infected individuals than lower expressed allotypes, except for the intracellularly retained *KIR3DL1*004*, which is low in expression but highly protective [62-64].

A total of 1110 human *KIR* alleles are catalogued in the Immuno Polymorphism Database (IPD-KIR, release 2.9.0), whereas the number of reported alleles for different non-human primate species ranges from 521 *KIR* alleles in rhesus macaques to 5 *KIR* alleles in Bornean orangutans (IPD-NHKIR, release 1.2.0.0). These allele numbers may give a distorted view of the actual levels of polymorphism within a species due to the differential number of individuals studied. The high level of allelic polymorphism appears to be at least comparable in humans and macaques. The thoroughly documented allelic polymorphism in humans and macaques reveals a varying number of alleles per *KIR* gene, with most nucleotide variation exhibited by the framework genes [7, 29, 44]. In addition, a high number of alleles were reported for certain *KIR* genes located on the telomeric haplotype region in humans (*KIR3DL1*, *KIR2DS4*) and the highly frequent inhibitory *KIR* genes in macaques (*KIR3DL01*, *KIR3DL07*). An expansion of the allele numbers for the frequently expressed *KIR* genes might indicate a continuous role in co-evolution with particular pathogens. The less common *KIR* genes, which include mostly activating KIR, vary in gene content rather than allelic polymorphism and therefore seem to execute more specialized functions and/or might be involved in the recognition of conserved ligands and peptides [7, 29, 44].

For humans, *KIR* alleles are also distinguished by SNPs in their introns (IPD-KIR, release 2.9.0) [65], which might impact, for instance, the expression level and post-transcriptional splicing. A total of 353 human *KIR* alleles can only be distinguished from the reference gene based on intronic variations (IPD-KIR, release 2.9.0), and this

number is likely to be underestimated [65]. Sequence data on the non-coding *KIR* gene regions are lacking for non-human primate species, but a similar extent of intronic variations might be feasible and may impact their receptor functionality. However, there are no disease or health associations reported for intronic polymorphisms within the *KIR* genes, but abundant pathological conditions are described for intronic variations in many other genes mapping elsewhere in the genome [66]. For example, a SNP in the human *CYP2D6* gene is linked to a decreased expression of the functional transcript and correlates with a lower metabolic activity [67]. For *HLA-DP*, a single nucleotide variation in the 3' UTR modulates the expression level of different allotypes, which impacts the susceptibility to chronic hepatitis B virus infection [68].

Allele variation is mainly generated by synonymous and non-synonymous point mutations, and only the latter ones will impact the composition of the gene products. In sharp contrast to MHC class I polymorphisms, the allelic nucleotide variations of the *KIR* genes are evenly distributed over the coding regions. The high concentration of CpG islands located in the *KIR* gene cluster might contribute to an elevated mutation rate, as these islands are in general more prone to promote nucleotide transitions [69-71]. In addition, chromosomal rearrangements are known as mutagenic events [69, 72-74]. In particular, the regions that surround genomic insertions and deletions display an increased mutation rate, which might be induced by error-prone DNA replication [69, 75-77]. The abundant recombination that is accompanied by insertions and deletions in the primate KIR cluster is likely to contribute to the extensive allelic KIR variation. Within two and three generations of human and macaque families studied, the birth of novel KIR alleles is described, which might further substantiate the rapid mutation rate in this gene cluster [29, 78]. To our knowledge, such an event has not been recorded for the highly polymorphic *MHC class I* genes.

The variation involving *KIR* genes at the allele level impacts the interactions with their highly polymorphic MHC class I ligands, and demonstrate that point mutations contribute to a diversified *KIR* gene system. The general lack of allele level characterization in the clinic might limit the number of associations reported for *KIR* allele heterogeneity and their functional and disease-related effects. Even intronic variations might impact the KIR receptor expression and function. These few associations highlight the need to further characterize the *KIR* gene content of humans and other primate species at an allele level resolution.

Alternative splicing as a mechanism for structural diversification

The complexity of the primate *KIR* gene cluster is further extended by alternative splicing (**Figure 2C**) [79-83]. This post-transcriptional mechanism can generate multiple messenger RNA (mRNA) transcripts from a single gene, which are translated into different receptor isoforms. Constitutive splicing excludes the intronic sequences from the precursor mRNA (pre-mRNA) and ligates the coding exons. Alternative splicing

deviates from this pattern by the use of alternative splice sites, the skipping of exons, and the retention of introns (**Figure 2C**) [84]. The alternative splice events for human and macaque KIR transcripts are well documented, and demonstrated that both in- and out-of-frame transcripts are generated [79-83]. The out-of-frame transcripts often have an early stop codon, and this results in early truncation of the transcript. Even though these out-of-frame transcripts appear as a redundant side effect of alternative splicing, it might reflect a regulatory pathway to rapidly downregulate receptor expression. The functional impact of the in-frame generated KIR isoforms may be diverse. The skipping of exons generates transcripts that encode modified KIR isoforms, which lack one or two extracellular domains, the stem region, or the transmembrane region. These KIR isoforms probably exhibit differential binding properties or are secreted as soluble receptors (**Figure 2C**) [85]. In-frame splice events that involve alternative splice sites might insert a partial intronic sequence into the transcript or delete a part of a coding exon. Although the functional and structural consequences of these KIR isoforms are harder to predict, they are likely to modify the receptor expression level, cellular localization, and ligand interactions.

Several splice events were frequently recorded or were defined for multiple *KIR* genes, and implicate the existence of conserved splice events that generate structurally and functionally distinct isoforms. For example, exon 4 (coding for the D1 domain) is frequently skipped from KIR3DL20 transcripts in macaques, thereby generating transcripts that encode both the complete receptor and receptors with a D0-D2 domain configuration [43, 57]. This macaque isoform is termed *KIR2DL05*, as it displays an 89.5% sequence similarity with human *KIR2DL5*. Moreover, it demonstrates that alternative splicing expands the macaque KIR repertoire by generating a second two-domain structure (KIR2DL) additional to KIR2DL04. The most frequent KIR splice event in humans involved the skipping of exon 6, which encodes the stem region. Other frequent events included the skipping of exon 5 (D2 domain) and partial deletions in exons 4 and 5. These events result in isoforms that are likely to display altered binding properties, but their exact activity and localization remains elusive. Another common splice event in humans might function as a regulatory switch for expression of the 9A and 10A *KIR2DL4* alleles by restoring or disrupting the open reading frame [79]. Less frequent alternative splicing events were often found to be gene specific, and were mainly out-of-frame events that encoded for truncated receptors. Except for most exon skipping events, only a single splice event was shared between humans and macaques. This event involved a partial deletion of exon 3 (D0 domain) mediated by an alternative 5' splice site [79]. Data on the alternative splicing in other hominoids are lacking, but a similar extent of alternative splicing is likely to diversify their KIR receptors and repertoire.

The splicing of pre-mRNA not only facilitates diversification of the KIR repertoire, but might also compensate for genomic alterations that result in out-of-frame transcripts. The expression of human and macaque lineage III *KIR* genes, for example, requires the constitutive skipping of exon 3 to maintain an open reading frame (ORF). This exon contains a deletion of 5 bp at the genomic DNA level, which would shift the

reading frame that introduces an early stop codon [79, 86]. The constitutive skipping of exon 3 suggests that the expanded repertoire of human KIR2D receptors evolved from a *KIR3D* gene. The absence of a conserved 33 bp sequence in intron 2 of all human and macaque lineage III *KIR* genes might relate to the constitutive exon skipping by, for example, disrupting the spliceosome recognition site [79].

The extensive levels of alternative splicing observed in humans and macaques defines another layer of complexity for the *KIR* gene cluster. This diversifying mechanism generates structurally and functionally distinct receptor isoforms, and might be involved in the regulation of receptor expression levels. Although not all isoforms might be functional, the frequency and consistency of several alternative splicing events suggest that alternative splicing is a rapid mechanism to diversify the KIR content in hominoids and Old World monkeys.

Differential NK cell populations due to variegated *KIR* gene expression

KIR gene plasticity is further reflected by the stochastic expression of a subset of *KIR* genes from the total gene repertoire in individual NK cells (**Figure 2D**). This selective transcriptional activation generates specialized NK cell populations, which express different numbers and combinations of *KIR* genes [87, 88]. The stochastic KIR expression is activated during NK cell maturation, and the transcriptional pattern is maintained by the methylation of silenced *KIR* genes [89, 90]. The different KIR receptor combinations are generated largely at random, but might be shaped by the individual *KIR* gene frequencies and the MHC class I repertoire. Therefore, *KIR* genes that are present on both chromosomes in heterozygous individuals, or genes that are present as two or more allotypes on a single haplotype (e.g., by duplication or gene insertion), could be expressed in a mono- and multi-allelic manner. This may generate NK cell subsets that transcribe two or more allelic copies of a certain *KIR* gene [28]. Divergent expression patterns are documented for human *KIR2DL4*, which is expressed in all NK cells, and for *KIR3DL3*, which is expressed at low levels [91, 92].

The molecular regulation of *KIR* gene expression is well studied in humans, and involves multiple promoter regions in the intergenic sequences that control gene demethylation and transcription [27, 91-97]. The proximal promoter is located directly in front of the first exon of a *KIR* gene and functions as a probabilistic switch (**Figure 2D**). Bi-directional transcription of this promoter generates forward and reverse transcripts that correlate with the activation and suppression of *KIR* gene transcription, respectively. Forward transcripts of a distal promoter are associated with activation of the proximal promoter region and appear to be required for eventual *KIR* gene expression. A third promoter upstream of the proximal promoter, also denoted as the intermediate promoter, modulates the bidirectional transcription of the proximal promoter directly or indirectly by mediating correct splicing of the forward proximal promoter transcripts [27, 95]. In all human *KIR* genes, the promoter regions are highly

conserved, with 91-99.6% sequence similarity. Exceptions are found for the promoters of *KIR2DL4* and *KIR3DL3*, which substantiates their diverged expression profile [95]. Three types of promoter regions are defined for human *KIR2DL5*, which display considerable differences in their nucleotide sequence and transcription factor binding sites. Types I and III control variegated expression, whereas transcripts of *KIR2DL5* alleles that exhibit the type II promoter are undetectable [98, 99]. These type II promoters are probably inactivated by a SNP in their Runt-related transcription factor (RUNX) transcription binding site, which is an important motif in the regulation of gene expression, and is generally conserved in all *KIR* genes [99]. An identical SNP is identified in the proximal promoter of the pseudogene *KIR3DP1*, and might indicate that the inactive type II promoter is swapped to particular *KIR2DL5* alleles by chromosomal recombination [99-101]. Within the *KIR* promoter regions, multiple other transcription factor binding sites are identified, which can vary per *KIR* gene and thereby contribute to differential gene expression. Allelic variations of the different transcription factor binding sites modulate the expression levels of *KIR* alleles [27, 93]. For example, a *KIR2DL1* allele displayed low expression, which was associated with three SNPs in the distal promoter that generated a binding site for the Zinc finger E-box-binding homeobox 1 (ZEB1) protein [27]. This transcription factor is associated with the downregulation of IL2 expression, and might have a similar impact on the expression of this specific *KIR2DL1* allele. Just like the variation in the *KIR* gene introns, the nucleotide polymorphisms in the promoter regions are grossly undervalued, despite the direct impact on the expression of KIR alleles.

The variegated expression pattern of the *KIR* genes defines NK cell subsets, of which several are tissue resident. These NK cell populations might execute specialized functions in particular tissues that could be mediated by specific sets of KIR receptors. For example, the KIR expression profile of NK cells that were derived from the lung, liver, and uterus deviates from the expression pattern observed in peripheral blood NK cells [102-104]. Expression of KIR was also established for subsets of T cells, in particular terminally differentiated CD8+ T cells, of which 30% exhibited KIR expression [105-107]. The majority of these T cells dominantly express a single inhibitory or activating *KIR* gene, which is generally distinct from the *KIR* gene expression pattern on NK cells within the same individual [105]. The expression pattern of NK cells and CD8+ T cells can be erased by *in vitro* treatment with a methylation inhibitor (5-azacytidine), and thereby induce the expression of formerly silenced *KIR* genes [97, 108, 109]. This demonstrates that the stochastic *KIR* gene expression is maintained by methylation in both types of lymphocytes.

The variability in the promoter regions that is mainly generated by point mutations and chromosomal recombination events contributes to the diversification of NK cell subsets by the stochastic methylation of *KIR* genes. The promoter regions and epigenetic regulation of the *KIR* gene cluster in non-human primate species are less well characterized, but their stochastic expression pattern indicates a similar genetic mechanism.

The different characters of diversification in the *KIR* and *MHC* clusters

The expansion of the primate *KIR* cluster was probably initiated by the integration of multiple retroviral elements near or in the founding *KIR* genes. Subsequent duplications were mediated by these transposable elements, and this process had an impact on the expansion of the *KIR* gene repertoire [35]. These recombination events might have enhanced the mutation rate within this genomic region that generated a diverse set of *KIR* alleles, and subsequently some of these were positively selected during evolution. In the case of exons, the point mutations may affect the receptor structure, function, localization, and expression, whereas polymorphisms in the introns may enhance the level of alternative splicing by affecting existing or generating alternative splice sites. In addition, the high level of point mutations caused variation within the promoter regions, and thereby modulated the variegated expression pattern and expression level of KIR receptors. It appears that all the different molecular mechanisms are intertwined and enhanced by each other, which multiplies their diversifying impact on the primate *KIR* gene system.

The *MHC class I* gene family is considered one of the most polymorphic genomic regions in primates, but displays a different nature of diversity as compared to its KIR ligands. In hominoids, the fixed number of *MHC-A*, *-B*, and *-C* genes on a haplotype indicate low levels of recent duplications and chromosomal recombination, which is substantiated by an exceptionally low recombination rate for the *MHC class I* region [110, 111]. Chromosomal rearrangements that are accompanied by the formation of an *MHC class I* fusion gene, as is determined for the *KIR* genes, is to our knowledge not known. In most hominoids, *MHC class I* polymorphism is mainly generated by point mutations in concert with a recombination of small segments. These genetic modifications are especially located in the exons encoding the peptide-binding site, and indicate a rigorous selection for a diverse array of allotypes. The functional impact is reflected in differential peptide presentation [18]. Additional modification of the MHC repertoire is reflected at the transcription level by alternative splicing, which is reported for human and macaque MHC transcripts [112-116]. Considering the high level of allelic polymorphism in the *HLA* genes, which may involve nucleotide substitutions that disrupt existing or generate novel alternative splice sites, one might expect abundant alternative splicing events in their transcripts. However, only a modest level of alternative splicing is demonstrated for several classical and non-classical *HLA class I* alleles, which mainly involved exon skipping that abrogated receptor surface expression [112]. Specific isoforms of the non-classical *HLA-G*, however, are well known and are associated with cancer and inflammatory diseases [117-120]. In contrast, alternative splicing in primate *KIR* was not limited to certain alleles, and also comprised conserved splice events that were common to multiple *KIR* genes and lineages [79]. The classical MHC class I allotypes are constitutively expressed on all nucleated cells, and thereby lack a variegated expression pattern [121, 122]. However, individual MHC allotypes may display a differential expression level, which is

affected by sequence variation, tissue distribution, and pathogenic encounters [122, 123]. In humans, the relative surface expression of HLA-A and-B is approximately ten times higher compared to HLA-C molecules [122, 124]. This suggests that the *HLA-C* gene might slowly shift its main function from classical antigen presentation into the modulation of NK cell responses during infection and reproductive biology. In addition, the expression levels of different HLA-C alleles display variation, in which highly expressed allotypes correlated with a beneficial control of HIV infection [125]. The differential expression pattern is also determined for the expanded MHC class I region in macaques, with only a few highly expressed MHC-A and-B allotypes [126, 127]. The MHC expression levels are, however, not strictly maintained and can be modulated during infection by immune regulators such as interferon and tumor necrosis factor (TNF) [122].

The primate *KIR* and *MHC* gene families are both reflected by great complexity, and seem to co-evolve to maintain a functional relationship. The *MHC class I* diversification mainly involved allelic polymorphism in the exons encoding the peptide binding site and recombination of small segments, which is driven by the arms race with rapidly evolving pathogens. The *KIR* genes, in contrast, are diverged by haplotype expansion and contraction, random point mutations, and the generation of novel fusion genes. The expression and structural variability of the KIR receptors are further modified at the epigenetic and post-transcriptional level, whereas a similar diversification of the MHC class I molecules is limited. The conjunction of different genetic mechanisms generates an extensive plasticity for the primate *KIR* gene cluster, which seems to exceed the diversity of the polymorphic *MHC class I* genes.

CD94:NKG2A- or KIR-dependent education in different primate species

A comparison of the *KIR* gene system in primate species illustrates a variable degree of gene expansion, reflected in the differential expansion of gene lineages (**Figure 1**). This might be largely due to co-evolution with their diverse MHC class I repertoire. The variable extent of expansion, however, is emphasized by the number of functional genes per *KIR* haplotype and by the overall size of the *KIR* gene repertoire documented for a certain primate species. The extremes are represented by the heavily contracted KIR haplotypes in bonobos versus the widely expanded set of *KIR* genes in macaques (**Figure 1**). The flexibility to expand and contract *KIR* haplotypes and repertoires, apparently without compromising sufficient and protective immune responses, might be closely related to the nature of NK cell education in different primate species.

NK cells require self-tolerance and a signal to activate, which are acquired through an educational process. NK cell education involves the recognition of self-MHC class I molecules or the presented peptides by at least one inhibitory NK cell receptor. Alternative educational pathways that are MHC-independent are reported, but their exact contribution to the acquiring of NK cell functions is elusive [128, 129]. The MHC-

dependent education is predominant and can be approached in two ways (**Figure 3**) [13, 130, 131]. One strategy of NK cell education involves the interaction of inhibitory CD94:NKG2A NK cell receptors with the non-polymorphic MHC-E molecules, which are complexed with conserved signal peptides derived from the diversified classical MHC class I molecules [132-134]. One could argue that this approach allows the immune system to scan in a crude way whether total MHC class I expression has been abrogated. In the complementary approach, however, NK cell education is established through interaction of the MHC class I molecules with polymorphic KIR receptors. This seems to reflect a more sophisticated strategy in which the immune system checks at the epitope level for a malfunctioning of MHC class I expression. KIR-dependent NK cell education is mainly conducted through the interactions of inhibitory KIR and MHC

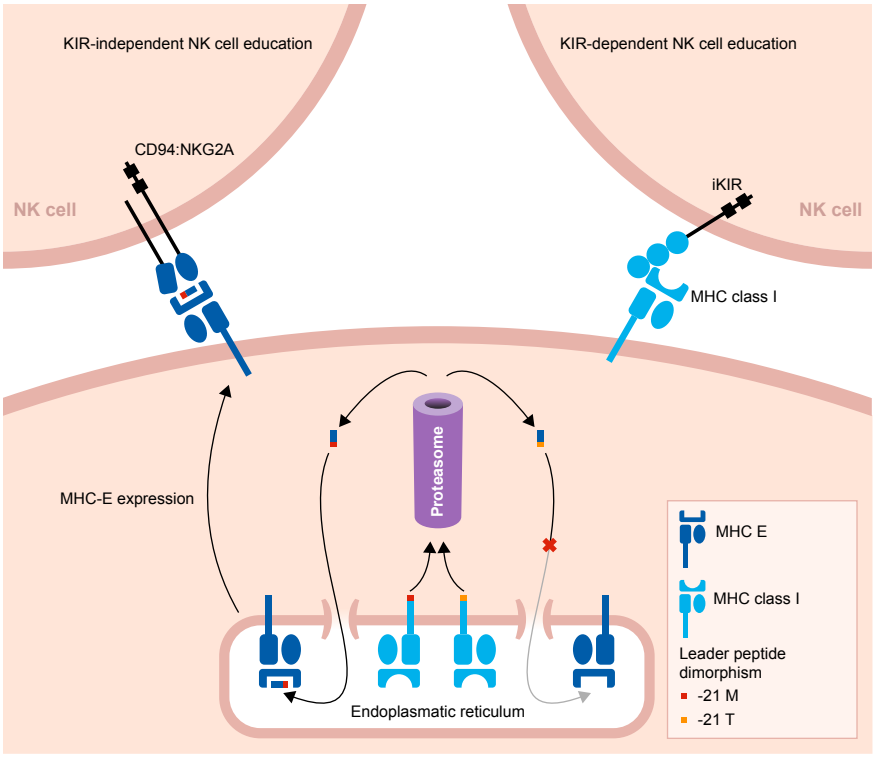


Figure 3. Two pathways to educate NK cells in primates. A schematic overview of two strategies to educate NK cells. The leader peptide of MHC class I molecules either contains a methionine (-21M) or threonine (-21T) residue. The-21M peptides strongly bind to MHC-E molecules, and these complexes display an increased cell surface level. Through the conserved CD94:NKG2A receptors, the MHC-E complexes educate and license NK cells. In contrast,-21T leader peptides, which are predominantly present in MHC-B allotypes of humans and chimpanzees, do not interact with MHC-E molecules. Therefore, in the presence of one or more-21T MHC class I allotypes, the cell surface level of MHC-E is decreased. In this case, NK cells are educated by MHC class I molecules that interact with their KIR ligands.

class I molecules. However, activating KIR contribute to the tuning of NK cell responsiveness by dampening NK cell activity upon MHC class I recognition [135]. Currently, only for KIR2DS1 the effect on NK cell education is described. In the following sections, we mainly consider the educational impact of inhibitory KIR.

Whether the NK cells are educated by the CD94:NKG2A or KIR pathway might depend on a single nucleotide dimorphism at position 21 of the MHC class I leader sequences. Most MHC-A and -C molecules in hominoids have a methionine (-21M) residue present at this position, whereas in general this position is occupied by threonine (-21T) in MHC-B molecules. The -21M peptides strongly bind to MHC-E molecules and promote cell surface expression of MHC-E complexes [136]. The presence of five or six classical MHC class I allotypes containing the-21M residue drives the NK cell education towards the more conserved MHC-E and CD94:NKG2A interactions. However, approximately 62% of human individuals display a -21T HLA-B homozygous genotype, with a variable distribution in different populations [130]. In chimpanzees, -21T is near fixed in their MHC-B allotypes [13]. The homozygous threonine genotype corresponds with a low MHC-E surface expression. As a consequence, human and chimpanzee NK cells are largely educated by their KIR repertoire [13, 130]. In contrast, in macaque MHC-A and-B allotypes, methionine is the predominant residue at position 21 of the leader sequence, which results in an NK cell education that mostly relies on the conserved CD94:NKG2A pathway [130].

In primate species with a KIR-dependent NK cell education, one can envision that an expanded KIR repertoire may compromise NK cell activity. This might drive selection for a limited KIR expansion, as we will discuss in the next section. If this reasoning is true, the KIR-independent education of NK cells in macaques might result in an extensive expansion of their *KIR* gene system. We think that the primary function of macaque KIR is focused on the recognition and elimination of infected or malignant cells. This defense mechanism relies on the recognition of Bw4 and Bw6 epitopes, but KIR interactions are also sensitive to non-self peptides that can be presented by MHC class I molecules [137-141]. A large genetic diversity of *KIR* genes provides a broader repertoire to scan all the variable MHC class I allotypes in combination with their peptides originating from pathogens. It has been proposed that up to seven distinct KIR receptors are required for successful peptide recognition [142]. This optimal receptor count might even be higher when the Bw4 and Bw6 epitope specificity is considered for the different KIR allotypes. The high level of chromosomal recombination and the relatively frequent formation of fusion genes in macaques might indicate selection for a widely diversified *KIR* gene system. Considering their KIR-independent NK cell education, KIR expansion in macaques might be exempted from potential negative selection on large *KIR* gene repertoires.

However, not all macaque *KIR* haplotypes contain a large number of genes, and they even display indications for contraction by chromosomal recombination events. The formation of novel gene entities by the shuffling of head- and tail-encoding exons is achieved by recombination events, which are coherently accompanied both by contractions and expansions of *KIR* haplotypes. There might be a trade-off between

the expansion of the overall KIR repertoire in a population by generating fusion genes and the contraction of KIR haplotypes in individuals. Rapid expansion and diversification generate a highly plastic macaque *KIR* gene system that appears to be maintained by selection to militate against rapidly evolving pathogens.

KIR haplotype expansion and contraction: finding the equilibrium

As compared to macaques, hominoids appear to have a more limited haplotype content and overall KIR repertoire (**Figure 1**, **Table I**). These limitations might be maintained by selective pressure on an efficient KIR-dependent NK cell education, but should be balanced with protection against infections. This balance might be reflected in the slightly variable *KIR* gene content per haplotype.

A large KIR repertoire is likely to provide a broad array of MHC class I specificities that may result in the education of an increased fraction of NK cells (**Figure 4A**). Moreover, the expression of multiple self-specific inhibitory KIR receptors by NK cell clones enhances the magnitude of their effector response [143]. Although only a small population of NK cells dominantly expresses more than one inhibitory KIR receptor, an expanded KIR repertoire might enlarge this NK cell population size and elevate the strength of the NK cell response (**Figure 4A** and **4C**). A potential detrimental effect of an expanded KIR haplotype might emerge if the repertoire comprises only a few or abundant self-specific receptors. On the one hand, the variegated expression of a large KIR repertoire that consists of few self-specific receptors might thin out the educated NK cell population and provide an inefficient immune surveillance (**Figure 4B**). Indications for a biased expression of self-specific KIR suggest modulation of the KIR expression by an individual's MHC class I repertoire [144-146], which would ensure a more robust immune response and might compensate for a large non-self-specific KIR expansion. On the other hand, a large repertoire of self-specific KIR might enlarge the fraction of educated NK cells that display increased activity, which might be protective in infections and cancer (**Figure 4C**). However, elevated NK cell activity, which might be further enhanced by the expression of multiple self-specific KIR on NK cell subsets, or excessive NK cell inhibition by abundant self-specific KIR interactions are also associated with implantation failure and recurrent miscarriages [147-150]. Furthermore, overactivation might desensitize NK cells and result in hyporeactivity [151], which might weaken subsequent immune responses. Therefore, a large KIR repertoire that is used in NK cell education might act as a double-edged sword that can both enhance and compromise an individual's immune response.

In contrast, individuals that have a limited KIR haplotype rely on only one or few self-specific KIR receptors to educate their NK cells (**Figure 4D**). Even though a sufficient percentage of NK cells might be educated by a limited KIR repertoire, the specificity is restricted, and specialized NK cell populations might lack. The complete absence of NK cell education occurs in MHC class I-deficient mice, which display a near normal NK cell

count with an overall reduced responsiveness [152, 153]. In humans and other hominoid species, individuals that completely lack self-specific KIR are not documented. This indicates that even minimal KIR haplotypes provide education, and suggests that framework KIR receptors could play a substantial role in the NK cell education of hominoids. In addition, the chance that an individual completely lacks self-specific KIR receptors is reduced by the heterozygous nature of the *KIR* gene cluster. As far as we know, only few human and no non-human primate individuals are documented that were homozygous for their *KIR* haplotypes at an allele level [154]. In a rhesus macaque family studied, one individual was assumed to be *KIR*-homozygous according to segregation. However, more detailed analysis illustrated that one *KIR* gene copy appeared to have gained point mutations that resulted in the haplotypes diverging at an allele level [29]. This individual macaque possessed a largely homozygous KIR

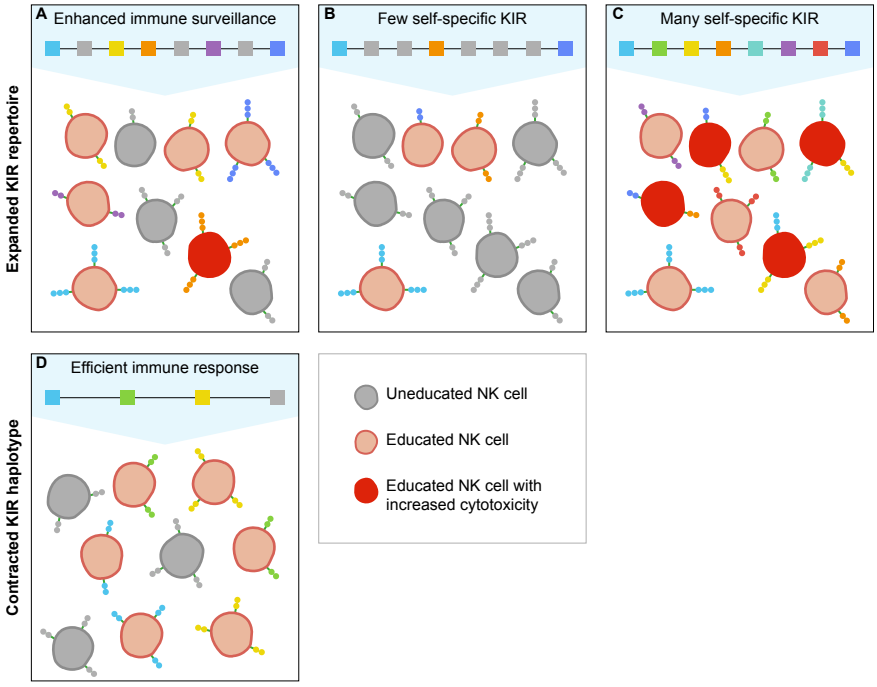


Figure 4. The education of NK cells by expanded and contracted KIR haplotypes. The proposed populations of NK cells that are educated (red cells) by self-specific KIR (colored boxes and receptors) on expanded and contracted haplotypes. The grey boxes represent non-self-specific KIR, and are displayed on the uneducated NK cell clones (grey cells). (A) An expanded KIR haplotype provides a broad MHC class I specificity, and might educate some NK cell clones through multiple self-specific receptors, which increases their cytotoxicity (dark red cells). (B) A large KIR repertoire with only a few self-specific KIR might lower the fraction of educated NK cells and thereby provide an insufficient immune response, (C) whereas abundant self-specific receptors increase the educated fraction and their cytotoxicity. (D) A small KIR repertoire might educate large fractions of NK cells with a limited MHC class I specificity, which might provide largely sufficient immune responses.

content, but did not display an impaired immune system; it also produced healthy offspring, which suggests that *KIR*-heterozygosity is not vital. However, *KIR* haplotype diversity might compensate for limited *KIR* haplotypes and improve the immune surveillance, as is also described for MHC heterozygosity [155-157].

In contrast to non-self-specific T lymphocytes, which are depleted upon a failed positive or negative selection in the thymus, uneducated NK cells are present in the peripheral blood. The relatively high level of uneducated NK cells in individuals with small or large non-self-specific KIR repertoires could affect their immune surveillance, but does not preclude an efficient immune response during infection or tumor formation. In fact, unlicensed NK cells appear to be more efficient at eradicating infected or malignant cells that persistently express MHC class I molecules or viral mimic ligands through their reactivation by cytokines or NKG2D receptors [158-160]. Therefore, a fraction of uneducated NK cells in combination with a largely educated NK cell population might be more protective than a completely educated NK cell pool with broad *MHC class I* specificity.

There could be another factor, however, that limits expansion of the KIR haplotypes and gene repertoire, in addition to their role in NK cell education. In orangutans, MHC-B allotypes contain a-21M leader peptide, which would suggest education via the conserved CD94:NKG2A pathway [130]. In contrast to macaques, the orangutan KIR system is not extensively expanded, and is more in line with other hominoids that display a KIR-dependent NK cell education. The emergence of MHC-C as a specialized ligand for KIR might override the dimorphism and coherent increase in MHC-E expression, and drive NK cell education via the KIR receptors. In addition, the number of characterized MHC-B molecules in orangutans is relatively low (IPD-MHC, release 3.4.0.1) [161]. A larger sample group of orangutans or additional functional studies would be required to test our hypothesis for the differential KIR expansion in primate species that exert a KIR-independent or-dependent NK cell education.

Nevertheless, the diverse *KIR* haplotype content and overall gene repertoire in hominoids and Old World monkeys are likely to affect the education, activity, and function of their NK cells, but the precise effect of the haplotype expansions and contractions remains ambiguous. The equal distribution of both small and large KIR repertoires in humans and macaques indicates a balancing selection, which might be an ongoing process to achieve a haplotype equilibrium that serves differential functions, such as fighting infections and promoting successful pregnancy.

Conclusion

The *KIR* gene system is well studied in humans, and reveals multiple mechanisms that contribute to the plasticity of this immunogenetic cluster (**Figure 2**). In other hominoid species, such as chimpanzees and orangutans, indications for a similar diversifying genetic toolset is evident, although robust data on some mechanisms are lacking, such as alternative splicing and variegated expression. The variability of the extensively diversified *KIR* gene cluster in macaques exceeds that observed in hominoids, with a prominent expansion of the lineage II *KIR* genes, which is largely mediated by recombination events. The rapid evolution of the *KIR* gene cluster may counteract the adaptive nature of pathogens. The species-specific diversification of the *KIR* gene cluster might be largely driven by co-evolution with their diversified MHC class I repertoire and thereby indirectly by the arms race with pathogens. In addition, a KIR-dependent or-independent NK cell education might impact the variable haplotype content and the extent of *KIR* gene expansion. Nevertheless, the different molecular mechanisms responsible for diversification of the *KIR* gene cluster are shared in Old World monkeys and hominoids, which suggests an evolutionary effort to diversify the *KIR* gene system.

Acknowledgements

We thank D. Devine for editing the manuscript and F. van Hassel for preparing the figures.

References

1. Ljunggren HG, Karre K: In search of the 'missing self': MHC molecules and NK cell recognition. *Immunology today* 1990, 11(7):237-244.
2. Vivier E, Tomasello E, Baratin M, Walzer T, Ugolini S: Functions of natural killer cells. *Nature immunology* 2008, 9(5):503-510.
3. Björkström NK, Riese P, Heuts F, Andersson S, Fauriat C, Ivarsson MA, Björklund AT, Flodström-Tullberg M, Michaëlsson J, Rottenberg ME *et al*: Expression patterns of NKG2A, KIR, and CD57 define a process of CD56dim NK-cell differentiation uncoupled from NK-cell education. *Blood* 2010, 116(19):3853-3864.
4. Brodin P, Karre K, Hoglund P: NK cell education: not an on-off switch but a tunable rheostat. *Trends in immunology* 2009, 30(4):143-149.
5. Kelley J, Walter L, Trowsdale J: Comparative genomics of natural killer cell receptor gene clusters. *PLoS genetics* 2005, 1(2):129-139.
6. Martin AM, Freitas EM, Witt CS, Christiansen FT: The genomic organization and evolution of the natural killer immunoglobulin-like receptor (KIR) gene cluster. *Immunogenetics* 2000, 51(4-5):268-280.
7. Bruijnesteijn J, de Groot NG, Otting N, Maccari G, Guethlein LA, Robinson J, Marsh SGE, Walter L, O'Connor DH, Hammond JA *et al*: Nomenclature report for killer-cell immunoglobulin-like receptors (KIR) in macaque species: new genes/alleles, renaming recombinant entities and IPD-NHKIR updates. *Immunogenetics* 2020, 72(1):37-47.
8. Marsh SG, Parham P, Dupont B, Geraghty DE, Trowsdale J, Middleton D, Vilches C, Carrington M, Witt C, Guethlein LA *et al*: Killer-cell immunoglobulin-like receptor (KIR) nomenclature report, 2002. *Immunogenetics* 2003, 55(4):220-226.
9. Robinson J, Guethlein LA, Maccari G, Blokhuis J, Bimber BN, de Groot NG, Sanderson ND, Abi-Rached L, Walter L, Bontrop RE *et al*: Nomenclature for the KIR of non-human species. *Immunogenetics* 2018, 70(9):571-583.
10. Sambrook JG, Bashirova A, Andersen H, Piatak M, Vernikos GS, Coggill P, Lifson JD, Carrington M, Beck S: Identification of the ancestral killer immunoglobulin-like receptor gene in primates. *BMC Genomics* 2006, 7(1):209.
11. Sanderson ND, Norman PJ, Guethlein LA, Ellis SA, Williams C, Breen M, Park SD, Magee DA, Babrzadeh F, Warry A *et al*: Definition of the cattle killer cell Ig-like receptor gene family: comparison with aurochs and human counterparts. *Journal of immunology (Baltimore, Md : 1950)* 2014, 193(12):6016-6030.
12. Guethlein LA, Abi-Rached L, Hammond JA, Parham P: The expanded cattle KIR genes are orthologous to the conserved single-copy KIR3DX1 gene of primates. *Immunogenetics* 2007, 59(6):517-522.
13. Wroblewski EE, Parham P, Guethlein LA: Two to Tango: Co-evolution of Hominid Natural Killer Cell Receptors and MHC. *Frontiers in immunology* 2019, 10:177.
14. de Groot NG, Heijmans CMC, van der Wiel MKH, Blokhuis JH, Mulder A, Guethlein LA, Doxiadis GGM, Claas FHJ, Parham P, Bontrop RE: Complex MHC Class I Gene Transcription Profiles and Their Functional Impact in Orangutans. *The Journal of Immunology* 2016, 196(2):750.
15. Karl JA, Bohn PS, Wiseman RW, Nimityongskul FA, Lank SM, Starrett GJ, O'Connor DH: Major histocompatibility complex class I haplotype diversity in Chinese rhesus macaques. *G3 (Bethesda)* 2013, 3(7):1195-1201.
16. Doxiadis GGM, de Groot N, Otting N, Blokhuis JH, Bontrop RE: Genomic plasticity of the MHC class I A region in rhesus macaques: extensive haplotype diversity at the population level as revealed by microsatellites. *Immunogenetics* 2011, 63(2):73-83.
17. Daza-Vamenta R, Glusman G, Rowen L, Guthrie B, Geraghty DE: Genetic divergence of the rhesus macaque major histocompatibility complex. *Genome Res* 2004, 14(8):1501-1515.
18. de Groot NG, Blokhuis JH, Otting N, Doxiadis GG, Bontrop RE: Co-evolution of the MHC class I and KIR gene families in rhesus macaques: ancestry and plasticity. *Immunological reviews* 2015, 267(1):228-245.
19. Otting N, Heijmans CMC, Noort RC, de Groot NG, Doxiadis GGM, van Rood JJ, Watkins DI, Bontrop RE: Unparalleled complexity of the MHC class I region in rhesus macaques. *Proc Natl Acad Sci U S A* 2005, 102(5):1626.
20. Wiseman RW, Karl JA, Bohn PS, Nimityongskul FA, Starrett GJ, O'Connor DH: Haplessly hoping: macaque major histocompatibility complex made easy. *ILAR J* 2013, 54(2):196-210.

21. Schafer JL, Colantonio AD, Neidermyer WJ, Dudley DM, Connole M, O'Connor DH, Evans DT: KIR3DL01 recognition of Bw4 ligands in the rhesus macaque: maintenance of Bw4 specificity since the divergence of apes and Old World monkeys. *Journal of immunology (Baltimore, Md : 1950)* 2014, 192(4):1907-1917.

22. Parham P, Moffett A: Variable NK cell receptors and their MHC class I ligands in immunity, reproduction and human evolution. *Nature reviews Immunology* 2013, 13(2):133-144.

23. Rosner C, Kruse PH, Hermes M, Otto N, Walter L: Rhesus macaque inhibitory and activating KIR3D interact with Mamu-A-encoded ligands. *Journal of immunology (Baltimore, Md : 1950)* 2011, 186(4):2156-2163.

24. Bimber BN, Evans DT: The killer-cell immunoglobulin-like receptors of macaques. *Immunological reviews* 2015, 267(1):246-258.

25. Hermes M, Weil S, Groth A, Dressel R, Koch J, Walter L: Characterisation of mouse monoclonal antibodies against rhesus macaque killer immunoglobulin-like receptors KIR3D. *Immunogenetics* 2012, 64(11):845-848.

26. Moreland AJ, Guethlein LA, Reeves RK, Broman KW, Johnson RP, Parham P, O'Connor DH, Bimber BN: Characterization of killer immunoglobulin-like receptor genetics and comprehensive genotyping by pyrosequencing in rhesus macaques. *BMC Genomics* 2011, 12:295.

27. Wright PW, Li H, Huehn A, O'Connor GM, Cooley S, Miller JS, Anderson SK: Characterization of a weakly expressed KIR2DL1 variant reveals a novel upstream promoter that controls KIR expression. *Genes and immunity* 2014, 15(7):440-448.

28. Chan HW, Kurago ZB, Stewart CA, Wilson MJ, Martin MP, Mace BE, Carrington M, Trowsdale J, Lutz CT: DNA methylation maintains allele-specific KIR gene expression in human natural killer cells. *The Journal of experimental medicine* 2003, 197(2):245-255.

29. Bruijnesteijn J, de Groot N, van der Wiel MKH, Otting N, de Vos-Rouweler AJM, de Groot NG, Bontrop RE: Unparalleled Rapid Evolution of KIR Genes in Rhesus and Cynomolgus Macaque Populations. *Journal of immunology (Baltimore, Md : 1950)* 2020.

30. Martin AM, Kulski JK, Witt C, Pontarotti P, Christiansen FT: Leukocyte Ig-like receptor complex (LRC) in mice and men. *Trends in immunology* 2002, 23(2):81-88.

31. Sambrook JG, Bashirova A, Palmer S, Sims S, Trowsdale J, Abi-Rached L, Parham P, Carrington M, Beck S: Single haplotype analysis demonstrates rapid evolution of the killer immunoglobulin-like receptor (KIR) loci in primates. *Genome Res* 2005, 15(1):25-35.

32. Wende H, Colonna M, Ziegler A, Volz A: Organization of the leukocyte receptor cluster (LRC) on human chromosome 19q13.4. *Mammalian genome : official journal of the International Mammalian Genome Society* 1999, 10(2):154-160.

33. Trowsdale J, Knight JC: Major histocompatibility complex genomics and human disease. *Annu Rev Genomics Hum Genet* 2013, 14:301-323.

34. Traherne JA, Martin M, Ward R, Ohashi M, Pellett F, Gladman D, Middleton D, Carrington M, Trowsdale J: Mechanisms of copy number variation and hybrid gene formation in the KIR immune gene complex. *Human molecular genetics* 2010, 19(5):737-751.

35. Guethlein LA, Older Aguilar AM, Abi-Rached L, Parham P: Evolution of killer cell Ig-like receptor (KIR) genes: definition of an orangutan KIR haplotype reveals expansion of lineage III KIR associated with the emergence of MHC-C. *Journal of immunology (Baltimore, Md : 1950)* 2007, 179(1):491-504.

36. Wilson MJ, Torkar M, Haude A, Milne S, Jones T, Sheer D, Beck S, Trowsdale J: Plasticity in the organization and sequences of human KIR/ILT gene families. *Proceedings of the National Academy of Sciences* 2000, 97(9):4778.

37. Bailey JA, Liu G, Eichler EE: An Alu transposition model for the origin and expansion of human segmental duplications. *American journal of human genetics* 2003, 73(4):823-834.

38. Sen SK, Han K, Wang J, Lee J, Wang H, Callinan PA, Dyer M, Cordaux R, Liang P, Batzer MA: Human genomic deletions mediated by recombination between Alu elements. *American journal of human genetics* 2006, 79(1):41-53.

39. Han K, Lee J, Meyer TJ, Wang J, Sen SK, Srikanta D, Liang P, Batzer MA: Alu recombination-mediated structural deletions in the chimpanzee genome. *PLoS genetics* 2007, 3(10):1939-1949.

40. Roe D, Vierra-Green C, Pyo CW, Eng K, Hall R, Kuang R, Spellman S, Ranade S, Geraghty DE, Maiers M: Revealing complete complex KIR haplotypes phased by long-read sequencing technology. *Genes and immunity* 2017, 18(3):127-134.

41. Rajalingam R, Gardiner CM, Canavez F, Vilches C, Parham P: Identification of seventeen novel KIR variants: fourteen of them from two non-Caucasian donors. *Tissue antigens* 2001, 57(1):22-31.

42. Pyo C-W, Wang R, Vu Q, Cereb N, Yang SY, Duh F-M, Wolinsky S, Martin MP, Carrington M, Geraghty DE: Recombinant structures expand and contract inter and intragenic diversification at the KIR locus. *BMC Genomics* 2013, 14(1):89.

43. Bruijnesteijn J, van der Wiel MKH, Swelsen WTN, Otting N, de Vos-Rouweler AJM, Elferink D, Doxiadis GG, Claas FHJ, Lardy NM, de Groot NG *et al*: Human and Rhesus Macaque KIR Haplotypes Defined by Their Transcriptomes. *The Journal of Immunology* 2018, 200(5):1692.

44. Vierra-Green C, Roe D, Hou L, Hurley CK, Rajalingam R, Reed E, Lebedeva T, Yu N, Stewart M, Noreen H *et al*: Allele-Level Haplotype Frequencies and Pairwise Linkage Disequilibrium for 14 KIR Loci in 506 European-American Individuals. *PLOS ONE* 2012, 7(11):e47491.

45. Norman PJ, Abi-Rached L, Gendzekhadze K, Hammond JA, Moesta AK, Sharma D, Graef T, McQueen KL, Guethlein LA, Carrington CV *et al*: Meiotic recombination generates rich diversity in NK cell receptor genes, alleles, and haplotypes. *Genome Res* 2009, 19(5):757-769.

46. Leaton LA, Shortt J, Kichula KM, Tao S, Nemat-Gorgani N, Mentzer AJ, Oppenheimer SJ, Deng Z, Hollenbach JA, Gignoux CR *et al*: Conservation, Extensive Heterozygosity, and Convergence of Signaling Potential All Indicate a Critical Role for KIR3DL3 in Higher Primates. *Frontiers in immunology* 2019, 10(24).

47. Rajalingam R, Hong M, Adams EJ, Shum BP, Guethlein LA, Parham P: Short KIR haplotypes in pygmy chimpanzee (Bonobo) resemble the conserved framework of diverse human KIR haplotypes. *The Journal of experimental medicine* 2001, 193(1):135-146.

48. de Groot NG, Heijmans CMC, Helsen P, Otting N, Pereboom Z, Stevens JMG, Bontrop RE: Limited MHC class I intron 2 repertoire variation in bonobos. *Immunogenetics* 2017, 69(10):677-688.

49. Maibach V, Vigilant L: Reduced bonobo MHC class I diversity predicts a reduced viral peptide binding ability compared to chimpanzees. *BMC Evolutionary Biology* 2019, 19(1):14.

50. Wroblewski EE, Guethlein LA, Norman PJ, Li Y, Shaw CM, Han AS, Ndjango J-BN, Ahuka-Mundeke S, Georgiev AV, Peeters M *et al*: Bonobos Maintain Immune System Diversity with Three Functional Types of MHC-B. *The Journal of Immunology* 2017, 198(9):3480.

51. de Groot NG, Stevens JMG, Bontrop RE: Does the MHC Confer Protection against Malaria in Bonobos? *Trends in immunology* 2018, 39(10):768-771.

52. Gonzalez-Galarza FF, McCabe A, Santos E, Jones J, Takeshita L, Ortega-Rivera ND, Cid-Pavon GMD, Ramsbottom K, Ghattaoraya G, Alfirevic A *et al*: Allele frequency net database (AFND) 2020 update: gold-standard data classification, open access genotype data and new query tools. *Nucleic Acids Res* 2020, 48(D1):D783-d788.

53. Nakimuli A, Chazara O, Farrell L, Hiby SE, Tukwasibwe S, Knee O, Jayaraman J, Traherne JA, Elliott AM, Kaleebu P *et al*: Killer cell immunoglobulin-like receptor (KIR) genes and their HLA-C ligands in a Ugandan population. *Immunogenetics* 2013, 65(11):765-775.

54. Abi-Rached L, Moesta AK, Rajalingam R, Guethlein LA, Parham P: Human-specific evolution and adaptation led to major qualitative differences in the variable receptors of human and chimpanzee natural killer cells. *PLoS genetics* 2010, 6(11):e1001192.

55. Guethlein LA, Norman PJ, Heijmans CMC, de Groot NG, Hilton HG, Babrzadeh F, Abi-Rached L, Bontrop RE, Parham P: Two Orangutan Species Have Evolved Different KIR Alleles and Haplotypes. *Journal of immunology (Baltimore, Md : 1950)* 2017, 198(8):3157-3169.

56. Macaca mulatta isolate AG07107 chromosome 19, whole genome shotgun sequence [https://www.ncbi.nlm.nih.gov/nuccore/CM014354]

57. Blokhuis JH, van der Wiel MK, Doxiadis GGM, Bontrop RE: The mosaic of KIR haplotypes in rhesus macaques. *Immunogenetics* 2010, 62(5):295-306.

58. VandenBussche CJ, Dakshanamurthy S, Posch PE, Hurley CK: A single polymorphism disrupts the killer Ig-like receptor 2DL2/2DL3 D1 domain. *Journal of immunology (Baltimore, Md : 1950)* 2006, 177(8):5347-5357.

59. Pando MJ, Gardiner CM, Gleimer M, McQueen KL, Parham P: The protein made from a common allele of KIR3DL1 (3DL1*004) is poorly expressed at cell surfaces due to substitution at positions 86 in Ig domain 0 and 182 in Ig domain 1. *Journal of immunology (Baltimore, Md : 1950)* 2003, 171(12):6640-6649.

60. Frazier WR, Steiner N, Hou L, Dakshanamurthy S, Hurley CK: Allelic variation in KIR2DL3 generates a KIR2DL2-like receptor with increased binding to its HLA-C ligand. *Journal of immunology (Baltimore, Md : 1950)* 2013, 190(12):6198-6208.

61. Hilton HG, Norman PJ, Nemat-Gorgani N, Goyos A, Hollenbach JA, Henn BM, Gignoux CR, Guethlein LA, Parham P: Loss and Gain of Natural Killer Cell Receptor Function in an African Hunter-Gatherer Population. *PLoS genetics* 2015, 11(8):e1005439.

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. *The Journal of experimental medicine* 2006, 203(3):633-645.

63. Martin MP, Qi Y, Gao X, Yamada E, Martin JN, Pereyra F, Colombo S, Brown EE, Shupert WL, Phair J *et al*: Innate partnership of HLA-B and KIR3DL1 subtypes against HIV-1. *Nature genetics* 2007, 39(6):733-740.

64. Gardiner CM, Guethlein LA, Shilling HG, Pando M, Carr WH, Rajalingam R, Vilches C, Parham P: Different NK cell surface phenotypes defined by the DX9 antibody are due to KIR3DL1 gene polymorphism. *Journal of immunology (Baltimore, Md : 1950)* 2001, 166(5):2992-3001.

65. Robinson J, Halliwell JA, McWilliam H, Lopez R, Marsh SGE: IPD--the Immuno Polymorphism Database. *Nucleic Acids Res* 2013, 41(Database issue):D1234-D1240.

66. Cooper DN: Functional intronic polymorphisms: Buried treasure awaiting discovery within our genes. *Hum Genomics* 2010, 4(5):284-288.

67. Toscano C, Klein K, Blievernicht J, Schaeffeler E, Saussele T, Raimundo S, Eichelbaum M, Schwab M, Zanger UM: Impaired expression of CYP2D6 in intermediate metabolizers carrying the *41 allele caused by the intronic SNP 2988G>A: evidence for modulation of splicing events. *Pharmacogenetics and genomics* 2006, 16(10):755-766.

68. Thomas R, Thio CL, Apps R, Qi Y, Gao X, Marti D, Stein JL, Soderberg KA, Moody MA, Goedert JJ *et al*: A novel variant marking HLA-DP expression levels predicts recovery from hepatitis B virus infection. *Journal of virology* 2012, 86(12):6979-6985.

69. Hodgkinson A, Eyre-Walker A: Variation in the mutation rate across mammalian genomes. *Nature Reviews Genetics* 2011, 12(11):756-766.

70. Hwang DG, Green P: Bayesian Markov chain Monte Carlo sequence analysis reveals varying neutral substitution patterns in mammalian evolution. *Proc Natl Acad Sci U S A* 2004, 101(39):13994-14001.

71. Nachman MW, Crowell SL: Estimate of the mutation rate per nucleotide in humans. *Genetics* 2000, 156(1):297-304.

72. Nachman MW: Single nucleotide polymorphisms and recombination rate in humans. *Trends in genetics : TIG* 2001, 17(9):481-485.

73. Lercher MJ, Hurst LD: Human SNP variability and mutation rate are higher in regions of high recombination. *Trends in genetics : TIG* 2002, 18(7):337-340.

74. Duret L, Arndt PF: The Impact of Recombination on Nucleotide Substitutions in the Human Genome. *PLoS genetics* 2008, 4(5):e1000071.

75. Tian D, Wang Q, Zhang P, Araki H, Yang S, Kreitman M, Nagylaki T, Hudson R, Bergelson J, Chen JQ: Single-nucleotide mutation rate increases close to insertions/deletions in eukaryotes. *Nature* 2008, 455(7209):105-108.

76. Zhu L, Wang Q, Tang P, Araki H, Tian D: Genomewide association between insertions/deletions and the nucleotide diversity in bacteria. *Molecular biology and evolution* 2009, 26(10):2353-2361.

77. McDonald MJ, Wang W-C, Huang H-D, Leu J-Y: Clusters of Nucleotide Substitutions and Insertion/Deletion Mutations Are Associated with Repeat Sequences. *PLOS Biology* 2011, 9(6):e1000622.

78. Maxwell LD, Williams F, Gilmore P, Meenagh A, Middleton D: Investigation of killer cell immunoglobulin-like receptor gene diversity: II. KIR2DS4. *Human immunology* 2004, 65(6):613-621.

79. Bruijnesteijn J, van der Wiel MKH, de Groot N, Otting N, de Vos-Rouweler AJM, Lardy NM, de Groot NG, Bontrop RE: Extensive Alternative Splicing of KIR Transcripts. *Frontiers in immunology* 2018, 9:2846.

80. Vilches C, Rajalingam R, Uhrberg M, Gardiner CM, Young NT, Parham P: KIR2DL5, a Novel Killer-Cell Receptor with a D0-D2 Configuration of Ig-Like Domains. *The Journal of Immunology* 2000, 164(11):5797.

81. Dohring C, Samaridis J, Colonna M: Alternatively spliced forms of human killer inhibitory receptors. *Immunogenetics* 1996, 44(3):227-230.

82. Wilson MJ, Torkar M, Trowsdale J: Genomic organization of a human killer cell inhibitory receptor gene. *Tissue antigens* 1997, 49(6):574-579.

83. Blokhuis JH, Doxiadis GG, Bontrop RE: A splice site mutation converts an inhibitory killer cell Ig-like receptor into an activating one. *Molecular immunology* 2009, 46(4):640-648.

84. Lee Y, Rio DC: Mechanisms and Regulation of Alternative Pre-mRNA Splicing. *Annual Review of Biochemistry* 2015, 84(1):291-323.

85. Goodridge JP, Lathbury LJ, Steiner NK, Shulse CN, Pullikotil P, Seidah NG, Hurley CK, Christiansen FT, Witt CS: Three common alleles of KIR2DL4 (CD158d) encode constitutively expressed, inducible and secreted receptors in NK cells. *European journal of immunology* 2007, 37(1):199-211.

86. Rajalingam R, Parham P, Abi-Rached L: Domain shuffling has been the main mechanism forming new hominoid killer cell Ig-like receptors. *Journal of immunology (Baltimore, Md : 1950)* 2004, 172(1):356-369.

87. Valiante NM, Uhrberg M, Shilling HG, Lienert-Weidenbach K, Arnett KL, D'Andrea A, Phillips JH, Lanier LL, Parham P: Functionally and structurally distinct NK cell receptor repertoires in the peripheral blood of two human donors. *Immunity* 1997, 7(6):739-751.

88. Andersson S, Malmberg JA, Malmberg KJ: Tolerant and diverse natural killer cell repertoires in the absence of selection. *Experimental cell research* 2010, 316(8):1309-1315.

89. Santourlidis S, Trompeter HI, Weinhold S, Eisermann B, Meyer KL, Wernet P, Uhrberg M: Crucial role of DNA methylation in determination of clonally distributed killer cell Ig-like receptor expression patterns in NK cells. *Journal of immunology (Baltimore, Md : 1950)* 2002, 169(8):4253-4261.

90. Chan H-W, Kurago ZB, Stewart CA, Wilson MJ, Martin MP, Mace BE, Carrington M, Trowsdale J, Lutz CT: DNA methylation maintains allele-specific KIR gene expression in human natural killer cells. *The Journal of experimental medicine* 2003, 197(2):245-255.

91. Stewart CA, Van Bergen J, Trowsdale J: Different and divergent regulation of the KIR2DL4 and KIR3DL1 promoters. *Journal of immunology (Baltimore, Md : 1950)* 2003, 170(12):6073-6081.

92. Trompeter HI, Gomez-Lozano N, Santourlidis S, Eisermann B, Wernet P, Vilches C, Uhrberg M: Three structurally and functionally divergent kinds of promoters regulate expression of clonally distributed killer cell Ig-like receptors (KIR), of KIR2DL4, and of KIR3DL3. *Journal of immunology (Baltimore, Md : 1950)* 2005, 174(7):4135-4143.

93. Li H, Pascal V, Martin MP, Carrington M, Anderson SK: Genetic Control of Variegated KIR Gene Expression: Polymorphisms of the Bi-Directional KIR3DL1 Promoter Are Associated with Distinct Frequencies of Gene Expression. *PLoS genetics* 2008, 4(11):e1000254.

94. Davies GE, Locke SM, Wright PW, Li H, Hanson RJ, Miller JS, Anderson SK: Identification of bidirectional promoters in the human KIR genes. *Genes and immunity* 2007, 8(3):245-253.

95. Li H, Wright PW, McCullen M, Anderson SK: Characterization of KIR intermediate promoters reveals four promoter types associated with distinct expression patterns of KIR subtypes. *Genes and immunity* 2016, 17(1):66-74.

96. Parham P: Immunogenetics of killer-cell immunoglobulin-like receptors. *Tissue antigens* 2003, 62(3):194-200.

97. Gomez-Lozano N, Trompeter HI, de Pablo R, Estefania E, Uhrberg M, Vilches C: Epigenetic silencing of potentially functional KIR2DL5 alleles: Implications for the acquisition of KIR repertoires by NK cells. *European journal of immunology* 2007, 37(7):1954-1965.

98. Cisneros E, Moraru M, Gómez-Lozano N, López-Botet M, Vilches C: KIR2DL5: An Orphan Inhibitory Receptor Displaying Complex Patterns of Polymorphism and Expression. *Frontiers in immunology* 2012, 3:289-289.

99. Vilches C, Gardiner CM, Parham P: Gene Structure and Promoter Variation of Expressed and Nonexpressed Variants of the KIR2DL5 Gene. *The Journal of Immunology* 2000, 165(11):6416.

100. Gomez-Lozano N, Estefania E, Williams F, Halfpenny I, Middleton D, Solis R, Vilches C: The silent KIR3DP1 gene (CD158c) is transcribed and might encode a secreted receptor in a minority of humans, in whom the KIR3DP1, KIR2DL4 and KIR3DL1/KIR3DS1 genes are duplicated. *European journal of immunology* 2005, 35(1):16-24.

101. van Bergen J, Stewart CA, van den Elsen PJ, Trowsdale J: Structural and functional differences between the promoters of independently expressed killer cell Ig-like receptors. *European journal of immunology* 2005, 35(7):2191-2199.

102. Marquardt N, Scharenberg M, Mold JE, Hård J, Kekäläinen E, Buggert M, Nguyen S, Wilson JN, Al-Ameri M, Ljunggren H-G *et al*: High-dimensional analysis reveals a distinct population of adaptive-like tissue-resident NK cells in human lung. *bioRxiv* 2019:2019.2012.2020.883785.

103. Marquardt N, Beziat V, Nystrom S, Hengst J, Ivarsson MA, Kekalainen E, Johansson H, Mjosberg J, Westgren M, Lankisch TO *et al*: Cutting edge: identification and characterization of human intrahepatic CD49a+ NK cells. *Journal of immunology (Baltimore, Md : 1950)* 2015, 194(6):2467-2471.

104. Ivarsson MA, Stiglund N, Marquardt N, Westgren M, Gidlof S, Bjorkstrom NK: Composition and dynamics of the uterine NK cell KIR repertoire in menstrual blood. *Mucosal immunology* 2017, 10(2):322-331.

105. Björkström NK, Béziat V, Cichocki F, Liu LL, Levine J, Larsson S, Koup RA, Anderson SK, Ljunggren H-G, Malmberg K-J: CD8 T cells express randomly selected KIRs with distinct specificities compared with NK cells. *Blood* 2012, 120(17):3455-3465.

106. Mingari MC, Moretta A, Moretta L: Regulation of KIR expression in human T cells: a safety mechanism that may impair protective T-cell responses. *Immunology today* 1998, 19(4):153-157.

107. Huard B, Karlsson L: KIR expression on self-reactive CD8+ T cells is controlled by T-cell receptor engagement. *Nature* 2000, 403(6767):325-328.

108. Liu Y, Kuick R, Hanash S, Richardson B: DNA methylation inhibition increases T cell KIR expression through effects on both promoter methylation and transcription factors. *Clin Immunol* 2009, 130(2):213-224.

109. Chan H-W, Kurago ZB, Stewart CA, Wilson MJ, Martin MP, Mace BE, Carrington M, Trowsdale J, Lutz CT: DNA Methylation Maintains Allele-specific KIR Gene Expression in Human Natural Killer Cells. *Journal of Experimental Medicine* 2003, 197(2):245-255.

110. Lam TH, Shen M, Chia JM, Chan SH, Ren EC: Population-specific recombination sites within the human MHC region. *Heredity* 2013, 111(2):131-138.

111. Miretti MM, Walsh EC, Ke X, Delgado M, Griffiths M, Hunt S, Morrison J, Whittaker P, Lander ES, Cardon LR *et al*: A High-Resolution Linkage-Disequilibrium Map of the Human Major Histocompatibility Complex and First Generation of Tag Single-Nucleotide Polymorphisms. *The American Journal of Human Genetics* 2005, 76(4):634-646.

112. Voorter CE, Gerritsen KE, Groeneweg M, Wieten L, Tilanus MG: The role of gene polymorphism in HLA class I splicing. *International journal of immunogenetics* 2016, 43(2):65-78.

113. Krangel MS: Secretion of HLA-A and-B antigens via an alternative RNA splicing pathway. *Journal of Experimental Medicine* 1986, 163(5):1173-1190.

114. Dunn PP, Hammond L, Coates E, Street J, Griner L, Darke C: A "silent" nucleotide substitution in exon 4 is responsible for the "alternative expression" of HLA-A*01:01:38L through aberrant splicing. *Human immunology* 2011, 72(9):717-722.

115. Dai Z-X, Zhang G-H, Zhang X-H, Xia H-J, Li S-Y, Zheng Y-T: The β2-Microglobulin–Free Heterodimerization of Rhesus Monkey MHC Class I A with Its Normally Spliced Variant Reduces the Ubiquitin-Dependent Degradation of MHC Class I A. *The Journal of Immunology* 2012:1100665.

116. Dai ZX, Zhang GH, Zhang XH, Zheng YT: Identification and characterization of a novel splice variant of rhesus macaque MHC IA. *Molecular immunology* 2013, 53(3):206-213.

117. Ishitani A, Geraghty DE: Alternative splicing of HLA-G transcripts yields proteins with primary structures resembling both class I and class II antigens. *Proceedings of the National Academy of Sciences* 1992, 89(9):3947.

118. Rouas-Freiss N, Bruel S, Menier C, Marcou C, Moreau P, Carosella ED: Switch of HLA-G alternative splicing in a melanoma cell line causes loss of HLA-G1 expression and sensitivity to NK lysis. *International Journal of Cancer* 2005, 117(1):114-122.

119. Kirszenbaum M, Moreau P, Gluckman E, Dausset J, Carosella E: An alternatively spliced form of HLA-G mRNA in human trophoblasts and evidence for the presence of HLA-G transcript in adult lymphocytes. *Proceedings of the National Academy of Sciences* 1994, 91(10):4209.

120. Rizzo R, Bortolotti D, Bolzani S, Fainardi E: HLA-G Molecules in Autoimmune Diseases and Infections. *Frontiers in immunology* 2014, 5:592-592.

121. Fleming KA, McMichael A, Morton JA, Woods J, McGee JO: Distribution of HLA class 1 antigens in normal human tissue and in mammary cancer. *Journal of clinical pathology* 1981, 34(7):779-784.

122. Wei X, Orr HT: HLA Class I. In: *Encyclopedia of Immunology (Second Edition).* Edited by Delves PJ. Oxford: Elsevier; 1998: 1108-1111.

123. Dellgren C, Nehlin JO, Barington T: Cell surface expression level variation between two common Human Leukocyte Antigen alleles, HLA-A2 and HLA-B8, is dependent on the structure of the C terminal part of the alpha 2 and the alpha 3 domains. *PLoS One* 2015, 10(8):e0135385.

124. Apps R, Meng Z, Del Prete GQ, Lifson JD, Zhou M, Carrington M: Relative expression levels of the HLA class-I proteins in normal and HIV-infected cells. *Journal of immunology (Baltimore, Md : 1950)* 2015, 194(8):3594-3600.

125. Apps R, Qi Y, Carlson JM, Chen H, Gao X, Thomas R, Yuki Y, Del Prete GQ, Goulder P, Brumme ZL *et al*: Influence of HLA-C Expression Level on HIV Control. *Science* 2013, 340(6128):87.

126. Otting N, Heijmans CM, Noort RC, de Groot NG, Doxiadis GG, van Rood JJ, Watkins DI, Bontrop RE: Unparalleled complexity of the MHC class I region in rhesus macaques. *Proc Natl Acad Sci U S A* 2005, 102(5):1626-1631.

127. Otting N, Heijmans CM, van der Wiel M, de Groot NG, Doxiadis GG, Bontrop RE: A snapshot of the Mamu-B genes and their allelic repertoire in rhesus macaques of Chinese origin. *Immunogenetics* 2008, 60(9):507-514.

128. He Y, Tian Z: NK cell education via nonclassical MHC and non-MHC ligands. *Cellular & molecular immunology* 2017, 14(4):321-330.

129. He Y, Peng H, Sun R, Wei H, Ljunggren H-G, Yokoyama WM, Tian Z: Contribution of inhibitory receptor TIGIT to NK cell education. *Journal of Autoimmunity* 2017, 81:1-12.

130. Horowitz A, Djaoud Z, Nemat-Gorgani N, Blokhuis J, Hilton HG, Béziat V, Malmberg KJ, Norman PJ, Guethlein LA, Parham P: Class I HLA haplotypes form two schools that educate NK cells in different ways. *Science immunology* 2016, 1(3).

131. Anfossi N, André P, Guia S, Falk CS, Roetynck S, Stewart CA, Breso V, Frassati C, Reviron D, Middleton D *et al*: Human NK Cell Education by Inhibitory Receptors for MHC Class I. *Immunity* 2006, 25(2):331-342.

132. Borrego F, Ulbrecht M, Weiss EH, Coligan JE, Brooks AG: Recognition of human histocompatibility leukocyte antigen (HLA)-E complexed with HLA class I signal sequence-derived peptides by CD94/NKG2 confers protection from natural killer cell-mediated lysis. *The Journal of experimental medicine* 1998, 187(5):813-818.

133. Braud VM, Allan DS, O'Callaghan CA, Soderstrom K, D'Andrea A, Ogg GS, Lazetic S, Young NT, Bell JI, Phillips JH *et al*: HLA-E binds to natural killer cell receptors CD94/NKG2A, B and C. *Nature* 1998, 391(6669):795-799.

134. Lee N, Llano M, Carretero M, Ishitani A, Navarro F, López-Botet M, Geraghty DE: HLA-E is a major ligand for the natural killer inhibitory receptor CD94/NKG2A. *Proc Natl Acad Sci U S A* 1998, 95(9):5199-5204.

135. Fauriat C, Ivarsson MA, Ljunggren HG, Malmberg KJ, Michaëlsson J: Education of human natural killer cells by activating killer cell immunoglobulin-like receptors. *Blood* 2010, 115(6):1166-1174.

136. Lee N, Goodlett DR, Ishitani A, Marquardt H, Geraghty DE: HLA-E surface expression depends on binding of TAP-dependent peptides derived from certain HLA class I signal sequences. *Journal of immunology (Baltimore, Md : 1950)* 1998, 160(10):4951-4960.

137. Malnati MS, Peruzzi M, Parker KC, Biddison WE, Ciccone E, Moretta A, Long EO: Peptide specificity in the recognition of MHC class I by natural killer cell clones. *Science* 1995, 267(5200):1016-1018.

138. Rajagopalan S, Long EO: The direct binding of a p58 killer cell inhibitory receptor to human histocompatibility leukocyte antigen (HLA)-Cw4 exhibits peptide selectivity. *The Journal of experimental medicine* 1997, 185(8):1523-1528.

139. 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. *European journal of immunology* 2004, 34(6):1673-1679.

140. Fadda L, Borhis G, Ahmed P, Cheent K, Pageon SV, Cazaly A, Stathopoulos S, Middleton D, Mulder A, Claas FHJ *et al*: Peptide antagonism as a mechanism for NK cell activation. *Proceedings of the National Academy of Sciences* 2010, 107(22):10160.

141. Li Y, Mariuzza RA: Structural basis for recognition of cellular and viral ligands by NK cell receptors. *Frontiers in immunology* 2014, 5:123.

142. Carrillo-Bustamante P, de Boer RJ, Keşmir C: Specificity of inhibitory KIRs enables NK cells to detect changes in an altered peptide environment. *Immunogenetics* 2018, 70(2):87-97.

143. Yu J, Heller G, Chewning J, Kim S, Yokoyama WM, Hsu KC: Hierarchy of the Human Natural Killer Cell Response Is Determined by Class and Quantity of Inhibitory Receptors for Self-HLA-B and HLA-C Ligands. *The Journal of Immunology* 2007, 179(9):5977.

144. Yawata M, Yawata N, Draghi M, Little A-M, Partheniou F, Parham P: Roles for HLA and KIR polymorphisms in natural killer cell repertoire selection and modulation of effector function. *The Journal of experimental medicine* 2006, 203(3):633-645.

145. Shilling HG, Young N, Guethlein LA, Cheng NW, Gardiner CM, Tyan D, Parham P: Genetic control of human NK cell repertoire. *Journal of immunology (Baltimore, Md : 1950)* 2002, 169(1):239-247.

146. Yu J, Heller G, Chewning J, Kim S, Yokoyama WM, Hsu KC: Hierarchy of the human natural killer cell response is determined by class and quantity of inhibitory receptors for self-HLA-B and HLA-C ligands. *Journal of immunology (Baltimore, Md : 1950)* 2007, 179(9):5977-5989.

147. Templer S, Sacks G: A blessing and a curse: is high NK cell activity good for health and bad for reproduction? *Human fertility (Cambridge, England)* 2016, 19(3):166-172.

148. Shakhar K, Rosenne E, Loewenthal R, Shakhar G, Carp H, Ben-Eliyahu S: High NK cell activity in recurrent miscarriage: what are we really measuring? *Human reproduction (Oxford, England)* 2006, 21(9):2421-2425.

149. Colucci F: The role of KIR and HLA interactions in pregnancy complications. *Immunogenetics* 2017, 69(8):557-565.

150. Hiby SE, Regan L, Lo W, Farrell L, Carrington M, Moffett A: Association of maternal killer-cell immunoglobulin-like receptors and parental HLA-C genotypes with recurrent miscarriage. *Human reproduction (Oxford, England)* 2008, 23(4):972-976.

151. Tripathy SK, Keyel PA, Yang L, Pingel JT, Cheng TP, Schneeberger A, Yokoyama WM: Continuous engagement of a self-specific activation receptor induces NK cell tolerance. *The Journal of experimental medicine* 2008, 205(8):1829-1841.

152. Liao NS, Bix M, Zijlstra M, Jaenisch R, Raulet D: MHC class I deficiency: susceptibility to natural killer (NK) cells and impaired NK activity. *Science* 1991, 253(5016):199-202.

153. Hoglund P, Ohlen C, Carbone E, Franksson L, Ljunggren HG, Latour A, Koller B, Karre K: Recognition of beta 2-microglobulin-negative (beta 2m-) T-cell blasts by natural killer cells from normal but not from beta 2m- mice: nonresponsiveness controlled by beta 2m- bone marrow in chimeric mice. *Proc Natl Acad Sci U S A* 1991, 88(22):10332-10336.

154. Norman PJ, Hollenbach JA, Nemat-Gorgani N, Marin WM, Norberg SJ, Ashouri E, Jayaraman J, Wroblewski EE, Trowsdale J, Rajalingam R *et al*: Defining KIR and HLA Class I Genotypes at Highest Resolution via High-Throughput Sequencing. *American journal of human genetics* 2016, 99(2):375-391.

155. Doherty PC, Zinkernagel RM: Enhanced immunological surveillance in mice heterozygous at the H-2 gene complex. *Nature* 1975, 256(5512):50-52.

156. McClelland EE, Penn DJ, Potts WK: Major histocompatibility complex heterozygote superiority during coinfection. *Infect Immun* 2003, 71(4):2079-2086.

157. Penn DJ, Damjanovich K, Potts WK: MHC heterozygosity confers a selective advantage against multiple-strain infections. *Proc Natl Acad Sci U S A* 2002, 99(17):11260-11264.

158. Tu MM, Mahmoud AB, Makrigiannis AP: Licensed and Unlicensed NK Cells: Differential Roles in Cancer and Viral Control. *Frontiers in immunology* 2016, 7(166).

159. Tarek N, Le Luduec JB, Gallagher MM, Zheng J, Venstrom JM, Chamberlain E, Modak S, Heller G, Dupont B, Cheung NK *et al*: Unlicensed NK cells target neuroblastoma following anti-GD2 antibody treatment. *The Journal of clinical investigation* 2012, 122(9):3260-3270.

160. Orr MT, Murphy WJ, Lanier LL: 'Unlicensed' natural killer cells dominate the response to cytomegalovirus infection. *Nature immunology* 2010, 11(4):321-327.

161. Maccari G, Robinson J, Ballingall K, Guethlein LA, Grimholt U, Kaufman J, Ho CS, de Groot NG, Flicek P, Bontrop RE *et al*: IPD-MHC 2.0: an improved inter-species database for the study of the major histocompatibility complex. *Nucleic Acids Res* 2017, 45(D1):D860-d864.

Chapter 7

General discussion and future prospects

Jesse Bruijnesteijn

General discussion and future prospects

Non-human primates are important models for many human diseases that involve different components of the immune system. For that reason, a detailed characterization of these components, and in particular the similarity to its equivalent in humans, is a necessity. During the course of this thesis, a comparative immunogenetic approach was chosen to unravel the complexity of a set of important immune receptors that are expressed by natural killer (NK) cells. NK cells are cytotoxic lymphocytes, which provide a rapid and non-specific immune response analogous to that of cytotoxic T cells (CTLs) [1]. The adaptive response of CTLs is far more specific, as their activation and clonal expansion depend on rearranged antigen-specific receptors that are educated in the thymus. Expression of different classes of germline-encoded activating and inhibitory receptors tightly control the education and activity of NK cells. The killer cell immunoglobulin-like receptor (KIR) family is one of the major components that is involved in this regulation, and mainly acts through interactions with major histocompatibility complex (MHC) class I molecules. In contrast to the T cell receptors, the KIR receptors are not subjected to rearrangement and have an innate character. Although KIR may recognize peptides in context of MHC, they primarily seem to interact with a limited set of polymorphic determinants on MHC class I molecules. The human *KIR* genes display a high degree of diversity, which is reflected by copy number variation (CNV), allelic polymorphism, and different haplotype configurations [2, 3]. This genetic diversification of KIR might be propelled by selective co-evolution with their MHC ligands, and both systems are in a continuous arms race with pathogens. For decades, the *MHC* region was considered the most polymorphic and variable gene system in primates, but the *KIR* cluster seems to exceed this plasticity [4, 5]. The diverse compositions and allelic variety of *KIR* genes at an individual and populational level is highly relevant in relation to disease susceptibility and protection, which advocates for a comprehensive characterization in humans [6-8].

The genetic *KIR* cluster is to some extent also characterized in several non-human primate species. We focused on the thorough characterization of *KIR* genes in rhesus and cynomolgus macaques, whereas others have defined the *KIR* cluster in chimpanzees, orangutans, and baboons [9-12]. Understanding the genetics and evolution of primate *KIR* provides a foundation to study their functional impact related to health and disease, and might also help to refine disease models. The NK cell biology of macaques and humans display great similarity, which is in sharp contrast to the largely diverged NK cells in rodents. Therefore, macaque species seem to be a powerful model to study NK cells and KIR receptor function in relation to health and disease.

High accuracy characterization of the human and rhesus macaque KIR genes

In recent years, the *KIR* gene cluster in humans is extensively examined using different conventional sequencing approaches. Most of these studies resolved the genomic *KIR* content based on the presence or absence of particular gene segments by amplification with sequence-specific primers [13-15]. This approach may miss out on essential information on allelic polymorphism, recombinant genes, copy number variation (CNV), gene expression levels, and post-transcriptional modifications, such as alternative splicing. Less frequent was the characterization of complete human *KIR* haplotypes, but has been demonstrated using Fosmid library sequencing and target-enriched illumuna sequencing [16, 17]. Contrasting to the mainly genomic strategy in humans, the *KIR* genes in macaques were initially characterized by Sanger sequencing of partial and full-length cDNA sequences that were amplified by sequence-specific primers [18, 19]. Several macaque *KIR* haplotype configurations were deduced by studying gene segregation in families, and revealed extensive expansions and contractions. Although these early studies were informative, the conventional characterization strategies often lack an allele level resolution, and might miss genes with low transcription levels. These shortcomings are mostly explained by a relatively low throughput and limited fragment length of the used techniques, such as Roche 454 and Sanger sequencing. The development of single-molecule real-time (SMRT) sequencing platforms, which are commercialized by Pacific Biosciences (PacBio) and Oxford Nanopore Technologies (ONT), overcome many of the conventional limitations.

The PacBio platform generates sequence information during real time synthesis of complement DNA using fluorescent nucleotides. This technique offers long read lengths and high throughput. The relatively high error rate compared to conventional sequencing methods is largely compensated by sequencing one strand multiple times, and subsequently allows one to generate a consensus sequence. The high accuracy and throughput of PacBio consensus sequencing allows complete definition of *KIR* transcriptomes in humans and rhesus macaques (**Chapter 2**). In this study, full-length *KIR* transcripts were characterized, which resolved and confirmed novel alleles. Even more, fusion gene transcripts, which comprised head and tail segments from different *KIR* genes, indicated recombination events in the human and rhesus macaque *KIR* cluster. Using a family-based study design, segregation analysis enabled the definition of *KIR* haplotype configurations. In humans, *KIR* haplotypes are largely conform a typical centromeric and telomeric region, with either a more inhibitory or activating gene content (group A or B haplotypes). More dynamic gene content variability is featured for rhesus macaque *KIR* haplotype configurations, as reflected by expansions, contractions, and the generation of substantial numbers of fusion genes. The comprehensive overview of human and rhesus macaque *KIR* transcriptomes demonstrates the power of the new characterization strategy, which is relatively fast, cost-efficient, and highly accurate [20].

Unparalleled rapid expansion of *KIR* genes in macaque species and their populations

Characterization of the *KIR* gene cluster in different distantly related primate species, such as humans, chimpanzees, orangutans, and macaques, provided a comparative overview of its evolution [21]. Extensive tandem duplications, deletions, and recombination events shaped the *KIR* region in these primates, and seems to be largely propelled by co-evolution with their MHC class I ligands. The dynamics and plasticity of the *KIR* gene cluster might, however, be better reflected by comparing closely related primate species that share similar MHC class I systems. Using the PacBio full-length transcriptome strategy, we characterized the *KIR* gene content in rhesus and cynomolgus macaques, and their geographically distinct populations (**Chapter 3**). These two macaque species are relatively closely related, with a shared common ancestor that lived approximately 1-3 million years ago.

A profound overview of the *KIR* gene system was defined in both macaque species, which recorded 34 rhesus and 55 cynomolgus macaque *KIR* genes. Of these genes, 24 were considered orthologous, thereby reflecting the close common ancestry, whereas the considerable number of species-specific entities indicate a rapid evolution. In both species, the number of fusion gene entities indicated abundant chromosomal recombination events. This was substantiated by the highly dynamic *KIR* haplotype content that involved expansions and contractions. Several *KIR* genes were not only found specific for one species, but were also restricted to a particular macaque population. These population-specific *KIR* genes suggest an unprecedented rapid evolution of the genetic cluster. Even more, the macaque *KIR* genes also feature extensive allelic polymorphism, with 285 rhesus and 260 cynomolgus macaque allotypes defined in the studied cohorts. Only two KIR allotypes were recorded the be shared in both species, and even the different macaque populations only shared one or two allelic variants.

The unpararelled rapid evolution of the macaque *KIR* region diversifies their genetic content at a species and populational level. This *KIR* diversity might modulate the outcome of preclinical studies using macaque models. A better understanding of the genetic profile of macaques, for instance based on their species- and population-specific *KIR* gene content, might help to refine biomedical experiments. These advanced insights might better the translation of results in macaque models to humans. This might shed a light on the different KIR and MHC class I combinations that have been associated with disease susceptibility and protection, and eventually help to develop medicines and therapies for humans.

A wide variety of KIR isoforms: functional relevance or transcriptional noise?

The full-length *KIR* transcriptomes derived from the PacBio studies not only defined constitutively spliced transcripts, but also resolved high numbers of different alternatively spliced variants. Some of these posttranscriptional modifications were previously documented [22-24], but a comprehensive overview of the *KIR* spliceosome was lacking. In a follow-up study, we used the PacBio protocol to examine the alternative splicing profiles in several human and rhesus macaque families (**Chapter 4**). In total, 18 and 29 distinct splice events were recorded in humans and rhesus macaques, respectively, and were confirmed in at least two individuals. These splice events, of which some were gene-specific, were facilitated by different splicing mechanisms, such as exon skipping and the use of alternative splice sites. At this stage, the structural and functional consequences of the different splice events is hard to predict. Several splice events introduce an early stop codon, and might thereby abandon receptor function. Such an event might be noise in the transcriptional process, but might also have biological relevance as they function as a rapid switch to modulate receptor expression levels. Human 9A *KIR2DL4* alleles, for example, encode constitutively truncated receptors, and its expression is not detectable [25]. The use of an alternative splice site includes an intronic stretch to generate 9A *KIR2DL4* transcripts that seem to encode complete receptor isoforms. In a transfection study, we demonstrated that these 9A KIR2DL4 isoforms were expressed in endosomes (manuscript in preparation). In contrast, the same splice event in the in-frame 10A KIR2DL4 allotypes results in undetectable isoforms, which suggests that alternative splicing might function as an expression modulator.

The selection and conservation of different alternative splice sites suggests a functional role for particular isoforms. A highly conserved splice event in macaques, for instance, involves their only framework gene, *KIR3DL20*. Transcripts of *KIR3DL20* encode KIR3D and KIR2D receptors, the latter of which are consistently generated by the exclusion of exon 4 from *KIR3D* transcripts. These alternatively spliced transcripts display sequence similarity to human *KIR2DL5*, and are therefore referred to as *Mamu-* and *Mafa-KIR2DL05*. The function of KIR2DL05 isoforms in macaques, or its equivalent in humans, is not known. Nevertheless, alternative splicing seems to expand the framework KIR repertoire in macaques, and might facilitate a form of convergent evolution to generate KIR2DL5-like receptors.

With a comprehensive overview of the *KIR* splicing profiles in humans and macaques, the main challenge remains to determine its biological relevance. A method to determine isoform quantities relative to the constitutively spliced transcripts might help in understanding isoform function. Most conventional sequencing and quantification methods, such as real-time qPCR, are limited by short reads, and are insufficient to determine isoform quantities. The development of SMRT sequencing platforms might help in resolving isoform expression levels. This would provide insights into the distribution of KIR isoforms in different tissues or under certain conditions, and

thereby indicate their functional relevance. Functional isoforms, and the regulation of its expression, has been documented in literature [26-28]. For instance, the induction of alternative splicing has been reported in NK cells for the tumor-recognition (NK-TR) gene, of which the constitutively spliced transcripts encode a truncated and non-functional receptor [26]. Upon activation of NK cells via IL-2 stimulation, the splicing pattern of NK cells changed, and functional NK-TR receptors were expressed. A similar mechanism might regulate the expression and function of KIR receptors and its isoforms. With the current technological sequencing developments and the profound knowledge on KIR splicing profiles, future experiments might aim at the understanding of isoform function.

Genomic haplotype organization

The characterization of the *KIR* genes at the transcription level fosters the knowledge on allelic variants, expression levels, and posttranscriptional modifications. Transcriptome studies lack, however, information on pseudogenes, the physical location of genes at the haplotype, and non-coding sequences, such as introns. Therefore, a genomic assembly of the complete multigenic region might provide additional insights. The complete genomic characterization of *KIR* haplotypes is a challenging and time-consuming enterprise, as the region is large (100-300 kb), contains a dynamic gene content that share high sequence similarity, and harbors substantial repetitive elements. Indeed, the relatively standard human *KIR* haplotype configurations are well defined at the genomic DNA level using time-consuming Fosmid library and long-range PCR sequencing approaches [17, 29]. These completely sequenced haplotypes revealed a genomic architecture, in which centromeric and telomeric regions are occasionally shuffled (e.g., cA01-tA01, cB01-tB01, cA01-tB01). The haplotype reorganizations seem to be mediated by abundant transposable elements that are present in the non-coding regions. The genomic characterization of complete *KIR* haplotypes helps to understand how the haplotype diversity is generated, and what functional consequences might result from that.

Knowledge on the macaque *KIR* haplotypes is limited, and, so far, only two completely characterized short haplotypes are documented [30, 31]. Transcriptome and segregation studies, however, suggest an extensive gene content diversity, indicating expanded and contracted haplotypes. With the continuous and rapid evolution, and only a single framework gene, the macaque *KIR* haplotypes are unlikely to follow relatively rigid organizations, as is documented for group A and B haplotypes in humans. The availability of more completely characterized macaque *KIR* haplotypes might improve the understanding and evolution of this highly dynamic gene cluster.

We developed a targeted enrichment protocol using Cas9 nuclease activity and Nanopore sequencing to characterize multigenic clusters, such as the *KIR* gene region (**Chapter 5**). Large and overlapping fragments are generated without amplification, which avoids errors by strand synthesis and retains modification information, like

methylation profiles. The overlaps allow phasing of haplotypes, even at regions with high sequence similarity and in the absence of a reference genome. With this novel technological approach, we resolved six human and six rhesus macaque *KIR* haplotypes at an allele level resolution. This fast and relatively cheap characterization of complex immune regions might improve clinical protocols, such as the detailed matching of donor and acceptor compatibility in solid organ transplantation. In addition, more insights into the dynamic and complex macaque *KIR* haplotypes might help to understand the mechanism driving their extensive diversification. A comprehensive haplotype study may, for instance, reveal a linkage disequilibrium for different macaque *KIR* gene combinations that is similar to some non-randomly associated human *KIR* genes, such as *KIR2DL2-KIR2DS3-KIR2DL5B*. The function of *KIR* alleles or genes that are consistently low expressed might be compensated by family members that display a genomic linkage. For example, in humans it is thought that the function of the poorly expressed KIR3DL1*004 allotype might be balanced by the presence of KIR3DL2 or/ and KIR2DL4 allotypes that display a strong linkage disequilibrium [32]. Therefore, defining linkage disequilibriums might help in understanding the concerted function of KIR receptors.

Overall, this strategy to define complex immune regions might provide a tool to improve clinical applications. In solid-organ transplantation, for instance, this tool might improve the resolution of *KIR* gene matching for the graft donor and acceptor, and thereby may contribute to an enhanced graft survival. The high resolution may also lift disease association studies from an allele to a haplotype/region level. In a functional perspective, the insights into the genomic *KIR* haplotypes and their epigenetic profiles might aid the understanding of receptor function. Furthermore, the characterization of complete KIR haplotypes helps to understand the evolution of highly dynamic gene clusters that are involved in modulating immune responses.

A diverse group of immune receptors: impact on health and disease

The primate *KIR* gene cluster might be one of the most plastic and dynamic regions of the genome, and is diversified by different molecular mechanisms (**Chapter 6**). The combination of these mechanisms, involving chromosomal recombination events, point mutations, alternative splicing, and variegated expression, generates an array of diversity that is reflected by haplotype contractions and expansions, recombinant genes, allelic polymorphism, structural distinct isoforms, and differential expression patterns. One might expect that this great diversity impacts NK cell education, activation, and cytotoxicity, and thereby modulate health and disease at an individual and populational level.

An important biological process in which NK cells are involved is pregnancy [33, 34]. In the early stages of pregnancy, an semi-allogenic fetal trophoblast is implanted in the maternal uterus, which requires modulation of the local immune components.

Specialized NK cells form the predominant population of lymphocytes during early pregnancy, and are suggested to be involved in trophoblast invasion and vascular remodelling. These NK cells display relatively low cytotoxic activity, and mainly produce cytokines and other regulatory factors that modulate the local immune environment. The variety of activating and inhibitory KIR receptors play an important role to balance the local NK cell activity. The expression of activating KIR, including KIR2DS1, KIR2DS4, and KIR2DS5, in combination with the presence of the HLA-C2 epitope, strongly enhance successful pregnancy [35-38]. The uterus-resident NK cells display increased expression of those activating receptors during different stages of pregnancy compared to periphial blood (pb) NK cells. The absence of activating KIR, their HLA-C2 ligands, or the presence of mainly inhibitory receptors on the uterus-resident NK cells might contribute to the opposite effect, and is associated with pre-eclampsia [39]. There are, however, more robust association studies required to determine the precise role of KIR (allotypes) and its HLA class I ligands in pregnancy, as contradicting results are documented [40].

As component of the immune system, NK cells also play a major role in disease, in which they are involved in the identification and lysis of infected cells, the surveillance of tumour cells, and the regulation of other immune components. The *KIR* gene content diversity and the extensive allelic polymorphism might impact susceptibility and protection to disease at an individual level, and increase survival at a populational level. These disease associations are often found in concert with the independently segregating MHC class I ligands. Several viral pathogeneses have been related to the genetic content of the *KIR* and *MHC* class I clusters. The diverse outcomes upon human immunodeficiency virus (HIV) infection, for example, is at least in part modulated by the polymorphisms of the *KIR* and *HLA* genes of the host [41, 42]. These HIV-related associations indicate susceptibility, with fast progression to acquired immunodeficiency syndrome (AIDS), or protection, with complete control of the infection without medicine. Other diseases associated with *KIR* gene diversity are, among others, tuberculosis, rheumatoid arthritis, and different types of cancer [43-45]. Disease associations studies are, however, only powerful when the involved genes are characterized in detail, the possible other interfering factors are considered, and the control group represents a substantial size. The difficulty to interpret these studies is demonstrated by potentially conflicting outcomes recorded in literature. A role for KIR2DS4, for instance, was associated with melanoma tumors through its interaction with non-classical HLA molecules [46], but this was not confirmed in a clinical study [47].

In the human population, the initial event or infection that proceeds disease is in many cases not known. The controlled study designs in animal models, however, allows one to study values for immunization or infection in an individual (t=0), and follow the onset and course of disease, which facilitates more validated experiments. This might provide more robust associations of *KIR* gene content and disease progress. For instance, infection of macaques with the simian immunodeficiency virus (SIV), which is similar to HIV-1 in humans, models the progression and symptoms of AIDS, and allowed

to study the highly active antiretroviral therapy (HAART) [48, 49]. These SIV studies demonstrated that the presence of particular *KIR* genes, including *KIR3DL02* and *KIR3DSW08*, were associated with protection to AIDS progression. A detailed characterization of the immune regions in macaques helps in selecting appropriate individuals for specific preclinical models, and thereby refine *in vivo* experiments.

Functional implications of differential KIR gene expression

The current studies on KIR receptors in humans, but also in other primate species, are mainly performed on periphial blood NK (pbNK) cells, where they account for approximately 10-15% of the circulating lymphocytes. The *KIR* transcriptomes and epigenetic profiles obtained from SMRT sequencing studies provide a proxy of the total *KIR* repertoire present in an individual. A large proportion of NK cells, however, reside within peripheral and lymphoid organs, such as the uterus, liver, lungs, gut, and thymus [50-54]. These tissue-resident NK (trNK) cells may display distinct surface receptors that retain them within tissues and prevent egress into the periphery [55]. Even more, trNK cells might have different development pathways and execute specialized tissue-related functions, as discussed above for uterus-resident NK cells in pregnancy [54, 55]. Knowledge on the *KIR* gene expression in these trNK cell populations is limited, but one might expect differential expression profiles compared to that observed in pbNK cells. This is for example demonstrated in lung- and liver-resident NK cells. In the liver, trNK cells exhibited an oligoclonal expression pattern of one, two, or three KIR receptors [50]. Samples from different liver donors did not indicate a clear dominance of a specific inhibitory KIR, but co-expression of two activating KIR was hardly observed. In the lungs, only the adaptive (or memory-like) NK cells that reside within the tissue displayed a distinct KIR expression profile contrasting to the profile in other lung-resident or pbNK cell subsets [51]. This study, however, determined the expression of only four *KIR* genes, and might reflect an incomplete KIR phenotype [51]. The most extensively studied trNK cell population is the one that resides in the uterus [33, 36, 52]. The enhancement of trophoblast migration and vascular remodelling by the release of chemokines and cytokines from activated uterus-resident NK cells is a clear example of a tissue-specific NK cell function mediated by differential KIR expression.

The reason why the KIR expression profiles of trNK cells is poorly examined has practical and ethical reasons. Tissue samples from human organs are mainly harvested during surgery, such as lobectomy or liver resection for the removal of primary or metastatic tumours. These tissues are derived from patients that already have a compromised immune system, and might not resemble NK cells in a healthy tissue. With informed consent, tissues from healthy donor organs that were not used for transplantation can be harvested after death, but gene expression might be affected by a substantial period of time without organ perfusion [56]. Other options to study trNK cells include tissue cell line experiments, which might give indications for specific *KIR* gene expression profiles, but one has to take into account that this is always affected

by *in vitro* circumstances. Animal models, however, might provide more insights into trNK cells and their specialized functions. As stated earlier, rodent animal models have a different NK cell biology, and tissue-specialized subpopulations might not translate to the human situation. Macaques might be the suitable species to examine the phenotype and function of trNK cells in a controlled and powerful model, which might broadly translate to human trNK cells. Even more, the many proven disease models in macaques allow association studies of phenotypical and functional NK cell variations, including differential KIR expression profiles, in regard to viral and bacterial infections. The profound insights on the macaque *KIR* genes gathered in the past decade might contribute to a better understanding of the specialized NK cell functions in tissues.

Quantification of KIR gene transcripts and their isoforms

The determination of differential KIR expression in distinct NK cell populations is only the first step to elucidate their biological role. Discrimination between high and low level expression of different *KIR* genes, and shifts in these expression levels, would provide additional insights. Fluctuations in expression level might be triggered by external factors, such as cell activation through cytokines, the presence of ligands, or pathogenic or tumorigenic conditions. Current quantification of *KIR* transcripts rely on short fragment techniques, such as real-time qPCR and RNAseq in combination with Illumina sequencing [51, 57, 58]. These methods are able to discriminate the expression level of most human *KIR* genes. Differential KIR expression was, for example, demonstrated in NK/T-cell lymphoma's, in which KIR2DL4 was overexpressed compared to healthy conditions, whereas all other KIR were downregulated [58]. Short-read sequencing techniques are, however, irrespective to allelic variation, thereby omitting variable expression levels of KIR allotypes. Differential expression has been demonstrated for multiple KIR3DL1 allotypes, ranging from high, low, or no surface expression for particular allotypes. High-resolution quantification of *KIR* gene expression is challenging due to their extensive polymorphisms, copy number variation, and gene similarity. In macaques, the current short-read methods are insufficient to quantify the highly expanded *KIR* repertoire at all, considering the abundant fusion genes and gene similarities.

Direct sequencing of full-length RNA and cDNA transcripts using SMRT sequencing platforms enable quantitative transcriptome studies. A high coverage of transcripts is, however, required, to overcome the relatively low accuracy of these techniques. The thorough characterization of human and macaque *KIR* at the transcription level (**Chapters 2** and **3**) might contribute to the implementation of transcript quantification studies. In future experiments, the technological breakthroughs of SMRT sequencing platforms should be explored to gather more insights on the differential KIR expression patterns. This would help in defining the (specialized) functions of the distinct KIR receptors, which so far remained largely elusive.

The evolution of primate *KIR* genes

This thesis, together with other characterization studies, provides a comprehensive overview of the extensive diversity and variation of the primate *KIR* genes. The *KIR* diversification might be largely driven by co-evolution with *MHC class I* genes, and thereby indirectly by the arms race with evolving pathogens. This co-evolution is substantiated by a balanced expansion of MHC and KIR in different primate species, with more expanded repertoires in macaques compared to hominoids. The plasticity of the *KIR* genes, however, exceeds that observed for the *MHC class I* genes, with more extensive allelic polymorphism, chromosomal recombination events, CNV, variegated expression, and alternative splicing. The wide variety of KIR receptors and isoforms, generated by a continuous evolutionary effort, regulate the NK cell responsiveness, which in turn modulates the elimination of infected or tumorigenic cells, the regulation of (adaptive) immune cells, and the trophoblast migration during different stages of pregnancy. The genomic and transcriptomic characterization of the primate *KIR* gene system is now largely elucidated, which paves the way to further explorer the function of KIR in different biological processes.

References

1. Vivier E, Raulet DH, Moretta A, Caligiuri MA, Zitvogel L, Lanier LL, Yokoyama WM, Ugolini S: Innate or adaptive immunity? The example of natural killer cells. *Science* 2011, 331(6013):44-49.
2. 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(6):753-763.
3. Middleton D, Gonzelez F: The extensive polymorphism of KIR genes. *Immunology* 2010, 129(1):8-19.
4. Hedrick PW: Evolutionary Genetics of the Major Histocompatibility Complex. *The American Naturalist* 1994, 143(6):945-964.
5. Sommer S: The importance of immune gene variability (MHC) in evolutionary ecology and conservation. *Frontiers in Zoology* 2005, 2(1):16.
6. Khakoo SI, Thio CL, Martin MP, Brooks CR, Gao X, Astemborski J, Cheng J, Goedert JJ, Vlahov D, Hilgartner M *et al*: HLA and NK cell inhibitory receptor genes in resolving hepatitis C virus infection. *Science* 2004, 305(5685):872-874.
7. Qi Y, Martin MP, Gao X, Jacobson L, Goedert JJ, Buchbinder S, Kirk GD, O'Brien SJ, Trowsdale J, Carrington M: KIR/HLA pleiotropism: protection against both HIV and opportunistic infections. *PLoS Pathog* 2006, 2(8):e79.
8. Kulkarni S, Martin MP, Carrington M: The Yin and Yang of HLA and KIR in human disease. *Seminars in Immunology* 2008, 20(6):343-352.
9. Rajalingam R, Hong M, Adams EJ, Shum BP, Guethlein LA, Parham P: Short KIR Haplotypes in Pygmy Chimpanzee (Bonobo) Resemble the Conserved Framework of Diverse Human KIR Haplotypes. *Journal of Experimental Medicine* 2001, 193(1):135-146.
10. Abi-Rached L, Moesta AK, Rajalingam R, Guethlein LA, Parham P: Human-specific evolution and adaptation led to major qualitative differences in the variable receptors of human and chimpanzee natural killer cells. *PLoS Genet* 2010, 6(11):e1001192.
11. Wroblewski EE, Parham P, Guethlein LA: Two to Tango: Co-evolution of Hominid Natural Killer Cell Receptors and MHC. *Frontiers in Immunology* 2019, 10(177).
12. Guethlein LA, Norman PJ, Heijmans CM, de Groot NG, Hilton HG, Babrzadeh F, Abi-Rached L, Bontrop RE, Parham P: Two Orangutan Species Have Evolved Different KIR Alleles and Haplotypes. *J Immunol* 2017, 198(8):3157-3169.
13. Vilches C, Castaño J, Gómez-Lozano N, Estefanía E: Facilitation of KIR genotyping by a PCR-SSP method that amplifies short DNA fragments. *Tissue Antigens* 2007, 70(5):415-422.
14. Hong HA, Loubser AS, de Assis Rosa D, Naranbhai V, Carr W, Paximadis M, Lewis DA, Tiemessen CT, Gray CM: Killer-cell immunoglobulin-like receptor genotyping and HLA killer-cell immunoglobulin-like receptor-ligand identification by real-time polymerase chain reaction. *Tissue Antigens* 2011, 78(3):185-194.
15. Kulkarni S, Martin MP, Carrington M: KIR genotyping by multiplex PCR-SSP. *Methods Mol Biol* 2010, 612:365-375.
16. Norman PJ, Hollenbach JA, Nemat-Gorgani N, Marin WM, Norberg SJ, Ashouri E, Jayaraman J, Wroblewski EE, Trowsdale J, Rajalingam R *et al*: Defining KIR and HLA Class I Genotypes at Highest Resolution via High-Throughput Sequencing. *Am J Hum Genet* 2016, 99(2):375-391.
17. Roe D, Vierra-Green C, Pyo CW, Eng K, Hall R, Kuang R, Spellman S, Ranade S, Geraghty DE, Maiers M: Revealing complete complex KIR haplotypes phased by long-read sequencing technology. *Genes & Immunity* 2017, 18(3):127-134.
18. Blokhuis JH, van der Wiel MK, Doxiadis GGM, Bontrop RE: The mosaic of KIR haplotypes in rhesus macaques. *Immunogenetics* 2010, 62(5):295-306.
19. Bimber BN, Moreland AJ, Wiseman RW, Hughes AL, O'Connor DH: Complete characterization of killer Ig-like receptor (KIR) haplotypes in Mauritian cynomolgus macaques: novel insights into nonhuman primate KIR gene content and organization. *J Immunol* 2008, 181(9):6301-6308.
20. Bruijnesteijn J, van der Wiel MKH, Swelsen WTN, Otting N, de Vos-Rouweler AJM, Elferink D, Doxiadis GG, Claas FHJ, Lardy NM, de Groot NG *et al*: Human and Rhesus Macaque KIR Haplotypes Defined by Their Transcriptomes. *J Immunol* 2018, 200(5):1692-1701.

21. Guethlein LA, Older Aguilar AM, Abi-Rached L, Parham P: Evolution of Killer Cell Ig-Like Receptor KIR Genes: Definition of an Orangutan KIR Haplotype Reveals Expansion of Lineage III KIR Associated with the Emergence of MHC-C. *The Journal of Immunology* 2007, 179(1):491.

22. Blokhuis JH, Doxiadis GG, Bontrop RE: A splice site mutation converts an inhibitory killer cell Ig-like receptor into an activating one. *Mol Immunol* 2009, 46(4):640-648.

23. Prall TM, Graham ME, Karl JA, Wiseman RW, Ericsen AJ, Raveendran M, Alan Harris R, Muzny DM, Gibbs RA, Rogers J *et al*: Improved full-length killer cell immunoglobulin-like receptor transcript discovery in Mauritian cynomolgus macaques. *Immunogenetics* 2017, 69(5):325-339.

24. Bimber BN, Evans DT: The killer-cell immunoglobulin-like receptors of macaques. *Immunol Rev* 2015, 267(1):246-258.

25. Bruijnesteijn J, van der Wiel MKH, de Groot N, Otting N, de Vos-Rouweler AJM, Lardy NM, de Groot NG, Bontrop RE: Extensive Alternative Splicing of KIR Transcripts. *Front Immunol* 2018, 9:2846.

26. Rinfret A, Anderson SK: IL-2 regulates the expression of the NK-TR gene via an alternate RNA splicing mechanism. *Molecular Immunology* 1993, 30(14):1307-1313.

27. Toei M, Saum R, Forgac M: Regulation and isoform function of the V-ATPases. *Biochemistry* 2010, 49(23):4715-4723.

28. Bhuiyan SA, Ly S, Phan M, Huntington B, Hogan E, Liu CC, Liu J, Pavlidis P: Systematic evaluation of isoform function in literature reports of alternative splicing. *BMC Genomics* 2018, 19(1):637.

29. Pyo C-W, Wang R, Vu Q, Cereb N, Yang SY, Duh F-M, Wolinsky S, Martin MP, Carrington M, Geraghty DE: Recombinant structures expand and contract inter and intragenic diversification at the KIR locus. *BMC Genomics* 2013, 14(1):89.

30. Graves T, Eichler,E.E. and Wilson,R.K.: Macaca mulatta isolate AG07107 chromosome 19, whole genome shotgun sequence. 2019.

31. Sambrook JG, Bashirova A, Palmer S, Sims S, Trowsdale J, Abi-Rached L, Parham P, Carrington M, Beck S: Single haplotype analysis demonstrates rapid evolution of the killer immunoglobulin-like receptor (KIR) loci in primates. *Genome research* 2005, 15(1):25-35.

32. Gourraud PA, Meenagh A, Cambon-Thomsen A, Middleton D: Linkage disequilibrium organization of the human KIR superlocus: implications for KIR data analyses. *Immunogenetics* 2010, 62(11-12):729-740.

33. Moffett-King A: Natural killer cells and pregnancy. *Nature Reviews Immunology* 2002, 2(9):656-663.

34. Faas MM, de Vos P: Uterine NK cells and macrophages in pregnancy. *Placenta* 2017, 56:44-52.

35. Hiby SE, Apps R, Sharkey AM, Farrell LE, Gardner L, Mulder A, Claas FH, Walker JJ, Redman CW, Morgan L *et al*: Maternal activating KIRs protect against human reproductive failure mediated by fetal HLA-C2. *J Clin Invest* 2010, 120(11):4102-4110.

36. Kennedy PR, Chazara O, Gardner L, Ivarsson MA, Farrell LE, Xiong S, Hiby SE, Colucci F, Sharkey AM, Moffett A: Activating KIR2DS4 Is Expressed by Uterine NK Cells and Contributes to Successful Pregnancy. *J Immunol* 2016, 197(11):4292-4300.

37. Hiby SE, Walker JJ, O'Shaughnessy K M, 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(8):957-965.

38. Nakimuli A, Chazara O, Hiby SE, Farrell L, Tukwasibwe S, Jayaraman J, Traherne JA, Trowsdale J, Colucci F, Lougee E *et al*: A KIR B centromeric region present in Africans but not Europeans protects pregnant women from pre-eclampsia. *Proceedings of the National Academy of Sciences* 2015, 112(3):845.

39. Hiby SE, Regan L, Lo W, Farrell L, Carrington M, Moffett A: Association of maternal killer-cell immunoglobulin-like receptors and parental HLA-C genotypes with recurrent miscarriage. *Hum Reprod* 2008, 23(4):972-976.

40. Colucci F: The role of KIR and HLA interactions in pregnancy complications. *Immunogenetics* 2017, 69(8):557-565.

41. Bashirova AA, Thomas R, Carrington M: HLA/KIR restraint of HIV: surviving the fittest. *Annu Rev Immunol* 2011, 29:295-317.

42. Alter G, Martin MP, Teigen N, Carr WH, Suscovich TJ, Schneidewind A, Streeck H, Waring M, Meier A, Brander C *et al*: Differential natural killer cell–mediated inhibition of HIV-1 replication based on distinct KIR/HLA subtypes. *Journal of Experimental Medicine* 2007, 204(12):3027-3036.

43. Pydi SS, Sunder SR, Venkatasubramanian S, Kovvali S, Jonnalagada S, Valluri VL: Killer cell immunoglobulin like receptor gene association with tuberculosis. *Hum Immunol* 2013, 74(1):85-92.

44. Majorczyk E, Pawlik A, Łuszczek W, Nowak I, Wiśniewski A, Jasek M, Kuśnierczyk P: Associations of killer cell immunoglobulin-like receptor genes with complications of rheumatoid arthritis. *Genes & Immunity* 2007, 8(8):678-683.

45. Kulkarni S, Martin MP, Carrington M: The Yin and Yang of HLA and KIR in human disease. *Semin Immunol* 2008, 20(6):343-352.

46. Katz G, Gazit R, Arnon TI, Gonen-Gross T, Tarcic G, Markel G, Gruda R, Achdout H, Drize O, Merims S *et al*: MHC Class I-Independent Recognition of NK-Activating Receptor KIR2DS4. *The Journal of Immunology* 2004, 173(3):1819.

47. Naumova E, Mihaylova A, Stoitchkov K, Ivanova M, Quin L, Toneva M: Genetic polymorphism of NK receptors and their ligands in melanoma patients: prevalence of inhibitory over activating signals. *Cancer Immunol Immunother* 2005, 54(2):172-178.

48. Clements JE, Gama L, Graham DR, Mankowski JL, Zink MC: A simian immunodeficiency virus macaque model of highly active antiretroviral treatment: viral latency in the periphery and the central nervous system. *Curr Opin HIV AIDS* 2011, 6(1):37-42.

49. Dinoso JB, Rabi SA, Blankson JN, Gama L, Mankowski JL, Siliciano RF, Zink MC, Clements JE: A Simian Immunodeficiency Virus-Infected Macaque Model To Study Viral Reservoirs That Persist during Highly Active Antiretroviral Therapy. *Journal of Virology* 2009, 83(18):9247.

50. Marquardt N, Beziat V, Nystrom S, Hengst J, Ivarsson MA, Kekalainen E, Johansson H, Mjosberg J, Westgren M, Lankisch TO *et al*: Cutting edge: identification and characterization of human intrahepatic CD49a+ NK cells. *J Immunol* 2015, 194(6):2467-2471.

51. Marquardt N, Scharenberg M, Mold JE, Hård J, Kekäläinen E, Buggert M, Nguyen S, Wilson JN, Al-Ameri M, Ljunggren H-G *et al*: Expansions of adaptive-like NK cells with a tissue-resident phenotype in human lung and blood. *bioRxiv* 2020:2019.2012.2020.883785.

52. Sojka DK, Yang L, Yokoyama WM: Uterine Natural Killer Cells. *Front Immunol* 2019, 10:960.

53. Poggi A, Benelli R, Venè R, Costa D, Ferrari N, Tosetti F, Zocchi MR: Human Gut-Associated Natural Killer Cells in Health and Disease. *Front Immunol* 2019, 10:961.

54. Sojka DK, Plougastel-Douglas B, Yang L, Pak-Wittel MA, Artyomov MN, Ivanova Y, Zhong C, Chase JM, Rothman PB, Yu J *et al*: Tissue-resident natural killer (NK) cells are cell lineages distinct from thymic and conventional splenic NK cells. *Elife* 2014, 3:e01659.

55. Dogra P, Rancan C, Ma W, Toth M, Senda T, Carpenter DJ, Kubota M, Matsumoto R, Thapa P, Szabo PA *et al*: Tissue Determinants of Human NK Cell Development, Function, and Residence. *Cell* 2020, 180(4):749-763. e713.

56. Ferreira PG, Muñoz-Aguirre M, Reverter F, Sá Godinho CP, Sousa A, Amadoz A, Sodaei R, Hidalgo MR, Pervouchine D, Carbonell-Caballero J *et al*: The effects of death and post-mortem cold ischemia on human tissue transcriptomes. *Nature Communications* 2018, 9(1):490.

57. Li Y, Wang T, Hu X, Zhang H, Chen L, Bao X, He J: Study of KIR gene expression at the mRNA level in specific donor-derived NK cells after allogeneic HSCT. *Immunogenetics* 2020, 72(3):135-141.

58. Küçük C, Hu X, Gong Q, Jiang B, Cornish A, Gaulard P, McKeithan T, Chan WC: Diagnostic and Biological Significance of KIR Expression Profile Determined by RNA-Seq in Natural Killer/T-Cell Lymphoma. *Am J Pathol* 2016, 186(6):1435-1441.

Appendices

Summary

Natural killer (NK) cells represent a major component of the immune system, and are involved in the protection against pathogens, the surveillance of tumor cells, and the regulation of other players in the immune response. The education, activation and functioning of NK cells is tightly modulated by two major sets of inhibitory and activating receptors. A conserved set of NK cell receptors include members of the CD94:NKG2 family that recognize MHC-E molecules. The other set comprise the killer cell immunoglobulin-like receptors (KIR), which represent a diverse group of structurally similar transmembrane molecules. Members of the KIR family co-evolved with their diversified MHC-A,-B and-C ligands that are in a continuous arms race with pathogens. The KIR receptors are encoded within the Leukocyte Receptor Complex (LRC) on chromosome 19q13.4, and segregate independent from their MHC ligands. The *KIR* genes are arranged in a polygenic cluster conform a head-to-tail tandem organization, which is characterized by extensive diversification generated by point mutations and chromosomal recombination. As such, the *KIR* gene content is highly unique at an individual level and displays differential gene distribution at a population level. The highly variable *KIR* gene content, in combination with their polymorphic MHC class I ligands, is associated with differential susceptibility and protection in health and disease.

In this thesis, we aimed to improve the *KIR* gene characterization in humans and macaques. The latter species is commonly used as model in preclinical studies to develop medicine and therapies. Their *KIR* gene system and NK cell biology display similarities to humans, with species-specific variation mainly reflected in the KIR receptor structure and haplotype organization. Initial transcriptome characterization studies using conventional sequencing techniques indicated a highly plastic *KIR* gene system in macaques that might exceed the complexity observed in humans. However, a comprehensive overview of the macaque *KIR* gene system is lacking.

Conventional sequencing techniques, such as Roche 454 and Sanger sequencing, are limited to relatively short fragments. These techniques might be used to determine the presence or absence of *KIR* genes, or to define partial and common full-length transcripts at an allele level resolution. Nevertheless, the conventional approaches have an insufficient sequencing depth to obtain a comprehensive characterization of complex systems, such as KIR. A more thorough characterization of the *KIR* gene system is achieved by the application of single-molecule real-time (SMRT) sequencing platforms, which are commercialized by Pacific Biosciences (PacBio) and Oxford Nanopore Technologies (ONT). The high accuracy and throughput of PacBio circular consensus sequencing allowed the complete definition of full-length *KIR* transcriptomes in humans and rhesus macaques at the allele level resolution (**Chapter 2**). A family-based study design and subsequent segregation analysis enabled the deduction of *KIR* haplotype configurations. In humans, these configurations were conforming the common centromeric and telomeric regions, with either more inhibitory or activating gene content (A or B). A more dynamic gene content variability is featured in rhesus

macaques, as reflected by expansions, contractions, and the generation of substantial numbers of fusion genes. The differences between the human and rhesus macaque *KIR* gene systems are in line with the evolutionary distance to a common ancestor, approximately 25 million years ago. The rapid pace of *KIR* gene evolution is also illustrated by the comparison of the closely related rhesus and cynomolgus macaques, and their distinct populations (**Chapter 3**). The presence of abundant species- and population-specific macaque *KIR* genes indicate an unparalleled rapid evolution, which is probably propelled by the rapid adaption to differential pathogenic selective pressures. The diversification is largely mediated through chromosomal recombination events, which may shuffle head and tail segments of *KIR* genes, but also complete *KIR* haplotype segments.

The *KIR* transcriptome studies not only resolved the *KIR* gene diversity and allelic polymorphism, but also unravelled a wide spectrum of alternative splicing profiles (**Chapter 4**). Even though some of the alternative splicing events might represent transcriptional noise, the conservation of different splice events and the relatively large number of spliced transcripts in the transcriptome studies, suggest that at least several isoforms may have functional relevance. The variety of structurally diverged receptors might, for example, alter their ligand interactions, cellular localization, expression level, and signalling properties. The wide range of KIR isoforms extends the plasticity of the *KIR* gene cluster in primates.

The *KIR* transcriptome studies revealed information on allelic variants, copy number variation, chromosomal recombination events, haplotype configurations, and alternative splicing. However, these studies lack insights on the genomic organization of this complex gene cluster. We developed a Cas9-mediated target enrichment protocol to characterize complete *KIR* haplotypes in humans and rhesus macaques (**Chapter 5**). Using this strategy, the *KIR* haplotype architecture, the non-coding stretches, such as promotor regions and introns, and the epigenetic profiles were resolved.

The dynamics of the primate *KIR* gene system relies on multiple molecular mechanisms that together generates a compound array of diversity at an individual, population, and species level (**Chapter 6**). This molecular toolbox that facilitates rapid evolution is shared in the different primate species, which demonstrates an evolutionary effort to diversify their *KIR* gene cluster. The improved insights into the genomic and transcriptomic *KIR* profiles, and the development of novel high-throughput techniques, pave the way to examine functional implications of the dynamic *KIR* cluster in macaque models for health and disease (**Chapter 7**). This knowledge might translate to a better understanding of the human NK cell biology and related disease phenotypes.

Nederlandse samenvatting

Natural killer (NK) cellen vormen een belangrijk onderdeel van het immuunsysteem, en zijn betrokken bij de bestrijding van pathogenen, de detectie van tumorcellen en de regulatie van andere immuun componenten. De educatie, de activatie en het functioneren van NK-cellen worden strikt gemodelleerd door twee verschillende receptor groepen, welke beide activerende en inhiberende signalen kunnen doorgeven. De geconserveerde groep bestaat uit leden van de CD94:NKG2 receptorfamilie, welke een interactie kunnen aangaan met MHC-E moleculen. De andere groep omvat de killer cell immunoglobulin-like receptoren (KIR), welke grote variatie vertonen, maar structureel verwant zijn aan elkaar. De KIR receptoren en hun zeer diverse MHC-A,-B, en -C liganden, die in een constante adaptieve strijd zijn met pathogenen, zijn onderhevig aan co-evolutie. De genen die voor de KIR receptoren coderen liggen op chromosoom 19q13.4 en segregeren onafhankelijk van hun MHC liganden. De *KIR* genen liggen in een kop-staart organisatie gerangschikt binnen een polygenetisch cluster. Deze gen regio ondergaat veelvuldige en uiteenlopende diversificatie, aangedreven door puntmutaties en chromosomale recombinaties. Hierdoor is de combinatie van aanwezige *KIR* genen tot op zekere hoogte uniek per individu, en zijn genen en allelische variaties verschillend verdeeld in geografische populaties. De verscheidenheid van *KIR* genen, in combinatie met hun polymorfe MHC klasse I liganden, wordt geassocieerd met de vatbaarheid voor en de bescherming tegen bepaalde ziektes. Daarnaast zijn verschillende KIR en MHC combinaties in verband te brengen met de regulatie van biologische processen, zoals zwangerschap.

In deze thesis streven wij naar een verbetering van de karakterisatie van de *KIR* genen in de mens en makaaksoorten. De laatstgenoemde diersoorten worden op een regelmatige basis gebruikt als model in preklinische studies om medicijnen en therapieën te ontwikkelen. Het makaak *KIR* gen systeem en hun NK cel biologie vertoont veel gelijkenissen met dat van de mens. Diersoort-specifieke kenmerken kunnen vooral gevonden worden in de receptor structuur en de haplotype organisatie. De eerste studies voor het in kaart brengen van KIR transcripten gebeurde met behulp van conventionele sequencing methodes en indiceerde een complex systeem met meer genetische diversiteit in de makaak dan in de mens. Echter, een omvangrijk overzicht van het *KIR* gen cluster in de makaak is niet beschikbaar.

Conventionele sequencing technieken, zoals Roche 454 en Sanger sequencing, zijn beperkt tot het sequensen van korte fragmenten. Daarmee zijn deze methodes geschikt om de aan- en afwezigheid van genen te bepalen of om partiele en veelvoorkomende gehele transcripten te identificeren op allel niveau. Desalniettemin zijn deze technieken niet toereikend genoeg om een uitgebreid overzicht van het gehele *KIR* gen systeem te geven. Een meer gedetailleerd beeld kan verkregen worden met behulp van single-molecule real-time (SMRT) sequencing platforms, zoals Pacific Biosciences (PacBio) en Oxford Nanopore Technologies (ONT). De relatief hoge accuraatheid en output van PacBio circular consensus sequencing (CCS) maakte het mogelijk om humane en makaak *KIR* genen volledig in kaart te brengen op allel niveau

resolutie (**Hoofdstuk 2**). Door een familie-gebaseerd onderzoeksopzet en segregatie analyses konden *KIR* haplotype configuraties worden afgeleid. In de mens volgen de configuraties een standaard organisatie, zowel op de centromere als telomere segmenten, met of meer inhiberende of meer activerende genen (A of B). Daarentegen kenmerken de KIR configuraties in de resus makaak een meer dynamische content, waarbij expansie, contractie, en de formatie van fusie genen voor variatie zorgen. Met een evolutionaire afstand tot een gemeenschappelijke voorouder van ongeveer 25 miljoen jaar zijn de verschillen tussen het *KIR* gen systeem in de mens en makaak in de lijn der verwachtingen. De snelheid waarmee het *KIR* gen systeem evolueert is ook duidelijk zichtbaar bij de vergelijking tussen twee nauwverwante makaaksoorten, de resus en cynomolgus makaak, en hun verschillende populaties (**Hoofdstuk 3**). De grote diversiteit aan diersoort- en populatie-specifieke makaak *KIR* genen tonen een ongekend snelle evolutie aan, welke waarschijnlijk wordt aangedreven door evolutionaire selectie onder een uiteenlopende pathogene druk. Deze diversificatie wordt grotendeels gemedieerd door chromosomale recombinatie, waarbij de kop en de staart van genen kunnen worden uitgewisseld, maar ook complete haplotype segmenten kunnen worden gereorganiseerd.

Het karakteriseren van de *KIR* transcriptie gaf echter niet alleen inzicht in de *KIR* gen diversiteit en de allelische variatie, maar ontrafelde ook een wijd spectrum aan alternatieve splicing profielen (**Hoofdstuk 4**). Een deel van de alternatieve splicing gebeurtenissen zijn mogelijk ruis van het transcriptie proces. Echter, de conservatie van bepaalde splicing gebeurtenissen, en de aanwezigheid van een relatief grote hoeveelheid aan alternatieve sequenties in de transcriptie studies, suggereren dat een deel van de isovormen een functionele relevantie hebben. De verscheidenheid aan structureel verschillende receptoren kunnen bijvoorbeeld van invloed zijn op ligand interacties, de lokalisatie in de cel, de expressie niveaus, en de signaaltransductie. Hiermee wordt de plasticiteit van de *KIR* genen in primaten verder verruimd.

De *KIR* transcriptie studies geven voornamelijk een beeld van het polymorfisme, copy nummer variatie, chromosomale recombinatie, haplotype configuraties, en alternatieve splicing, maar de genomische structuur van het cluster blijft hiermee vooralsnog onduidelijk. Daarom hebben wij een Cas9-gemedieerde methode ontwikkelt om specifiek de *KIR* regio te kunnen verrijken en daarmee de genomische organisatie van *KIR* haplotypes in de mens en de makaak te kunnen bepalen (**Hoofdstuk 5**). Deze strategie maakte het mogelijk om de volgorde van *KIR* genen op een haplotype, de niet-coderende segmenten, zoals de intronen en promotor regio's, en de epigenetische profielen in kaart te brengen.

De dynamiek van de *KIR* regio in primaten berust op verschillende moleculaire mechanismes, welke tezamen een uitgebreid pallet aan diversiteit genereerd op individu, populatie en diersoort niveau (**Hoofdstuk 6**). Deze set aan mechanismes faciliteert een snelle evolutie en is geconserveerd in primaten. Dit wijst erop dat er een evolutionaire selectiedruk is om de *KIR* regio divers te houden. Met de in dit proefschrift verworven inzichten op het *KIR* gen profiel in primaten en met de verdere doorontwikkeling van nieuwe sequencing technieken, ontstaan er mogelijkheden om

de functionele implicaties van de dynamische KIR receptor familie gedetailleerder te onderzoeken (**Hoofdstuk 7**). Dit brengt een mogelijke vertaalslag van de kennis die wordt verkregen uit makaak studies naar een beter begrip van de menselijke NK cel biologie en gerelateerde ziektefenotypes.

Dankwoord

Meer dan vijf jaar aan werk is opgeschreven in dit proefschrift, wat tot stand is gekomen door het isoleren van DNA uit miljoenen humane en makaak cellen, het op volgorde zetten van zeker een miljard nucleotiden, en het gebruik maken van ontelbare uren aan rekenkracht. Al dit werk zou absoluut onmogelijk zijn geweest zonder de directe en indirecte steun van mensen uit mijn omgeving. Met dit dankwoord wil ik iedereen bedanken die mij in de afgelopen jaren heeft bijgestaan met raad en daad. Maar uiteraard is dit ook de plaats waar ik een aantal personen in het bijzonder kan uitlichten.

Allereerst wil ik jou bedanken, Ronald. Nog altijd vertel jij vol bewondering mooie verhalen over jouw toenmalige promotor, Prof. Dr. Jon van Rood. Als mijn promotor ben jij in zijn voetsporen getreden en kan ik op eenzelfde manier teren op jouw levendige anekdotes, wijze lessen en academische visie. Het vertrouwen dat je in mij stelde sinds de eerste dag waardeer ik enorm. Elke keer wanneer ik uit wetenschappelijke nieuwsgierigheid een nieuwe techniek of methode voorstelde, gaf je mij de ruimte om het in praktijk te brengen. Ondanks dat er in theorie geen verschil zit tussen theorie en praktijk, was dit in de praktijk vaak wel het geval. Dit hield het onderzoek in ieder geval uitdagend. Ondanks je drukke schema benadrukte je altijd de “open deur” cultuur, waarin ik kon binnen lopen om resultaten te bespreken of voor een simpel praatje over voetbal. Het werk op jouw afdeling kreeg nog extra glans door de diverse congressen, waar je mij introduceerde aan wetenschappers in het veld en waar ik ons werk mocht presenteren voor een groot publiek. In de afgelopen jaren heb ik onder jouw begeleiding mijn grenzen kunnen verleggen en ben ik ontwikkeld op persoonlijk en wetenschappelijk vlak. Met het vertrouwen dat je in mij uitspreekt door de aanstelling als vaste medewerker in de Bontrop-groep hoop ik op nog vele wetenschappelijke discussies, welke leiden tot mooie publicaties voor onze afdeling. Wie schrijft, die blijft…

Minstens zo belangrijk tijdens mijn promotie was jij, Natasja. Als de eerste PhD student die jij hebt begeleid kan ik je vertellen dat je trots mag zijn. In de eerste maanden heb jij met geduld mij wegwijs gemaakt in de moleculaire technieken en ook in de daaropvolgende jaren heb je mij altijd voorzien van praktisch advies. Elke keer als we weer een stapje dichterbij een mooi resultaat kwamen, kon ik bij jou binnen lopen om de data te bespreken en de vervolgstappen te bediscussiëren. Ik waardeer de kritische en gedetailleerde aanpak waarmee jij mijn manuscripten van commentaar hebt voorzien, wat vervolgens heeft geleid tot mooie publicaties. Sinds het vertrek van Gaby heb je laten zien dat je een goede manager bent op de afdeling, waar ik een voorbeeld kan nemen aan jouw gestructureerde en rustige manier van leiding geven. Gedurende het PhD traject zijn we steeds meer naar eenzelfde niveau gegroeid en ik kijk ernaar uit om samen met jou de afdeling verder door te ontwikkelen.

De grote hoeveelheid data dat is verwerkt in dit proefschrift heb ik natuurlijk nooit alleen kunnen genereren en verwerken. Daarvoor is een geolied en saamhorig team nodig dat voor elkaar klaar staat. Ik ben enorm dankbaar dat ik me bij dit team heb mogen voegen en dat ik heb mogen ervaren hoe jullie met plezier aan de slag gaan. Jullie bijdrage was niet louter wetenschappelijk… Met alle verhalen over jullie kinderen ben ik de meest ervaren kinderloze opvoedkundige geworden!

Marit, in de afgelopen jaren heb je veel bijgedragen aan het KIR onderzoek; eerst bij Jeroen en daarna met mij. Je beheerst de complexiteit van het gen systeem tot in detail en hierdoor kon je altijd kritisch meedenken over de praktische uitvoering van onze projecten. Ik bewonder hoe snel je nieuwe technieken oppakt en zelf aan de slag gaat met ingewikkelde data-analyses. Samen hebben we veel opgehelderd over de ingewikkelde wirwar aan genen en je mag trots zijn op jouw grote bijdrage aan dit proefschrift!

Het verbaast me dat jij niet duizelig wordt achter de computer, Nanine. De grote tabellen om de KIR haplotypes te definiëren zijn zeer waardevol en daarmee heb ook jij een grote bijdrage geleverd aan verschillende hoofdstukken van dit proefschrift. Er is geen analist die zo snel en zo nauwkeurig de PacBio data kan verwerken. Heel veel dank voor dit werk!

Op onze afdeling is het eigenlijk vanzelfsprekend dat de voorraad aan cellen en geïsoleerd DNA/RNA op orde is. Maar dit ligt er natuurlijk niet zomaar. Daarom wil ik ook jou bedanken, Annemiek, voor alle opwerkingen die jij in de afgelopen jaren hebt verricht.

Hoe zeer ik stilte en rust waardeer tijdens het schrijven van artikelen, was het fijn om toch een vast kamergenootje te krijgen. Lisanne, ik waardeer de gesprekken op de kamer en de discussies tijdens de PhD overleggen. Ook had ik zonder jou waarschijnlijk na vijf jaar nog steeds in een kale lege kamer gezeten, dus ik ben blij met de plantjes die ons zijn komen vergezellen! Nu nog de taak om ze in leven houden…

Corinne, ondanks ons werk misschien pas recentelijk wat meer overlap heeft gekregen nu je ook onderzoek doet naar het KIR systeem in de chimpansees, wil ik ook jou bedanken voor je input tijdens mijn onderzoek.

En dan nog de twee ex-CGR'ers. Allereerst Nel, je nuchtere blik op vraagstukken kon ik altijd erg waarderen en soms ook wel om lachen. Ik ben blij dat je me hebt geholpen met de vele allel submissies naar de database en ik hoop dat je inmiddels geniet van je pensioen! En ook jou, Gaby, wil ik bedanken voor de adviezen in de eerste jaren van mijn PhD. Als initiator van het PacBio sequencen op het BPRC heb je toch een hoeksteen gelegd van dit manuscript.

Al met al dus een heel waardevol team op de genetica afdeling, en zonder jullie was dit onderzoek nooit zo ver gekomen, en was het werken ook nooit zo fijn geweest. Bedankt! En hopelijk volgen er nog vele mooie onderzoeken samen als team.

Ondanks dat ik mijn promotietraject afsluit in een bijzondere tijd, waarbij contact met medewerkers van andere afdelingen minimaal was, wil ik ook deze collega's bedanken. Als eerst de groep post-docs die de tijd nam om actief mee te doen met de

Journal Club. Ondanks dat mijn cel-immunologische kennis soms niet even toereikend was, heb ik veel geleerd van deze bijeenkomsten. Dus dank daarvoor Ed, Frank, Bart, Gerrit, Michel, Marieke, Kirsta en Jeffrey. En daarmee wil ik ook meteen mijn dank richten tot mijn mede-lotgenoten: Karin, Jordon, Raissa, Lisanne, Annemarie, Kinga, Aafke, Astrid, en Tina, voor de gezellige etentjes in Delft en het delen van hun ervaringen. Het KIR onderzoek heeft veel gevraagd van onze computer systemen. Ik wil Doxx, en dan in het bijzonder Billy, enorm bedanken voor hun ondersteuning in het opzetten van de Linux/Windows computer waarop het Nanopore platform draait en alle support die ze daarbij hebben verleend. Verder nog mijn dank voor een aantal mensen die misschien wel het meeste inzicht hebben van het hele instituut. Want ondanks dat het BPRC gelegen is in een regio waar het slecht is gesteld met de kennis over voetbal, zijn er toch collega's die beter weten. Daarom ook mijn dank aan Dirk, Jeffrey en Paul, die altijd de tijd namen om de wedstrijden van onze landskampioen te analyseren. Uiteraard ook nog een dank aan Francisca. Je hebt met veel geduld al mijn figuren ontworpen, waarbij ik vaak terugkwam om nog de kleinste details te laten aanpassen. Heel erg bedankt voor de tijd die jij hebt gestoken in het ontwerpen van de figuren en het vormgeven van dit proefschrift.

The many congresses in the most beautiful places introduced me to inspiring scientists in the field of immunogenetics and NK cells. It is amazing to be part of this community and I am thankful for the warm welcome. Specifically, I would like to thank prof. Dr. Peter Parham, prof. Dr. Lutz Walter and Prof. Dr. Jim Kaufman for taking the time to read my thesis and to take part in the defense-committee. Uiteraard wil ik ook de andere twee professoren die mijn proefschrift willen beoordelen bedanken, prof. Dr. Edwin Cuppen en prof. Dr. Rob de Boer. It is an honor to defend my thesis in front of these well-established scientists.

Naast de zeer gewaardeerde collega's zou ik ook graag mijn dank willen richten aan vrienden en familie. Ondanks dat vele hun directe en indirecte bijdrage hebben gehad in het behalen van mijn doctoraat, wil ik een select aantal in het bijzonder noemen.

Remi, jij bent iemand die altijd voor iedereen klaar staat en je bent van grote waarde geweest in mijn periode als PhD'er. Ondanks dat je zelf moeilijke tijden hebt doorgemaakt liet je je nooit uit het veld slaan, en kon ik altijd bouwen op je realistische en positieve kijk op verschillende onderwerpen. De avonden met lekker eten, bordspelletjes, drank en je zwarte humor zijn geweldig. En hier hoort Sanne natuurlijk ook bij. Jouw kookkunsten, enthousiaste verhalen en zorgzame karakter maken je een bijzonder persoon. Dank dat ik altijd op jullie kan rekenen! En als vriend die ik al bijna 25 jaar ken, Dwayne, ben jij ook een belangrijk persoon geweest in de afgelopen jaren. Met je uitsloverij tijdens de wintersport en je soms wat rare streken ben je nog altijd het jochie van de Fliere fluiter. Bedankt voor de gezellige momenten bij jou en Esther thuis, en mooi om te zien hoe de kwajongen nu doorgaat als trotse vader. Ruben, een vriendschap die een gekke oorsprong kent, maar zich wel ontwikkelt tot iets waardevols. Als nieuwe mug in Haarlem is het fijn om soms een biertje te kunnen doen en simpelweg te kunnen zeuren op onzinnige dingen.

Soms gaan dingen in het leven ook heel anders dan verwacht... Toch wil ik jou, Esther, bedanken voor de grote steun die je bent geweest tijdens het doen van mijn onderzoek. Inmiddels weet je zelf hoe het is om een promotie traject te doorlopen en zal je begrijpen dat het soms lekker is om even op het onderzoek te kunnen zeuren. Het is fijn dat je in die periode een luisterend oor kon zijn en dat je me altijd motiveerde om door te gaan. De reisjes naar zonnige stranden en etentjes bij de voor ons welbekende restaurants waren mooie momenten om het werk te vergeten. Met jouw ambitie weet ik zeker dat je ook je eigen dromen zal najagen en uiteindelijk als zeergeleerde vrouw door het leven zal gaan.

Als laatste wil ik mijn familie bedanken. Mads en Demi, jullie beeld van mij, werkende in een witte jas op een laboratorium in een apencentrum, is natuurlijk al heel accuraat. Hopelijk kunnen jullie met het lezen van dit manuscript dat beeld nog wat verfijnen! Lieve papa en mama, jullie hebben altijd duidelijk laten blijken trots te zijn op mijn werk en vaak gevraagd naar nieuwe publicaties. Dit heeft mij erg gemotiveerd. Zoals jullie graag willen, en in tegenstelling tot mijn universitaire diploma's, zal ik deze mijlpaal wel vieren!

Curriculum Vitae

Jesse Bruijnesteijn was born on the 28th of February 1991 in Hoorn, the Netherlands. In 2009 he obtained his VWO diploma (Copernicus, Atlas College, Hoorn). In the same year he started his bachelor's in Biomedical Sciences at the University of Amsterdam (UvA). He complemented these years with extracurricular courses in Chemistry and World politics. Following his Bachelor's degree, he started the master Drugs Discovery and Safety: Analytical Chemistry at the Vrije Universiteit (VU) in Amsterdam. This programme involved an internship at the VU medical center (Amsterdam) during which he developed a method to define and quantify corticosteroids in the faeces of preterm neonates using a liquid chromatography and mass spectrometry (LC-MS) protocol. Next, he wrote a literature thesis in collaboration with MSD-Merck on the two commercially available PD-1 blockers (Pembrolizumab and Nivolumab) to treat advanced melanomas and non-small-cell lung cancer.

After his graduation, he started at the Biomedical Primate Research Centre (BPRC; Rijswijk, the Netherlands) as PhD candidate at the department of Comparative Genetics and Refinement. This study involved the comprehensive characterization of killer cell immunoglobulin-like receptors (KIR) in macaque species and humans.

List of publications

Jesse Bruijnesteijn, Marit K. H. van der Wiel, Wendy T. N. Swelsen, Nel Otting, Annemiek J. M. de Vos-Rouweler, Diënne Elferink, Gaby G. Doxiadis, Frans H. J. Claas, Neubury M. Lardy, Natasja G. de Groot and Ronald E. Bontrop. (2018) Human and Rhesus Macaque KIR Haplotypes Defined by Their Transcriptomes. J Immunol, 200 (5) 1692-1701.

Jesse Bruijnesteijn, Marit K. H. van der Wiel, Nanine de Groot, Nel Otting, Annemiek J. M. de Vos-Rouweler, Neubury M. Lardy, Natasja G. de Groot and Ronald E. Bontrop. (2018) Extensive Alternative Splicing of KIR Transcripts. Front Immunol 9, 2846.

Nanine de Groot; Rens Groen; Vaneesha Orie; **Jesse Bruijnesteijn**; Natasja G. de Groot; Gaby G.M. Doxiadis; Ronald E. Bontrop (2019). Analysis of macaque BTN3A genes and transcripts in the extended MHC : conserved orthologs of human γδ T cell modulators. Immunogenetics, volume 71, issue 8-9, pp. 545 – 559.

Jesse Bruijnesteijn, Natasja G. de Groot, Nel Otting, Giuseppe Maccari, Lisbeth A. Guethlein, James Robinson, Steven G. E. Marsh, Lutz Walter, David H. O'Connor, John A. Hammond, Peter Parham & Ronald E. Bontrop. (2019) Nomenclature report for killer-cell immunoglobulin-like receptors (KIR) in macaque species: new genes/alleles, renaming recombinant entities and IPD-NHKIR updates. Immunogenetics volume 72, issue 1-2, pp. 37-47.

Jesse Bruijnesteijn, Nanine de Groot, Marit K. H. van der Wiel, Nel Otting, Annemiek J. M. de Vos-Rouweler, Natasja G. de Groot and Ronald E. Bontrop. (2020) Unparalleled Rapid Evolution of KIR Genes in Rhesus and Cynomolgus Macaque Populations. J Immunol, 204 (7) 1770-1786.

Jesse Bruijnesteijn, Natasja G. de Groot and Ronald E. Bontrop. (2020) The Genetic Mechanisms Driving Diversification of the KIR Gene Cluster in Primates. Front. Immunol 11, 582804.

Natasja G. de Groot, Corrine M. C. Heijmans, Suzanne Bezstarosti, **Jesse Bruijnesteijn**, Geert W. Haasnoot, Arend Mulder, Frans H. J. Claas, Sebastiaan Heidt and Ronald E. Bontrop. (2021) Two Human Monoclonal HLA-Reactive Antibodies Cross-React with Mamu-B*008, a Rhesus Macaque MHC Allotype Associated with Control of Simian Immunodeficiency Virus Replication. J Immunol 206 (8), 1957-1965.

Jesse Bruijnesteijn, Marit van der Wiel, Natasja G. de Groot, and Ronald E. Bontrop. (2021) Rapid characterization of complex killer cell immunoglobulin-like (KIR) regions using Cas9 enrichment and Nanopore sequencing. Accepted for publication in Front. Immunol.