IPD—the Immuno Polymorphism Database

James Robinson[1], Jason A. Halliwell[1], Hamish McWilliam[2], Rodrigo Lopez[2] and Steven G. E. Marsh[1,3,*]

[1]Anthony Nolan Research Institute, Royal Free Hospital, Pond Street, Hampstead, London NW3 2QG, [2]External Services, EMBL Outstation–European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SD and [3]UCL Cancer Institute, University College London, Royal Free Campus, Pond Street, Hampstead, London NW3 2QG, UK

ABSTRACT

The Immuno Polymorphism Database (IPD), http://www.ebi.ac.uk/ipd/ is a set of specialist databases related to the study of polymorphic genes in the immune system. The IPD project works with specialist groups or nomenclature committees who provide and curate individual sections before they are submitted to IPD for online publication. The IPD project stores all the data in a set of related databases. IPD currently consists of four databases: IPD-KIR, contains the allelic sequences of killer-cell immunoglobulin-like receptors, IPD-MHC, a database of sequences of the major histocompatibility complex of different species; IPD-HPA, alloantigens expressed only on platelets; and IPD-ESTDAB, which provides access to the European Searchable Tumour Cell-Line Database, a cell bank of immunologically characterized melanoma cell lines. The data is currently available online from the website and FTP directory. This article describes the latest updates and additional tools added to the IPD project.

INTRODUCTION

The study of the immune system constitutes many different complex areas of research. The Immuno Polymorphism Database (IPD) is a set of specialist databases related to the study of polymorphic genes of the immune system. The IPD project works with specialist groups and/or nomenclature committees, which each curate a different section of the project. IPD currently consists of four databases: IPD-KIR, currently contains the allelic sequences of human killer-cell immunoglobulin-like receptors (KIR); IPD-MHC, a database of sequences of the major histocompatibility complex (MHC) of different species; IPD-HPA, alloantigens expressed only on platelets; and IPD-ESTDAB, which provides access to

the European Searchable Tumour Cell-Line Database (ESTDAB), a cell bank of immunologically characterized melanoma cell lines. By providing a centralized resource for the work of different groups, it is hoped we can bring together similar data to aid in the study and analysis of this area. The IPD project stores all the data in a set of related databases. Those sections with similar data, IPD-KIR and IPD-MHC share the same database structure. The sharing of a common database structure makes it easier to implement common tools for data submission and retrieval. Other unrelated sections like IPD-ESTDAB currently have their own unique structure.

IPD provides a large number of tools for the analysis of KIR and non-human MHC sequences. These tools are either custom written for the database or are incorporated into existing tools on the European Bioinformatics Institute (EBI) website, see Figure 1 (1,2).

These tools include the following:

- Sequence alignments: access to alignment tool, which filters pre-generated alignments to the users' specification. Provides alignments at the protein and level.
- Allele queries: access to detailed information on any allele, including information on database cross-references and seminal publications. This information is also available through integration in the Sequence Retrieval System service at EBI (3).
- Sequence similarity search tools: integration into EBI's suite of search tools including FASTA (4) and BLAST (5).
- Downloads: access to a directory containing all the data from the current and previous releases in a variety of commonly used formats like FASTA, MSF and PIR.

IPD-MHC

The MHC sequences of many different species have been reported (6–17), along with different nomenclature systems used in the naming and identification of new genes and alleles in each species (18). The sequences of

Figure 1. The IPD Homepage provides access to the different IPD projects.

the MHC from a number of different species are highly conserved between species (19). The nomenclature for MHC genes and alleles in species other than humans (20,21) and mice (22,23) has historically been overseen either informally by groups generating sequences, or by formal nomenclature committees set up by the International Society for Animal Genetics (ISAG) (24). This work is now overseen by the Comparative MHC Nomenclature Committee and is supported by ISAG and the Veterinary Immunology Committee of the International Union of Immunological Societies (25). By bringing the work of different nomenclature committees and the sequences of different species together, it is hoped to provide a central resource that will facilitate further research on the MHC of each species and on their comparison (26).

The first version of the IPD-MHC database involved the work of groups specializing in non-human primates (NHP) (16), canines (DLA) (12) and felines (FLA) (27) and incorporated all data previously available in the IMGT/MHC database (26). Since the first version, we have been able to add sequences from cattle (BoLA) (17), teleost fish (28), rats (RT1) (29), sheep (OLA) (15) and swine (SLA) (14). In 2012, the nomenclature used to

describe the alleles of NHP was extensively revised and updated (16). This was accompanied by updating the IPD-MHC NHP section to complement the publication; IPD-MHC NHP currently contains >4000 alleles covering 47 species of Apes, Old World and New World Monkeys. The management of the sequences within IPD-MHC and the provision of an online submission tool have enabled these databases to grow, the number of sequences increasing by at least 10% each year, and the nomenclature to expand since the inclusion of a species within IPD. This has resulted in regular publications reporting updates or changes to the nomenclature (15–17,30).

IPD-KIR

The KIRs are members of the immunoglobulin super family formerly called killer-cell inhibitory receptors. KIRs have been shown to be highly polymorphic both at the allelic and haplotypic levels (31). They are composed of two or three Ig-domains, a transmembrane region and cytoplasmic tail, which can in turn be short (activatory) or long (inhibitory). The leukocyte receptor complex, which encodes KIR genes, has been shown to be polymorphic, polygenic and complex in a manner similar to the MHC. Because of the complexity in the KIR region and KIR sequences, a KIR Nomenclature Committee was established in 2002, to undertake the naming of human KIR allele sequences. The first KIR Nomenclature report was published in 2003 (32), which coincided with the first release of the IPD-KIR database. The number of officially named human KIR alleles has increased since the initial release, which contained 89 alleles. As of September 2012, there are >600 alleles, which code for >320 unique protein sequences.

A recent development on the IPD-KIR website is online tools to assist in the prediction of transplant outcome in an unrelated haematopoetic stem cell transplant based on the KIR content of the individuals involved. In 2008, a tool was added to the website to help predict NK cell alloreactivity based on the KIR ligands present in the patient and donor, as transplant strategies based on KIR-ligand mismatches had been shown to influence relapse, graft versus host disease and survival in patients with acute myeloid leukaemia (33). In 2010, with the goal of developing a donor selection strategy to improve transplant outcome, Cooley *et al.,* (34) compared the contribution of KIR gene motifs with the clinical benefit conferred by donors with a particular haplotype. Donor KIR genotype influenced transplantation outcome for some forms of leukaemia, after a T-cell replete unrelated donor transplant. KIR genotyping several HLA-matched potential donors could substantially increase the frequency of transplants using unrelated donor grafts with favorable KIR gene content. To implement this strategy, the IPD-KIR database was asked to provide an online version of the algorithm described in the article, see Figure 2. The B-Content calculator (http://www.ebi.ac.uk/ipd/kir/donor_b_content.html) allows the user to enter the KIR genotypes for up to five prospective donors, and receive their B-Content assignments, and a prediction result of the effect of the KIR genotype on transplant outcome. To ensure only valid KIR genotypes are submitted, all genotypes submitted are compared with a list of predicted genotypes based on known KIR haplotypes. In addition, this list has been supplemented with a number of additional KIR genotypes that have been defined in routine KIR typing. If a prospective donor's KIR typing does not match any of the genotypes on this list, a warning is issued.

In addition to allelic polymorphism, there is haplotypic variability within the KIR region due to the different number and kind of KIR genes. Haplotypic diversity is a major contributor to the population diversity of KIR and of cell repertoires. There are a number of fully sequenced KIR haplotypes that have been reported (35,36). The IPD-KIR website collates this data into a graphic displaying the gene and allele content of these, Figure 3. Where possible the alleles sequenced at the individual KIR genes are also listed. Some genes are only partially sequenced and for these an allele designation is not provided. The graphic provides interactive links to the individual allele entries, as well as details on the cell source used for each haplotype.

The IPD-KIR database is also being expanded to include the KIR sequences from other species, and most recently work has begun on including the sequences of KIR alleles found in Rhesus Macaques (*Macaca mulatta*) (37,38). Work is also underway on a number of other species of Macaques as well as Chimpanzees, Orangutans and Domestic Cattle (39–42). The non-human KIR sequences will be included into the IPD-KIR section and be accessible using the same tools as the human KIR sequences. Submission tools are already available for depositing non-human KIRs into the database.

IPD-HPA

Human platelet antigens (HPA) are alloantigens expressed only on platelets, specifically on platelet membrane glycoproteins. These platelet-specific antigens are immunogenic and can result in pathological reactions to transfusion therapy. The HPA nomenclature system was adopted in 1990 (43,44) to overcome problems with the previous nomenclature. Since then, more antigens have been described and the molecular basis of many has been resolved. As a result, the nomenclature was revised in 2003 (45) and included in the IPD project. The IPD-HPA section contains nomenclature information and additional background material. The different genes in the HPA system have not been sequenced to the same level as some of the other projects and so currently only single-nucleotide polymorphisms are used to determine alleles. This information is presented in a grid of single-nucleotide polymorphisms for each gene. The IPD and HPA nomenclature committee hope to expand this to provide full sequence alignments when possible.

IPD-ESTDAB

IPD-ESTDAB is a database of immunologically characterized melanoma cell lines. The database works in conjunction with the European Searchable Tumour Cell Line Database (ESTDAB) cell bank (46,47), which

Figure 2. The B-Content Calculator allows you to enter the genotypes for up to five prospective donors and receive their assignments to one of three groups based on KIR B-Content. The groups refer to the associated relapse protection seen in a T-cell replete Unrelated Donor Hematopoietic Cell Transplant for Acute Myeloid Leukaemia.
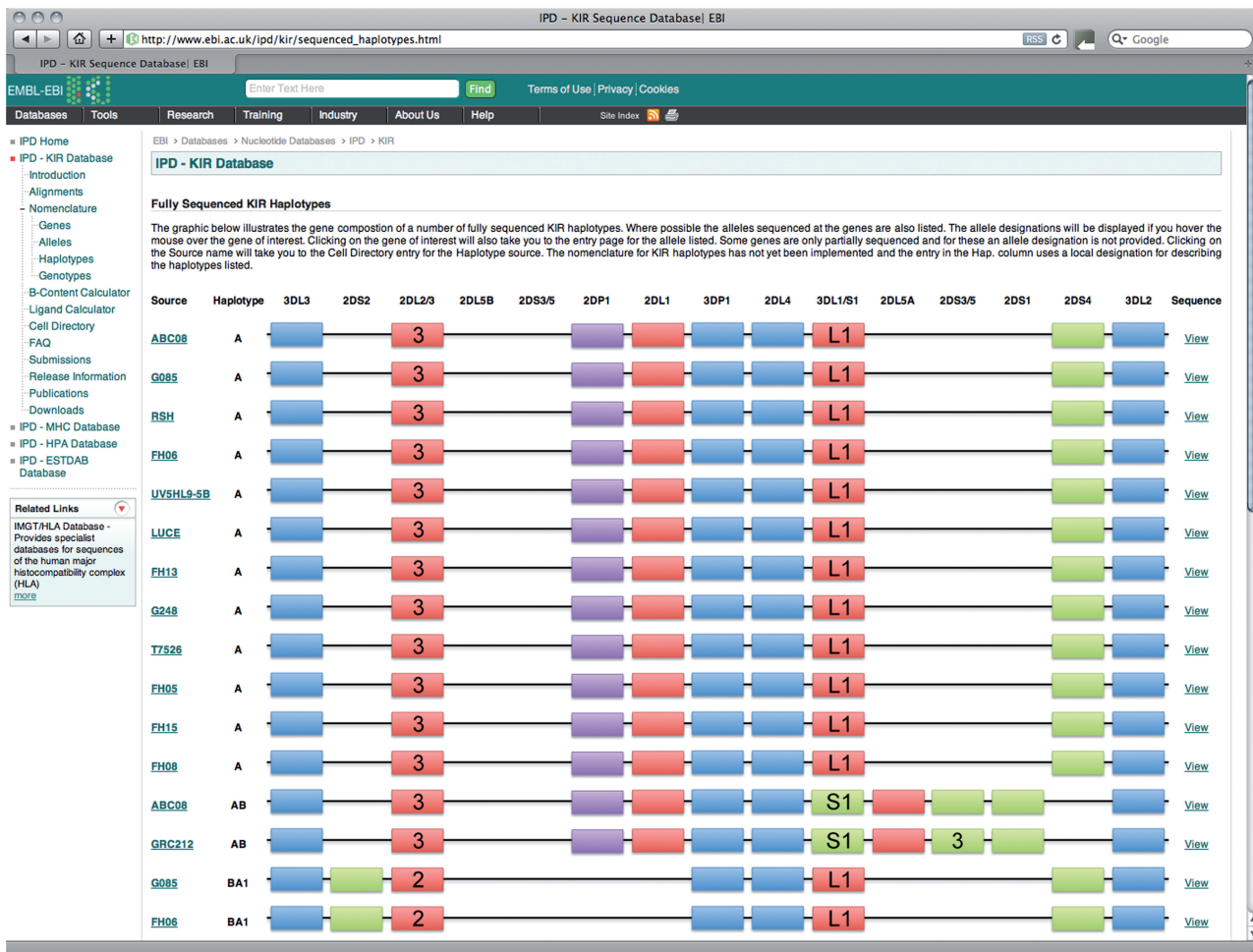
Figure 3. A selection of fully sequenced KIR haplotypes, the IPD-KIR website provides an interactive graphic that illustrates the gene composition of a number of fully sequenced KIR haplotypes.

is housed in Tübingen, Germany and provides immunologically characterized tumour cells. The IPD-ESTDAB section of the website provides an online search facility for cells stored in this cell bank. This enables investigators to identify cells possessing specific parameters important for studies of immunity, immunogenetics, gene expression, metastasis, response to chemotherapy and other tumour biological experimentation. The search tool allows for searches based on a single parameter, or clusters of parameters on >250 different markers for each cell. The detailed reports produced can then be used to identify cells of interest, which can be obtained from the cell bank.

OTHER DEVELOPMENTS

In 2012, the IMGT/HLA Database added an Extensible Markup Language (XML) export to the data formats available (20). XML is a simple but flexible language that defines a set of rules for encoding documents in a format that is both human-readable and machine-readable. Designed to meet the challenges of large-scale electronic publishing, XML is playing an increasingly important role in the exchange of scientific data. Work

is currently underway to also produce XML formatted data files for the IPD-KIR and IPD-MHC database.

DISCUSSION

The IPD project provides a resource for those interested in the study of polymorphic sequences in the immune system. By accommodating related systems in a single database, data can be made available in common formats aiding use and interpretation. As the projects grow and more sections are added, the benefit of having expertly curated sequences from related areas stored in a single location is becoming more apparent. This is particularly true of the IPD-MHC project, where cross-species studies are able to use the high-quality sequences provided by the different nomenclature committees in a common standardized format, ready for use. The initial release of the IPD Database contained only four sections and a small number of tools; however, as the database has grown and more sections and species have been added, more tools have been added to the website. We plan to use the existing database structures to house data for new sections of the IPD project as they become available. The files will also be made available in different formats to download from

the website, FTP server and included different web services at the EBI (1).

AVAILABILITY

ACKNOWLEDGEMENTS

FUNDING

REFERENCES

1. McWilliam,H., Valentin,F., Goujon,M., Li,W., Narayanasamy,M., Martin,J., Miyar,T. and Lopez,R. (2009) Web services at the European Bioinformatics Institute-2009. *Nucleic Acids Res.*, **37**, W6–W10.
2. Goujon,M., McWilliam,H., Li,W., Valentin,F., Squizzato,S., Paern,J. and Lopez,R. (2010) A new bioinformatics analysis tools framework at EMBL-EBI. *Nucleic Acids Res.*, **38**, W695–W699.
3. Etzold,T., Ulyanov,A. and Argos,P. (1996) SRS: information retrieval system for molecular biology data banks. *Methods Enzymol.*, **266**, 114–128.
4. Pearson,W.R. and Lipman,D.J. (1988) Improved tools for biological sequence comparison. *Proc. Natl Acad. Sci. USA*, **85**, 2444–2448.
5. Altschul,S.F., Madden,T.L., Schaffer,A.A., Zhang,J., Zhang,Z., Miller,W. and Lipman,D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, **25**, 3389–3402.
6. Longenecker,B.M. and Mosmann,T.R. (1981) Nomenclature for chicken MHC (B) antigens defined by monoclonal antibodies. *Immunogenetics*, **13**, 25–28.
7. Briles,W.E., Bumstead,N., Ewert,D.L., Gilmour,D.G., Gogusev,J., Hala,K., Koch,C., Longenecker,B.M., Nordskog,A.W., Pink,J.R. et al. (1982) Nomenclature for chicken major histocompatibility (B) complex. *Immunogenetics*, **15**, 441–447.
8. (1991) Leukocyte antigens in cattle, sheep and goats. Nomenclature. *Vet. Immunol. Immunopathol.*, **27**, 15–16.
9. Davies,C.J., Andersson,L., Joosten,I., Mariani,P., Gasbarre,L.C. and Hensen,E.J. (1992) Characterization of bovine MHC class II polymorphism using three typing methods: serology, RFLP and IEF. *Eur. J. Immunogenet.*, **19**, 253–262.
10. Naessens,J. (1993) Leukocyte antigens of cattle and sheep Nomenclature. *Vet. Immunol. Immunopathol.*, **39**, 11–12.
11. Kennedy,L.J., Altet,L., Angles,J.M., Barnes,A., Carter,S.D., Francino,O., Gerlach,J.A., Happ,G.M., Ollier,W.E., Polvi,A. et al. (2000) Nomenclature for factors of the dog major histocompatibility system (DLA), 1998: first report of the ISAG DLA Nomenclature Committee. *Anim. Genet.*, **31**, 52–61.
12. Kennedy,L.J., Angles,J.M., Barnes,A., Carter,S.D., Francino,O., Gerlach,J.A., Happ,G.M., Ollier,W.E., Thomson,W. and Wagner,J.L. (2001) Nomenclature for factors of the dog major histocompatibility system (DLA), 2000: second report of the ISAG DLA Nomenclature Committee. *Anim. Genet.*, **32**, 193–199.
13. Miller,M.M., Bacon,L.D., Hala,K., Hunt,H.D., Ewald,S.J., Kaufman,J., Zoorob,R. and Briles,W.E. (2004) 2004 Nomenclature for the chicken major histocompatibility (B and Y) complex. *Immunogenetics*, **56**, 261–279.
14. Smith,D.M., Lunney,J.K., Ho,C.S., Martens,G.W., Ando,A., Lee,J.H., Schook,L., Renard,C. and Chardon,P. (2005) Nomenclature for factors of the swine leukocyte antigen class II system, 2005. *Tissue Antigens*, **66**, 623–639.
15. Ballingall,K.T., Herrmann-Hoesing,L., Robinson,J., Marsh,S.G.E. and Stear,M.J. (2011) A single nomenclature and associated database for alleles at the major histocompatibility complex class II DRB1 locus of sheep. *Tissue Antigens*, **77**, 546–553.
16. de Groot,N.G., Otting,N., Robinson,J., Blancher,A., Lafont,B.A., Marsh,S.G.E., O'Connor,D.H., Shiina,T., Walter,L., Watkins,D.I. et al. (2012) Nomenclature report on the major histocompatibility complex genes and alleles of Great Ape, Old and New World monkey species. *Immunogenetics*, **64**, 615–631.
17. Hammond,J.A., Marsh,S.G.E., Robinson,J., Davies,C.J., Stear,M.J. and Ellis,S.A. (2012) Cattle MHC nomenclature: is it possible to assign sequences to discrete class I genes? *Immunogenetics*, **64**, 475–480.
18. Klein,J., Bontrop,R.E., Dawkins,R.L., Erlich,H.A., Gyllensten,U.B., Heise,E.R., Jones,P.P., Parham,P., Wakeland,E.K. and Watkins,D.I. (1990) Nomenclature for the major histocompatibility complexes of different species: a proposal. *Immunogenetics*, **31**, 217–219.
19. Parham,P. (1999) Virtual reality in the MHC. *Immunol. Rev.*, **167**, 5–15.
20. Robinson,J., Halliwell,J.A., McWilliam,H., Lopez,R., Parham,P. and Marsh,S.G.E. (2013) The IMGT/HLA Database. *Nucleic Acids Res.*, **41**, D1222–D1227.
21. Marsh,S.G.E., Albert,E.D., Bodmer,W.F., Bontrop,R.E., Dupont,B., Erlich,H.A., Fernandez-Vina,M., Geraghty,D.E., Holdsworth,R., Hurley,C.K. et al. (2010) Nomenclature for factors of the HLA system, 2010. *Tissue Antigens*, **75**, 291–455.
22. Rodgers,J.R., Levitt,J.M., Cresswell,P., Lindahl,K.F., Mathis,D., Monaco,J.T., Singer,D.S., Ploegh,H.L. and Bryant,P.W. (1999) A nomenclature solution to mouse MHC confusion. *J. Immunol.*, **162**, 6294.
23. Eppig,J.T., Blake,J.A., Bult,C.J., Kadin,J.A. and Richardson,J.E. (2012) The Mouse Genome Database (MGD): comprehensive resource for genetics and genomics of the laboratory mouse. *Nucleic Acids Res.*, **40**, D881–D886.
24. Ellis,S.A., Bontrop,R.E., Antczak,D.F., Ballingall,K., Davies,C.J., Kaufman,J., Kennedy,L.J., Robinson,J., Smith,D.M., Stear,M.J.

et al. (2006) ISAG/IUIS-VIC Comparative MHC Nomenclature Committee report, 2005. *Immunogenetics*, **57**, 953–958.

25. Ballingall,K.T. (2012) Progress of the Comparative MHC Committee and a summary of the Comparative MHC Workshops held at the 32nd ISAG, Edinburgh and the 9th IVIS, Tokyo, 2010. *Vet. Immunol. Immunopathol.*, **148**, 202–208.

26. Robinson,J., Waller,M.J., Parham,P., de Groot,N., Bontrop,R., Kennedy,L.J., Stoehr,P. and Marsh,S.G.E. (2003) IMGT/HLA and IMGT/MHC: sequence databases for the study of the major histocompatibility complex. *Nucleic Acids Res.*, **31**, 311–314.

27. Drake,G.J., Kennedy,L.J., Auty,H.K., Ryvar,R., Ollier,W.E., Kitchener,A.C., Freeman,A.R. and Radford,A.D. (2004) The use of reference strand-mediated conformational analysis for the study of cheetah (*Acinonyx jubatus*) feline leucocyte antigen class II DRB polymorphisms. *Mol. Ecol.*, **13**, 221–229.

28. Lukacs,M.F., Harstad,H., Bakke,H.G., Beetz-Sargent,M., McKinnel,L., Lubieniecki,K.P., Koop,B.F. and Grimholt,U. (2010) Comprehensive analysis of MHC class I genes from the U-, S-, and Z-lineages in Atlantic salmon. *BMC Genomics*, **11**, 154.

29. Fujii,H., Kakinuma,M., Yoshiki,T. and Natori,T. (1991) Polymorphism of the class II gene of rat major histocompatibility complex, RT1: partial sequence comparison of the first domain of the RT1.B beta 1 alleles. *Immunogenetics*, **33**, 399–403.

30. Ho,C.S., Lunney,J.K., Ando,A., Rogel-Gaillard,C., Lee,J.H., Schook,L.B. and Smith,D.M. (2009) Nomenclature for factors of the SLA system, update 2008. *Tissue Antigens*, **73**, 307–315.

31. Garcia,C.A., Robinson,J., Guethlein,L.A., Parham,P., Madrigal,J.A. and Marsh,S.G. (2003) Human KIR sequences 2003. *Immunogenetics*, **55**, 227–239.

32. Marsh,S.G., Parham,P., Dupont,B., Geraghty,D.E., Trowsdale,J., Middleton,D., Vilches,C., Carrington,M., Witt,C., Guethlein,L.A. *et al.* (2003) Killer-cell immunoglobulin-like receptor (KIR) nomenclature report, 2002. *Tissue Antigens*, **62**, 79–86.

33. Ruggeri,L., Capanni,M., Casucci,M., Volpi,I., Tosti,A., Perruccio,K., Urbani,E., Negrin,R.S., Martelli,M.F. and Velardi,A. (1999) Role of natural killer cell alloreactivity in HLA-mismatched hematopoietic stem cell transplantation. *Blood*, **94**, 333–339.

34. Cooley,S., Weisdorf,D.J., Guethlein,L.A., Klein,J.P., Wang,T., Le,C.T., Marsh,S.G., Geraghty,D., Spellman,S., Haagenson,M.D. *et al.* (2010) Donor selection for natural killer cell receptor genes leads to superior survival after unrelated transplantation for acute myelogenous leukemia. *Blood*, **116**, 2411–2419.

35. Mungall,A.J., Palmer,S.A., Sims,S.K., Edwards,C.A., Ashurst,J.L., Wilming,L., Jones,M.C., Horton,R., Hunt,S.E., Scott,C.E. *et al.* (2003) The DNA sequence and analysis of human chromosome 6. *Nature*, **425**, 805–811.

36. Pyo,C.W., Guethlein,L.A., Vu,Q., Wang,R., Abi-Rached,L., Norman,P.J., Marsh,S.G.E., Miller,J.S., Parham,P. and Geraghty,D.E. (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.

37. Blokhuis,J.H., van der Wiel,M.K., Doxiadis,G.G. and Bontrop,R.E. (2009) Evidence for balancing selection acting on KIR2DL4 genotypes in rhesus macaques of Indian origin. *Immunogenetics*, **61**, 503–512.

38. 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**, 640–648.

39. 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**, 6301–6308.

40. 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**, 491–504.

41. Guethlein,L.A., Abi-Rached,L., Hammond,J.A. and Parham,P. (2007) The expanded cattle KIR genes are orthologous to the conserved single-copy KIR3DX1 gene of primates. *Immunogenetics*, **59**, 517–522.

42. 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**, 135–146.

43. von dem Borne,A.E. and Decary,F. (1990) Nomenclature of platelet-specific antigens. *Hum. Immunol.*, **29**, 1–2.

44. von dem Borne,A.E. and Decary,F. (1990) ICSH/ISBT Working Party on platelet serology. Nomenclature of platelet-specific antigens. *Vox Sang.*, **58**, 176.

45. Metcalfe,P., Watkins,N.A., Ouwehand,W.H., Kaplan,C., Newman,P., Kekomaki,R., De Haas,M., Aster,R., Shibata,Y., Smith,J. *et al.* (2003) Nomenclature of human platelet antigens. *Vox Sang.*, **85**, 240–245.

46. Pawelec,G. and Marsh,S.G. (2006) ESTDAB: a collection of immunologically characterised melanoma cell lines and searchable databank. *Cancer Immunol. Immunother.*, **55**, 623–627.

47. Robinson,J., Roberts,C.H., Dodi,I.A., Madrigal,J.A., Pawelec,G., Wedel,L. and Marsh,S.G. (2009) The European searchable tumour line database. *Cancer Immunol. Immunother.*, **58**, 1501–1506.