# Studies on polymorphic Alu insertions and genomic diversity within the Major Histocompatibility Complex

**David Suliman Dunn BSc (Hons), MSc.**

**This thesis is presented for the degree of Doctor of Philosophy of Murdoch University, 2005.**

## Declaration

I declare that this thesis is my own account of my research and contains as its main content work which has not previously been submitted for a degree at any tertiary education institution.

.....

David Suliman Dunn

## Dedication

In loving memory of my late mother (MARIAM), the light of my academic achievements, and to my late father (Dixon Suliman) **.**

## Acknowledgments

The work described in this thesis has only been possible through the continued support of my colleagues and collaborators to whom I am extremely grateful.

I am indebted to my supervisors Professor Jerzy (Yurek) Kulski and Professor Matthew Bellgard for their intellectual support, guidance and encouragement and for the opportunity to pursue the work described in this thesis.

I am also grateful for the assistance from the many collaborators; Professor Roger Dawkins, Dr Silvana Gaudieri, Dr Guan Tay, Professor Chanvit Leelayuwat, Mr Joseph Williamson and Dr Sonia Cattely for the initial studies on MICA, Professor Frank Christiansen and Dr Campbell Witt for their willingness to share DNA samples and HLA data, Dr Shiina, Professor Hiditoshi Inoko, Dr Naruse, Dr M Ota, Dr Munkhbat, Dr Maude Phipps, Dr Brian Tait and Dr Himladevi Soodyall for supplying DNA and related data.

My sincere gratitude to my colleagues in the various research groups: Jennie Hui, Steve Iaschi, Nikitas Economou, David Sayer, David Scibechi, Adam Hunter, William Kenworthy, Paula Moolhauzjen and Kim Carter for their friendship and assistance.

This work was undertaken while holding an Australian Postgraduate Award Scholarship (January 2001 to December 2003) and a supplement research award from the Centre for Biological Computing and Bioinformatics (January 2001 to September 2003). The laboratory work was carried out in the laboratory of the State Agricultural Biological Centre (SABC) at Murdoch University.

**Abstract**

After the initiation of the human genome sequencing project and the introduction of the field of 'bioinformatics', interest in human genetic diversity studies has been increased. Sequence diversity has helped define differences between genes and genomic regions that were previously unknown or difficult to determine. In this thesis I have undertaken to study sequence diversity in the human genome in three areas; 1) investigated diversity in the MHC as represented by the MICA alleles with respect to the known HLA alleles, 2) investigated the structure and diversity in the intergenic region from an MHC related (paralogous) genomic region and related the structural and diversity findings to the knowledge available on the MHC and the wider genome, and 3) described the identification of three and characterization of five new MHC class I polymorphic markers (Alu) and their polymorphic characteristics in worldwide populations and their associations with skin cancer.

1. Phylogenetic analysis of MICA alpha-domain (extracellular) sequences demonstrated relationships with HLA-B cross-reactive serogroups. The HLA-B and MICA loci are in linkage disequilibrium. The data indicated that MICA and HLA-B have evolved in concert from their common ancestors and that the transmembrane polymorphisms have arisen independently and more recently.
2. Sequence analysis of the CD1 genomic region confirmed the presence of five CD1 genes and revealed that there are four unrelated intergenic regions (IGRs). The IGRs are composed mostly of retroelements including five full-length L1 PA sequences and various pseudogenes. Genomic and phylogenetic analyses support the view that the human CD1 gene copies were duplicated prior to the evolution of primates and the bulk of the HLA class I genes found in humans.
3. Five polymorphic Alu insertions (POALINs) were identified (two from previous studies) and located within the 1.8 megabase of the MHC class I genomic region. All five POALINs are polymorphic, and are positively associated with the HLA-A and HLA-B alleles. The AluyHJ insertion was found most frequently associated with *HLA-A1* or *A24*, AluyHG with *HLA-A2*, AluyHF with *HLA-A2*, *A-10* or *-A26* and AluyTF showed a marginal association with *HLA-A29*. The *AluyMICB* insertion was strongly associated with *HLA-B17* (*HLA-B57, HLA-B58*) and *HLA-B13*. The presence of three Alu insertions

(AluyHJ, AluyHG and AluyHF) was found in only one HLA class I haplotype (*HLA-A1, -B57, -Cw6*) in the 10th IHW cell lines. A novel positive association between the presence of AluyMICB and the '*MICAdel/MICBnull/HLA-B48'* haplotype was determined. The AluyMICB insertion was also associated with at least three different MICB alleles (**0102, *0107N* and **0105*) and three different HLA-B alleles (*B13, B48* and *B57*). Based on the analysis of associations between different polymorphic markers within the beta block, the *MICB*0102* allele was inferred to be the ancestral form of the *MICB*0105* and *MICB*0107N* alleles. The AluyMICB polymorphism can be used to further investigate haplotype relationship and consequently their lineage origins. Some of the MHC POALINs are haplospecific and associate strongly with certain groups of HLA class I alleles and MHC ancestral haplotypes. The AluyTF frequency was significantly associated with skin cancer ($p<0.005$).

MICA gene diversity is derived from two different evolving paths, therefore one or the other alone cannot reliably mark an ancestral haplotype. The CD1 duplicons originated well before the HLA class I duplicons. The MHC POALINs provide new lineage and linkage markers for the fine mapping study of different haplotypes and variations in linkage groups across 1.8 Mb of the MHC class I region. The POALINs may also prove useful in investigating the origins and history of human populations and in determining the role of human genetic diversity in disease risk.

**TABLE OF CONTENTS**

**ABBREVIATIONS**

| | |
|---|---|
| 4AOH | Fourth Asia Oceania Histocompatibility Workshop |
| 10thIHW | Tenth International Hitocompatibility Workshop |
| AH | Ancestral Haplotype |
| Alu | *Arthrobacter luteus* restriction enzyme (*AluI*) cleaved conserved insertion sequence |
| BLAST | Basic local alignment tool |
| BLUR | Bam-linked ubiquitous repeat |
| bp | base pair |
| CD1 | Cluster domain 1 |
| CEH | Conserved extended haplotype |
| CEPH | Centre d'Etude du Polymorphisme Humain |
| CREG | Cross-reactive group |
| DDR | Discoidin domain receptor |
| DNA | Deoxyribonucleic acid |
| EBV | Epstein-Barr virus |
| FcRn | Neonatal Fc receptor molecule |
| GvHD | Graft-versus-host disease |
| HERV | Human endogenous retrovirus |
| HSP | Heat shock protein |
| HLA | Human leucocyte antigen |
| H-W | Hardy-Weinberg equilibrium |
| IGRS | Intervening genomic region |
| Indel | Insertion/deletion |
| IVS | intervening sequence (region) |
| Kb | Kilobase |
| LD | Linkage disequilibrium |
| LINE | Long interspersed element |
| LTR | Long terminal repeat |
| MEGA | Molecular evolutionary genome analysis |
| MER | Medium reiterated frequency repeat |
| MHC | Major histocompatibility complex |

| | |
|---|---|
| MIC | MHC class I chain related |
| MIR | Medium interspersed repeat |
| MLR | Mixed lymphocyte reaction |
| mya | Million years ago |
| PCR | Polymerase chain reaction |
| POALIN | Polymorphic Alu insertion |
| RFLP | Restriction fragment length polymorphism |
| RNA | Ribonucleic acid |
| SBT | Sequence-based typing |
| SINE | Short interspersed elements |
| SSCP | Single strand conformational polymorphism |
| SSOP | Sequence-specific oligonucleotide probing |
| SSP | Sequence-specific priming |
| SVA | composite retrotransposon of SINE, VNTR and Alu |
| TE | Transposable element |
| TNF | Tumor necrosis factor |
| tRNA | Transfer ribonucleic acid |
| VNTR | Variable number tandem repeat |

## Glossary

| | |
|---|---|
| Allo-antisera | Different human or different animal sera containing antibodies that are specific for one or more antigens. |
| Diversity | The presence of a wide range of variation in the qualities or attributes under discussion. |
| Haplotype | Combination of alleles at several linked loci on the same chromosome which are inherited *en block* from a parent. |
| Isozyme | A different electrophoretic form of the same multi-subunit enzyme. Unlike allozymes, isozymes are due to differing subunit configurations rather than allelic differences. |
| Panmixis | Random mating within a breeding population. |
| Paralogous | Two genes or clusters of genes at different chromosomal locations in the same organism that have structural similarities indicating that they derived from a common ancestral gene and has since diverged from the parent copy by mutation and selection or drift. |
| Paralogy | Relationship between two genes related through gene duplication. |
| Retroelement | A genetic element that transposes via an RNA intermediate. |
| Submetacentric | Having the centromere near the center of a chromosome but not in the middle, so that one arm is shorter than the other. |
| Transposable element | A segment of genetic material that is capable of changing its location in the genome |