Hui Yee Greenaway, Computational Biology Group, Centre for Vascular Research, University of New South Wales, NSW, Australia
Benedict Ng, Computational Biology Group, Centre for Vascular Research, University of New South Wales, NSW, Australia
David A. Price, Institute of Infection and Immunity, Cardiff University School of Medicine, Cardiff, Wales, UK
Daniel C. Douek, Human Immunology Section, Vaccine Research Center, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland, USA.
Miles P. Davenport, Complex Systems in Biology Group, Centre for Vascular Research, University of New South Wales, NSW, Australia
Vanessa Venturi, Computational Biology Group, Centre for Vascular Research, University of New South Wales, NSW, Australia

A computational biology approach predicts that NKT and MAIT invariant TCRα sequences can be produced efficiently by VJ gene recombination.

T cells express receptors on their surface that enable the recognition of pathogen peptides. These T cell receptors (TCRs) are comprised of two polypeptide chains, the α- and β-chains. A diverse repertoire of TCR α- and β-chains is produced in the thymus by recombination of an individual's original (i.e. germline) TCR genes. The germline genes involved in this process include the variable (V), diversity (D; for β-chain only), and joining (J) genes. The cleavage of nucleotides from, and the addition of nucleotides to, the end of the gene segments generates additional TCR diversity. The process of gene recombination can generate an enormous potential diversity of TCRs in the thymus (eg. $>10^{18}$ in humans), which greatly exceeds the number of T cells found in the peripheral T cell repertoire of an individual at any given time (eg. $\sim10^{12}$ in humans). The amino acid sequence across the V-(D)-J gene segment junction, referred to as the CDR3, is an important determinant of the TCR's ability to interact with a pathogen peptide.

Natural killer T (NKT) and mucosal-associated invariant T (MAIT) cells are specialized, highly effective subsets of T cells with various roles in immunity. Both NKT and MAIT cells typically express semi-invariant TCRs that are comprised of an invariant TCR α-chain. NKT and MAIT TCR α-chains use invariant V and J gene combinations and feature ubiquitous canonical CDR3α amino acid sequences across the VJ junction that are dominant in a majority of individuals and highly similar across species. The prevalence and evolutionary conservation of the NKT and MAIT TCRs suggest that they play an important role in the immune system and thus it is surprising that their production is left to chance by the largely random gene recombination process. In this study, we use a computational biology approach, involving bioinformatics analysis and computer simulations of the gene recombination process, to investigate whether the efficiency of the production of the NKT and MAIT TCR α-chains could explain their prevalence across individuals and species. We surveyed studies reporting NKT and MAIT TCRα sequences for a variety of species. For all reported species, the NKT and MAIT invariant TCRα amino acid sequences can be encoded by at least one germline-derived nucleotide sequence, requiring no random nucleotide additions. Moreover, an "overlap" between the Vα and Jα genes enables nucleotides from either of the Vα or Jα genes to contribute to the formation of codons at the VJ junction. Consequently, the invariant TCRα amino acid sequences can be produced by a large variety of recombination mechanisms through a process of convergent recombination. In computer simulations of a random recombination process involving the invariant NKT and MAIT TCRα gene combinations, the human and mouse NKT and MAIT invariant TCRα amino acid sequences were the most generated of all sequences conforming to the CDR3α length restrictions associated with NKT and MAIT cells. These results suggest that the highly efficient production of the NKT and MAIT invariant TCRα sequences is an important determinant of their prevalence within individuals, across individuals, and across species.