Distinguishing functional polymorphism from random variation in the sequences of $> 10,000$ $HLA-A,-B$, and $-C$ alleles
Notes

Robinson, J. Guithlein, L. Cereb, N. Yang, S. Norman, P. Marsh, S. Parham, P.

March 1, 2020

1 Introduction

Recent clinical application of sequence based HLA typing has uncovered an unprecedented number of novel HLA class I alleles. Most alleles of HLA class I (¿80%) are very rare, often identified in one person or family, and differ by pointmutation from older and more common alleles. 3 regions within the HLA region, $HLA-A$, $HLA-B$, and $HLA-C$ encode highly polymorphic HLA class I molecules. There exists a large HLA sequence database that contains over 10,000 alleles.

2 Set of core alleles represents HLA polymorphism

Removing SNP and recombinant alleles reduces HLA class 1 variability to 11 $HLA-A$, 17 $HLA-B$, and 14 $HLA-C$ alleles that hold all significant variation in exons 2 and 3 of $HLA-A,-B,-C$.