HLA Genetics and Nomenclature

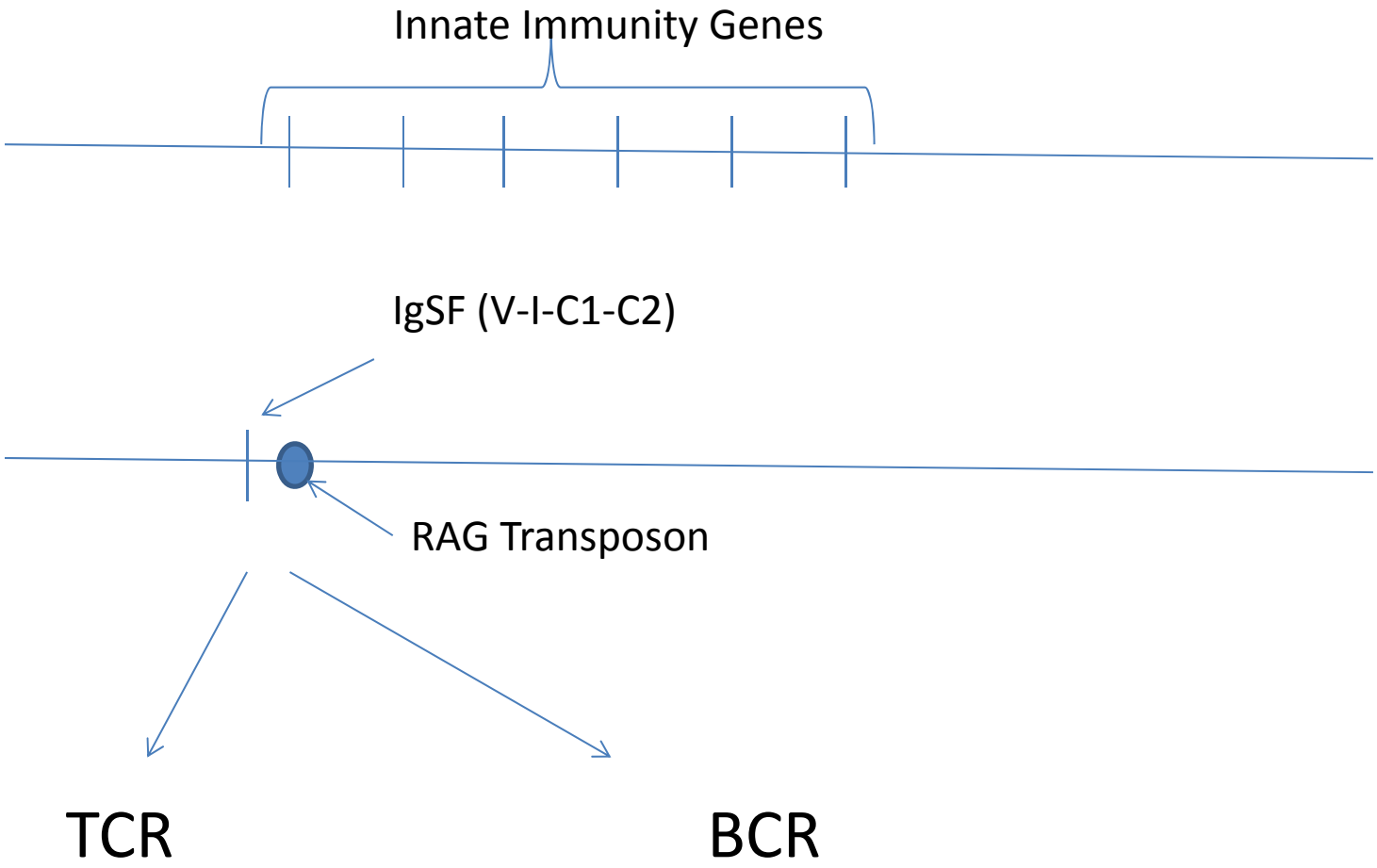
10th INTERNATIONAL SUMMER SCHOOL ON IMMUNOGENETICS

Stintino, Sardinia, 16 September, 2013

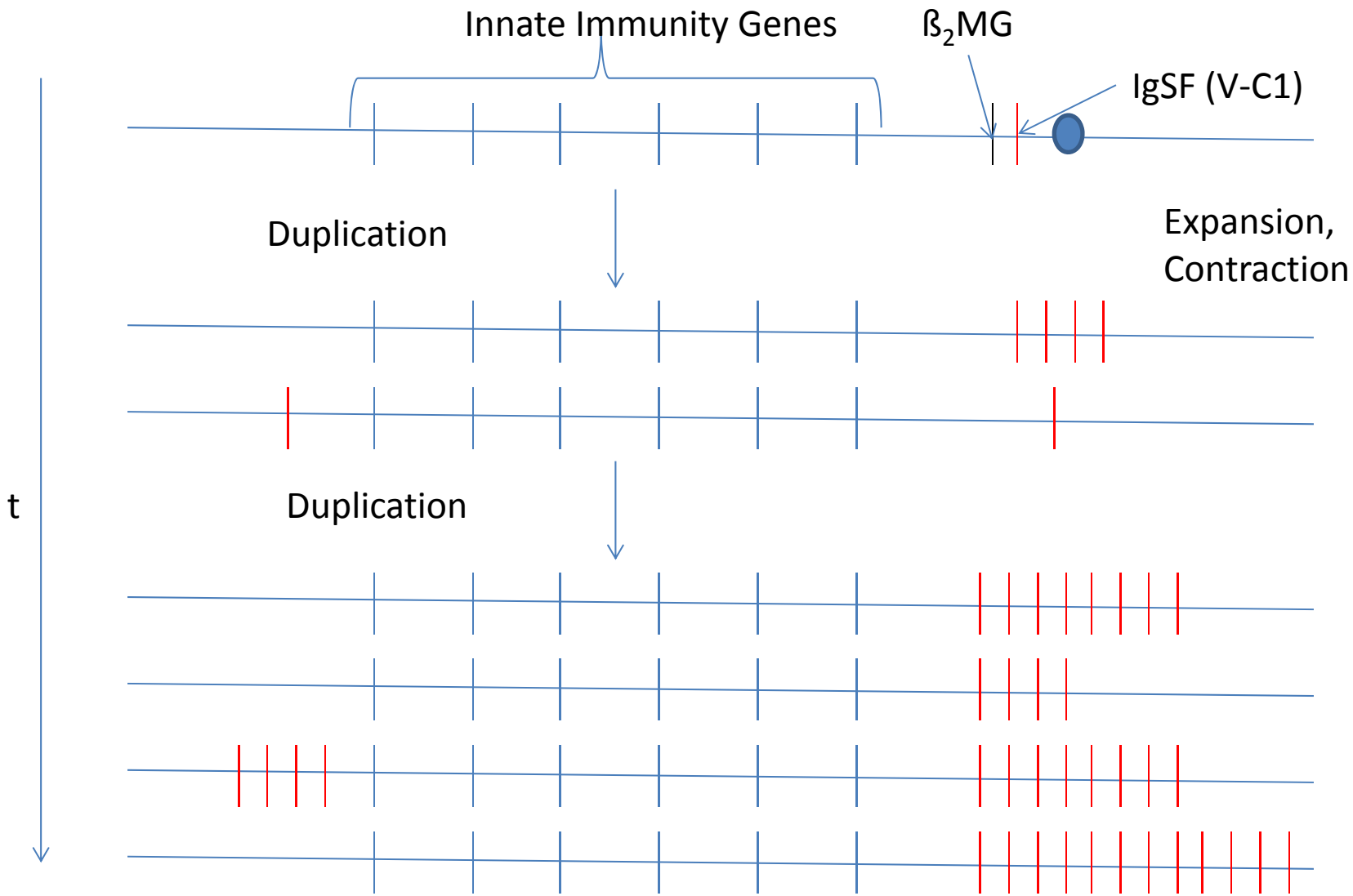
Outline

- The annotated human MHC
- Nomenclature
- What is left?

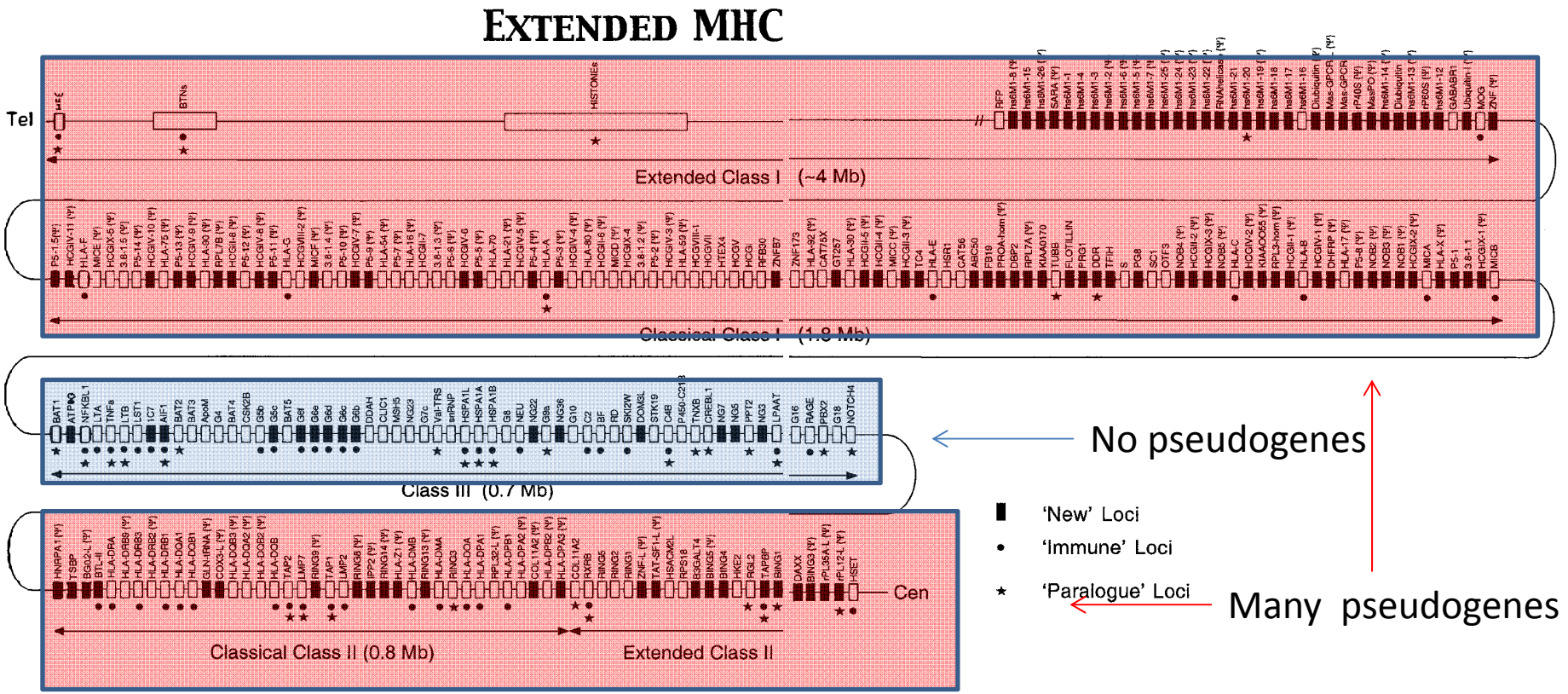
Once upon a time (t=-450 mya)



About the same time



Chromosomal Organisation



Beck and Trowsdale, Annual Reviews Genomics Human Genetics 2000

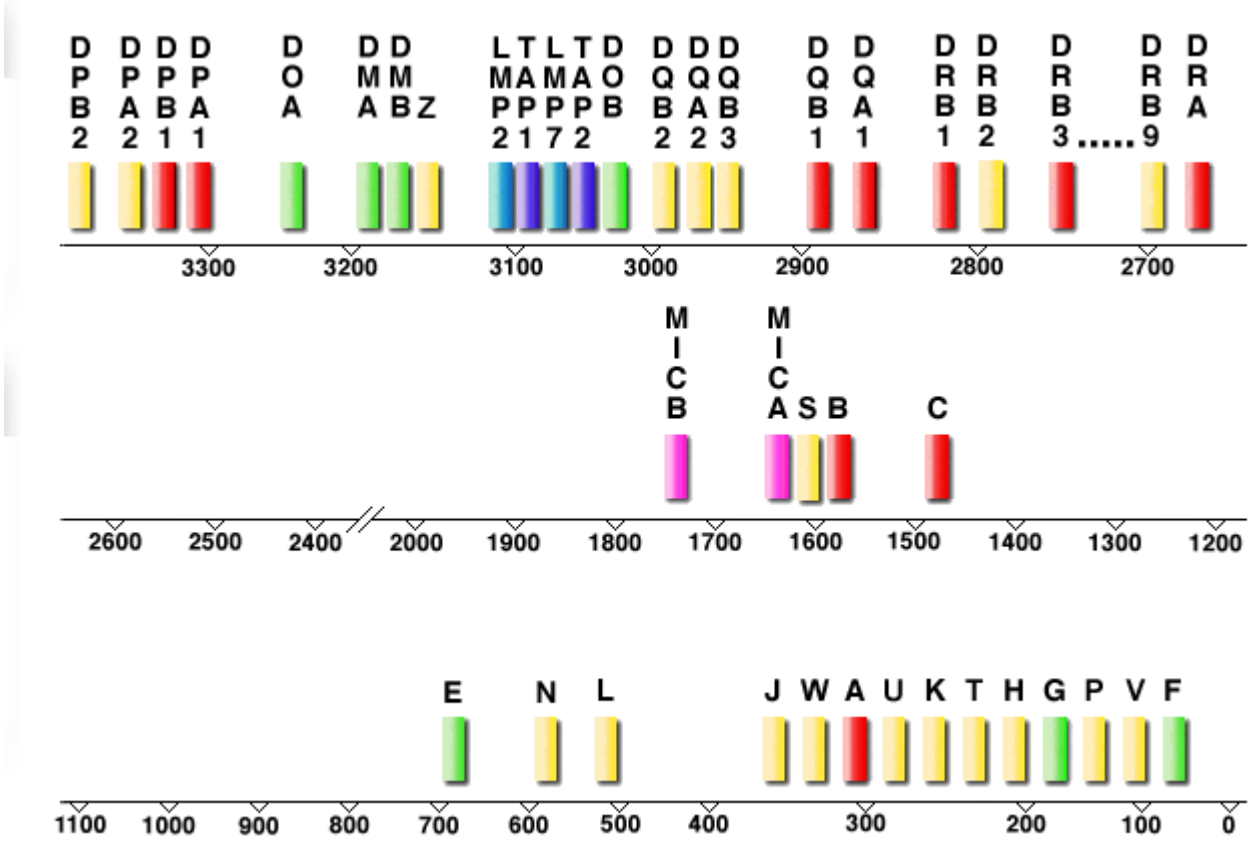
Quantitative, Qualitative

- 224 Loci (128 expressed)
- Functions
  - Immune related
    - Ligands for T cell and NK cell receptors
    - ABC (ATP Binding Cassette) transporter
    - Proteasome-related sequence
  - Olfactory receptors
  - Transcription factors (Zinc-/RING finger)
  - DNA repair
  - Development

Proteins with an MHC-Fold

- Classical
  polymorphic
  TCR-ligands (A, B, C, DR, DQ,DP)
  non orthologous
- Non-Classical
  partly orthologous
  less polymorphic, NK ligands (E, G, MICA/B, DM, DO)
- (Class I like
  Outside the MHC e.g. CD1)
- (Pseudogenes)

www.alleleles.org

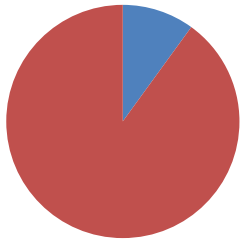


More functional loci … more peptides

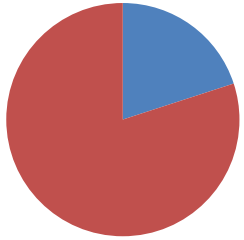
…

MHC-Loci (n)

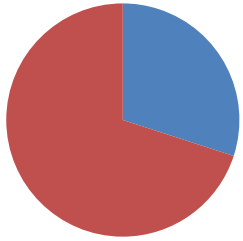
1



2



3



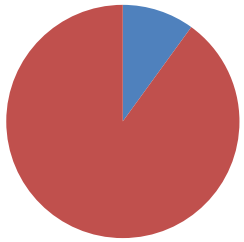
Peptides presented

Non Self Peptides

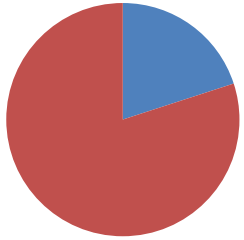
… but the T cells see less

HLA-Loci (n)

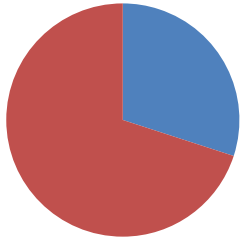
1



2



3



Self Peptides

TCR Repertoire

Polymorphism - Mechanisms

- Recombination (exon shuffling, hybrid genes)
- Double Recombination/Gene conversion
- Point mutation

Recombination

N.Holmes and **P.Parham**

DOMAIN 1

```
                          20                  40                  60                  80
Aw68    GSHSMRYFYTSVSRPGRGEPRFIAVGYVDDTQFVRFDSDAASQRMEPRAPWIEQEGPEYWDRNTRNVKAQSQTDRVDLGTLRGYYNQSEA
Aw69    --------------------------------------------------------------------------------
A2      --------F-----------------------------------------------------GE--K---H---H----------
```

DOMAIN 2

```
                  100                 120                 140                 160                 180
Aw68    GSHTIQMMYGCDVGSDGRFLRGYRQDAYDGKDYIALKEDLRSWTAADMAAQTTKHKWEAAHVAEQWRAYLEGTCVEWLRRYLENGKETLQRT
Aw69    ----V-R---------W------H-Y-----------------------------------------L-------------------------
A2      ----V-R---------W------H-Y-----------------------------------------L-------------------------
```

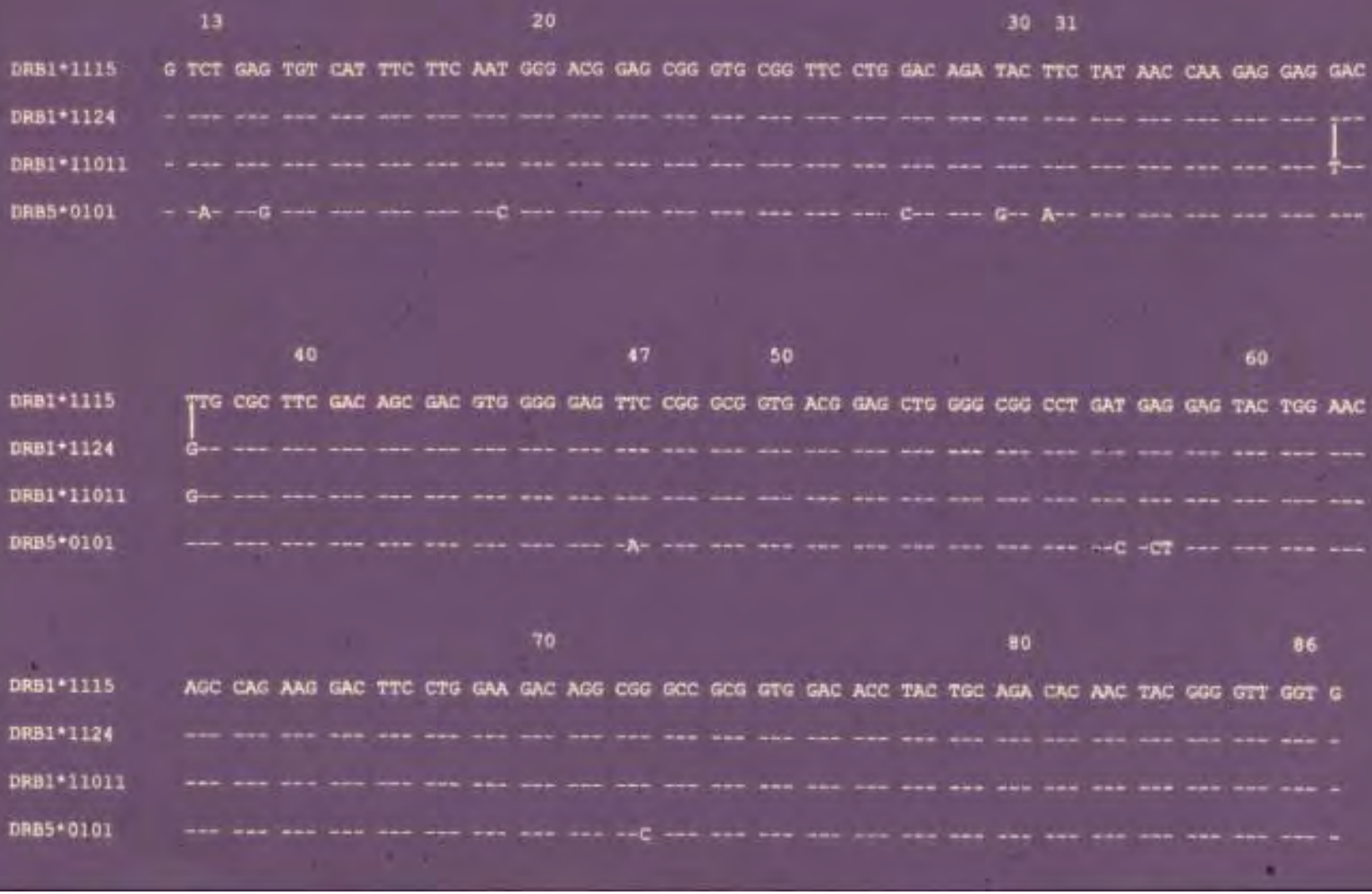
DOMAIN 3

```
                          200                 220                 240                 260
Aw68    DAPKTHMTHHAVSDHEATLRCWALSFYPAEITLTWQRDGEDQTQDTELVETRPAGDGTFQKWVAVVVPSGQEQRYTCHVQHEGLPKPLTLRW
Aw69    ----------------------------------------------------------A-------------------------------
A2      ----------------------------------------------------------A-------------------------------
```

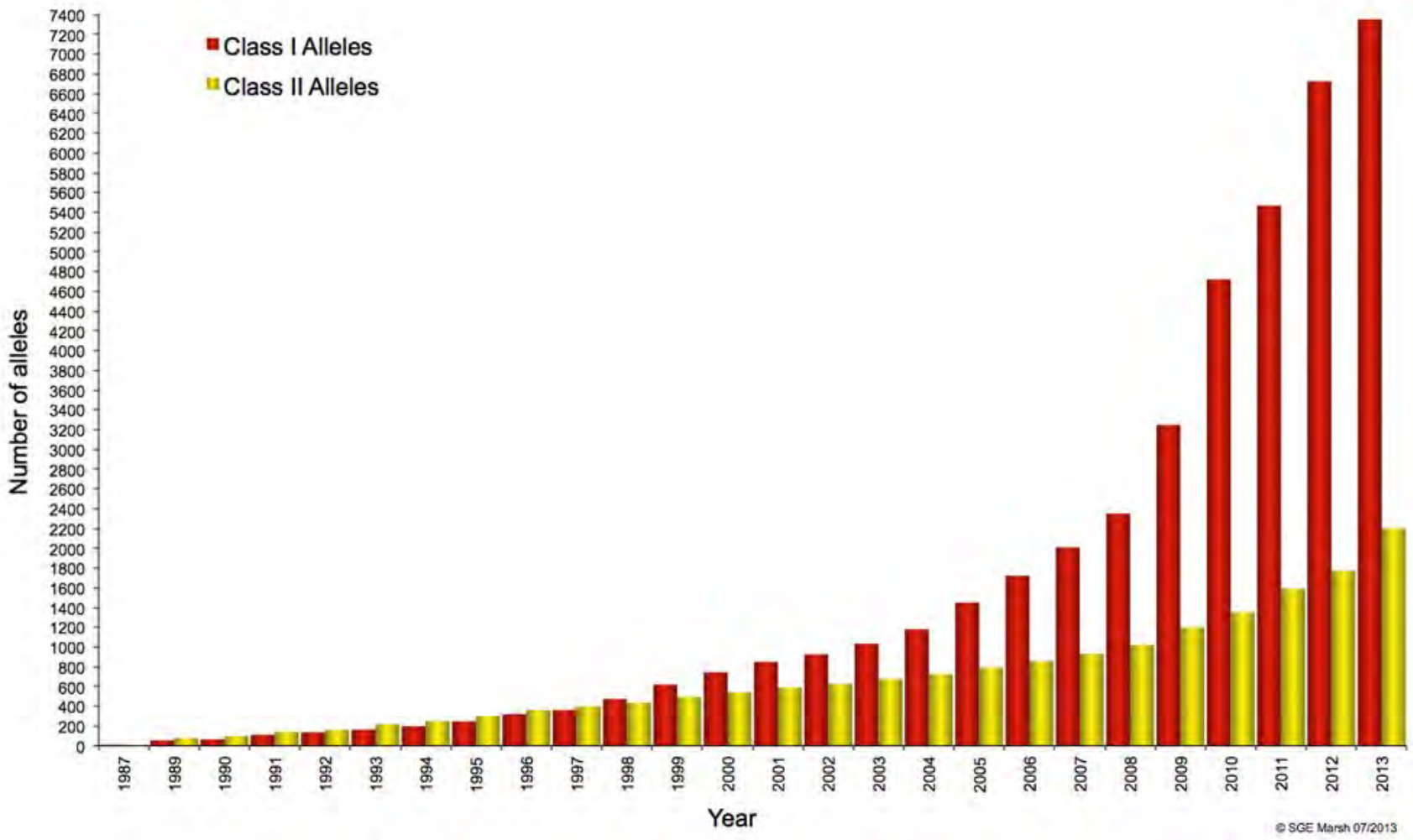
Fig. 1. Comparison of the deduced protein sequences for the first 274 residues of HLA-Aw68 from the genomic clone LBA5 of HLA-Aw69 from the genomic clones IDFA1, BJA1 and ZMA1 and of HLA-A2 from the LCL 721 cell line (Koller and Orr, 1985). Identity of HLA-Aw69 and/or HLA-A2 with HLA-Aw68 is denoted by (-). The standard one-letter amino acid code is used.

gene conversion ...

... or point mutation?

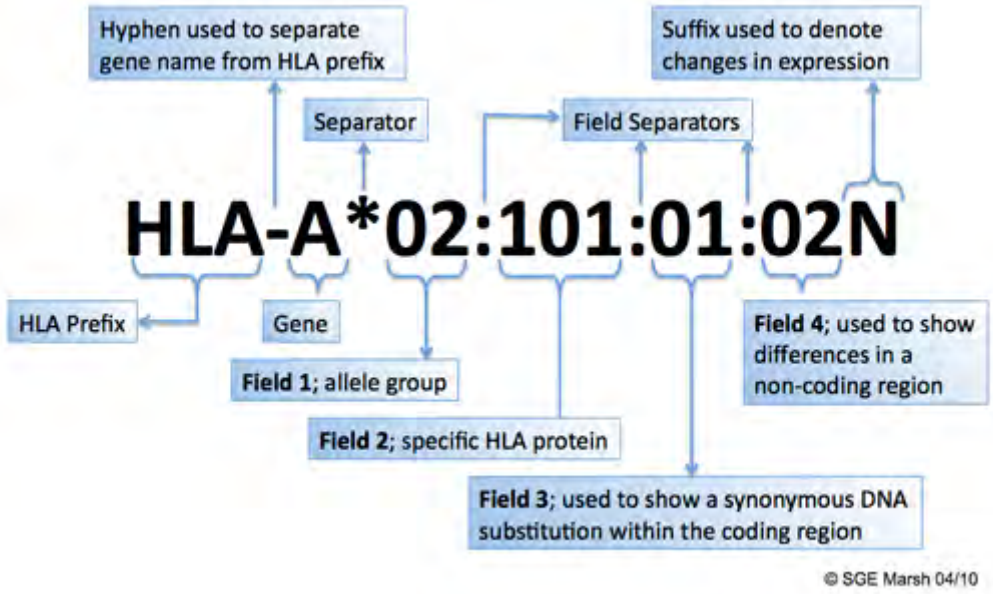


Curated Alleles, IMGT/HLA database



(**http://www.ebi.ac.uk/imgt/hla**)

www.alleles.org



Ambiguous Typings

- String of Alleles
  - e.g. A*02:01/02/07/20
- P Groups
  - HLA alleles having nucleotide sequences that encode the same protein sequence for the peptide binding domains
  - e.g. A*02:01P
- G Groups
  - HLA alleles that share identical nucleotide sequences for the exons encoding the peptide binding domains
  - e.g. A*02:01:01G

Allelic Resolution

The DNA-based typing result is consistent with a single allele as defined in a given version of the World Health Organization (WHO) HLA Nomenclature ….

…an allele is defined as a unique nucleotide sequence for a gene as defined by the use of all of the digits in a current allele name.

Examples include A*01:01:01:01 and A*02:07

Low Resolution

A DNA-based typing result at the level of the digits comprising the first field in the DNA-based nomenclature.

Examples include A*01; A*02. If the resolution corresponds to a serologic equivalent, this typing result should also be called low resolution.

High Resolution

A high resolution typing result is defined as a set of alleles that encode the same protein sequence for the region of the HLA molecule called the antigen binding site and that excludes alleles that are not expressed as cell-surface proteins. The antigen binding site includes domain 1 and domain 2 of the class I α polypeptides, and domain 1 of the class II α and domain 1 of the class II β polypeptide chains.

Other Levels of Resolution

If high resolution cannot be obtained, or if the laboratory's agreement with the entity requesting the testing limits the typing efforts to a subset of alleles, the laboratory may report its results at a level of resolution that falls between high resolution and low resolution.

Examples are to consider only those alleles that are expected to be found in the local population or that are designated as common and well defined. A third example is typing that assigns a G group designation (*e.g.*, A*02:01:01G).

Nomenclature

- constant competition between official WHO nomenclature and ‚pragmatic‘
- ambiguous typing results must be traceable (e.g. string of alleles)
- agreement between the tissue typing lab and its customer necessary

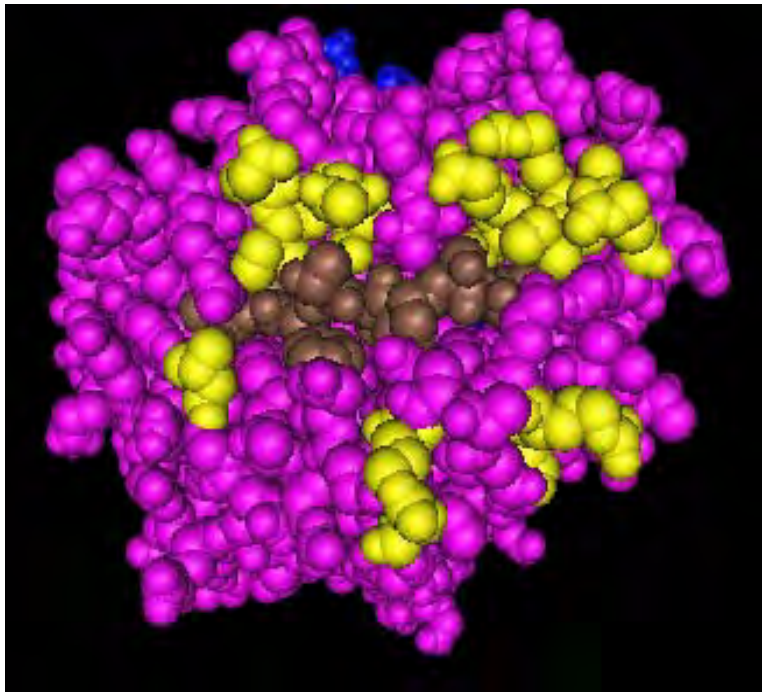
Do we understand the role of HLA in disease and Tx better …

… now, that we know the genes and alleles and their names.

?

- Restrictions of the nomenclature
- The typing legacy
- Still white spots on the landscape
- Back to the protein ?

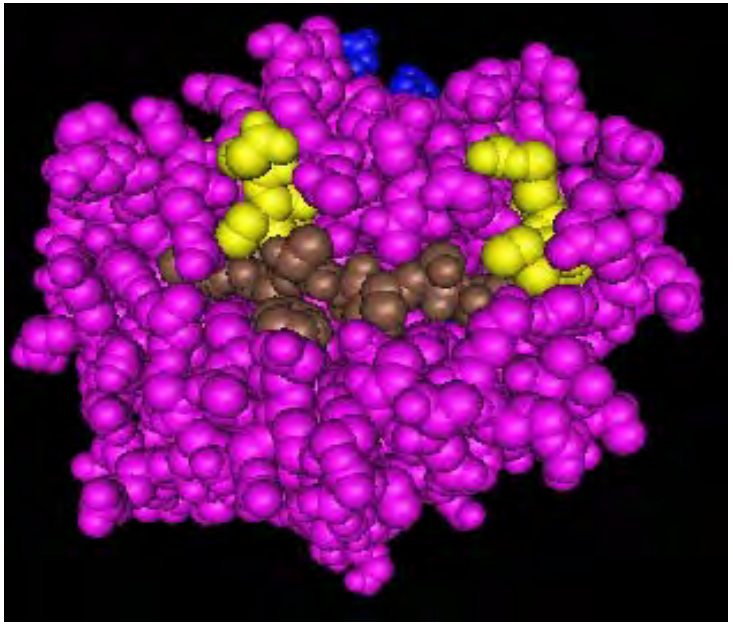
HLA-B51 Polymorphic Residues



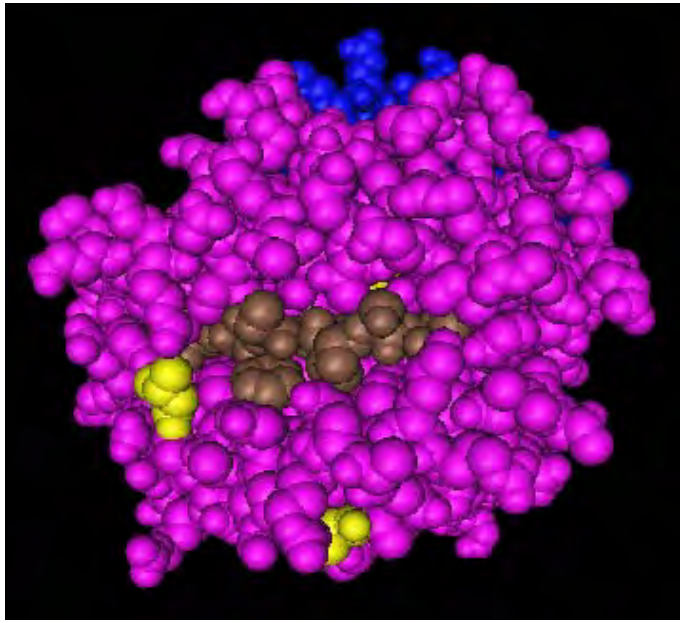
R. Duquesnoy

HLA-B51 Polymorphic Residues

"Seen" by A2,A68; B27,B44

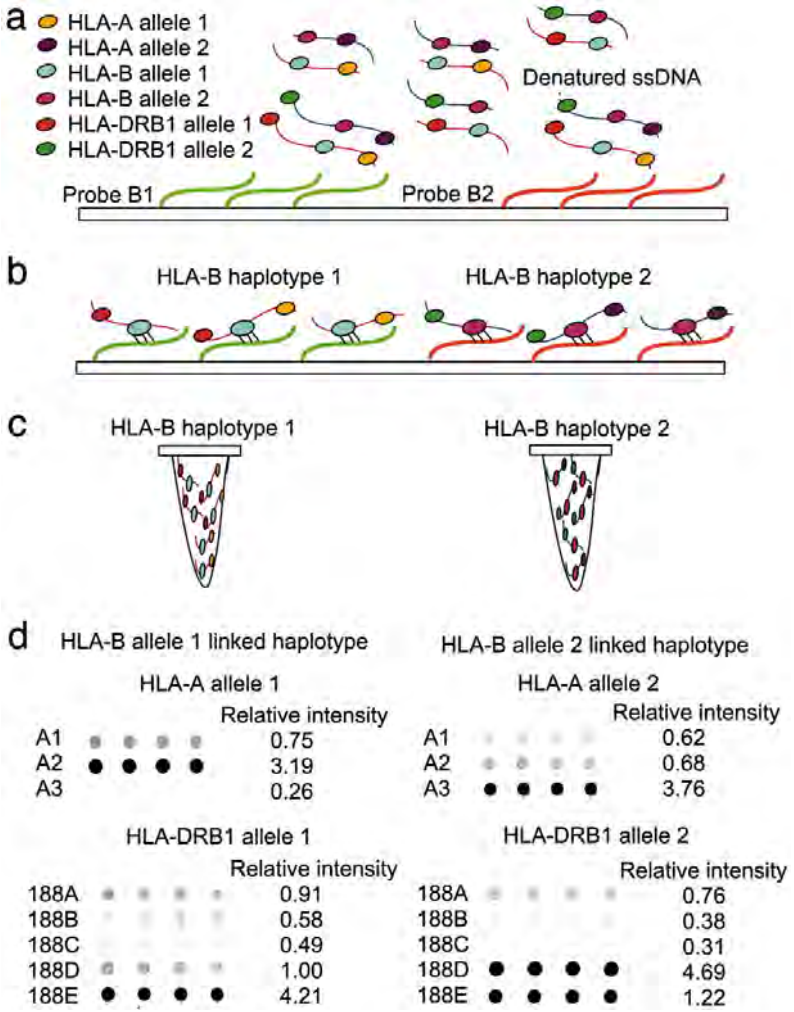


"Seen" by A2,A24; B7,B8



R. Duquesnoy

Fig. 1. Procedure for haplotyping 2-Mbp-long human genomic DNA



Guo, Zhen et al. (2006) Proc. Natl. Acad. Sci. USA 103, 6964-6969

The Genomic Landscape

- Gene annotation is not the end
  - Transcription – various cell types
  - Chromatin State – DNAse Hypersensitivity
  - Recombination
  - Vertebrate Conserved
  - SNPs
  - Structural Genomic variants - InDels
- Epigenetics – Hypermethylation of CpG Islands

The ENCODE Project (*Nature* 489:57, 2012)

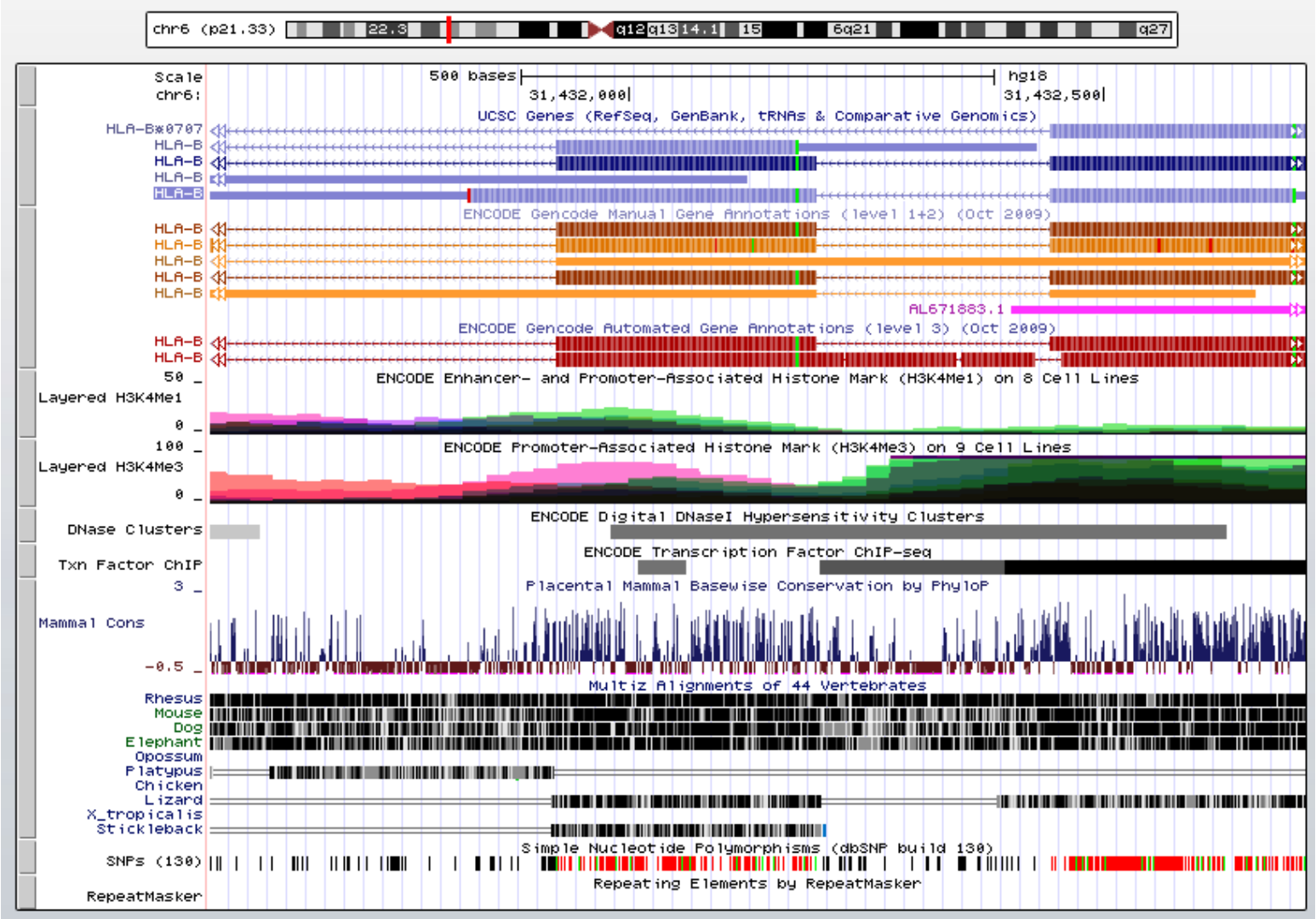
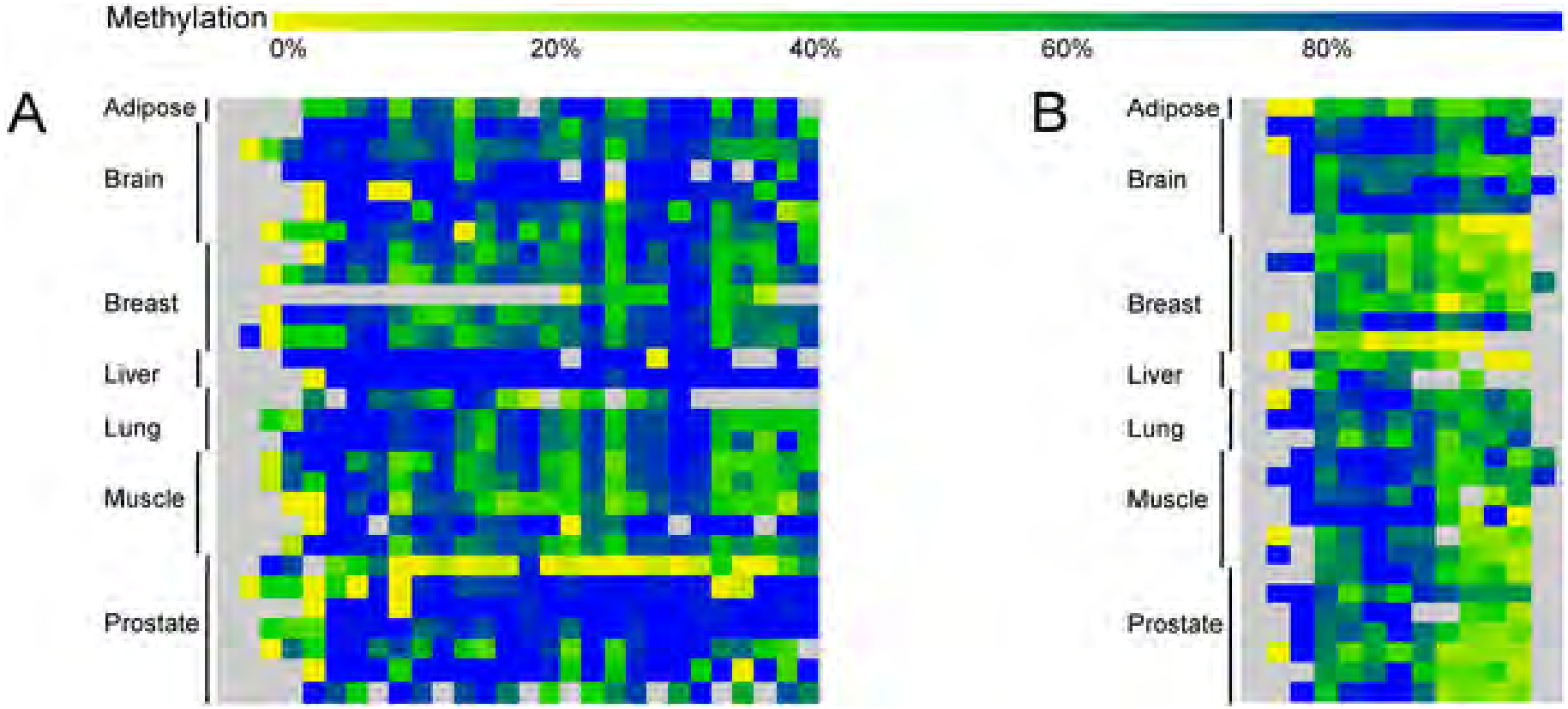


Figure 6. Example of METHANE Output Showing Regions That Display Inter-Individual Variation of Methylation Profiles



Rakyan VK, Hildmann T, Novik KL, Lewin J, et al. (2004) DNA Methylation Profiling of the Human Major Histocompatibility Complex: A Pilot Study for the Human Epigenome Project. PLoS Biol 2(12): e405. doi:10.1371/journal.pbio.0020405
http://www.plosbiology.org/article/info:doi/10.1371/journal.pbio.0020405

Polymorphism

The ultimate definition of polymorphism is through the characterisation of the amino acid sequence of the HLA - molecule.

A.M. Little, 1992

rs9264942

SNP 35 kb upstream of HLA-C

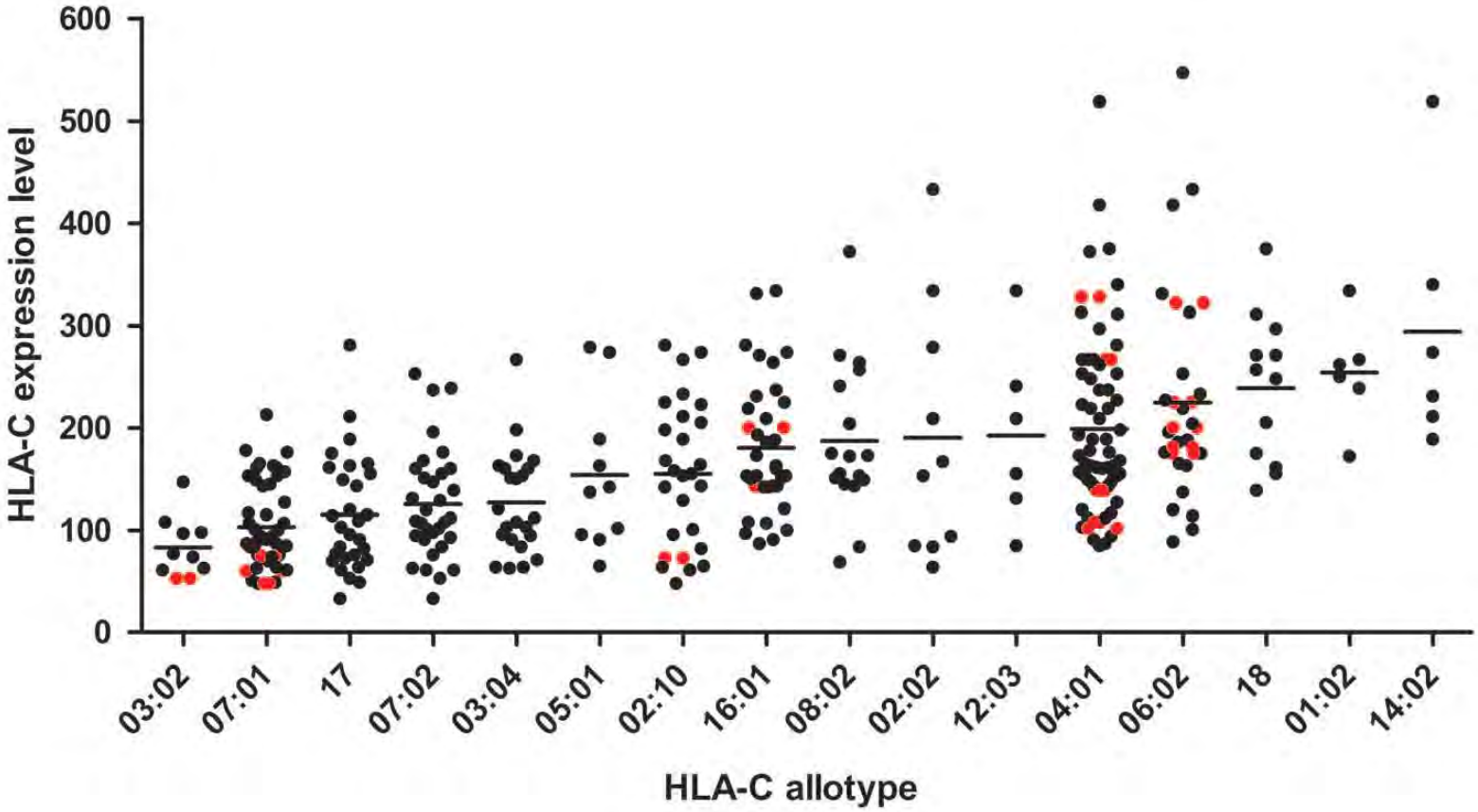
found in GWAS HIV load control

shows the highest correlation with HIV load control in European Americans

Correlation with HLA-C expression levels population specific

Science 330, 1551, 2010

Expression Levels of HLA-C



Science 340, 87, 2013

HLA-C expression level affects progression to early AIDS outcomes

Covariate	*P*	HR	95% CI	
HLA-C expression	**1×10^{-6}**	**0.67**	**0.55**	**0.74**
B*57:01	2×10^{-5}	0.30	0.17	0.52
C*06:02	7×10^{-4}	1.75	1.26	2.41
A*26:01	2×10^{-3}	0.43	0.25	0.73
B*35:03	2×10^{-3}	2.11	1.3	3.41
A*11:01	3×10^{-3}	0.62	0.45	0.85
A*74	4×10^{-3}	0.26	0.11	0.65
A*32:01	4×10^{-3}	0.53	0.34	0.82
A*31:01	3×10^{-2}	0.56	0.33	0.95
B*15:01	4×10^{-2}	0.72	0.52	0.99
C*04:01	4×10^{-2}	1.30	1.01	1.69

Science 340, 87, 2013

Summary

- HLA genetics has led to a clear picture of the human MHC
  - confusing typing results fall into place
- Technologies and nomenclature have evolved to capture and name the polymorphic variants efficiently
  - room for improvement
- Understanding of HLA-disease association
  - LD still a hurdle
  - many open questions