Sequence Ontology Workshop for Immunology.
MHC/HLA, and VDJ
Notes compiled by Karen Eilbeck and Suzanna Lewis.

12th June 2007.

Presentations:
Steve Marsh. Anthony Nolan Trust.
HLA Nomenclature
The WHO set up an HLA nomenclature committee in 1968. It is used predominantly to categorize tissue typing for transplants.
HLA genes are polymorphic. They make no distinction between allele and variant.
HLA database categorizes different alleles. -2714.

A note, the DNA sequence is not enough to predict a successful match – also need to look at expression.

The main take home about the nomenclature is that they encode meaning into the string they use to describe an allele. Information that goes into the string is serology, and location of variation, and finally expression.

Extremely polymorphic, e.g. 920 alleles for HLA-B (frequency of >10% is a "variation"), but just 49 serological variants. Release 2.17 of database. 2,714 HLA alleles, 1830 Class I 875 Class II alleles, 102 other alleles.

How the nomenclature works…
HLA-A*24 02 01 01 L
HLA-A gene name
* separator
24 allele family (serological)
02 amino acid difference
01 non-coding (silent) polymorphism
01 intron, 3', or 5' difference (e.g. expressed at low level)
N, L, S, A, Q: null, low, sec, abr, quest.

11M people have been HLA typed. Transplant people use it to figure out matches between donors and people in need.
High-resolution typing is now being done in volume. Database there since 1998, hosted at EBI. Lots of errors in early sequence data.
Need a system that copes with: expandability, useable, ambiguity

Richard Scheuermann – UT Southwestern

MHC/HLA data in the Immunology and Analysis Portal. (ImmPort) Database funded by NIAID. He is PI on Bioinformatics Integration Support Contract with Northrup-Grumman.

Data interoperability == minimum data standards + standardized ontologies + extensible data model. Effie Petersdorf is the driving force behind HLA ontology. Want to answer questions like:
Is there association between any HLA allele and disease/response?
Is there association between any HLA genotype and disease/response?
Is there association between any HLA haplotype and disease/response?
Is there association between any HLA group and disease/response?

(Group is a collection of different alleles of the same gene where as a haplotype is a genomic block).

Terms Richard wants at the top level of SO:
Genotype,
Allele,
Variant
Polymorphsim
Haplotype
Wildtype/normal/reference

Lindsay Cowell – Duke.
V(D)J recombination
How we create our diverse repertoire of antigen receptors

1. V(D)J recombination
2. Somatic hypermutation
3. Gene Conversion

Types of things: Gene segments, rearrangements, regulatory sequences
Related by: part of the same receptor type, part of the same protein chain, same genomic location, same regulatory.
Constant region, highly variable region (1, 2, and 3), antigen-binding site, complementarity defining regions, framework region, D-J joined rearranged DNA, V-J or V –DJ joined rearranged DNA, V gene segment, germline DNA, leader peptide, … palindrome ends (sticky ends), non-templated nucleotides, spacers (either 23 bases or 12 bases), recombination signals. Names composed of receptor type + chain + region + gene family: IGLV1 Immunoglobulin receptor, light chain, variable region, gene family 1.
Binding pocket, anchor residue.

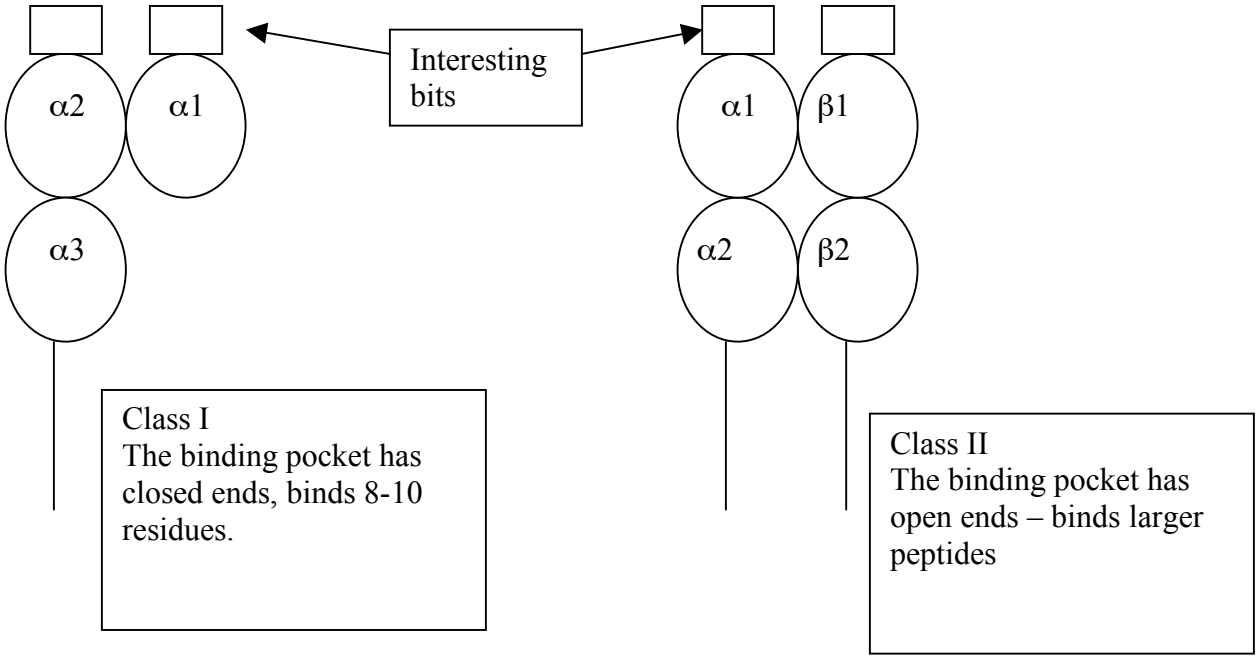
Darren Natale – Georgetown.
How to build a protein Ontology. Darren showed how the current OBO ontologies do not adequately allow the representation of all of the different states of a protein. E.g., Is it phosphorylated? He pointed out that Go annotations to a gene can be misleading as the protein products of a single gene can have different and sometimes opposing qualities.

12th June 11, 2007

Richard presented his ideas for Foundations for an HLA ontology.

Hierarchy needs to reflect more than the sequence similarity.
Two things to consider:
Serology – immune reaction to alloantisera – a reflection of the immune response.
Sequence difference (amino acid) - part of MHC that is interacting with the peptide.

How do we begin to classify alleles?



Changes in some residues are more interesting than others. For example, changes in the binding cleft are more interesting than in the membrane domain.

Questions about where we are going.
Is this going to be an ontology in its own right or a combinations of terms from other ontologies?
What are the entities? Sequence? Chemicals? Gene variants?

Think about what we are going to annotate. There is a tendency to mirror the nomenclature but that is not going to work.

Plan of work

I. Develop ontology of domains and motifs in SO. This will use the biosapiens terms for protein parts. Relate these terms using mereotopology. Get together in the fall to discuss progress/integration
II. Develop ontology of sequence variants – SO
III. Define motif variant types – hierarchy of relations between allele sequence of variants

IV. Annotate with motif features and variant terms.

Example of motif types
Leader peptide, cytoplasmic domain, membrane spanning domain, extracellular domain, α3 domain, α2 domain, α1 domain.

New relationships to be added to the relations ontology.
(Barry wants to make it clear that sequences are biological reality – not representations.)

1. Variant_of
   The process of mutation. A' is a mutation of A = definition every instance of A' is either an immediate mutation of some instance of A, or there is a chain of immediate mutation processes linking A' to some instance of A.

   A is an allele of B = definition there is some C such that A and B are both mutations of C.

2. Spliced_variant_of
   Splice variants are different from allelic variants.
   A is a splice product of B means that A is an mRNA and B is a type of primary transcript. Every instnce of A arises through splicing from some instance of B.
   A is a spliced_variant_of B = definition there is some C such that A is a spliceproduct of C and B is a splice product of C and A and C are structurally different.

Changes to SO.

Created object of type obo:TYPE with id variant_of
Copy_number_variation
Changed name of SO:0001019 from "<new term>" to "copy_number_variation"
Set definition of SO:0001019 to "A variation that increases or decreases the copy number of a given region."
Added definition dbxref SO:ke to SO:0001019
Added synonym "CNP" to SO:0001019
Changed synonym scope of SO:0001019 from "Related" to "Exact"
Added synonym "CNV" to SO:0001019
Changed synonym scope of SO:0001019 from "Related" to "Exact"
Added synonym "copy number polymorphism" to SO:0001019
Changed synonym scope of SO:0001019 from "Related" to "Exact"
Added synonym "copy number variation" to SO:0001019
Changed synonym scope of SO:0001019 from "Related" to "Exact"
Copied SO:0001019 to SO:0000109 with type OBO_REL:is_a
Assigned namespace "sequence" to SO:0001019

"mutation_affecting_copy_number"
Copied SO:0001020 to SO:1000132 with type OBO_REL:is_a
Assigned namespace "sequence" to SO:000102

"chromosome_breakpoint"
Copied SO:0001021 to SO:0000830 with type OBO_REL:is_a
Assigned namespace "sequence" to SO:0001021

"inversion_breakpoint"
Copied SO:0001022 to SO:0001021 with type OBO_REL:is_a
Assigned namespace "sequence" to SO:0001022

"allele"
Set definition of SO:0001023 to "An allele is one of a set of coexisting sequence variants of a gene."
Added definition dbxref SO:immuno_workshop to SO:0001023
Copied SO:0001023 to SO:0000109 with type OBO_REL:is_a
Copied SO:0001023 to SO:0000704 with type variant_of
Assigned namespace "sequence" to SO:0001023

"haplotype"
Set definition of SO:0001024 to "A haplotype is one of a set of coexisting sequence variants of a haplotype block."
Added definition dbxref SO:immuno_workshop to SO:0001024
Copied SO:0001024 to SO:0000355 with type variant_of
Copied SO:0001024 to SO:0000109 with type OBO_REL:is_a
Assigned namespace "sequence" to SO:0001024

"polymorphic_sequence_variant"
Set definition of SO:0001025 to "A sequence variant that is segregating in one or more natural populations of a species."
Added definition dbxref SO:immuno_workshop to SO:0001025
Copied SO:0001025 to SO:0000109 with type OBO_REL:is_a
Assigned namespace "sequence" to SO:0001025

"genome"
Set definition of SO:0001026 to "A genome is the sum of genetic material within a cell or virion."
Added definition dbxref SO:immuno_workshop to SO:0001026
Copied SO:0001026 to SO:0000000 with type OBO_REL:is_a
Assigned namespace "sequence" to SO:0001026
Populated new object SO:0001027 (macro)

"genotype"

Set definition of SO:0001027 to "A genotype is a variant genome, complete or incomplete."
Added definition dbxref SO:immuno_workshop to SO:0001027
Copied SO:0001027 to SO:0001026 with type variant_of
Copied SO:0001027 to SO:0000109 with type OBO_REL:is_a
Assigned namespace "sequence" to SO:0001027
Populated new object SO:0001028 (macro)

"diplotype"
Set definition of SO:0001028 to "A diplotype is a pair of haplotypes from a given individual. It is a genotype where the phase is known."
Added definition dbxref SO:immuno_workshop to SO:0001028
Copied SO:0001028 to SO:0001027 with type OBO_REL:is_a
Assigned namespace "sequence" to SO:0001028

variant_of
Added comment "Added to SO during the immunology workshop, June 2007. This relationship was approved by Barry Smith."
Set definition of variant_of to "A' is a variant (mutation) of A = definition every instance of A' is either an immediate mutation of s=ome instance of A, or there is a chain of immediate mutation processes linking A' to some instance of A."
Added definition dbxref SO:immuno_workshop to variant_of
Assigned namespace "sequence" to variant_of

Changed name of SO:0000207 from "simple_sequence_length_polymorphism" to **"simple_sequence_length_variation"**
Added synonym "simple sequence length variation" to SO:0000207
Changed synonym scope of SO:0000207 from "Related" to "Exact"
Added synonym "simple sequence length polymorphism" to SO:0000207

Changed definition of SO:0000355 from "A region of the genome which in which markers are co-inherited as the result of the lack of historic recombination between them due to their close proximity." to "**A region of the genome which is co-inherited as the result of the lack of historic recombination within it**."
Changed definition of SO:0000667 from "A region of sequence identified as having been inserted." to "**A region of sequence that has been inserted**."
Changed definition of SO:0000691 from "The space between two bases in a sequence which marks the position where a translocation has occurred." to "**The boundary in a sequence which marks the position where a translocation has occurred**."
Changed definition of SO:0000699 from "A junction refers to an interbase location of zero in a sequence." to "**A junction is a boundary between regions. A boundary has an extent of zero**."
Changed comment from "This term is mapped to MGED. Do not obsolete without consulting MGED ontology." to "This term is mapped to MGED. Do not obsolete without consulting MGED ontology.
A gene may be considered as a unit of inheritance."

Changed definition of SO:0000704 from "A locatable region of genomic sequence, corresponding to a unit of inheritance, which is associated with regulatory regions, transcribed regions and/or other functional sequence regions." to "**A region (or regions) that has the potential to encode a functional transcript. A gene may include regulatory regions, transcribed regions and/or other functional sequence regions**."
Deleted definition dbxref SO:rd from SO:0000704
Added definition dbxref SO:immuno_workshop to SO:0000704
Changed name of SO:0000830 from "chromosome_region" to "chromosome_part"
Changed definition of SO:0000941 from "A recombinationally rearranged gene of the vertibrate immune system." to "**A recombinationally rearranged gene of the vertebrate immune system**."
Changed definition of SO:1000008 from "A mutation event where a single DNA nucleotide changes into another nucleotide." to "**A single nucleotide change which has occurred at the same position of a corresponding nucleotide in a reference sequence**."
Deleted definition dbxref http://www.ebi.ac.uk/mutations/recommendations/mutevent.html from SO:1000008
Added definition dbxref SO:immuno_workshop to SO:1000008
Changed name of SO:1000050 from "no_change_in_transcript" to "**mutation_causing_no_change_in_transcript**"
Changed name of SO:1000052 from "complex_change_in_transcript" to "**mutation_affecting_complex_change_in_transcript**"
Changed name of SO:1000055 from "initiator_codon_change_in_transcript" to "**mutation_causing_initiator_codon_change_in_transcript**"
Changed name of SO:1000056 from "amino_acid_coding_codon_change_in_transcript" to "**mutation_causing_amino_acid_coding_codon_change_in_transcript**"
Changed name of SO:1000057 from "synonymous_codon_change_in_transcript" to "**mutation_causing_synonymous_codon_change_in_transcript**"
Changed name of SO:1000058 from "non_synonymous_codon_change_in_transcript" to "**mutation_causing_non_synonymous_codon_change_in_transcript**"
Changed name of SO:1000059 from "missense_codon_change_in_transcript" to "**mutation_causing_missense_codon_change_in_transcript**"
Changed name of SO:1000060 from "conservative_missense_codon_change_in_transcript" to "**mutation_causing_conservative_missense_codon_change_in_transcript**"
Changed name of SO:1000061 from "nonconservative_missense_codon_change_in_transcript" to "**mutation_causing_nonconservative_missense_codon_change_in_transcript**"
Changed name of SO:1000062 from "nonsense_codon_change_in_transcript" to "**mutation_causing_nonsense_codon_change_in_transcript**"
Changed name of SO:1000063 from "terminator_codon_change_in_transcript" to "**mutation_causing_terminator_codon_change_in_transcript**"
Changed name of SO:1000089 from "no_change_of_translational_product" to "**mutation_causing_no_change_of_translational_product**"
Changed name of SO:1000090 from "uncharacterised_change_of_translational_product" to "**mutation_causing_uncharacterised_change_of_translational_product**"

Changed name of SO:1000091 from
"partially_characterised_change_of_translational_product" to
"**mutation_causing_partially_characterised_change_of_translational_product**"
Changed name of SO:1000092 from "complex_change_of_translational_product" to
"**mutation_causing_complex_change_of_translational_product**"
Changed name of SO:1000093 from "amino_acid_substitution" to
"**mutation_causing_amino_acid_substitution**"
Changed name of SO:1000094 from "conservative_amino_acid_substitution" to
"**mutation_causing_conservative_amino_acid_substitution**"
Changed name of SO:1000095 from "nonconservative_amino_acid_substitution" to
"**mutation_causing_nonconservative_amino_acid_substitution**"
Changed name of SO:1000096 from "amino_acid_insertion" to
"**mutation_causing_amino_acid_insertion**"
Changed name of SO:1000097 from "amino_acid_deletion" to
"**mutation_causing_amino_acid_deletion**"
Changed name of SO:1000098 from "polypeptide_truncation" to
"**mutation_causing_polypeptide_truncation**"
Changed name of SO:1000099 from "polypeptide_elongation" to
"**mutation_causing_polypeptide_elongation**"
Changed name of SO:1000100 from "polypeptide_N_terminal_elongation" to
"**mutation_causing_polypeptide_N_terminal_elongation**"
Changed name of SO:1000101 from "polypeptide_C_terminal_elongation" to
"**mutation_causing_polypeptide_C_terminal_elongation**"
Changed name of SO:1000106 from "inframe_polypeptide_N_terminal_elongation" to
"**mutation_causing_inframe_polypeptide_N_terminal_elongation**"
Changed name of SO:1000107 from
"out_of_frame_polypeptide_N_terminal_elongation" to
"**mutation_causing_out_of_frame_polypeptide_N_terminal_elongation**"
Changed name of SO:1000108 from "inframe_polypeptide_C_terminal_elongation" to
"**mutaton_causing_inframe_polypeptide_C_terminal_elongation**"
Changed name of SO:1000109 from
"out_of_frame_polypeptide_C_terminal_elongation" to
"**mutation_causing_out_of_frame_polypeptide_C_terminal_elongation**"
Changed name of SO:1000112 from "no_3D_structural_change" to
"**mutation_causing_no_3D_structural_change**"
Changed name of SO:1000113 from "uncharacterised_3D_structural_change" to
"**mutation_causing_uncharacterised_3D_structural_change**"
Changed name of SO:1000114 from "partially_characterised_3D_structural_change" to
"**mutation_causing_partially_characterised_3D_structural_change**"
Changed name of SO:1000115 from "complex_3D_structural_change" to
"**mutation_causing_complex_3D_structural_change**"
Changed name of SO:1000116 from "conformational_change" to
"**mutation_causing_conformational_change**"
Changed name of SO:1000118 from "loss_of_function_of_polypeptide" to
"**mutation_causing_loss_of_function_of_polypeptide**"

Changed name of SO:1000119 from "inactive_ligand_binding_site" to "**mutation_causing_inactive_ligand_binding_site**"
Changed name of SO:1000120 from "inactive_catalytic_site" to "**mutation_causing_inactive_catalytic_site**"
Changed name of SO:1000121 from "polypeptide_localization_affected" to "**mutation_causing_polypeptide_localization_change**"
Changed name of SO:1000122 from "polypeptide_post_translational_processing_affected" to "**mutation_causing_polypeptide_post_translational_processing_change**"
Changed name of SO:1000124 from "partial_loss_of_function_of_polypeptide" to "**mutation_causing_partial_loss_of_function_of_polypeptide**"
Changed name of SO:1000125 from "gain_of_function_of_polypeptide" to "**mutation_causing_gain_of_function_of_polypeptide**"
Changed name of SO:1000127 from "compensatory_transcript_secondary_structure_mutation" to "**mutation_causing_compensatory_transcript_secondary_structure_mutation**"
Changed name of SO:1000132 from "consequences_of_mutation" to "mutation"
Changed name of SO:1000134 from "polypeptide_fusion" to "**mutation_causing_polypeptide_fusion**"
Changed name of SO:1000177 from "uncharacterised_change_in_transcript" to "**mutation_causing_uncharacterised_change_in_transcript**"
Changed name of SO:1000179 from "partially_characterised_change_in_transcript" to "**mutation_causing_partially_characterised_change_in_transcript**"
Changed name of SO:1000181 from "gene_fusion" to "**mutation_causing_gene_fusion**"