# Assessing the risk of alloimmune responses by scanning the human proteome

Adrian Shepherd

ISMB & Biological Sciences, Birkbeck, University of London

Immunogenicity and Tolerance Seminar, Amsterdam

15th November 2019

# Overview

- Birkbeck and my research group
- Missense mutation haemophilia A
- Proteome scanning: our strategy for identifying when self tolerance is likely to be broken
- Other applications, notably to organ transplantation

# About Birkbeck

Founded in 1823

University of London's only specialist provider of part-time higher education

Nobel Laureates Aaron Klug and Derek Barton worked in the School of Crystallography

Biological Sciences is part of a joint research institute with UCL: ISMB

*Motto: In nocte consilium*

# Research overview

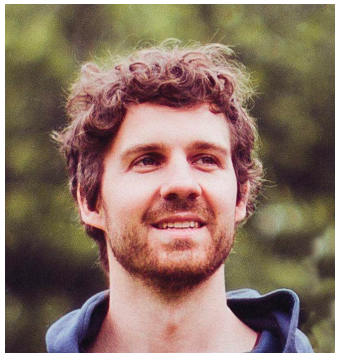
## Immunoinformatics:

- Antibodies (repertoire analytics, MD simulation, deep learning)
- B-cell epitopes (predicting antigenic escape)
- T-cell epitopes (allo-immunity, heterologous immunity)

## Antigens of interest:

- Protein therapeutics (e.g. tFVIII)
- Viruses (e.g. HCV, HIV, IAV)
- Bacteria (e.g. B. anthracis)
- Cancer (HCC)



# The AIRR Community

Established in 2015 at first of 4 Community Meetings:

- Devising relevant standards + guidelines about good practice
- Obtaining, analyzing, curating and comparing/sharing NGS AIRR datasets
- Validating tools for AIRR data analysis
- Legal and ethical issues involving the use and sharing of AIRR data sets derived from humans

A chapter of the Antibody Society since 2018.

AIRR Community

# AIRR Community recommendations

frontiers in Immunology

TECHNOLOGY REPORT
published: 28 September 2018
doi: 10.3389/fimmu.2018.02206

Check for updates

## AIRR Community Standardized Representations for Annotated Immune Repertoires

OPEN ACCESS

*Jason Anthony Vander Heiden[1†], Susanna Marquez[2], Nishanth Marthandan[3], Syed Ahmad Chan Bukhari[2], Christian E. Busse[4], Brian Corrie[5], Uri Hershberg[6,7,8], Steven H. Kleinstein[2,9], Frederick A. Matsen IV[10], Duncan K. Ralph[10], Aaron M. Rosenfeld[6], Chaim A. Schramm[11], The AIRR Community[‡], Scott Christley[12*†] and Uri Laserson[13*]*

*Vander Heiden et al., Frontiers in Immunology, 2018*

## Adaptive Immune Receptor Repertoire Community recommendations for sharing immune-repertoire sequencing data

Florian Rubelt[1,21], Christian E Busse[2,21], Syed Ahmad Chan Bukhari[3,21], Jean-Philippe Bürckert[4], Encarnita Mariotti-Ferrandiz[5], Lindsay G Cowell[6], Corey T Watson[7], Nishanth Marthandan[8], William J Faison[9], Uri Hershberg[10], Uri Laserson[11], Brian D Corrie[12,13], Mark M Davis[1,14], Bjoern Peters[15], Marie-Paule Lefranc[16], Jamie K Scott[8,12,17], Felix Breden[12,13], The AIRR Community[18], Eline T Luning Prak[19,22] & Steven H Kleinstein[3,20,22]

**High-throughput sequencing of B and T cell receptors is routinely being applied in studies of adaptive immunity. The Adaptive Immune Receptor Repertoire (AIRR) Community was formed in 2015 to address issues in AIRR sequencing studies, including the development of reporting standards for the sharing of data sets.**

*Rubelt et al., Nature Immunology, 2017*

# AIRR Community resources

| | |
|---|---|
| OGRDB | https://ogrdb.airr-community.org/<br>Lees et al., *NAR*, 2019 |
| VDJbase | https://www.vdjbase.org/<br>Omer et al., *NAR*, 2019 |
| sumrep | https://github.com/matsengrp/sumrep<br>Olson et al., *Frontiers in Immunology*, 2019 |

OGRDB

VDJbase

# Germline reference set errors

Immunology & Cell Biology

Australian and New Zealand SOCIETY FOR IMMUNOLOGY INC.

Outstanding Observation

## Many human immunoglobulin heavy-chain IGHV gene polymorphisms have been reported in error

Yan Wang, Katherine J L Jackson, William A Sewell, Andrew M Collins

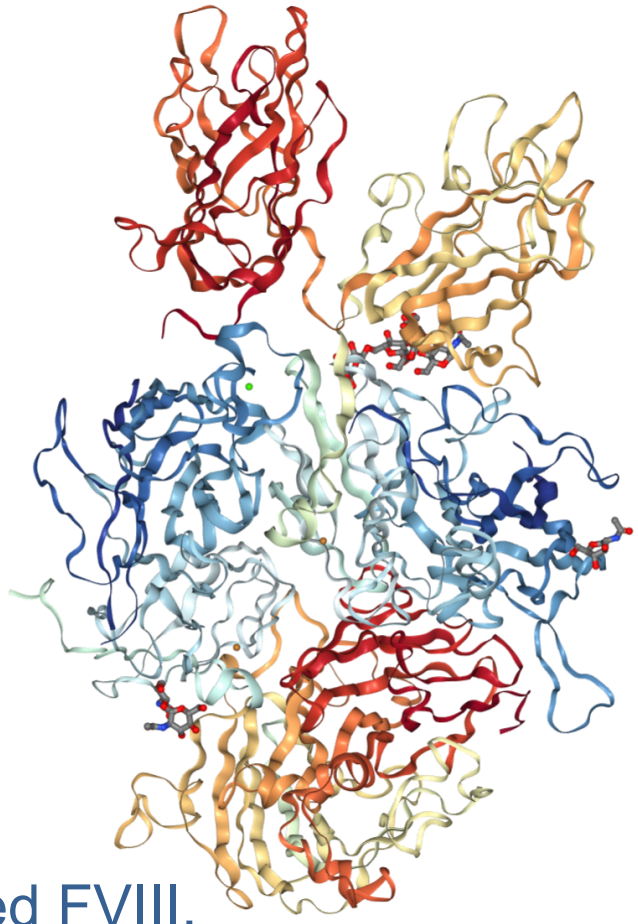
First published: 27 November 2007 | https://doi.org/10.1038/sj.icb.7100144 | Citations: 30

Of the 226 human IGHV genes in the IMGT database in 2007, at least 104 were likely to contain sequencing errors. Almost all of these genes are still there.

# Haemophilia and Inhibitors

# Haemophilia A

- X-linked disease affecting 1 in 5000 men
- Caused by defect in clotting factor VIII (FVIII), part of the coagulation cascade
- Standard treatment is replacement FVIII (very expensive recombinant therapeutic)



B-domain-deleted FVIII, PDB 3CDZ

# Haemophilia A gene therapy trial

NEWS

Home | UK | World | Business | Politics | Tech | Science | Health | Family & Education

Health

## Haemophilia A trial results 'mind-blowing'

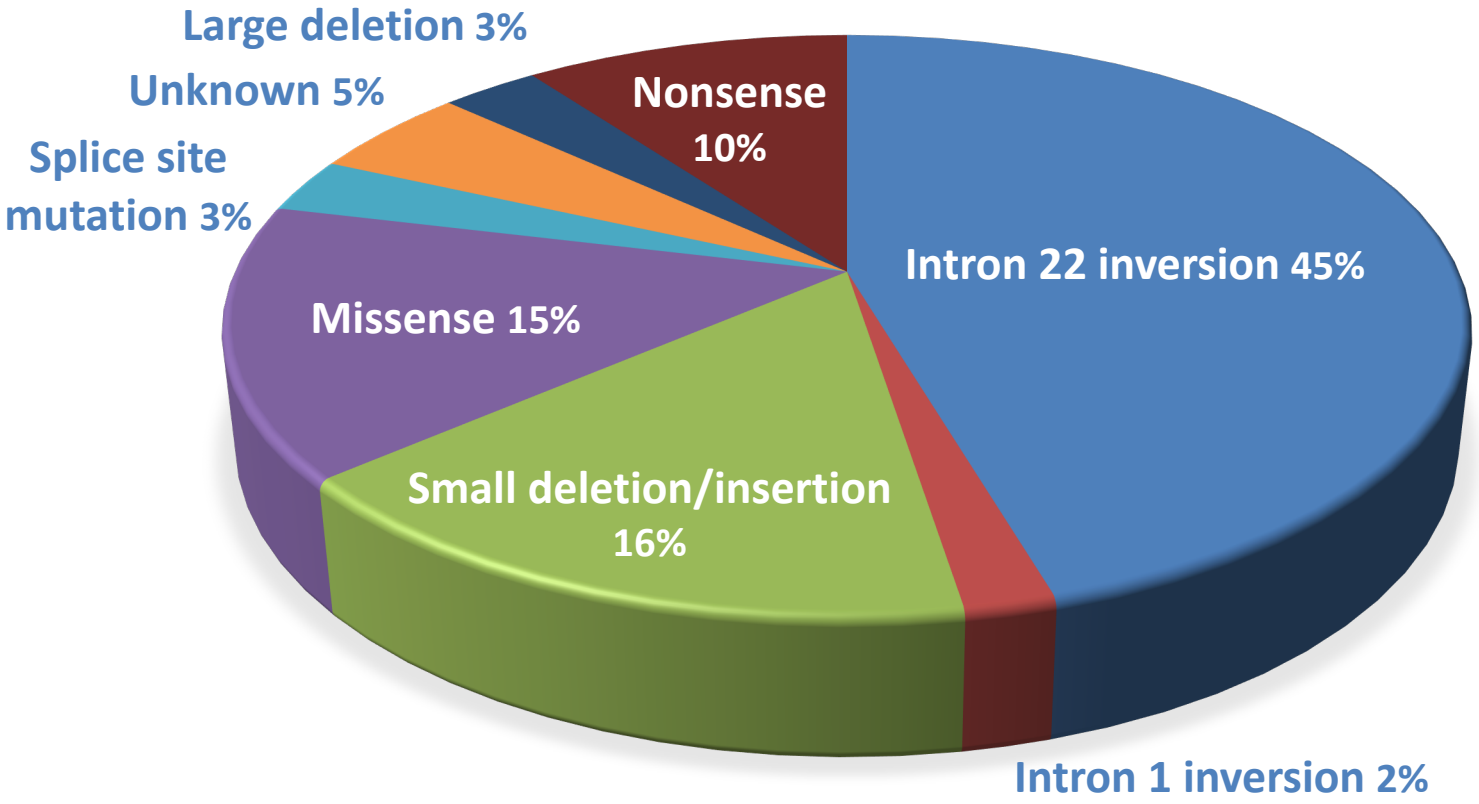
By James Gallagher
Health and science correspondent, BBC News

14 December 2017 Share

Father-of-two Jake Omer was born with haemophilia A

*BBC News, December 2017*

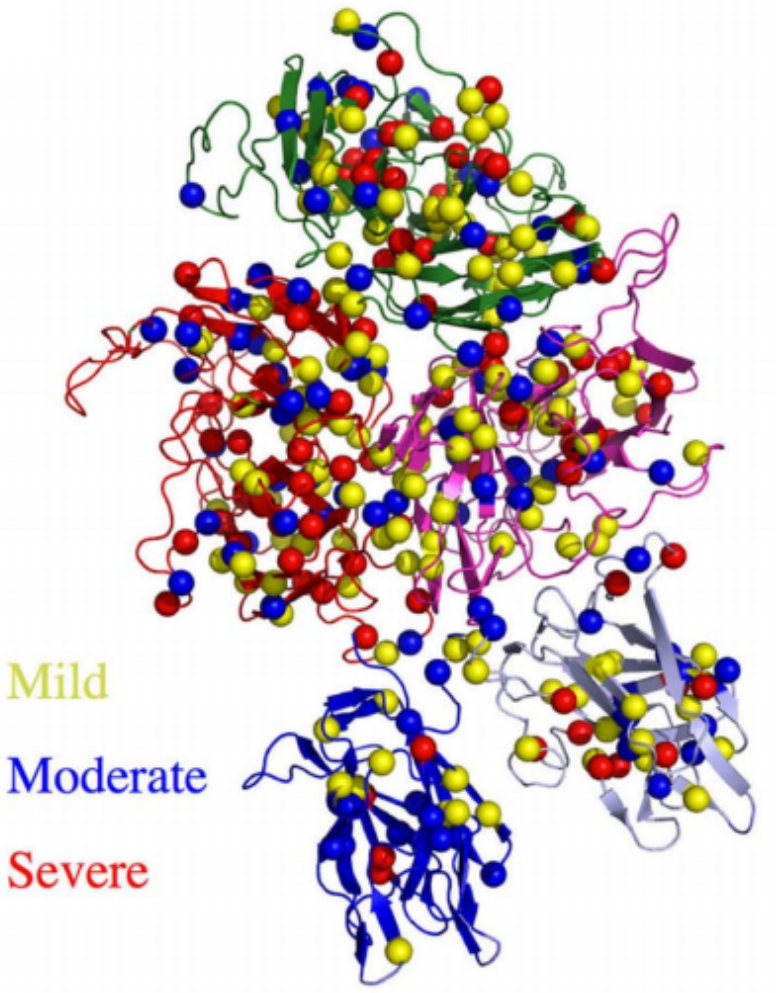
# F8 genotypes associated with HA




*Adapted from Gouw et al., Blood, 2012*

# FVIII missense mutations

- Total 977 missense mutations (March 2018)
- Associated with mild, moderate and severe haemophilia
- ~10-15% develop **inhibitors** (anti-replacement FVIII antibodies)
- Accurate risk stratification has potential benefits (e.g. can inform choices about alternative therapeutic options)



# At the self/non-self boundary

A single amino-acid substitution difference will be detected by the immune system provided it occurs within at least one spanning peptide that:

a) Binds to an MHC molecule long enough to be seen

b) Forms a novel peptide-MHC surface

c) There are T cells with TCRs capable of binding to that surface.

Note, however, that detection is not sufficient for the generation of inhibitors.

# Example missense mutation: R593C

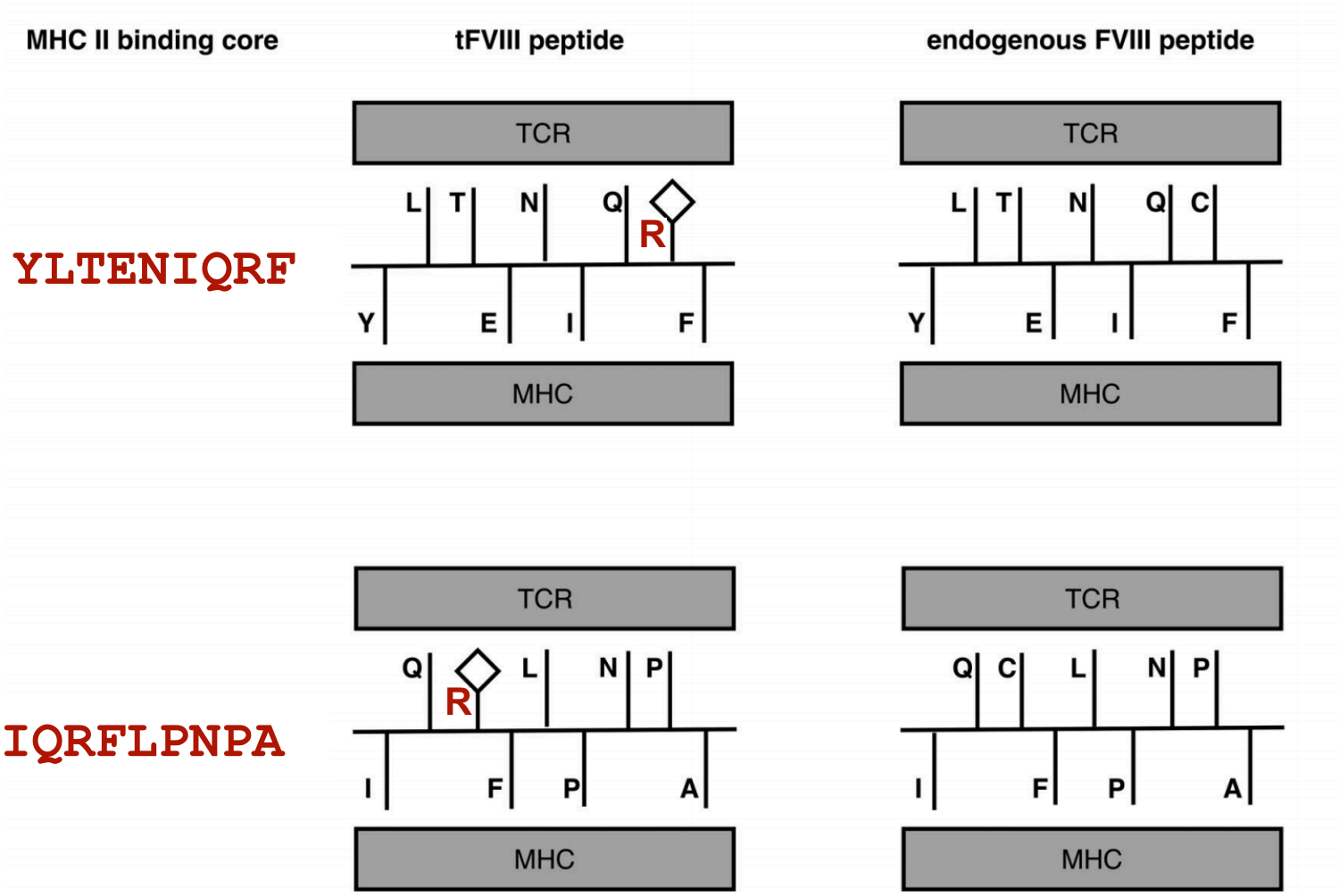
**tFVIII sequence spanning 593:** ...FDEBRSWYLTENIQRFLPNPAGVQLEDPE...

| 15-mer | MHC II binding core | predicted binding affinity, $IC_{50}$ (nmol/l) |
|---|---|---|
| FDEBRSWYLTENIQR | FDEBRSWYL | 601.9 |
| DEBRSWYLTENIQRF | YLTENIQRF | 118.9 |
| EBRSWYLTENIQRFL | YLTENIQRF | 33.3 |
| BRSWYLTENIQRFLP | YLTENIQRF | **39.4** |
| RSWYLTENIQRFLPN | YLTENIQRF | 28.4 |
| SWYLTENIQRFLPNP | YLTENIQRF | 75.8 |
| WYLTENIQRFLPNPA | IQRFLPNPA | 103.9 |
| YLTENIQRFLPNPAG | IQRFLPNPA | 54.8 |
| LTENIQRFLPNPAGV | IQRFLPNPA | 21.3 |
| TENIQRFLPNPAGVQ | IQRFLPNPA | **11.9** |
| ENIQRFLPNPAGVQL | FLPNPAGVQ | 9.0 |
| NIQRFLPNPAGVQLE | FLPNPAGVQ | 8.9 |
| IQRFLPNPAGVQLED | FLPNPAGVQ | 9.6 |
| QRFLPNPAGVQLEDP | FLPNPAGVQ | 16.7 |
| RFLPNPAGVQLEDPE | FLPNPAGVQ | 33.8 |

multiple predicted binding peptides spanning R593 with **HLA-DRB1*0101**

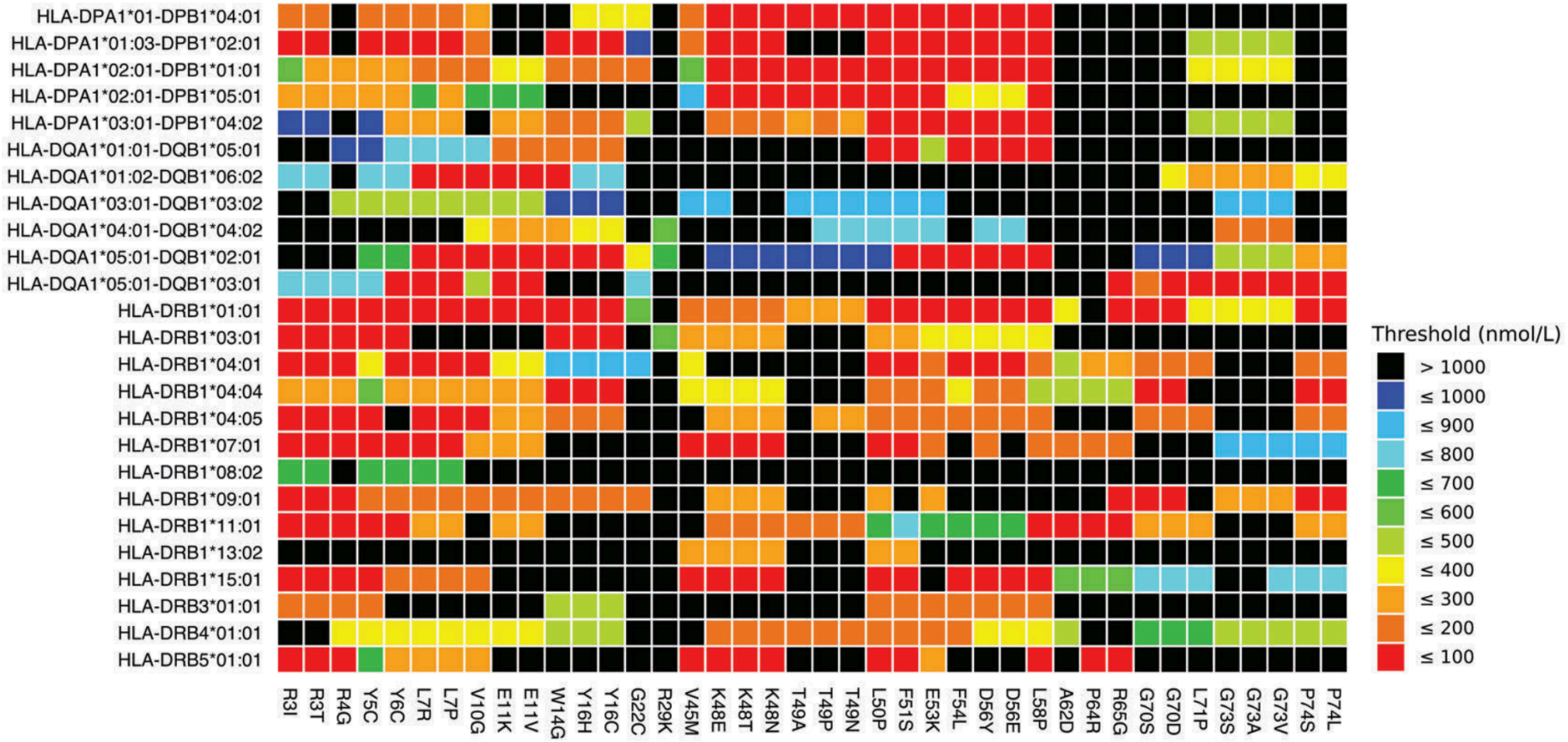
*Source: Hart et al., Haematologica, 2019*

# pMHC surface novelty with R593C




*Source: Hart et al., Haematologica, 2019*

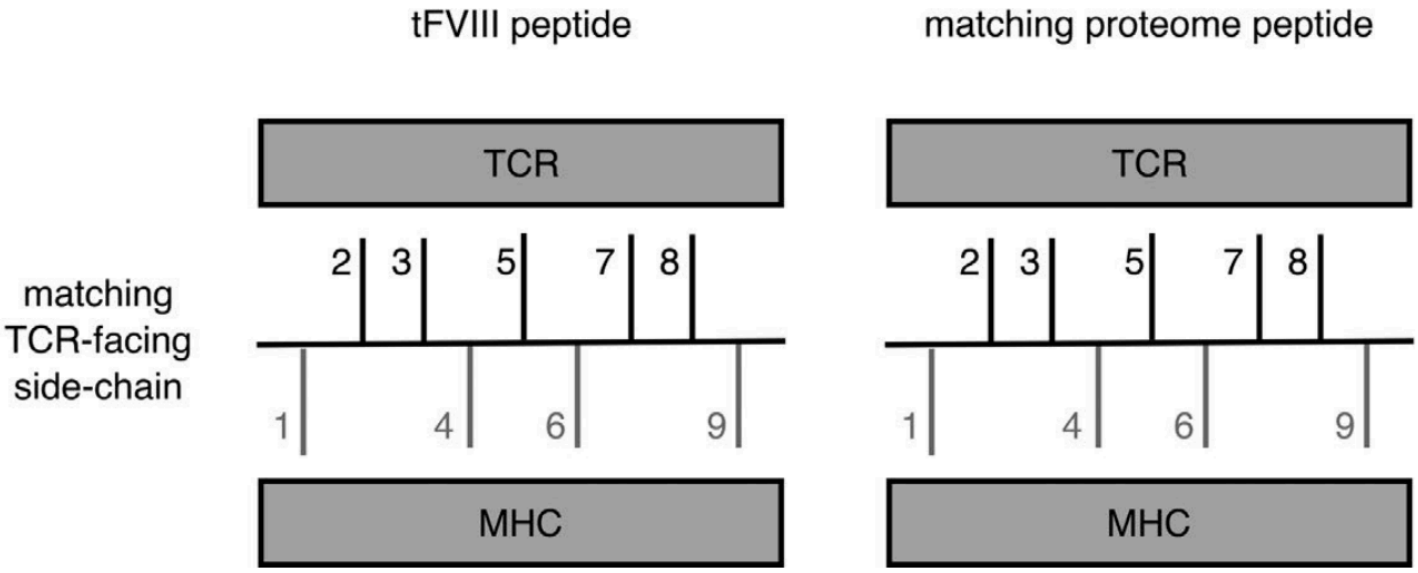
# Heatmap of inhibitor risk (1)




Percentage of risk-associated combinations = 49%

*Source: Hart et al., Haematologica, 2019*

# Proteome scanning




Complete human proteome from Ensembl contains over 100,000 proteins (including alternative isoforms) and 11,272,502 unique 9-mers

*Source: Hart et al., Haematologica, 2019*

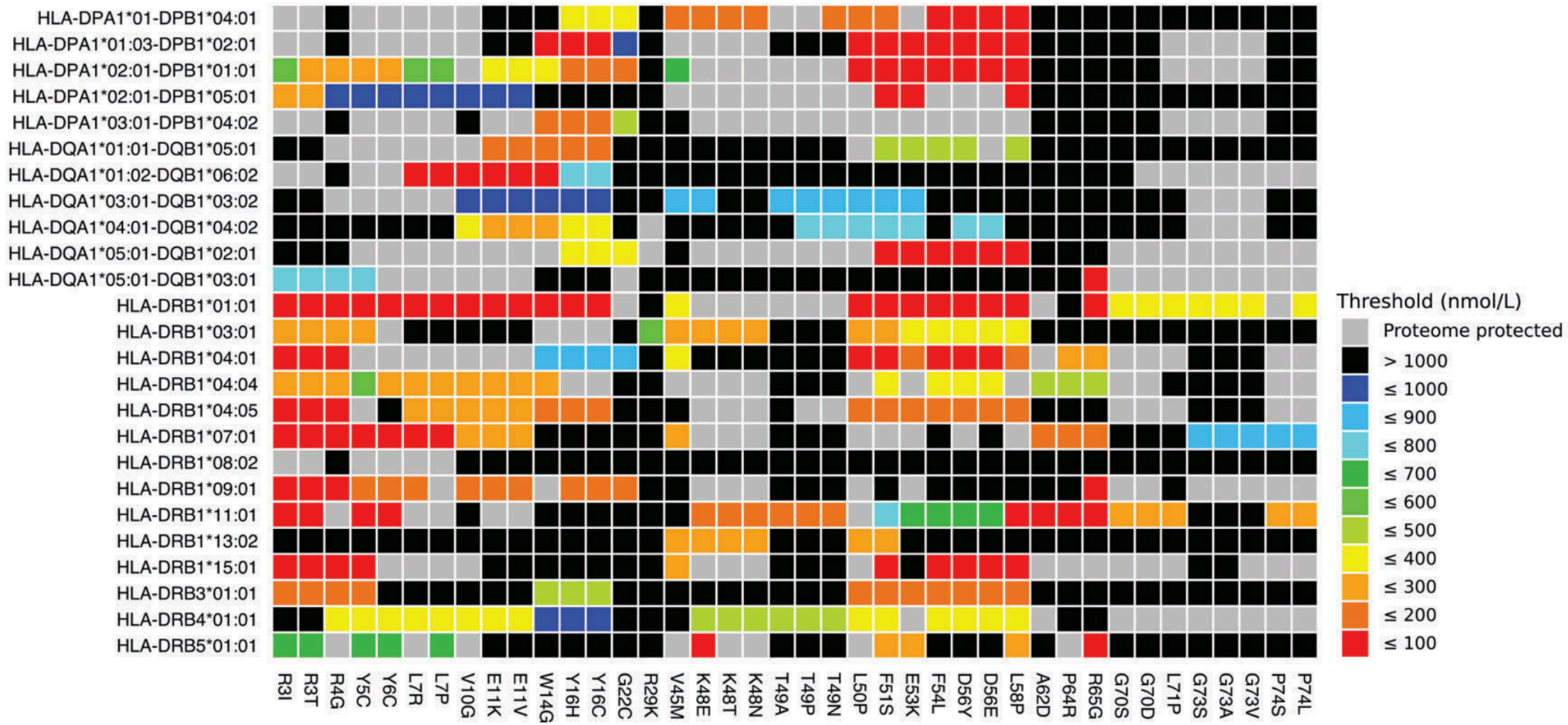
# Proteome scanning with R593C

Peptide **IQRFLPNPA** has the following TCR-facing pattern:
**XQRXLXNPX**

| 15-mer | MHC II binding core | predicted binding affinity, $IC_{50}$ (nmol/l) |
|---|---|---|
| **tubulin polyglutamylase matches to pattern XQRXLXNPX :** | | |
| SGRAASFQRELNNPL | FQRELNNPL | 26.3 |
| GRAASFQRELNNPLK | FQRELNNPL | 13.8 |
| RAASFQRELNNPLKR | FQRELNNPL | 7.9 |
| AASFQRELNNPLKRM | FQRELNNPL | 5.3 |
| ASFQRELNNPLKRMK | FQRELNNPL | 6.4 |
| SFQRELNNPLKRMKE | FQRELNNPL | 8.2 |
| FQRELNNPLKRMKEE | FQRELNNPL | 11.3 |

*Source: Hart et al., Haematologica, 2019*

# Heatmap with proteome scanning




Percentage of risk-associated combinations = 31%

*Source: Hart et al., Haematologica, 2019*

# Proteins with most cross-matches

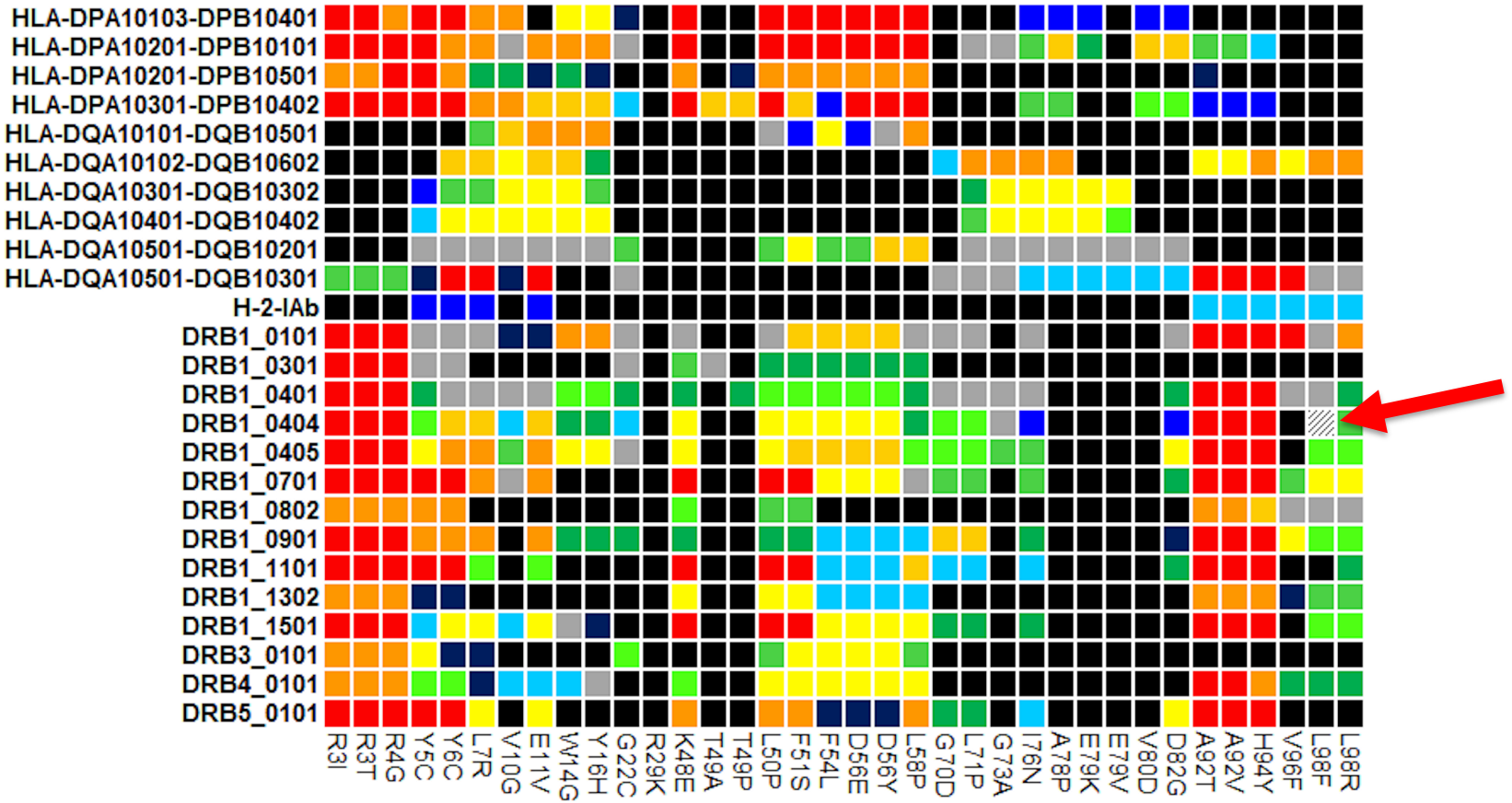
| UniProt ID | Protein name | Protected peptide count |
|---|---|---|
| P12259 | Coagulation factor V | 640 |
| Q6MZM0 | Hephaestin-like protein 1 | 457 |
| P00450 | Ceruloplasmin | 437 |
| Q9BQS7 | Hephaestin | 389 |
| P00451 | Coagulation factor VIII [match to different, but homologous, location within the protein] | 251 |
| O75445 | Usherin | 150 |
| Q14585 | Zinc finger protein 345 | 142 |
| Q14587 | Zinc finger protein 268 | 134 |

*

* all proteins with copper-binding sites

*Source: Hart et al., Haematologica, 2019*

# Taking SNPs into account




SNPs with >5% population frequency reduced predicted risk by only 0.1%

*MRes student Jacob Household*

# Computational Challenges

# Antigen presentation and binding

What can we predict? We have always used the "Denmark tools" from DTU (now part of the Department of Health Technology):

Class I:

NetMHCpan 4.0: predicts binding strength

NetCTLpan 1.1: predicts cleavage, TAP transport + binding strength

Class II:

NetMHCIIpan 3.2: predicts binding strength + binding register

# Antigen binding thresholds (class I)

There is controversy about absolute vs. relative (percentile rank) thresholds. This is reflected in the recommendations of the IEDB website. MHC class I:

1) "Currently recommends" making selections based on percentile rank of <= 1% for each MHC allele/length combination (wrt to reference set of 1 million random natural sequences)

2) "Alternatively" use $IC_{50}$ binding threshold of 500 nM

3) Recently, "a paper from our group showed that… MHC-specific thresholds should be used."

IMMUNE EPITOPE DATABASE AND ANALYSIS RESOURCE

# Antigen binding thresholds (class 2)

These are the IEDB recommendations for MHC class II:

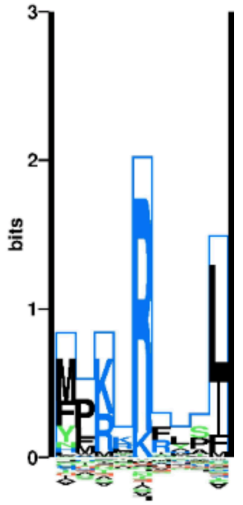
1) percentile rank of top 10%
2) $IC_{50}$ threshold of 1,000 nM

Note the following advice for both class I and class II under Interpreting predicted results:

"As a rough guideline, peptides with $IC_{50}$ values <50 nM are considered high affinity, <500 nM intermediate affinity and <5000 nM low affinity."

IMMUNE EPITOPE DATABASE AND ANALYSIS RESOURCE

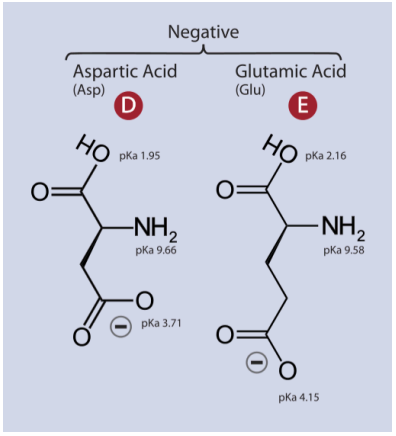
# Peptide anchors

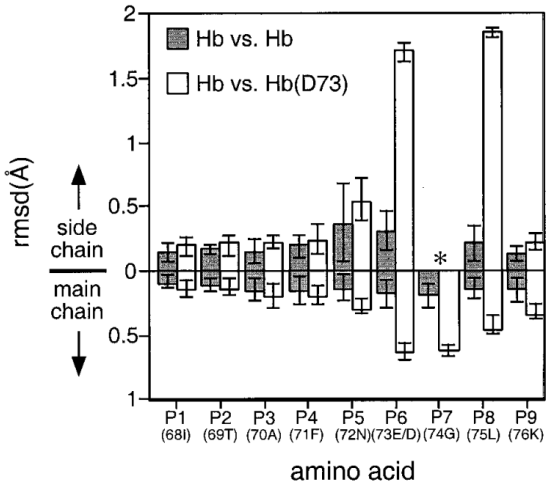



Which are the anchoring positions?
"Canonical" MHC class I & II molecules have pockets at positions 2 + 9 and 1 + 4 + 6 + 9 respectively. But some have additional/ alternative pockets.

HLA-B*08:01 using NetMHCpan v2 *(Source: MHC Motif Viewer)*

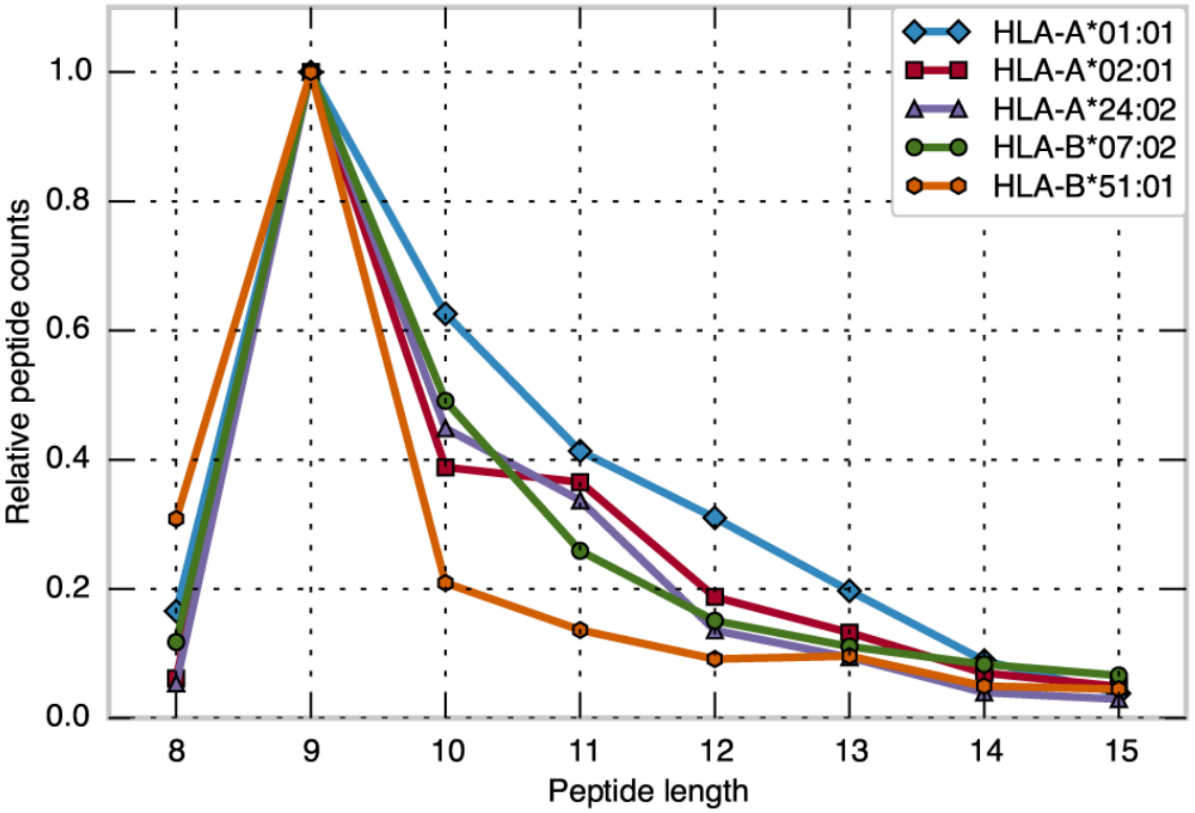
Evidence that mutations to anchoring residues may introduce novelty via main chain displacement
*(Kersh et al., Jol, 2001)*



# Challenges: class I peptide lengths

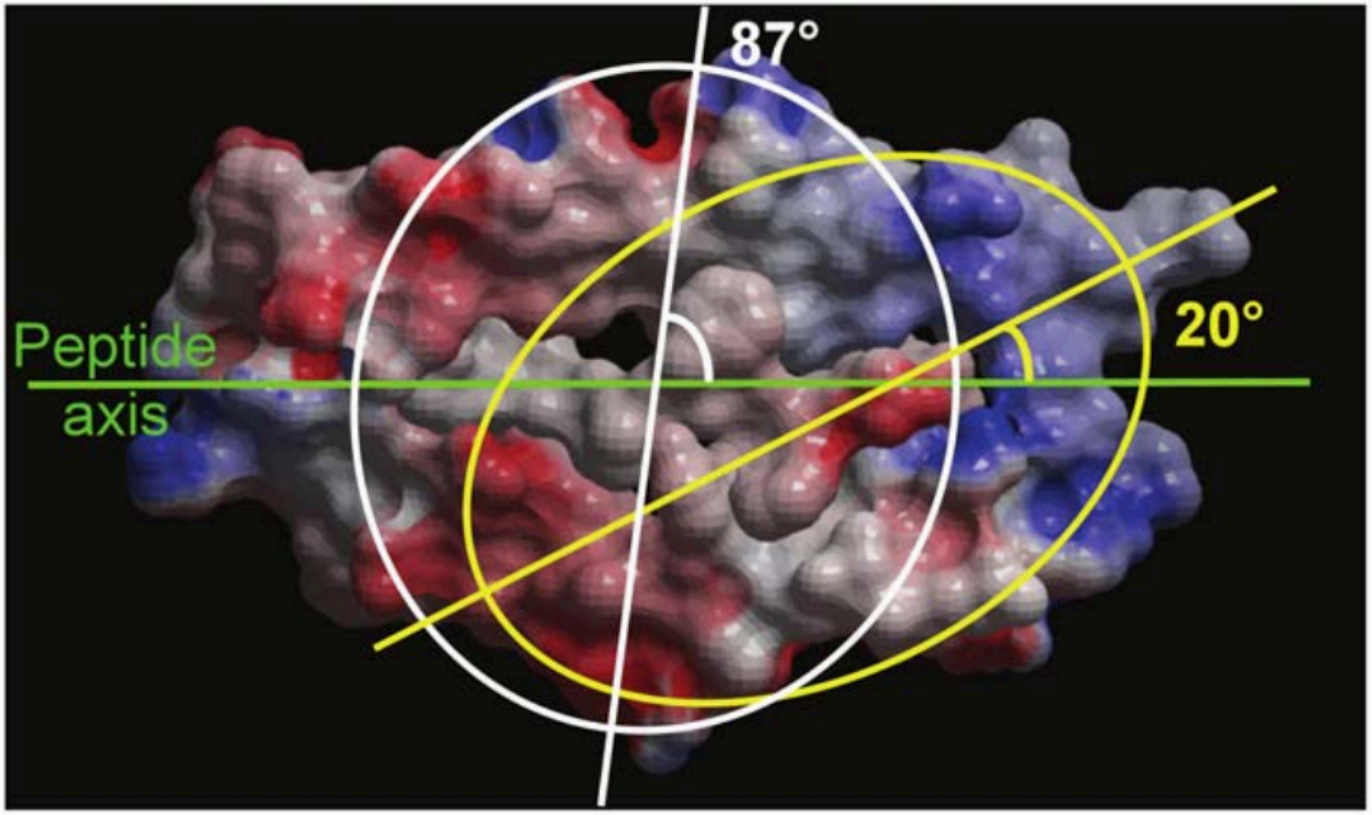
Most people (including members of my group) focus on 9-mers.




Length of naturally-presented peptides for 5 HLAs

*Trolle et al., JoI, 2016*

# Challenges: TCR binding orientation




*Source: Khan & Ranganathan, PLoS ONE, 2011*

# Acknowledgements


## My Group

Naz Uzun
Stuart Skelton
Jacob Househam
Alison Kakoschke
Zainab Aziz
William Lees

## Collaborators

### Haemophilia A

Dr Dan Hart (QMUL, Barts)

### HCC

Dr Shilpa Chokshi (Inst Hepatology)

The AIRR Community