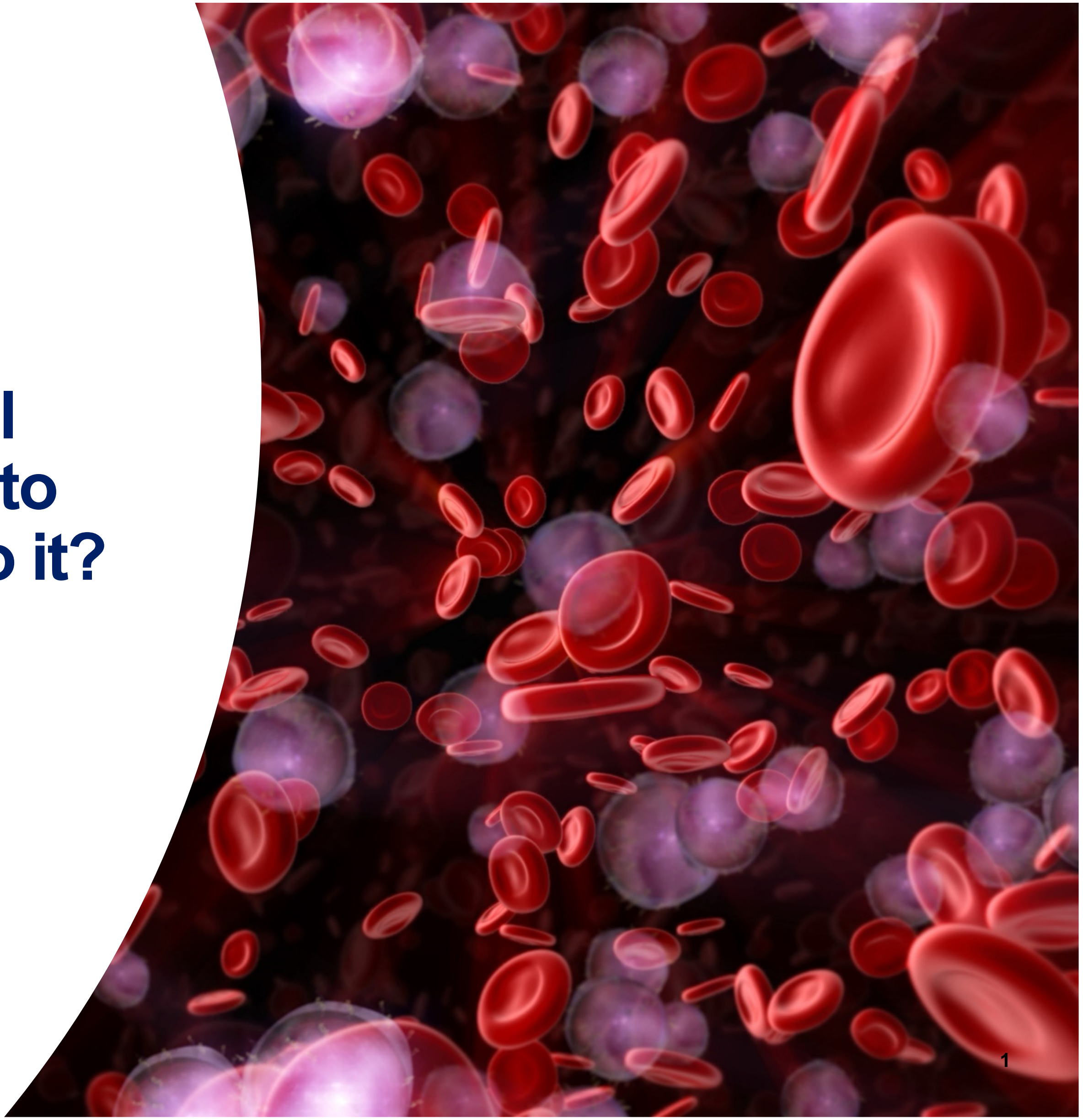
How to translate pre-clinical *in silico* IG profiling results to clinical IG rates – can we do it?

Dr. Michael Gutknecht
Lisbon
March 2026

NOVARTIS | Reimagining Medicine



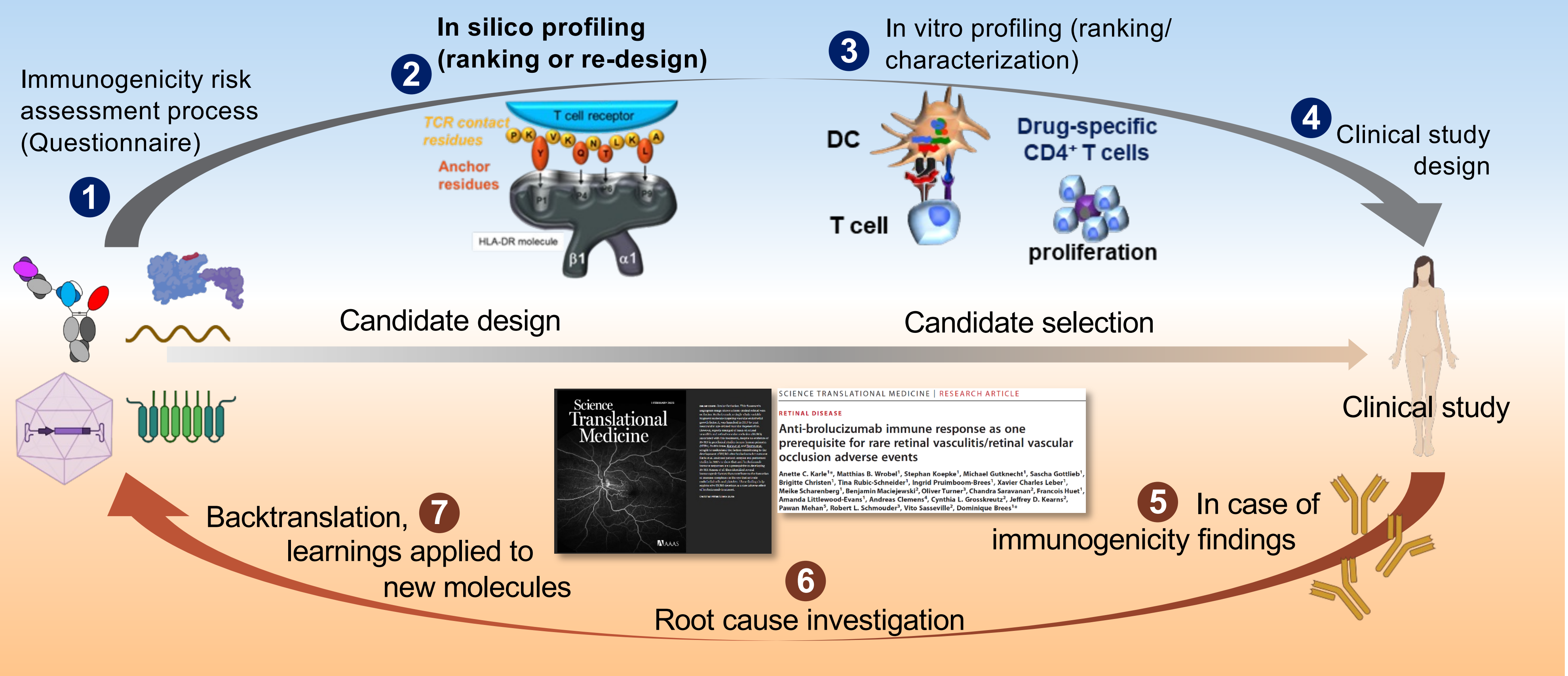
in silico Immunogenicity Profiling at Novartis

NOVARTIS | Reimagining Medicine

in silico Immunogenicity Profiling - advantages and limitations

- IG profiling should be started as timely as possible in the biotherapeutic development process to inform necessary de-immunization approaches early on and to avoid resource spending on candidates with a high inherent IG potential in later stages.
- Oftentimes, this is only possible using in silico tools, since in early drug development, high-quality candidate material is not available in the quantities necessary for most in vitro assays.
- Additionally, high cost and long timelines of in vitro assays are also factors that can be hurdles for pharma and biotech companies alike.
- Limiting factors are still the prediction accuracy, especially for B cell epitopes, and that additional aspects like aggregation, PTMs, change in structure upon grafting and endolysosomal processing can’t be assessed.

Immunogenicity strategy for biotherapeutics



1 Immunogenicity risk assessment process (Questionnaire)

2 **In silico profiling (ranking or re-design)**

3 In vitro profiling (ranking/ characterization)

Drug-specific CD4⁺ T cells

proliferation

4 Clinical study design

Candidate design

Candidate selection

Clinical study

5 In case of immunogenicity findings

6 Root cause investigation

7 Backtranslation, learnings applied to new molecules

SCIENCE TRANSLATIONAL MEDICINE | RESEARCH ARTICLE

RETINAL DISEASE

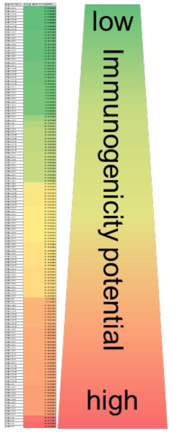
Anti-brolucizumab immune response as one prerequisite for rare retinal vasculitis/retinal vascular occlusion adverse events

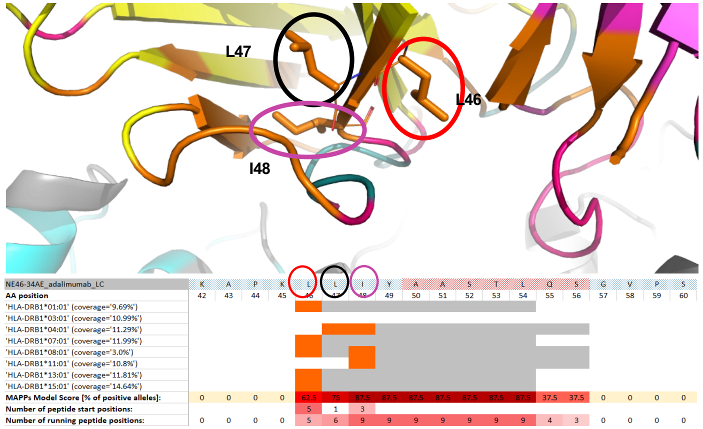
Anette C. Karle[1]*, Matthias B. Wrobel[1], Stephan Koepke[1], Michael Gutknecht[1], Sascha Gottlieb[1], Brigitte Christen[1], Tina Rubic-Schneider[1], Ingrid Pruimboom-Brees[1], Xavier Charles Leber[1], Meike Scharenberg[1], Benjamin Maciejewski[2], Oliver Turner[3], Chandra Saravanan[2], Francois Huet[1], Amanda Littlewood-Evans[1], Andreas Clemens[4], Cynthia L. Grosskreutz[2], Jeffrey D. Kearns[2], Pawan Mehan[5], Robert L. Schmouder[3], Vito Sasseville[2], Dominique Brees[1]*

in silico Immunogenicity Profiling at Novartis

☐ Prediction of HLA class II binding hotspots based on a PSSM

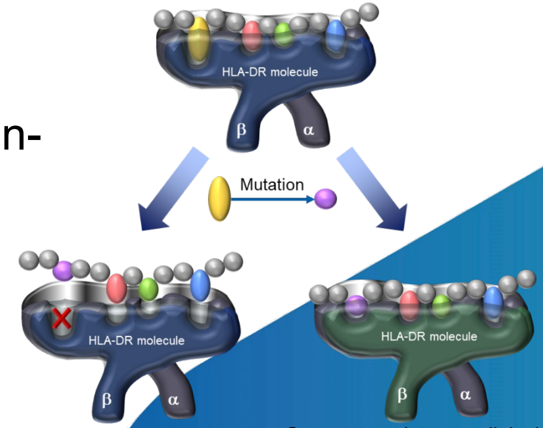
- **IG profiling of large candidate sets**
  early in the development process based on hotspots and CDR overlap





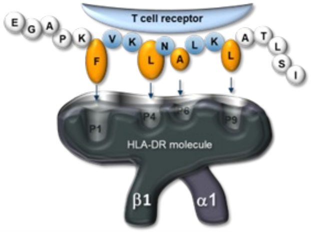
NE46-34AE_adalimumab_LC	K	A	P	K	L	L	I	Y	A	A	S	T	L	Q	S	G	V	P	S
AA position	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60
'HLA-DRB1*01:01' (coverage='9.69%')					🟧	▨	▨	▨	▨	▨	▨	▨	▨						
'HLA-DRB1*03:01' (coverage='10.99%')																			
'HLA-DRB1*04:01' (coverage='11.29%')						🟧	🟧	▨	▨	▨	▨	▨	▨	▨	▨				
'HLA-DRB1*07:01' (coverage='11.99%')					🟧	▨	▨	▨	▨	▨	▨	▨	▨						
'HLA-DRB1*08:01' (coverage='3.0%')					🟧	▨	🟧	▨	▨	▨	▨	▨	▨	▨	▨				
'HLA-DRB1*11:01' (coverage='10.8%')							🟧	▨	▨	▨	▨	▨	▨	▨	▨				
'HLA-DRB1*13:01' (coverage='11.81%')					🟧	▨	▨	▨	▨	▨	▨	▨	▨						
'HLA-DRB1*15:01' (coverage='14.64%')					🟧	▨	▨	▨	▨	▨	▨	▨	▨						
MAPPs Model Score [% of positive alleles]:	0	0	0	0	62.5	75	87.5	87.5	87.5	87.5	87.5	87.5	87.5	37.5	37.5	0	0	0	0
Number of peptide start positions:					5	1	3												
Number of running peptide positions:	0	0	0	0	5	6	9	9	9	9	9	9	9	4	3	0	0	0	0

- **De-immunization of hotspots**
  via sequences randomization to find non-binders that are confirmed via MAPPs assay

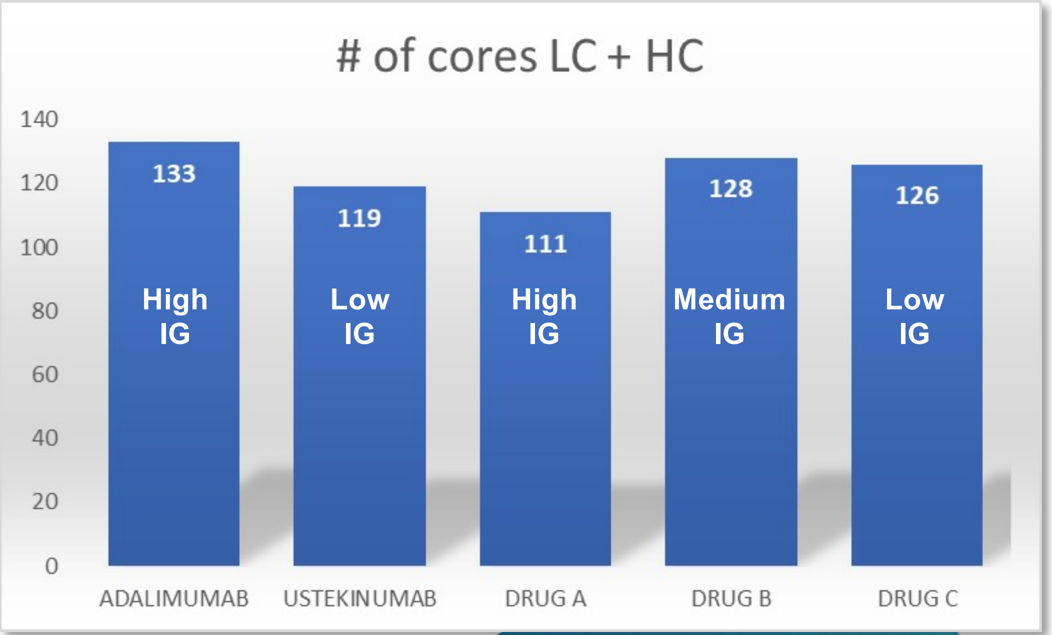


Can we translate pre-clinical in silico IG profiling results to clinical IG rates?

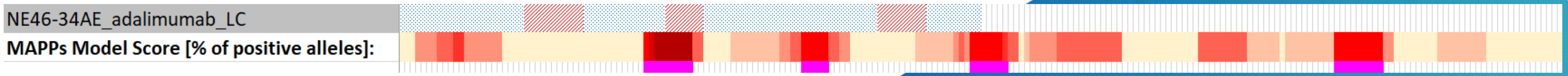
How we started –
Candidate ranking based on # of cores



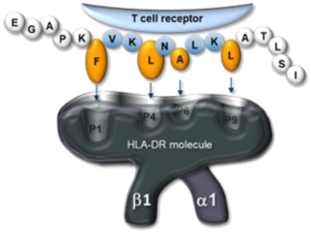
- ☐ **The simplest way to rank candidates is just using the # of predicted binding peptide cores in each sequence**
- ☐ **But this is only a very high-level analysis, which offers no option to address sequence specific questions!**
  - ➢ Like generation of neo-epitopes by introduction of Fc modifications, non-natural junctions, etc.



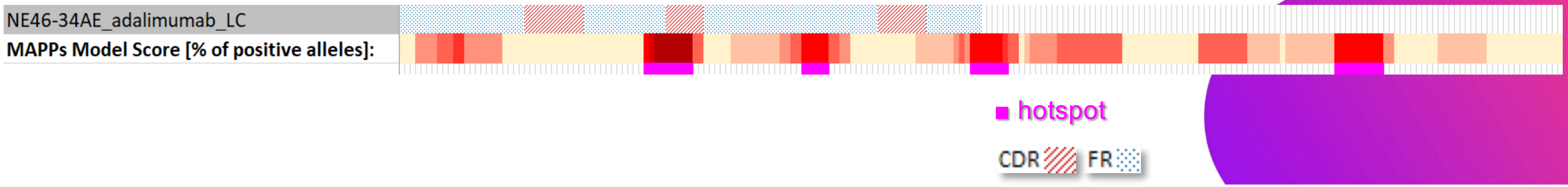
High / Low IG = relatively high / low clinical IG rate (% of ADA)



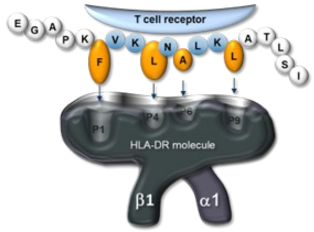
How we improved –
Introduction of a weigthing matrix



- ☐ **Only considering the # of predicted peptides is not enough to rank candidates properly!**
- ☐ **The quality of the hits is even more important and enables a better differentiation between candidates.**
  - ➢ Highly presented sequence regions harbor a greater risk – **hotspot ranking**
  - ➢ Sequence regions that the immune system does not “know” harbor a greater risk – **CDR overlap**

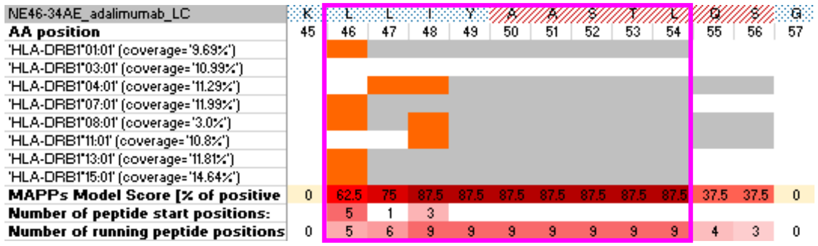


Hotspot ranking



What are hotspots?

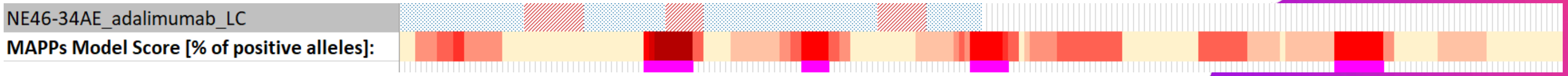
- Hotspots are sequence regions that show peptide binding to at least 5 out the 8 most frequent European HLA class II alleles



NE46-34AE_adalimumab_LC	K	L	L	I	Y	A	A	S	T	L	Q	S	G
AA position	45	46	47	48	49	50	51	52	53	54	55	56	57
'HLA-DRB1*01:01' (coverage='9.69%')													
'HLA-DRB1*03:01' (coverage='10.99%')													
'HLA-DRB1*04:01' (coverage='11.29%')													
'HLA-DRB1*07:01' (coverage='11.99%')													
'HLA-DRB1*08:01' (coverage='3.0%')													
'HLA-DRB1*11:01' (coverage='10.8%')													
'HLA-DRB1*13:01' (coverage='11.81%')													
'HLA-DRB1*15:01' (coverage='14.64%')													
MAPPs Model Score [% of positive	0	62.5	75	87.5	87.5	87.5	87.5	87.5	87.5	87.5	37.5	37.5	0
Number of peptide start positions:		5	1	3									
Number of running peptide positions	0	5	6	9	9	9	9	9	9	9	4	3	0

Hypothesis behind the hotspot ranking

- Would these sequence regions be recognized by T cells, then a larger proportion of the population could develop immunogenicity



NE46-34AE_adalimumab_LC

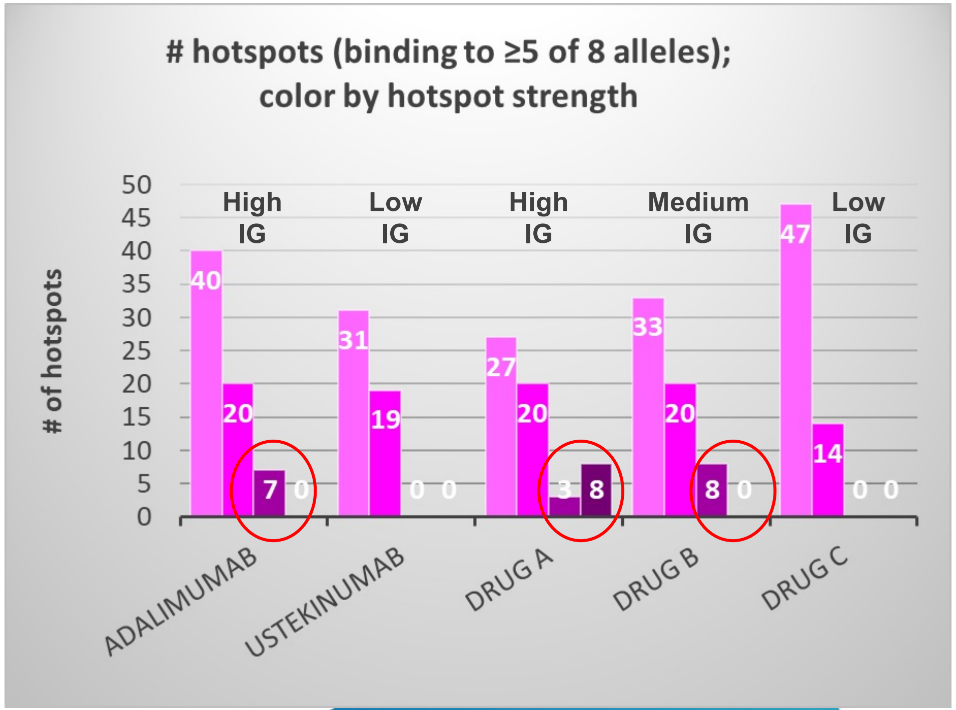
MAPPs Model Score [% of positive alleles]:

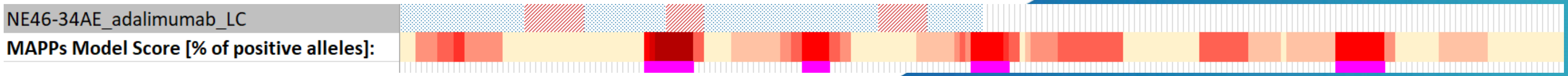
■ hotspot

CDR ▨ FR ░

Candidate ranking based on # of hotspots

- ❑ Hotspot ranking seems to contribute to candidate differentiation and shows a better correlation to known IG rates than only counting binding cores
- ❑ But can we do more?

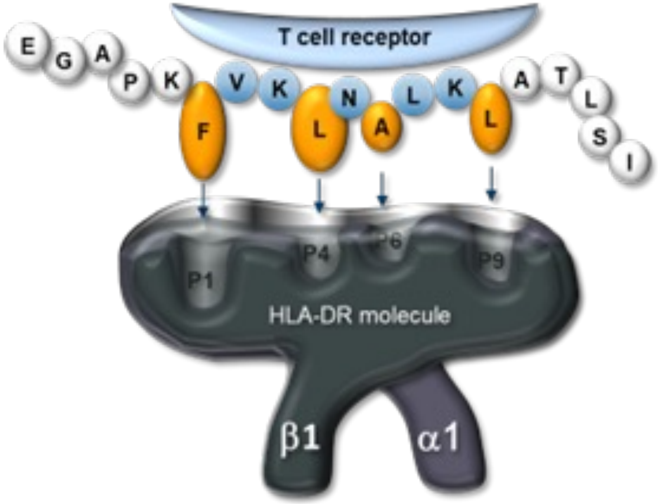




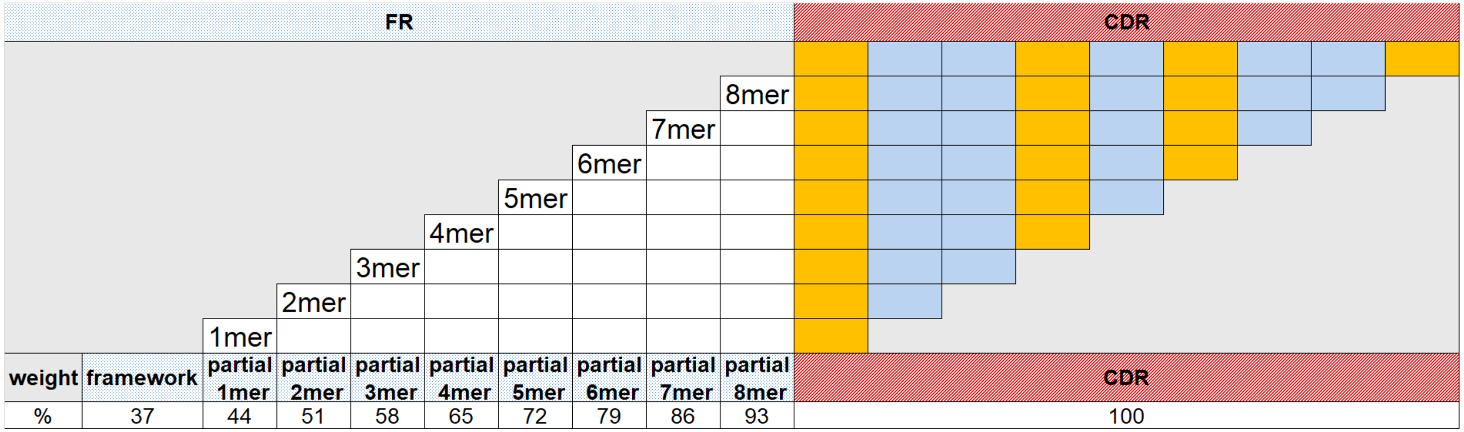
CDR overlap ranking

❑ Hypothesis behind the CDR overlap ranking

- ➢ T cell frequencies against foreign sequence regions (CDRs) are expected to be higher than for conserved human sequences in the framework (FR)



- ➢ Introduction of a "weighting system"

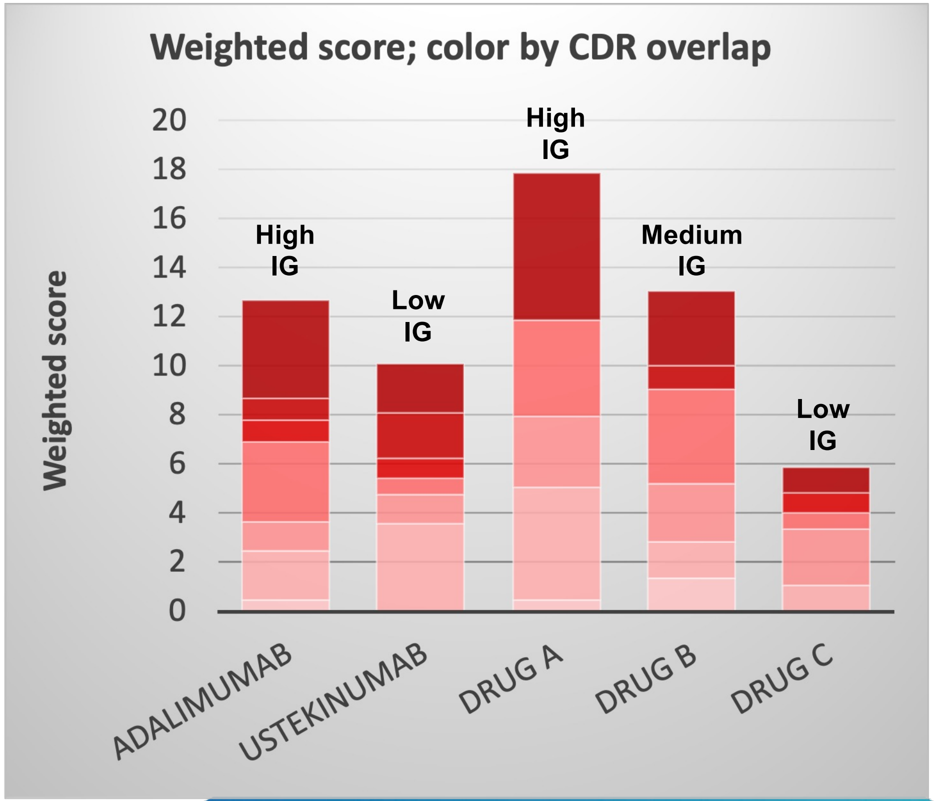


FR										CDR
									8mer	
								7mer		
							6mer			
						5mer				
					4mer					
				3mer						
			2mer							
		1mer								
weight	**framework**	**partial 1mer**	**partial 2mer**	**partial 3mer**	**partial 4mer**	**partial 5mer**	**partial 6mer**	**partial 7mer**	**partial 8mer**	**CDR**
%	37	44	51	58	65	72	79	86	93	100

Candidate ranking based on CDR overlap

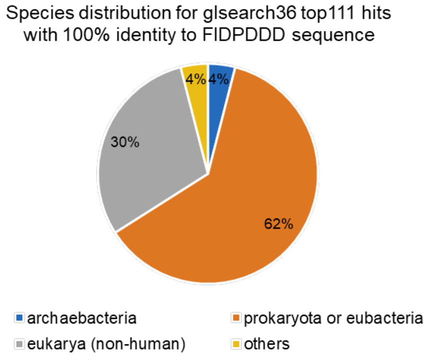
- ❑ Taking the CDR overlap into account, an even better differentiation between the candidates is possible!

- ➢ Are there additional parameters that we can include in our assessment to improve candidate ranking?

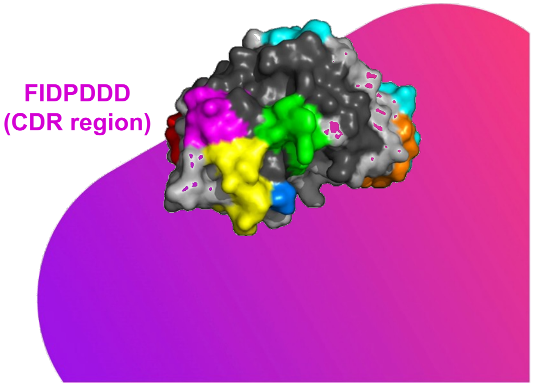


Additional parameters to improve candidate ranking

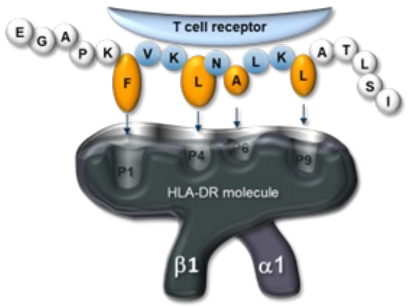
- Most in silico tools used in biotherapeutic development are predicting peptide binding to HLA class II molecules (e.g. NetMHCIIpan), frequently with the option to apply a weighting matrix, based on the **hypothesis that self-peptides and germline sequences have a lower IG potential.**



- Based on our experience during root cause analysis of adverse events in the clinic, we started to explore additional options to improve this weighting matrix. We could show that **biotherapeutic sequences can bear analogs to pathogen sequences**, which may result in a **high number of memory T cells** that are cross-specific to the biotherapeutic, as well as a **high prevalence of pre-existing anti-drug antibodies**.



New *in silico* immunogenicity profiling approach based on biotherapeutic / pathogen analogy



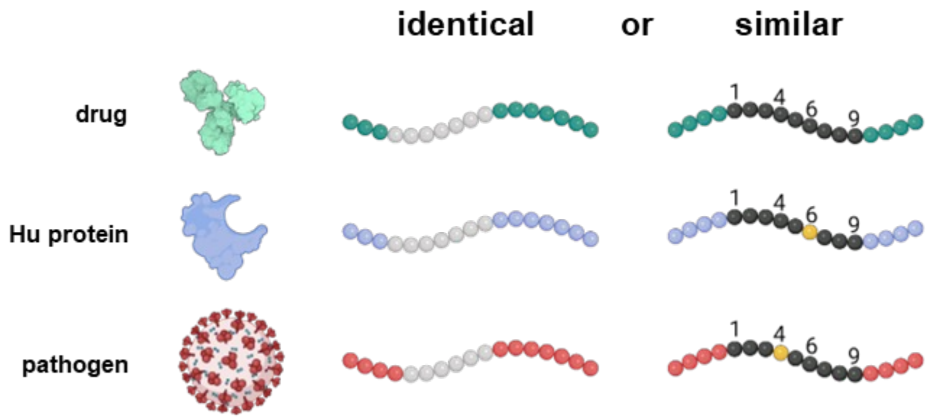
Picture created with BioRender.com

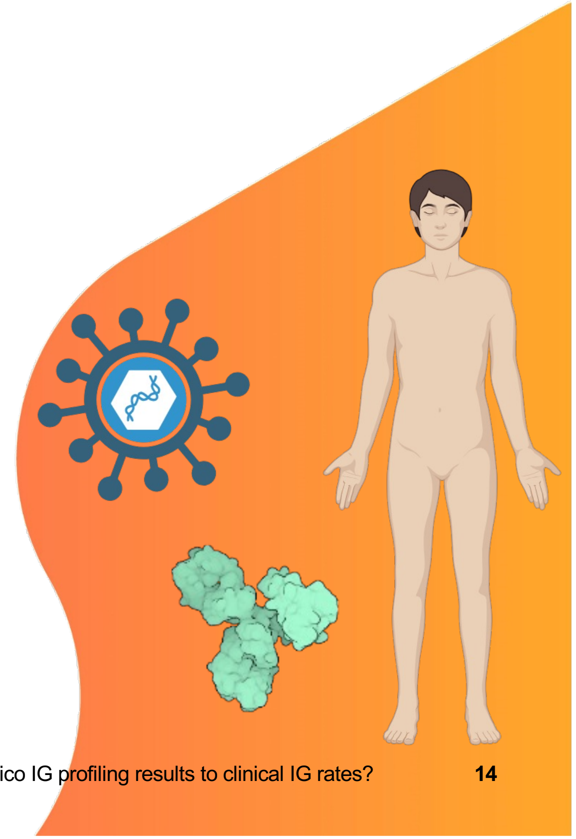
Hypothesis:

Biotherapeutic sequences can bear analogues to pathogen sequences.

These potential cross-reactive T & B cell epitopes may induce a strong immunogenicity response in a large proportion of the patient population.

New tools for the identification of:



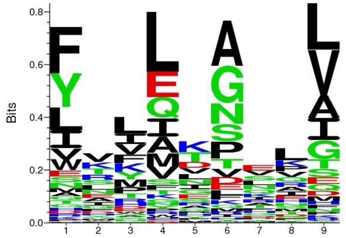


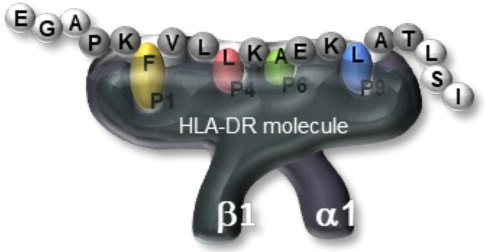
in silico toolbox

iSHAPe

(*in silico* HLA agretope prediction)

Predictive model for HLA class II binding peptides (potential T cell epitopes)

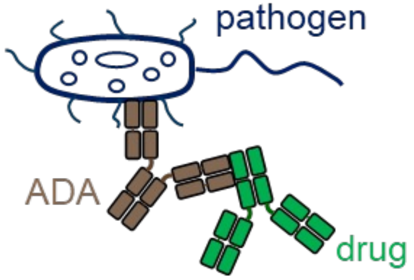




AP-BLAST

(Antigen vs Pathogen Blast)

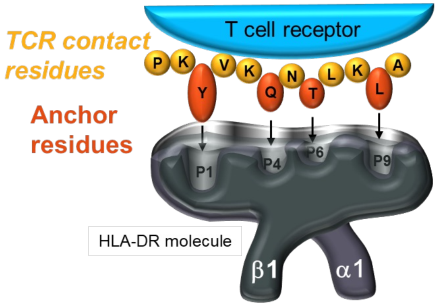
Alignment of drug sequence with other proteins to find analogous sequences which could lead to ADA cross-reactivity if surface exposed



MASE

(MASE: MAPPs Agretope Similarity Evaluation)

Predictive model for HLA class II binding peptides that could be recognized by cross-reactive T cells

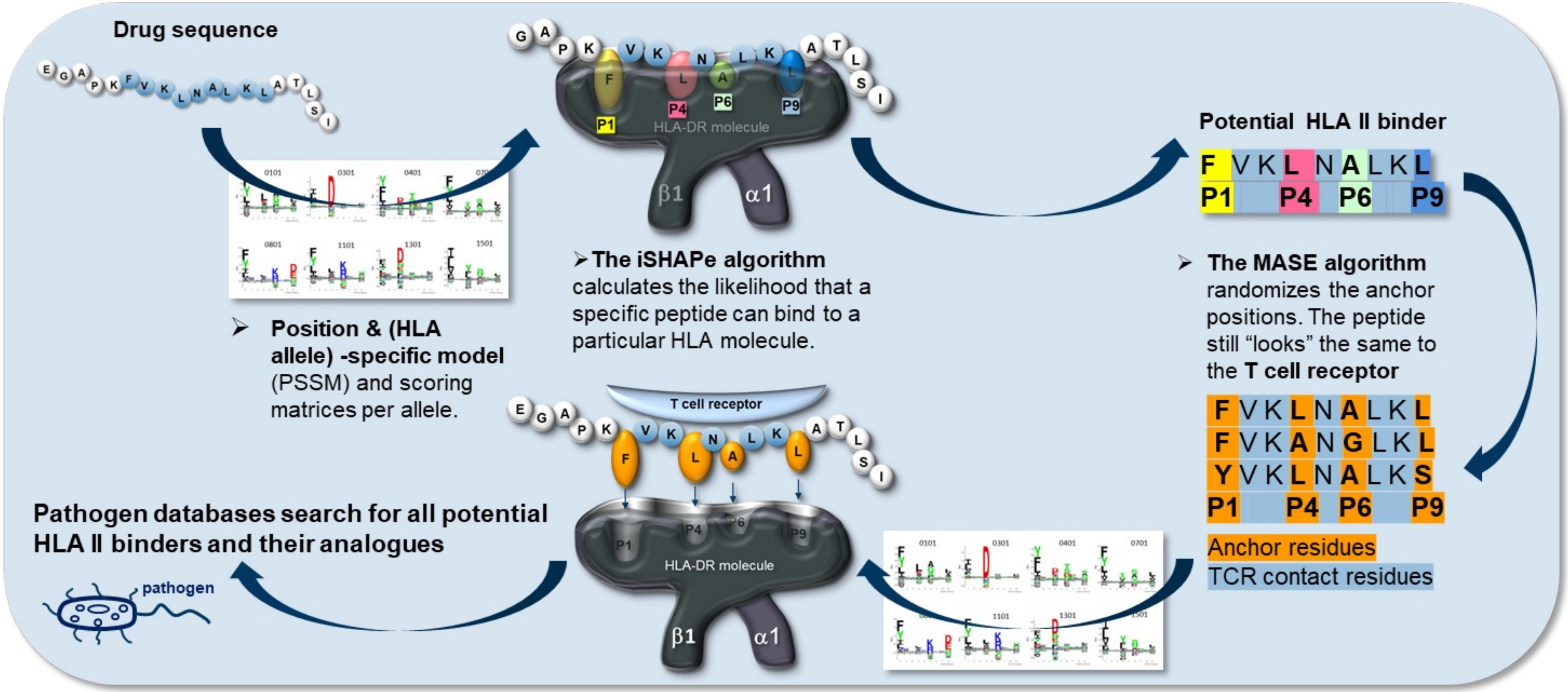


Different anchors but peptide “looks” the same to the T cell

PKYVKQNTLKLA
PKYVKHNTLKLA
PKYVKQNTLKIA

The principle of MASE (MAPPs Aggretope Similarity Evaluation)

MASE predicts peptides that may show T cell cross-reactivity due to sequence identity or similarity between drug sequences and pathogen sequences



Drug sequence

- **Position & (HLA allele) -specific model** (PSSM) and scoring matrices per allele.

- **The iSHAPe algorithm** calculates the likelihood that a specific peptide can bind to a particular HLA molecule.

Potential HLA II binder

F V K L N A L K L
P1 P4 P6 P9

- **The MASE algorithm** randomizes the anchor positions. The peptide still “looks” the same to the **T cell receptor**

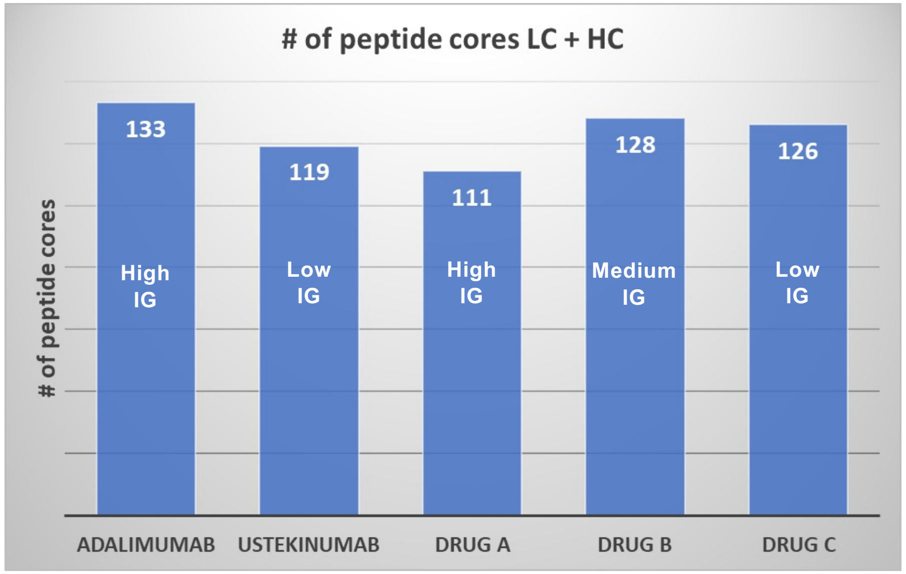
F V K L N A L K L
F V K A N G L K L
Y V K L N A L K S
P1 P4 P6 P9

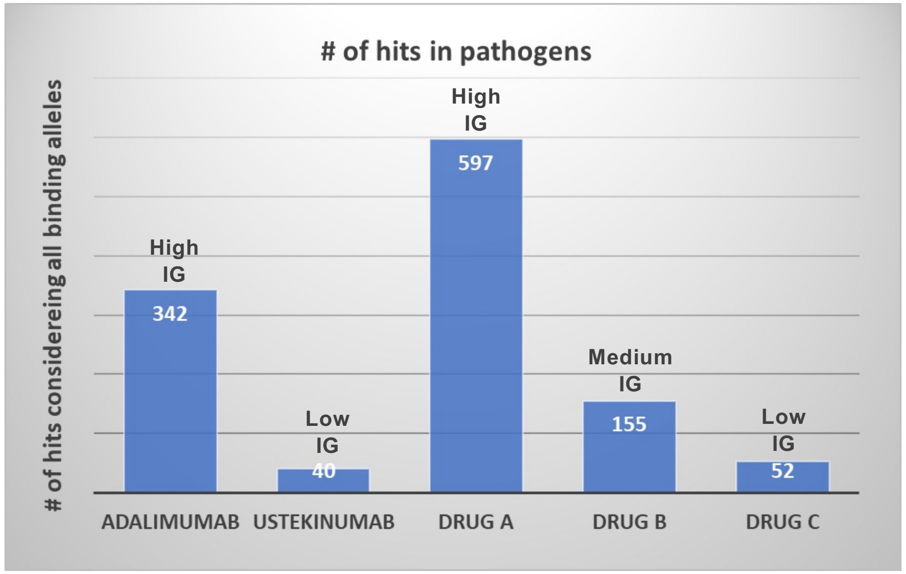
Anchor residues
TCR contact residues

Pathogen databases search for all potential HLA II binders and their analogues

Immunogenicity (IG) Profiling

Comparison between counting cores and biotherapeutic / pathogen similarity profiling



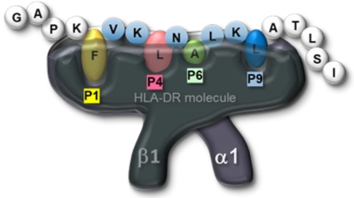


- ➢ **Implementation of biotherapeutic / pathogen analog profiling is a clear improvement for candidate ranking and shows better correlation to clinical IG rates!**

IGMotifFinder Pipeline: Identifying risky motifs in biotherapeutics

Brunno Rocha

- ☐ **Analysis of 200+ biotherapeutics (Marks et al.) with the goal to identify epitopes (Motifs) that correlate with higher clinical immunogenicity rates.**



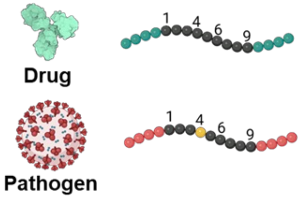
Therapies Dataset* → **HLA Binding Prediction** → **HLA Binders**

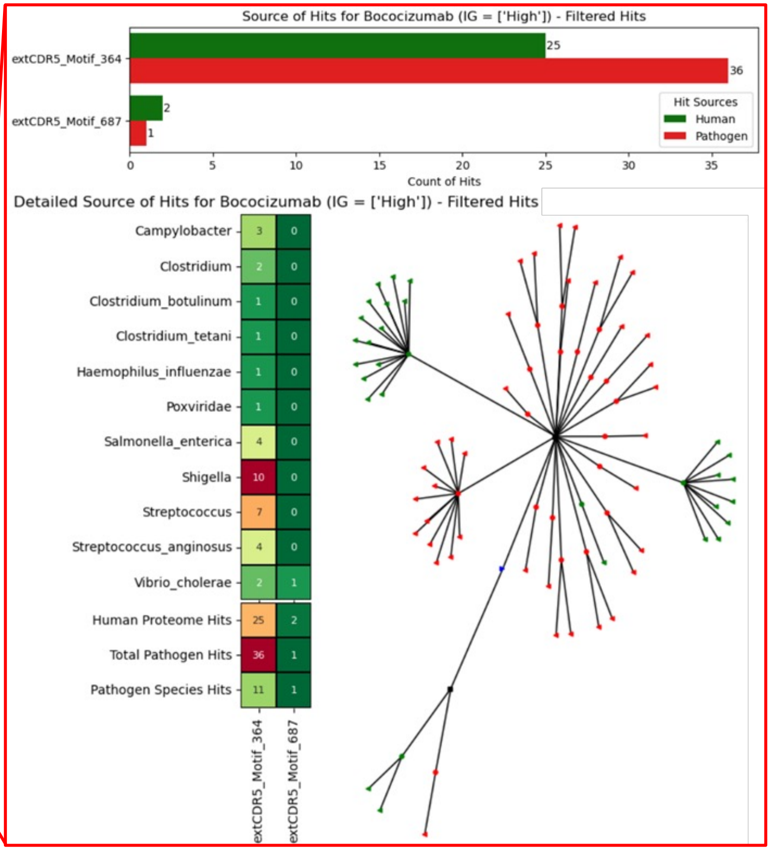
Pathogenic HLA Binders → **Cross-Reactive Epitopes Prediction**

HLA Binders ↓ **Cross-Reactive Epitopes Prediction** ↓ **Risky Motifs**

```
>Abagovomab_VH
QVKLQESGAELARPGASVKLSCKASGYTFTNYWMQWVKQRP
>Abrilumab_VH
QVQLVQSGAEVKKPGASVKVSCKVSGYTLSDLSIHWVRQAP
>Actoxumab_VH
QVQLVESGGGVVQPGRSLRLSCAASGFSFSNYGMHWVRQAP
>Adalimumab_VH
EVQLVESGGGLVQPGRSLRLSCAASGFTFDDYAMHWVRQAP
```

*Source: Marks et al. *Bioinformatics* 2021, *37*, 22, 4041-4047.





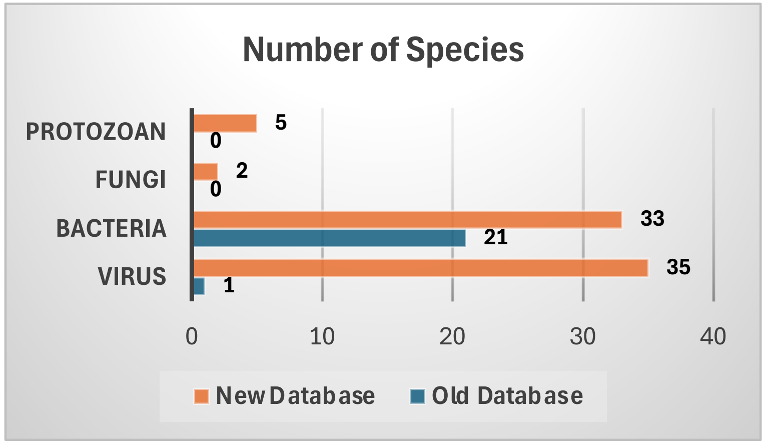
NOVARTIS | Reimagining Medicine

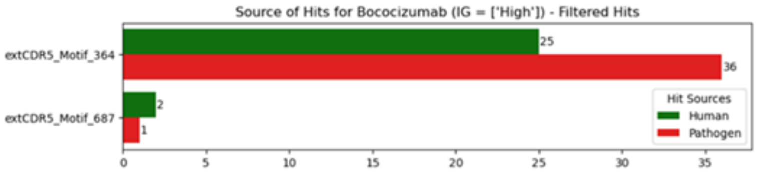
IGMotifFinder Pipeline: Identifying risky motifs in biotherapeutics

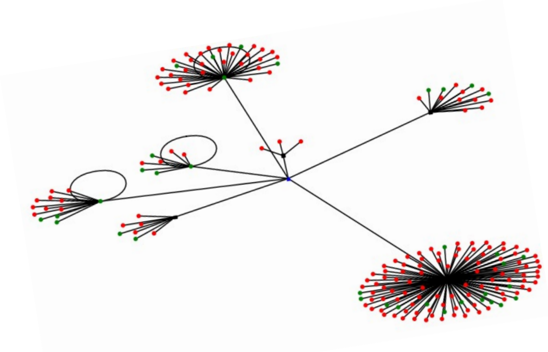
Brunno Rocha

Advantages of IGMotifFinder vs MASE:

- contains 3x the number of pathogen sequences in its database (75 vs 22)
- operates 5x faster than MASE, thanks to its indexed pathogen database
- Automatic generation of report ready figures



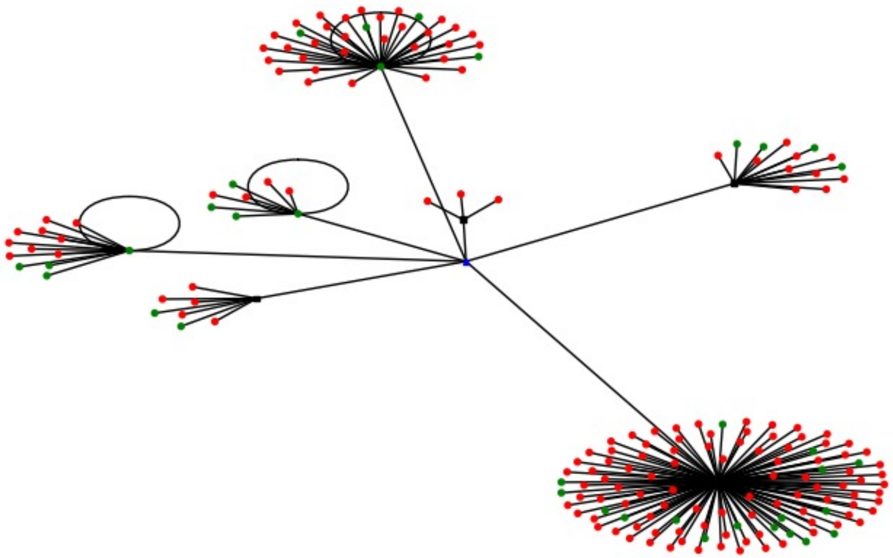




Post Doc Project: IGMotifFinder – Brunno Rocha

Brunno Rocha

- ☐ Network plots of MASE motifs in the **extCDR** can help distinguishing High/Moderate vs Low IG therapies by highlighting **pathogen** and **human** hits.

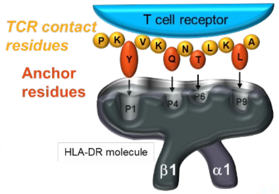


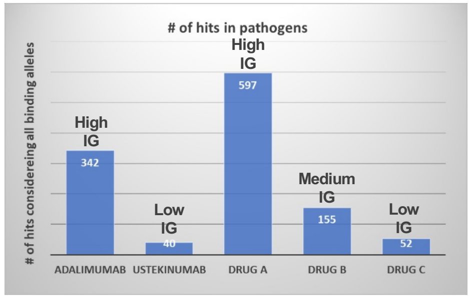
The nodes in the plots represent:

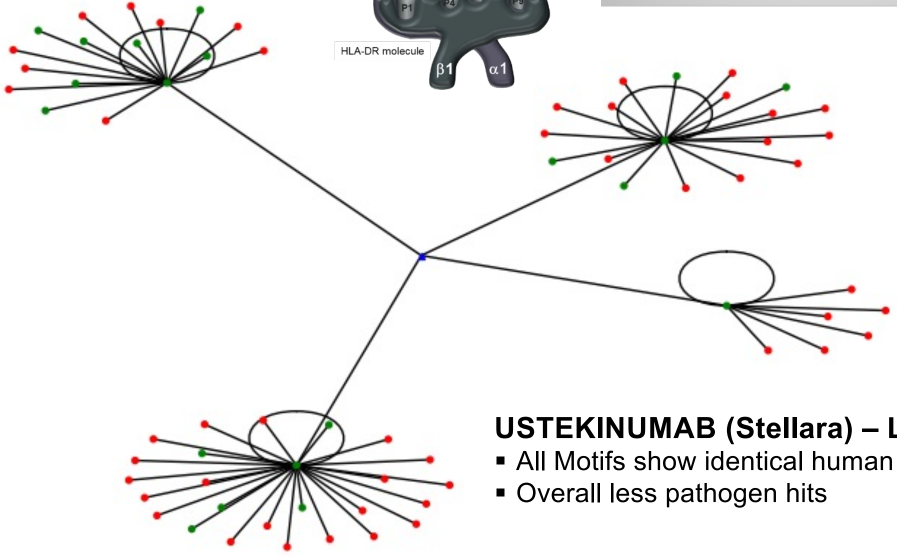
▲ 1st level = Biotherapeutic sequence

2nd level = HLA binding peptide core;
■ black square is an analog,
○ and the circle is an identical match

●● 3rd level = MASE-identified analogs







ADALIMUMAB (Humira) – High IG

- Less then half of the Motifs have an identical human hit
- Motifs with many pathogen hits

USTEKINUMAB (Stellara) – Low IG

- All Motifs show identical human hits
- Overall less pathogen hits

Can we translate pre-clinical in silico IG profiling results to clinical IG rates?

Post Doc Project: IGMotifFinder – Brunno Rocha

Brunno Rocha

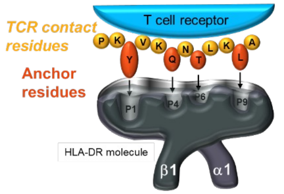
❑ Network plots of MASE motifs in the **extCDR** can help distinguishing High/Moderate vs Low IG therapies by highlighting **pathogen** and **human** hits.

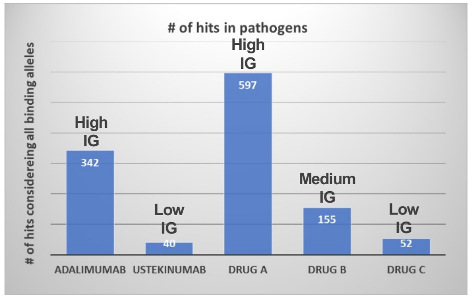
The nodes in the plots represent:

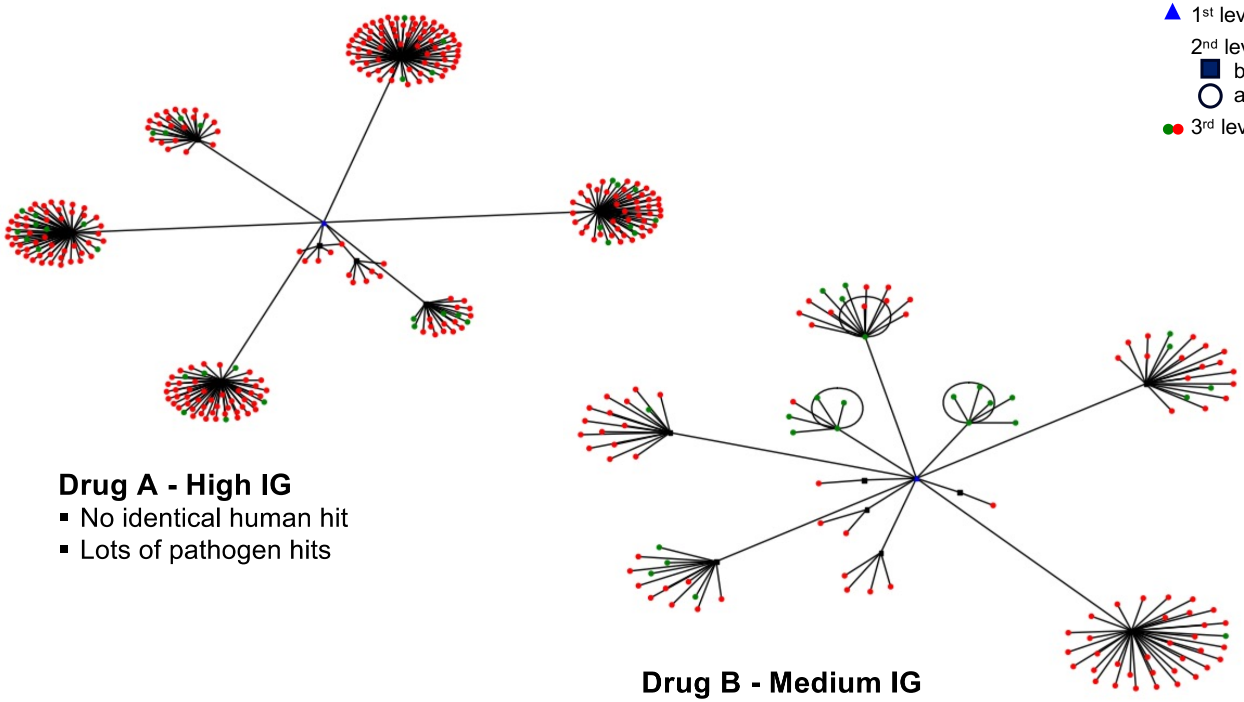
▲ 1st level = Biotherapeutic sequence

2nd level = HLA binding peptide core;
■ black square is an analog,
○ and the circle is an identical match

●● 3rd level = MASE-identified analogs

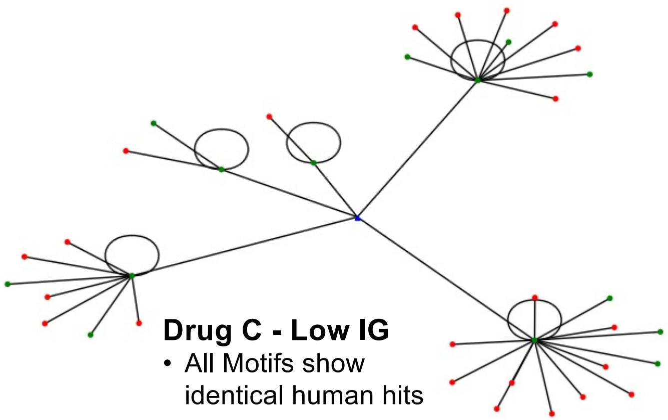






Drug A - High IG
- No identical human hit
- Lots of pathogen hits

Drug B - Medium IG
- Most Motifs show no identical human hit



Drug C - Low IG
- All Motifs show identical human hits

NOVARTIS | Reimagining Medicine

Can we translate pre-clinical in silico IG profiling results to clinical IG rates?

Post Doc Project: IGMotifFinder – Brunno Rocha

Brunno Rocha

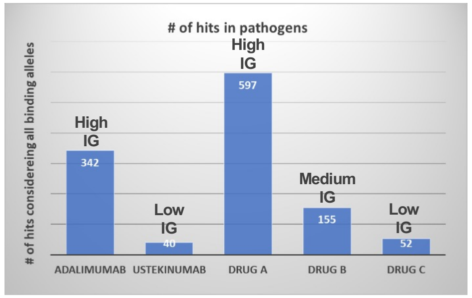
- ☐ Network plots of MASE motifs in the **full sequence** can help distinguishing High/Moderate vs Low IG therapies by highlighting **pathogen** and **human** hits.

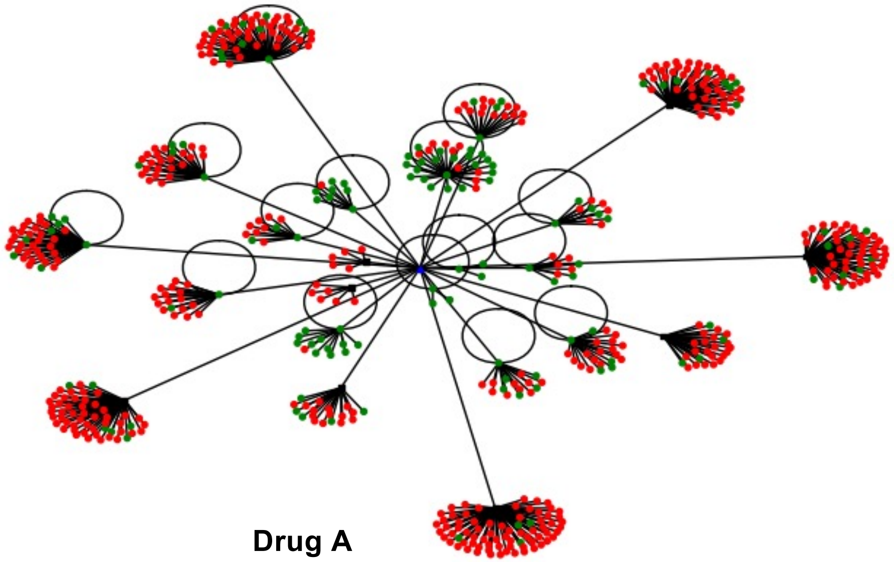
The nodes in the plots represent:

▲ 1st level = Biotherapeutic sequence

2nd level = HLA binding peptide core;
■ black square is an analog,
○ and the circle is an identical match

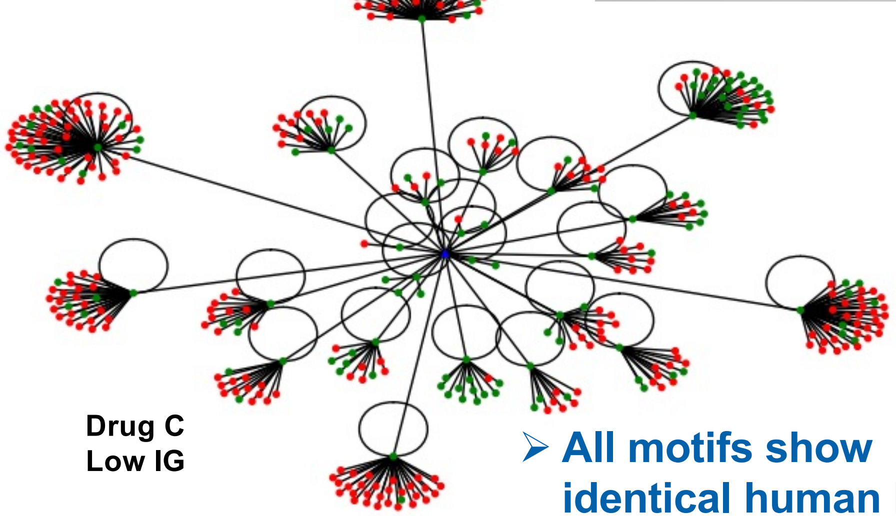
●● 3rd level = MASE-identified analogs





Drug A
High IG

- ➢ **all motifs w/o an identical human hit are located in the extCDR region**



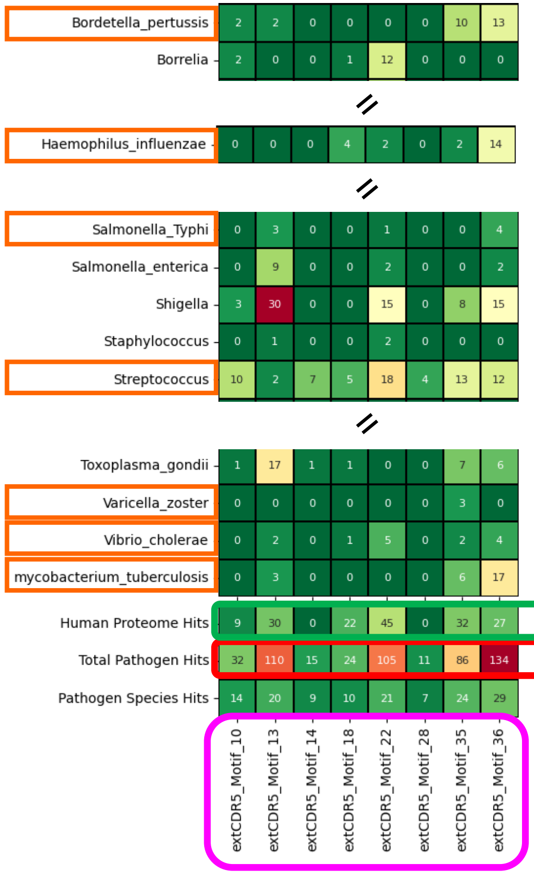
Drug C
Low IG

- ➢ **All motifs show identical human hits**

Parameters to include in a weighting matrix

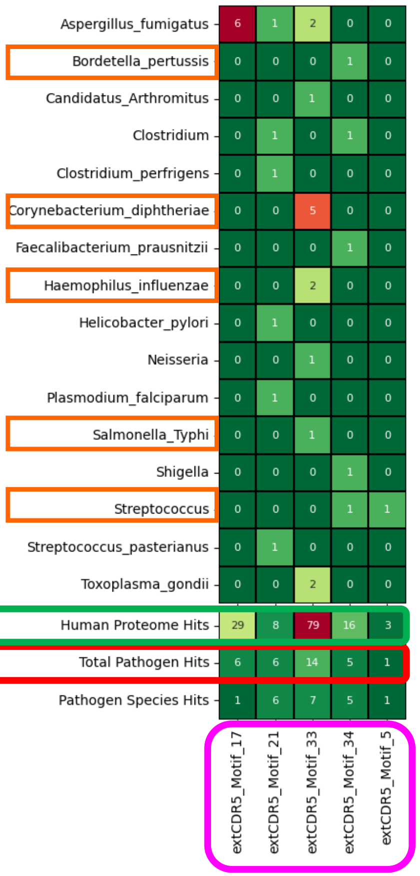

Brunno Rocha

Drug A - High IG



	extCDR5_Motif_10	extCDR5_Motif_13	extCDR5_Motif_14	extCDR5_Motif_18	extCDR5_Motif_22	extCDR5_Motif_28	extCDR5_Motif_35	extCDR5_Motif_36
Bordetella_pertussis	2	2	0	0	0	0	10	13
Borrelia	2	0	0	1	12	0	0	0
Haemophilus_influenzae	0	0	0	4	2	0	2	14
Salmonella_Typhi	0	3	0	0	1	0	0	4
Salmonella_enterica	0	9	0	0	2	0	0	2
Shigella	3	30	0	0	15	0	8	15
Staphylococcus	0	1	0	0	2	0	0	0
Streptococcus	10	2	7	5	18	4	13	12
Toxoplasma_gondii	1	17	1	1	0	0	7	6
Varicella_zoster	0	0	0	0	0	0	3	0
Vibrio_cholerae	0	2	0	1	5	0	2	4
mycobacterium_tuberculosis	0	3	0	0	0	0	6	17
Human Proteome Hits	9	30	0	22	45	0	32	27
Total Pathogen Hits	32	110	15	24	105	11	86	134
Pathogen Species Hits	14	20	9	10	21	7	24	29

Drug C - Low IG



	extCDR5_Motif_17	extCDR5_Motif_21	extCDR5_Motif_33	extCDR5_Motif_34	extCDR5_Motif_5
Aspergillus_fumigatus	6	1	2	0	0
Bordetella_pertussis	0	0	0	1	0
Candidatus_Arthromitus	0	0	1	0	0
Clostridium	0	1	0	1	0
Clostridium_perfrigens	0	1	0	0	0
Corynebacterium_diphtheriae	0	0	5	0	0
Faecalibacterium_prausnitzii	0	0	0	1	0
Haemophilus_influenzae	0	0	2	0	0
Helicobacter_pylori	0	1	0	0	0
Neisseria	0	0	1	0	0
Plasmodium_falciparum	0	1	0	0	0
Salmonella_Typhi	0	0	1	0	0
Shigella	0	0	0	1	0
Streptococcus	0	0	0	1	1
Streptococcus_pasterianus	0	1	0	0	0
Toxoplasma_gondii	0	0	2	0	0
Human Proteome Hits	29	8	79	16	3
Total Pathogen Hits	6	6	14	5	1
Pathogen Species Hits	1	6	7	5	1

- ☐ # of identified analogs (Motifs)
- ☐ Ratio of pathogen vs human analogs
- ☐ Quality of the hits (e.g. vaccination)
  - ➢ Quantity and Quality of the hits is important! E.g. do we have an identical human hit?

- ☐ Brunno has a lot more work to do ;-)
  - ➢ Weight / group pathogens
  - ➢ Include protein expression levels
  - ➢ Cross-check with known epitopes
  - ➢ …

IGMotifFinder pipeline – real life example humanization variants (filtered extCDR hits)

Brunno Rocha

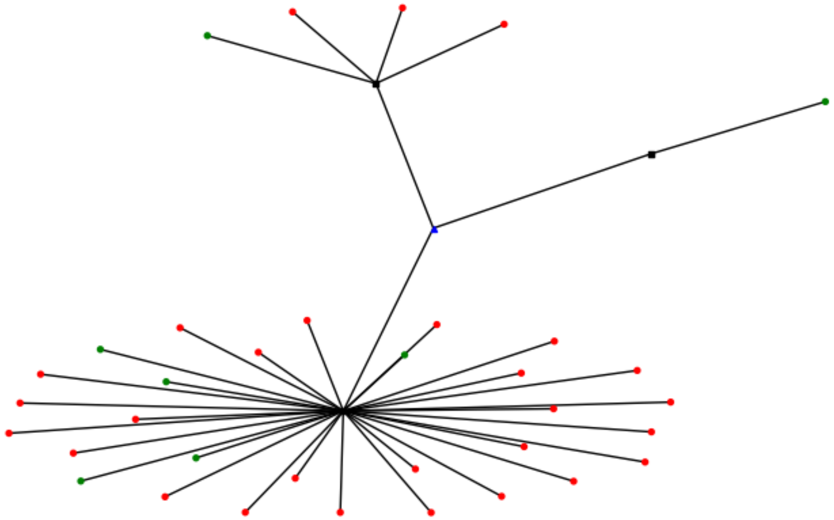
The nodes in the plots represent:

▲ 1st level = Biotherapeutic sequence

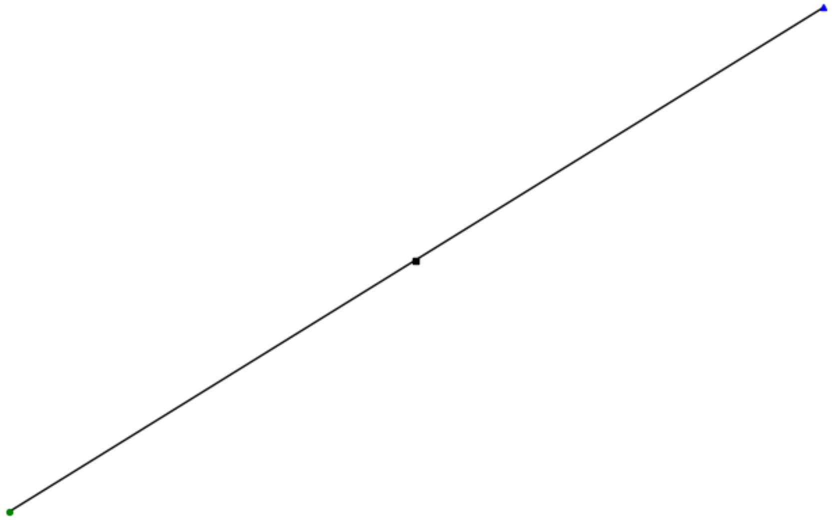
2nd level = HLA binding peptide core;
■ black square is an analog,
○ and the circle is an identical match

●● 3rd level = MASE-identified analogs
(red = pathogen; green = human)

parental



variant



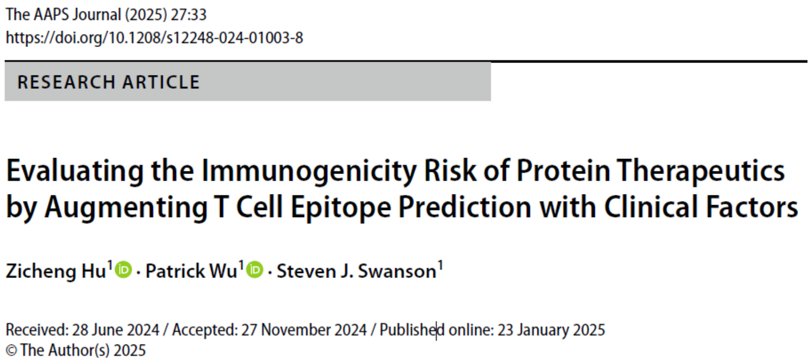
- ➢ filtered hits = all motifs with an identical human match are filtered out

Can we translate pre-clinical *in silico* IG profiling results to clinical IG rates?

NOVARTIS | Reimagining Medicine

Integrating Therapies’ Clinical Features for Immunogenicity Prediction

Brunno Rocha



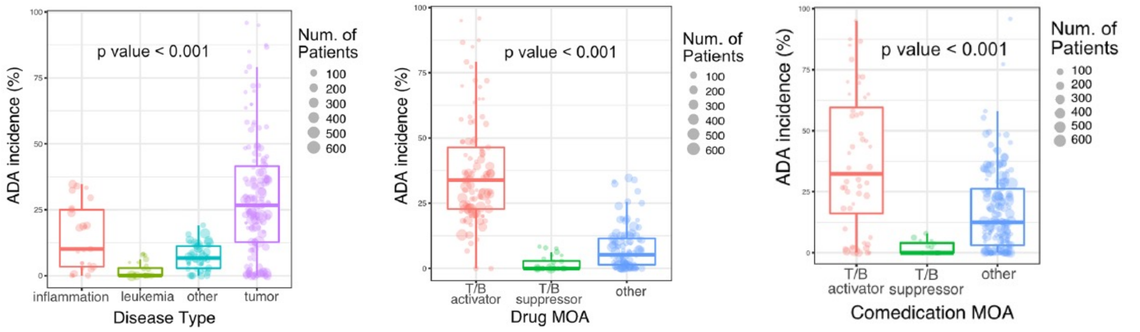
The AAPS Journal (2025) 27:33
https://doi.org/10.1208/s12248-024-01003-8

RESEARCH ARTICLE

Evaluating the Immunogenicity Risk of Protein Therapeutics by Augmenting T Cell Epitope Prediction with Clinical Factors

Zicheng Hu[1] · Patrick Wu[1] · Steven J. Swanson[1]

Received: 28 June 2024 / Accepted: 27 November 2024 / Published online: 23 January 2025
© The Author(s) 2025



- ❑ **Augmenting sequence-based prediction with clinical factors substantially improves model performance** (cross-validated AUC increased from 0.72 to 0.93)

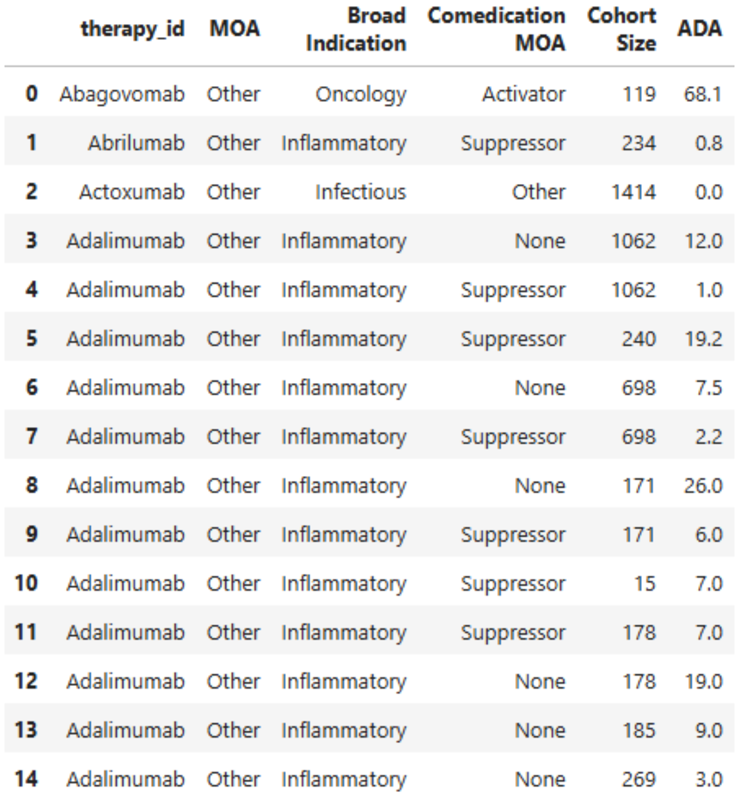
  - ➢ **For a proper correlation with clinical IG rates, the dataset needs to be large enough, bolstered with clinical features and curated!**
    - ➢ **The Marks dataset needs to be augmented!**

Integrating Therapies’ Clinical Features for Immunogenicity Prediction

➢ The Marks dataset needs to be augmented!

Name	**Immunogenicity**
3F8	100.0
Abagovomab	68.1
Abciximab	35.5
Abrilumab	0.4
Actoxumab	0.0
Adalimumab	14.0
Aducanumab	5.0
Afutuzumab	6.6
Alemtuzumab	5.1
Alirocumab	3.9
Amatuximab	29.2
Andecaliximab	21.2
Anifrolumab	3.3
Arcitumomab	0.5
Atezolizumab	28.9

*Source: Marks et al. *Bioinformatics* 2021, *37*, 22, 4041-4047.



	therapy_id	**MOA**	**Broad Indication**	**Comedication MOA**	**Cohort Size**	**ADA**
0	Abagovomab	Other	Oncology	Activator	119	68.1
1	Abrilumab	Other	Inflammatory	Suppressor	234	0.8
2	Actoxumab	Other	Infectious	Other	1414	0.0
3	Adalimumab	Other	Inflammatory	None	1062	12.0
4	Adalimumab	Other	Inflammatory	Suppressor	1062	1.0
5	Adalimumab	Other	Inflammatory	Suppressor	240	19.2
6	Adalimumab	Other	Inflammatory	None	698	7.5
7	Adalimumab	Other	Inflammatory	Suppressor	698	2.2
8	Adalimumab	Other	Inflammatory	None	171	26.0
9	Adalimumab	Other	Inflammatory	Suppressor	171	6.0
10	Adalimumab	Other	Inflammatory	Suppressor	15	7.0
11	Adalimumab	Other	Inflammatory	Suppressor	178	7.0
12	Adalimumab	Other	Inflammatory	None	178	19.0
13	Adalimumab	Other	Inflammatory	None	185	9.0
14	Adalimumab	Other	Inflammatory	None	269	3.0

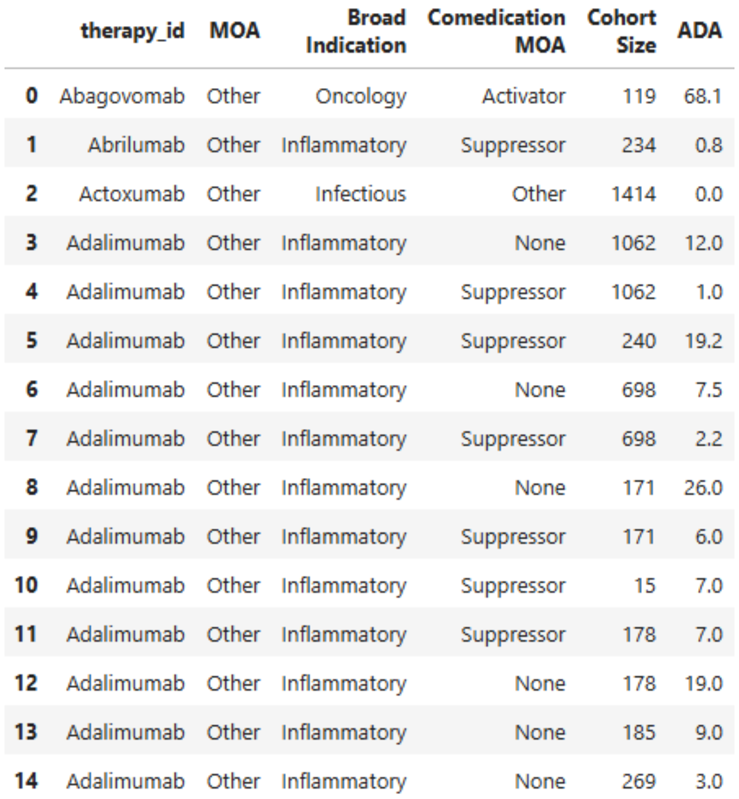
Source: Maximiliano Hess Cancino’s Dataset

Integrating Therapies’ Clinical Features for Immunogenicity Prediction

- Features like indication, drug MOA & comedication MOA are associated with each clinical trial for a given therapy.

Feature	Values
Broad Indication	• Cardiovascular • Infectious • Inflammatory • Neurological • Oncology • Other
MOA	• T/B Cell Activator • T/B Cell Suppressor • Other
Comedication MOA	• Activator • Suppressor • None (Monotherapy) • Other • Unknown

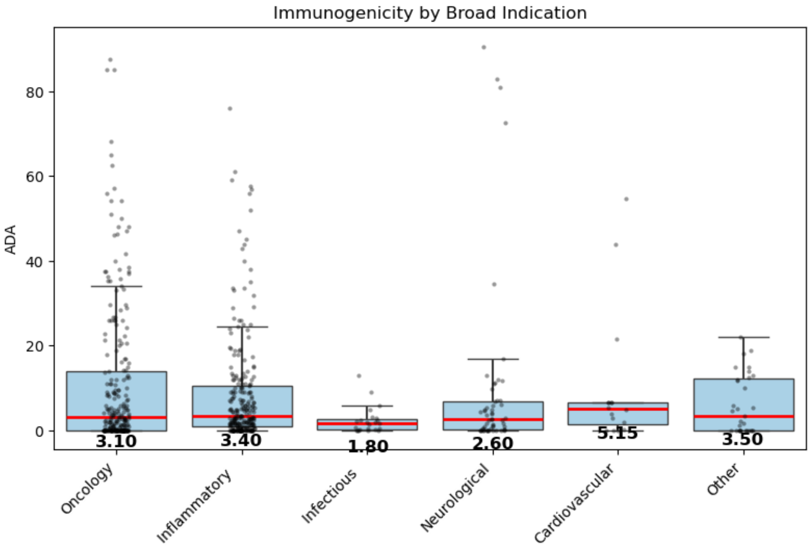
Dataset dimension: 190 Therapies (617 Datapoints)

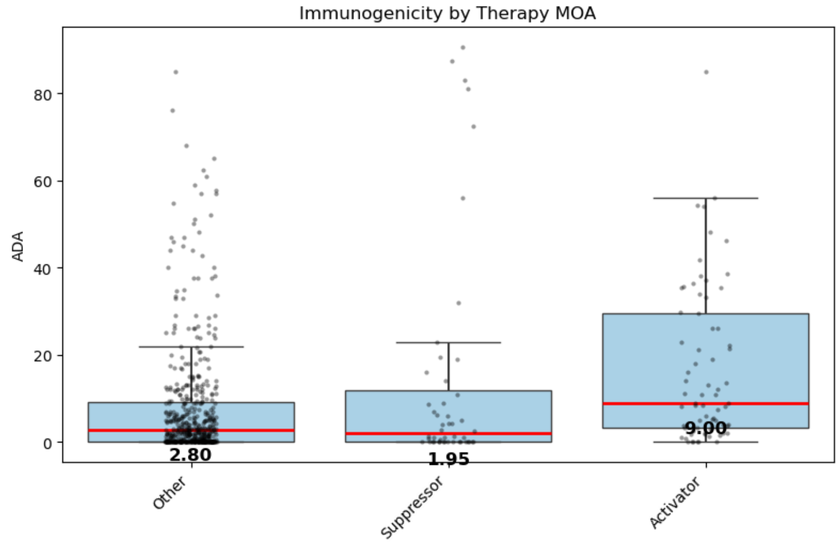


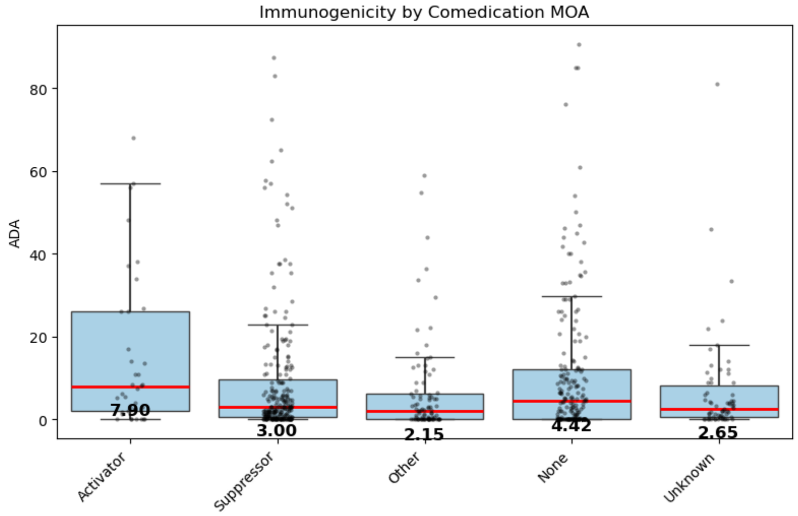
	therapy_id	MOA	Broad Indication	Comedication MOA	Cohort Size	ADA
0	Abagovomab	Other	Oncology	Activator	119	68.1
1	Abrilumab	Other	Inflammatory	Suppressor	234	0.8
2	Actoxumab	Other	Infectious	Other	1414	0.0
3	Adalimumab	Other	Inflammatory	None	1062	12.0
4	Adalimumab	Other	Inflammatory	Suppressor	1062	1.0
5	Adalimumab	Other	Inflammatory	Suppressor	240	19.2
6	Adalimumab	Other	Inflammatory	None	698	7.5
7	Adalimumab	Other	Inflammatory	Suppressor	698	2.2
8	Adalimumab	Other	Inflammatory	None	171	26.0
9	Adalimumab	Other	Inflammatory	Suppressor	171	6.0
10	Adalimumab	Other	Inflammatory	Suppressor	15	7.0
11	Adalimumab	Other	Inflammatory	Suppressor	178	7.0
12	Adalimumab	Other	Inflammatory	None	178	19.0
13	Adalimumab	Other	Inflammatory	None	185	9.0
14	Adalimumab	Other	Inflammatory	None	269	3.0

Source: Maximiliano Hess Cancino’s Dataset

Distribution of ADA rates according to indication and MOA


Immunogenicity by Broad Indication


Immunogenicity by Therapy MOA


Immunogenicity by Comedication MOA

ADA distribution follows expected patterns for MOA

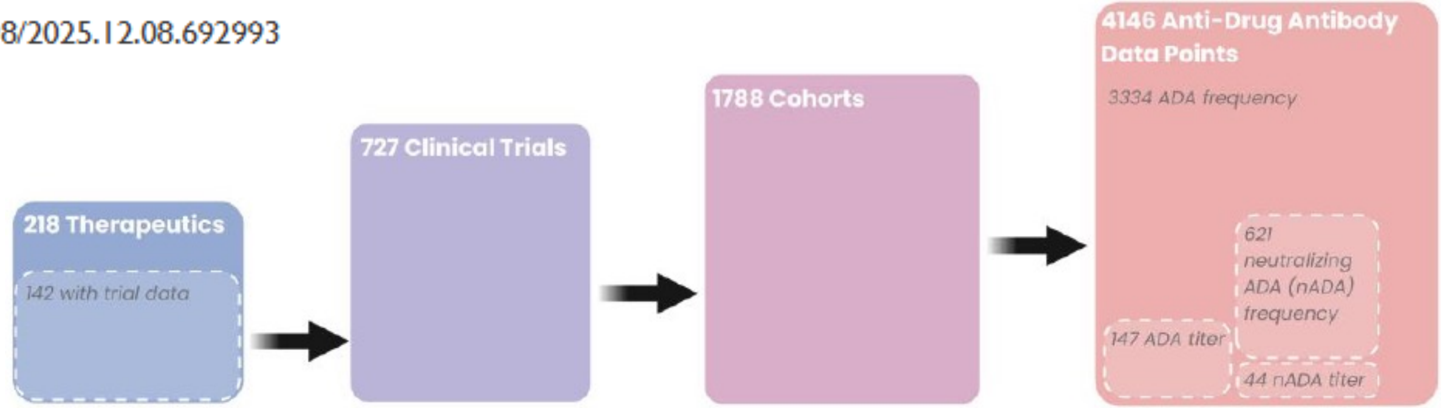
Source: Maximiliano Hess Cancino's Dataset

Integrating Therapies’ Clinical Features for Immunogenicity Prediction

The Immunogenicity Database Collaborative (IDC): A Standardized, Publicly Available Database for Clinical Immunogenicity Observations and Insights

Sudhanshu Agnihotri, Bruno Gonzalez-Nolasco, Brinda Monian, Sofie Pattijn, Chloe Ackaert, Patrick Wu, Hubert Kettenberger, Sophie Tourdot, Timothy Hickling, Zicheng Hu, Richard E. Higgs, Daniel S. Leventhal

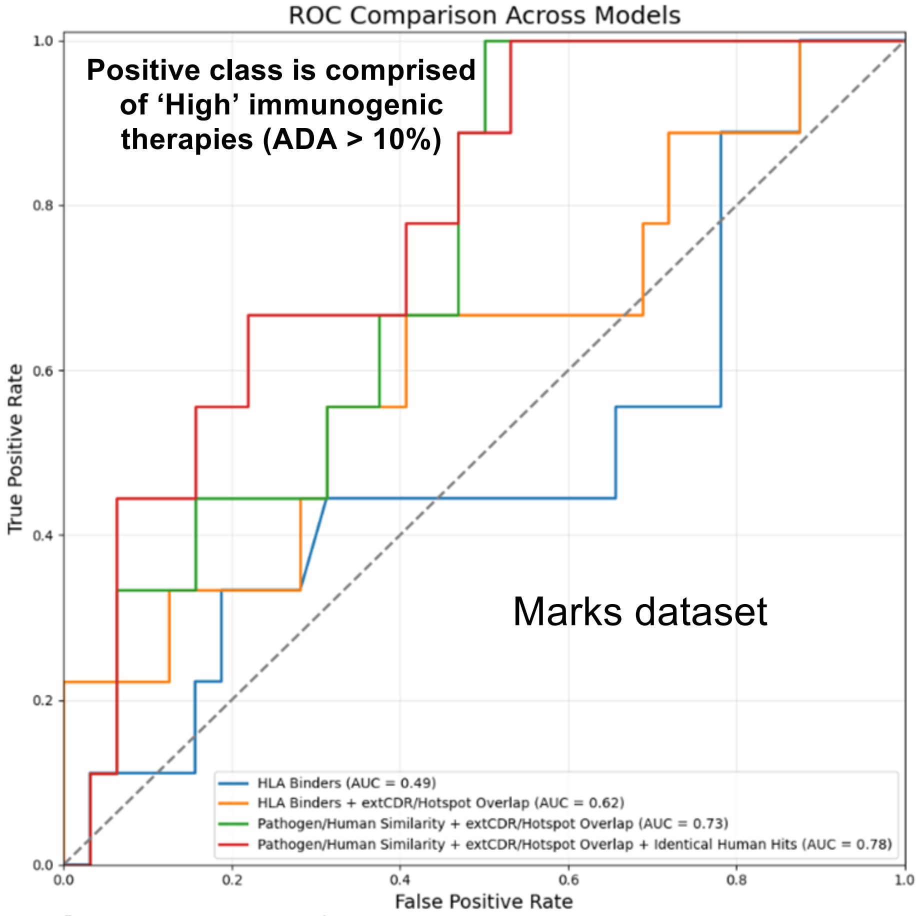
doi: https://doi.org/10.64898/2025.12.08.692993



IDC
ADA frequency **vs** **Novartis**
ADA incidence

(# of pos. samples) **(# of pos. patients)**

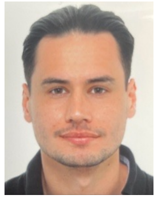
Random Forest models indicate a correlation between ADA incidence, # and quality of HLA binders, pathogen similarity and clinical features



Positive class is comprised of 'High' immunogenic therapies (ADA > 10%)

Marks dataset

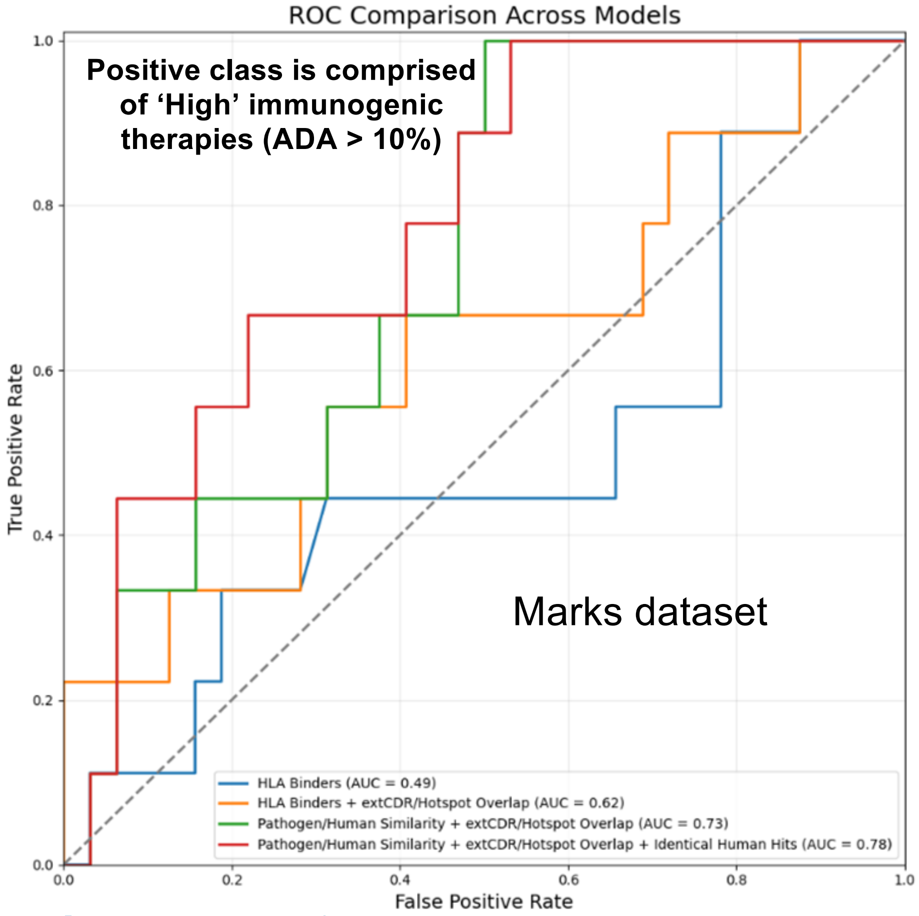
Brunno
Rocha



Maximiliano
Hess Cancino

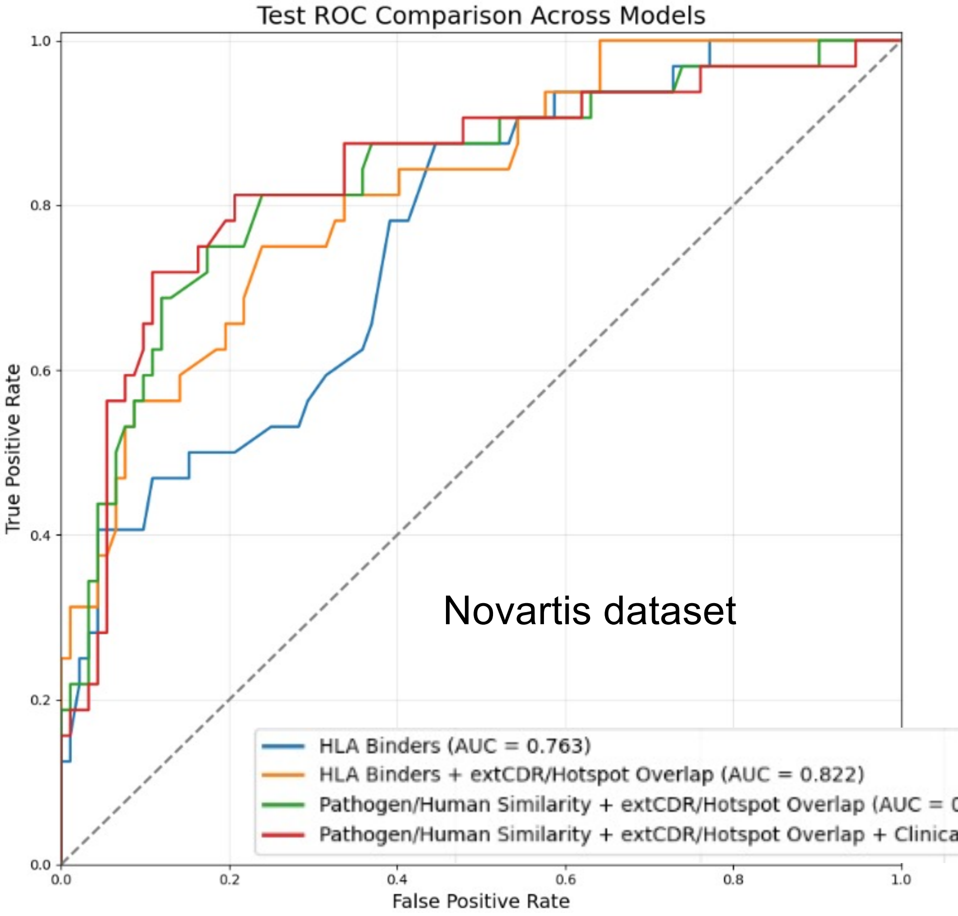
NOVARTIS | Reimagining Medicine

Random Forest models indicate a correlation between ADA incidence, # and quality of HLA binders, pathogen similarity and clinical features



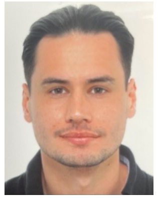
Positive class is comprised of ‘High’ immunogenic therapies (ADA > 10%)

Marks dataset



Novartis dataset

Brunno
Rocha



Maximiliano
Hess Cancino

Conclusion

Can we translate pre-clinical *in silico* IG profiling results to clinical IG rates?

- ☐ **Yes**, even though we are working with a “noisy” dataset, our random forest models show that there is an obvious correlation between clinical ADA incidence, the number and quality of HLA binders, pathogen similarity, and clinical features.
- ☐ **However**, we are still unable to predict the impact of ADA!
    - ➢ **Translation is limited** to the likelihood of ADA emergence.

Next steps:

- ☐ **Improve** the pathogen similarity scoring matrix (weight & group pathogens, add expression levels of proteins to better judge the possibility of tolerance, etc.)
- ☐ **Expand** the Novartis dataset by incorporating additional information (e.g. via incorporation of IDC data)
- ☐ **Enhance** performance of the random forest models (tweak feature ranking)
- ☐ **Confirm** “risky motifs” via an *in vitro* T cell assays (potential external collaboration opportunity) or via GenAI tool IGCheck (cross-check literature to identify *in vitro* confirmed epitopes)

Acknowledgement

Key team members

Brunno Rocha
Maximiliano Hess Cancino
Anette Karle
Miriam Fuhlendorf
Barbara Brannetti
Martin Jockel
Guillaume Roellinger
Jason Marchese
Shreosi Ghosh
Nicole DeCampo
Elizaveta Solovyeva
Smita Mohanty

Special thanks for funding Maxi

Lydia Michaut
Sebastian Hörsch

Michael Gutknecht
michael.gutknecht@novartis.com

Thank you