

Neo Epitope Prediction for personalized Cancer Immunotherapy

MASTER DEFENSE

Matthias Leitner
19.05.2020

OUTLINE

- 1** Motivation & Background
- 2** Analytical Workflow
- 3** Immunogenicity Prediction
- 4** Application & Summary

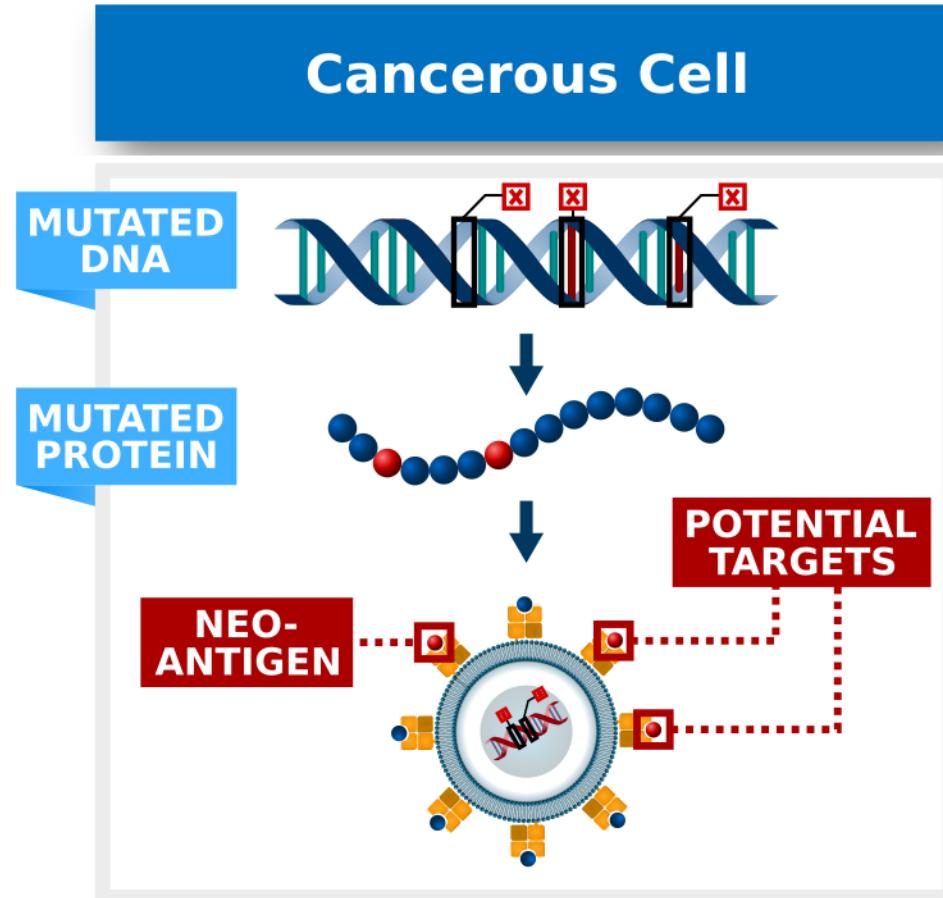
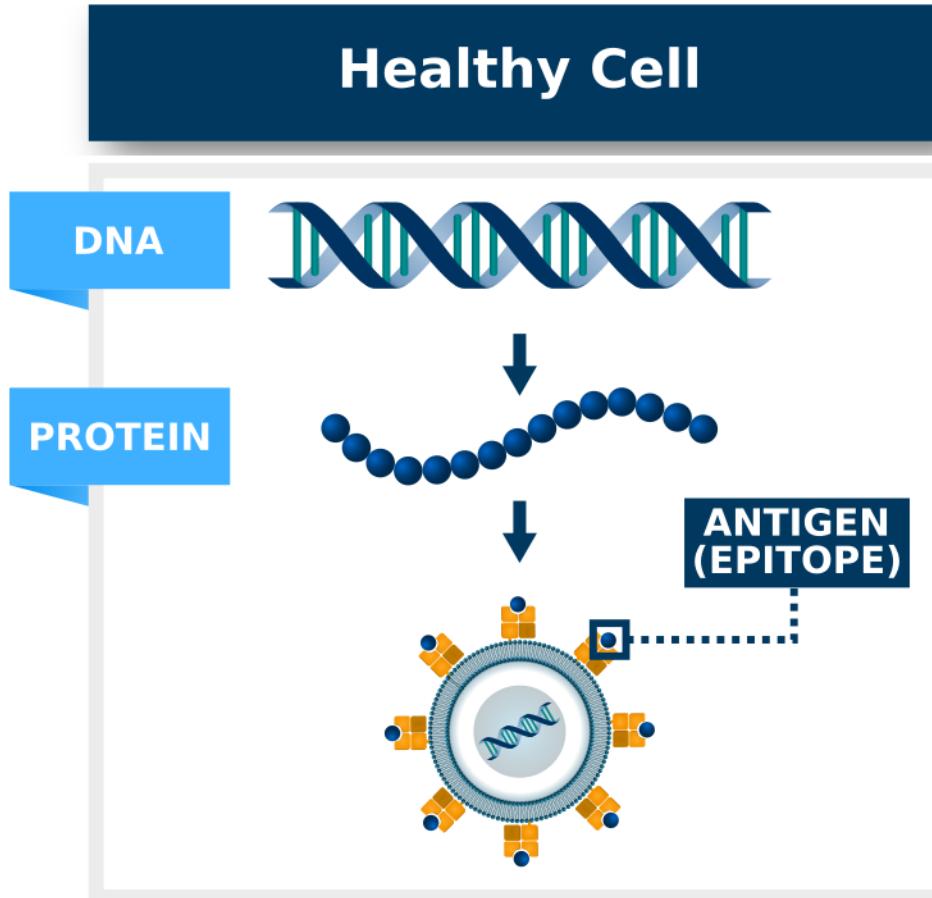
MOTIVATION & BACKGROUND

Motivation

- Traditional cancer treatments such as surgery, radiation or chemo therapy are usually **very expensive** and incur **heavy side effects**.
- Immunotherapy enhances very specific parts of a patient's immune system to fight cancer, and as such is usually **less side effect prone**.
- The immune system can identify cancerous cells by probing their surface for molecules not present on healthy cells (**neo antigens/epitopes**).
- Identifying these cancer specific neo antigens allows for the creation of personalized treatments such as **cancer vaccines**.

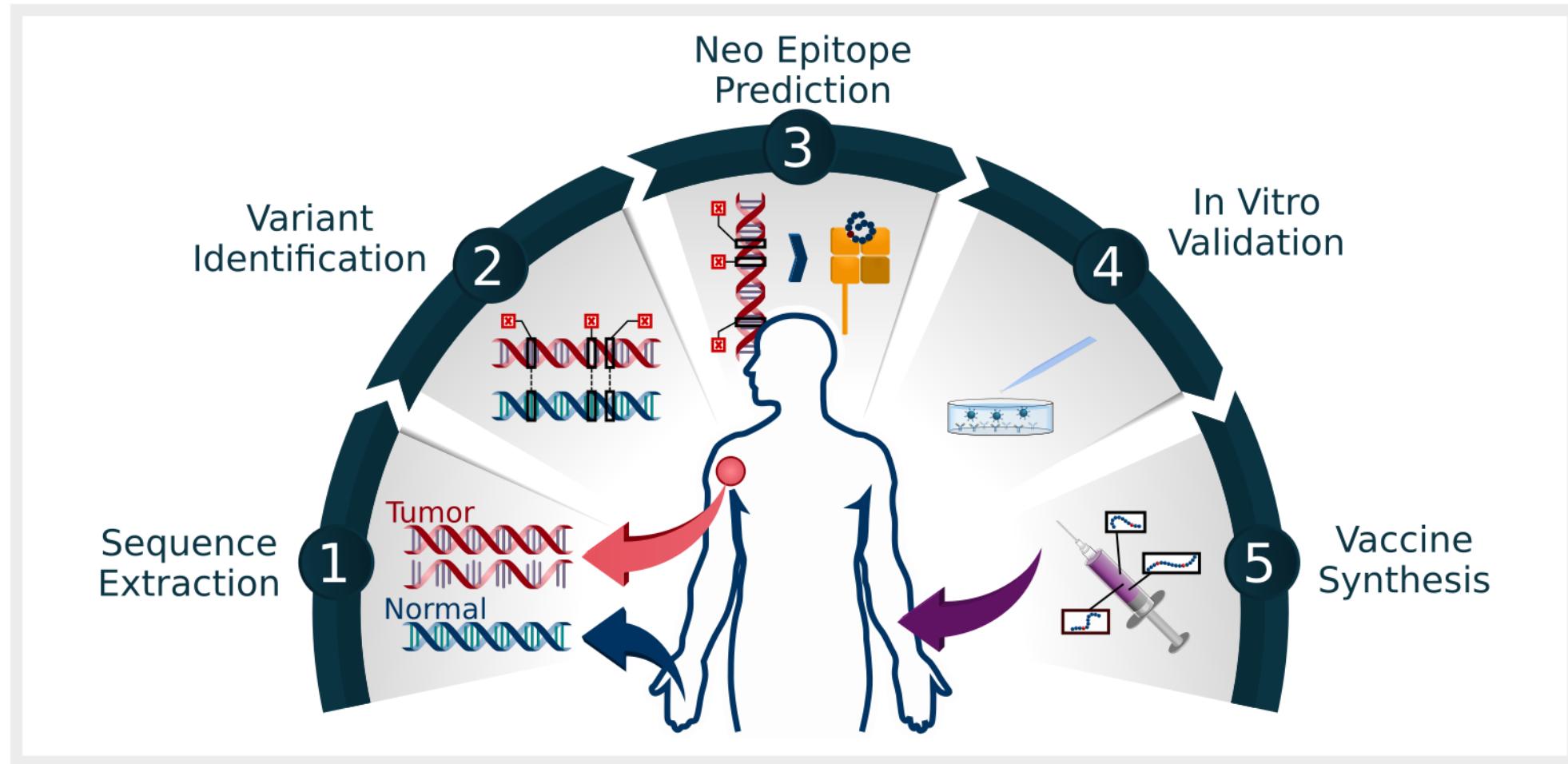
MOTIVATION & BACKGROUND

Antigen Presentation Mechanism



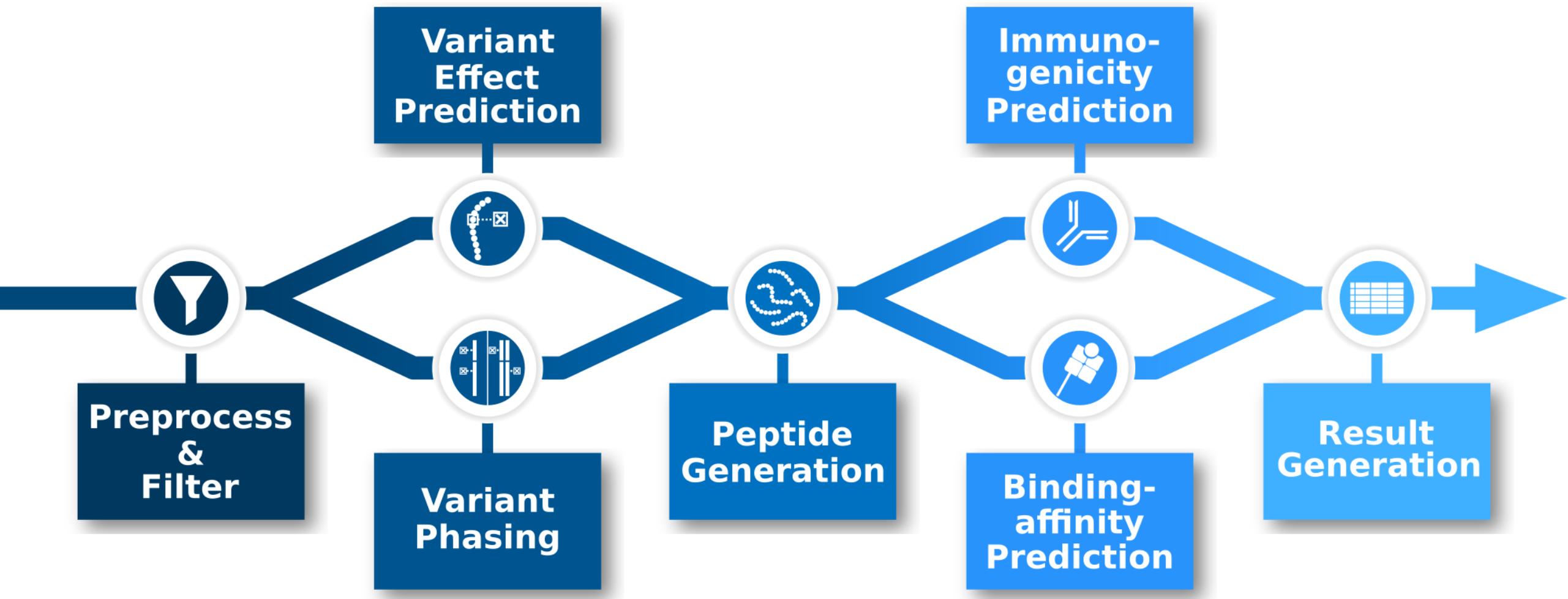
ANALYTICAL WORKFLOW

Global Overview



ANALYTICAL WORKFLOW

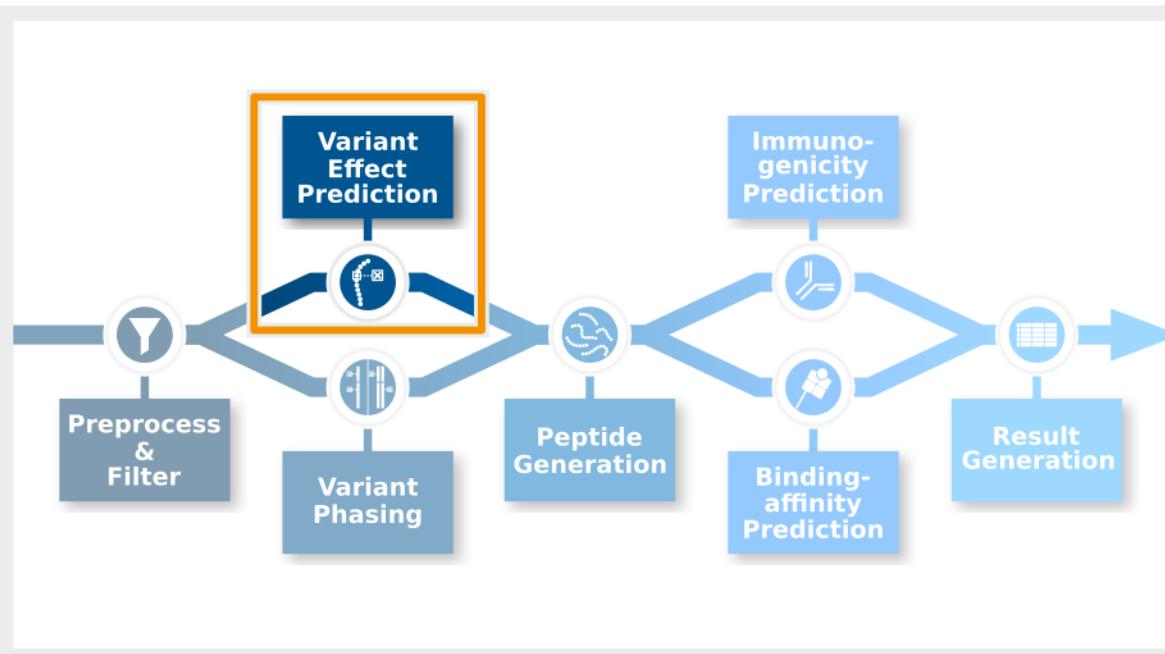
From Variants to Neo Epitopes



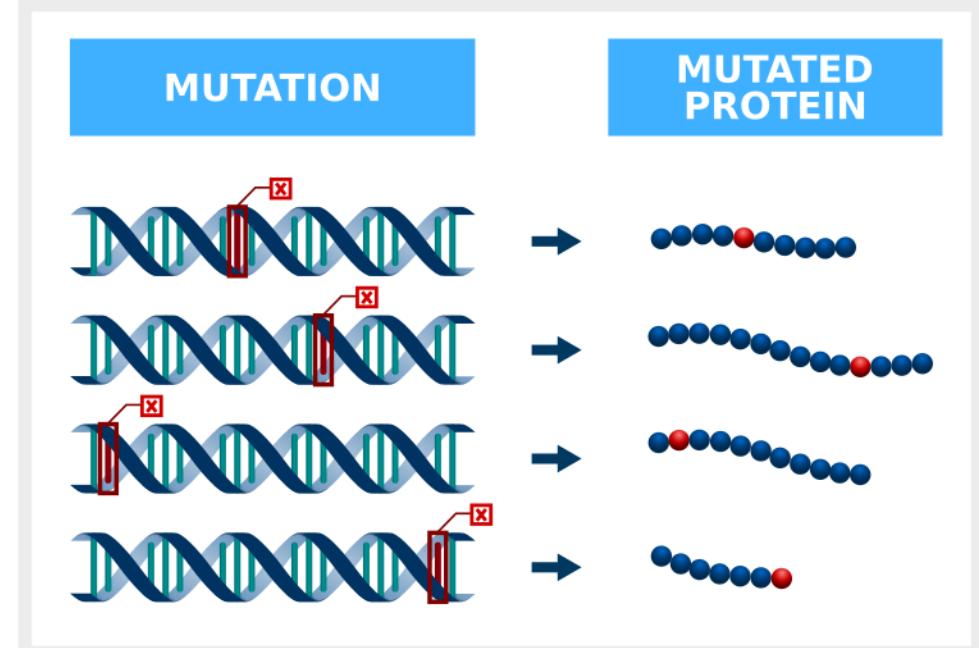
ANALYTICAL WORKFLOW

From Variants to Neo Epitopes

Workflow



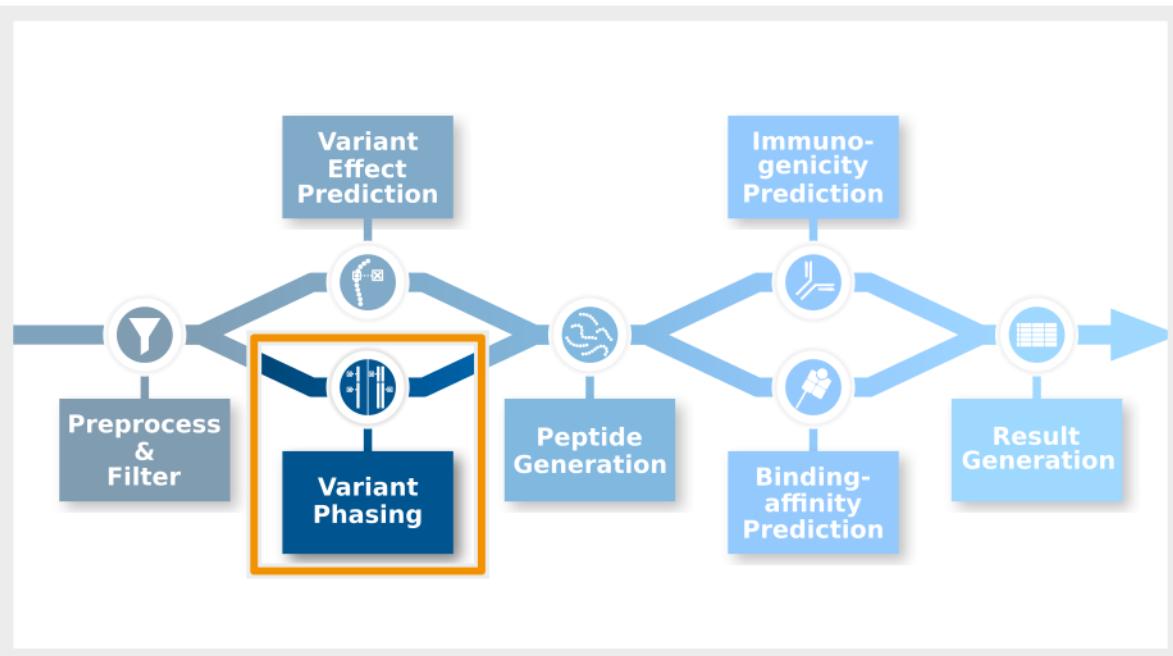
Variant Effect Prediction



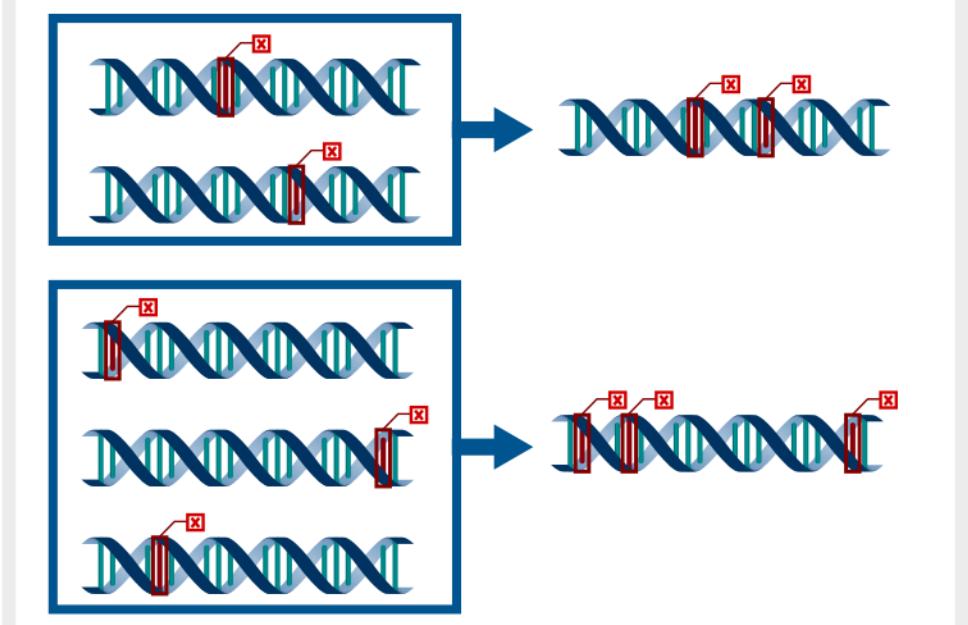
ANALYTICAL WORKFLOW

From Variants to Neo Epitopes

Workflow



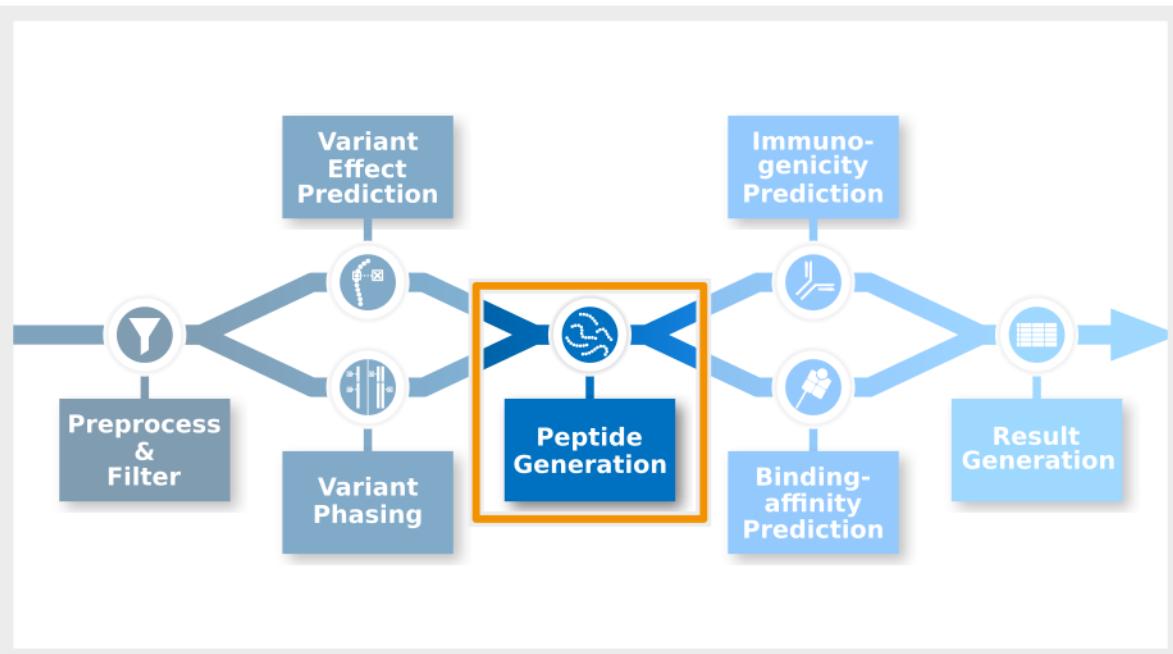
Variant Phasing



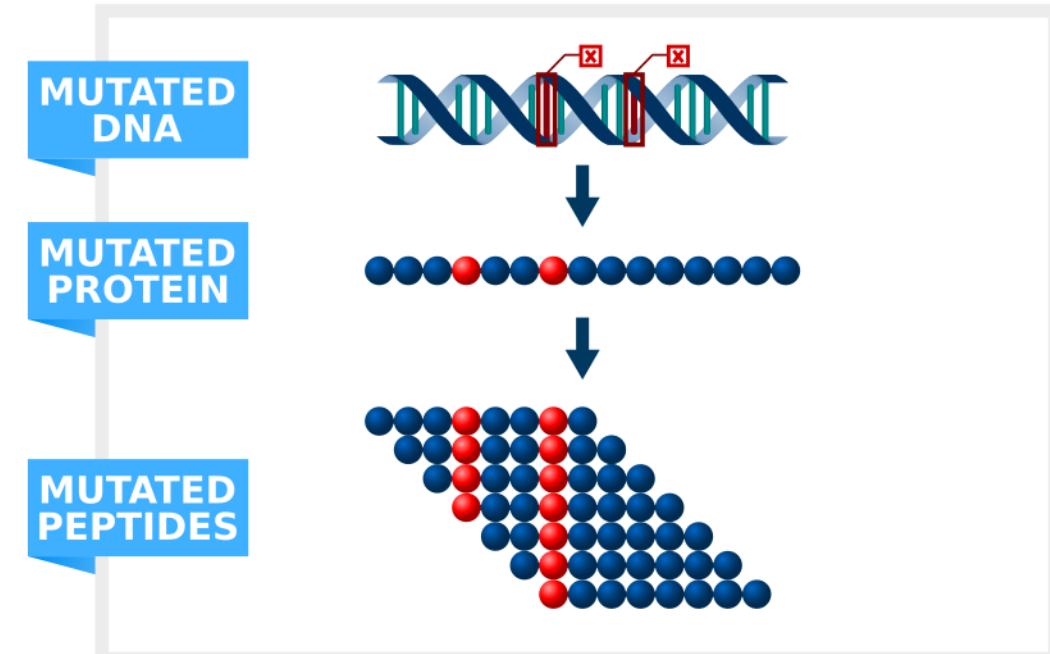
ANALYTICAL WORKFLOW

From Variants to Neo Epitopes

Workflow



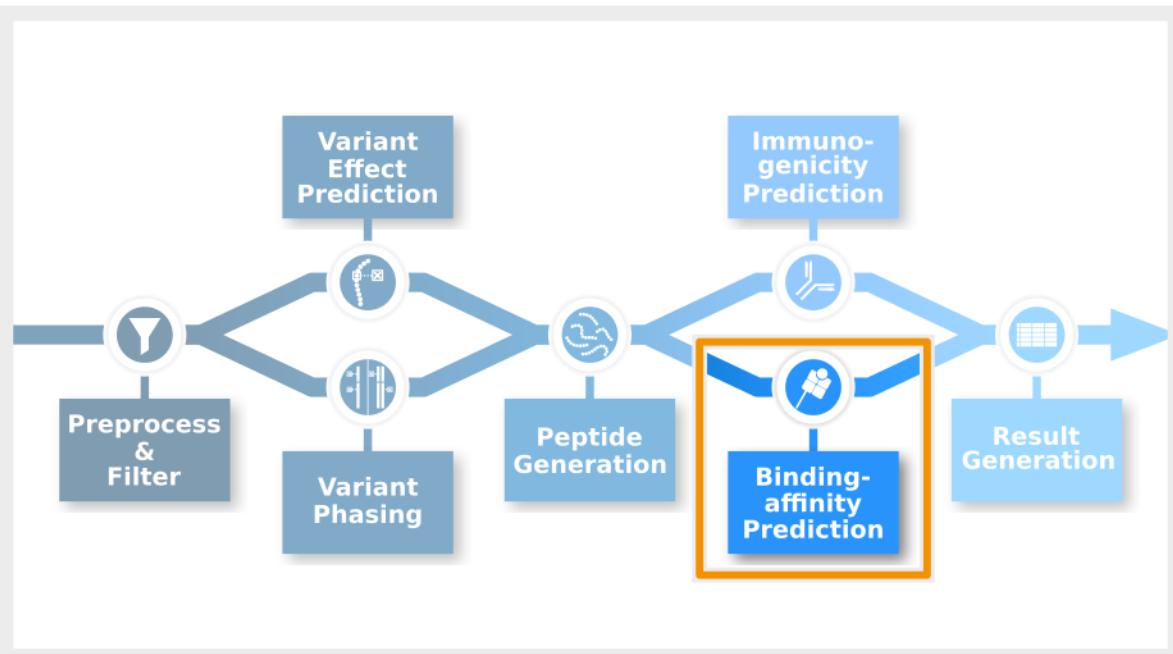
Peptide Generation



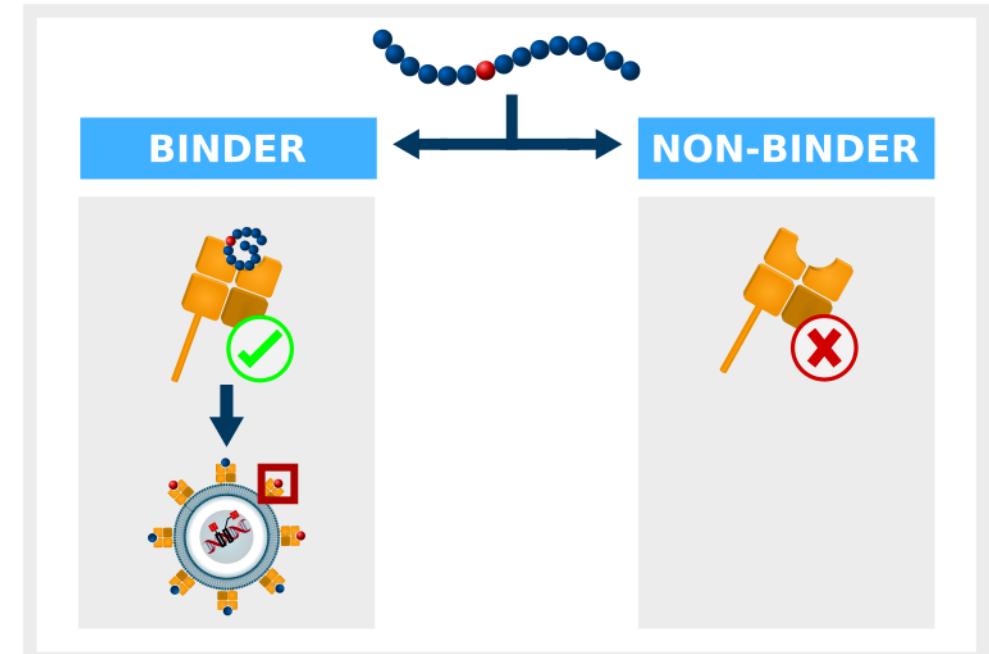
ANALYTICAL WORKFLOW

From Variants to Neo Epitopes

Workflow



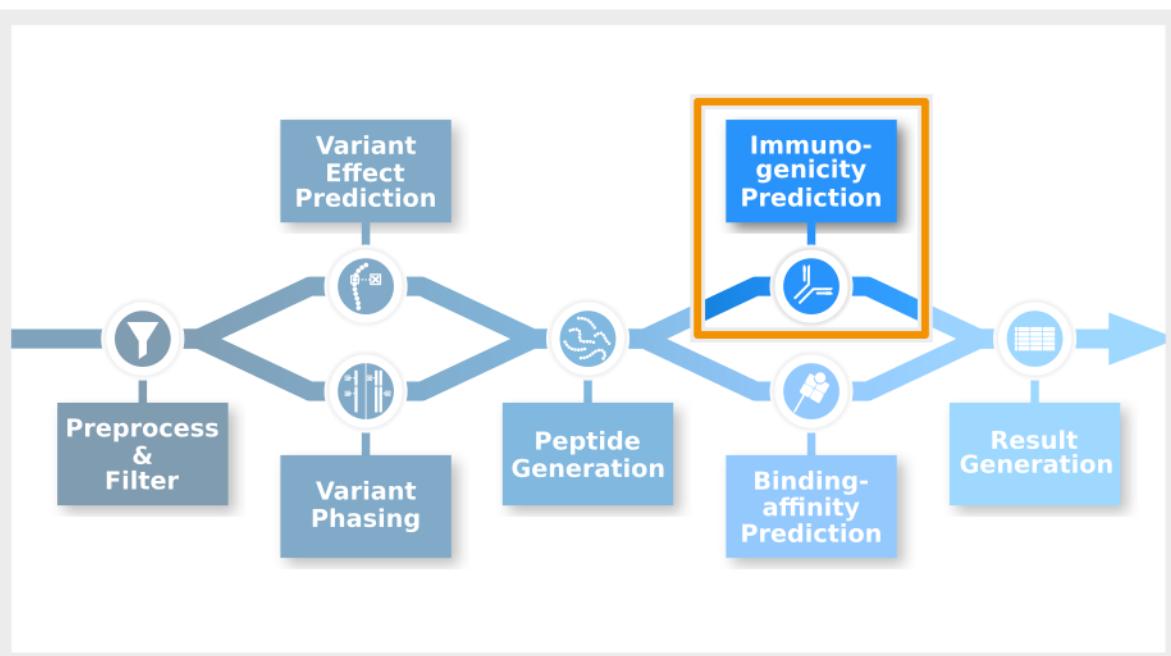
Binding Affinity Prediction



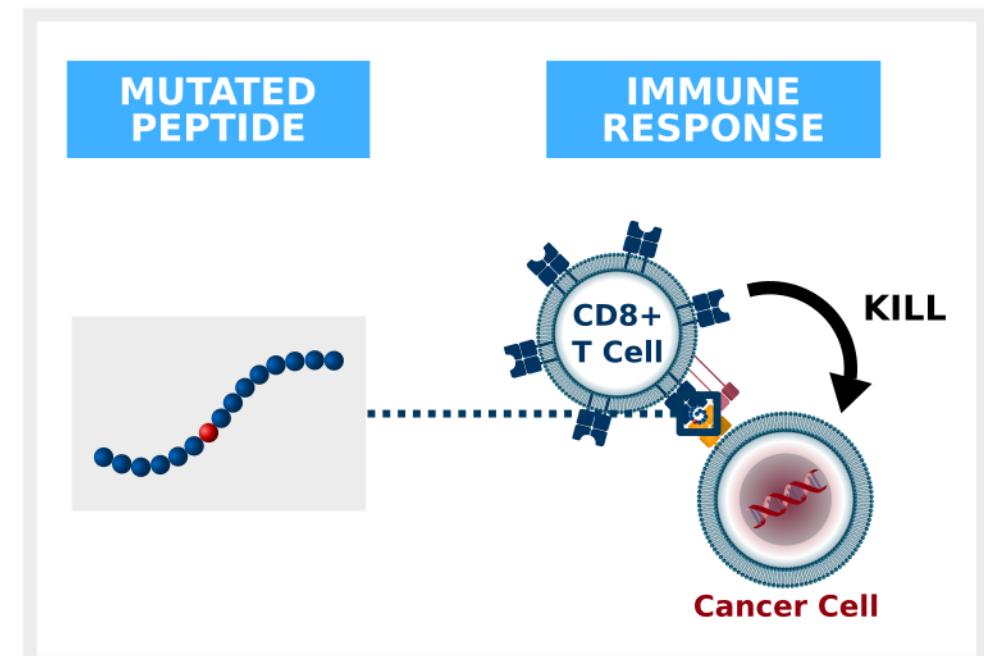
ANALYTICAL WORKFLOW

From Variants to Neo Epitopes

Workflow



Immunogenicity Prediction



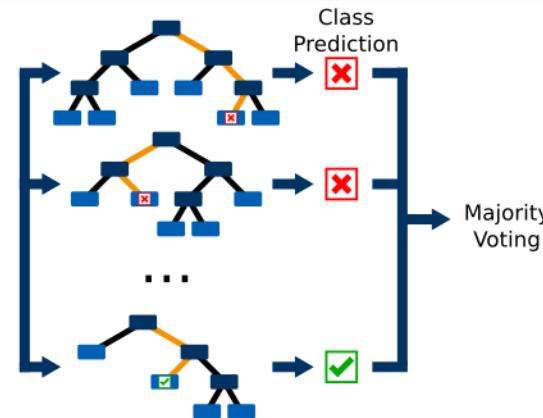
IMMUNOGENICITY PREDICTION

Problem Statement

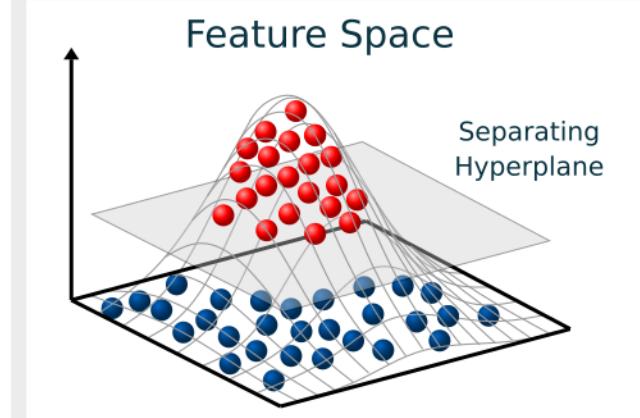


Not all identified epitopes are recognized by the immune system as foreign!

Random Forest



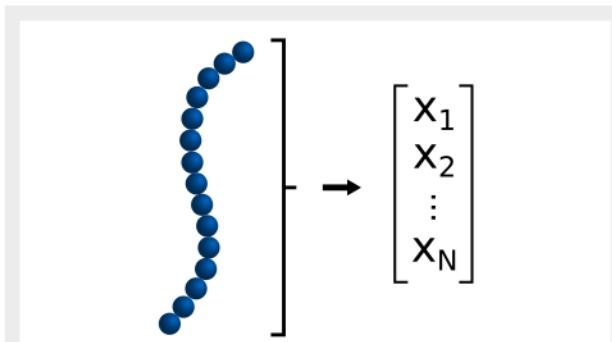
KSVM



IMMUNOGENICITY PREDICTION

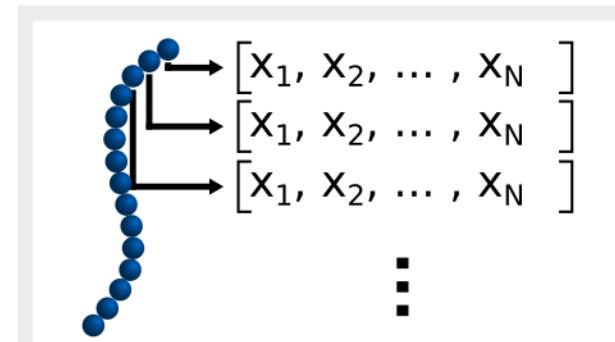
Peptide Encoding

Peptide - Based



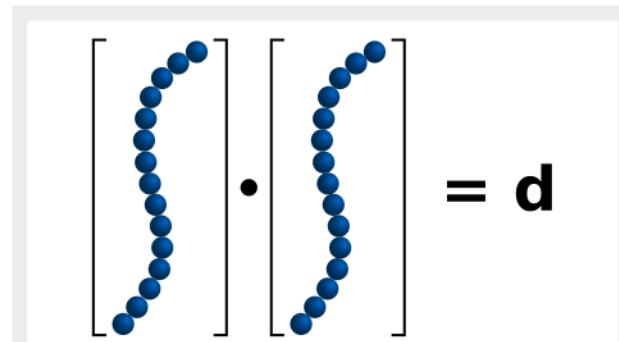
- AAC
- APseAAC
- PseAAC
- QSO

Amino Acid - Based



- Factor
- Hydro
- BLOMAP

Similarity - Based



- Local Alignment
- Wen

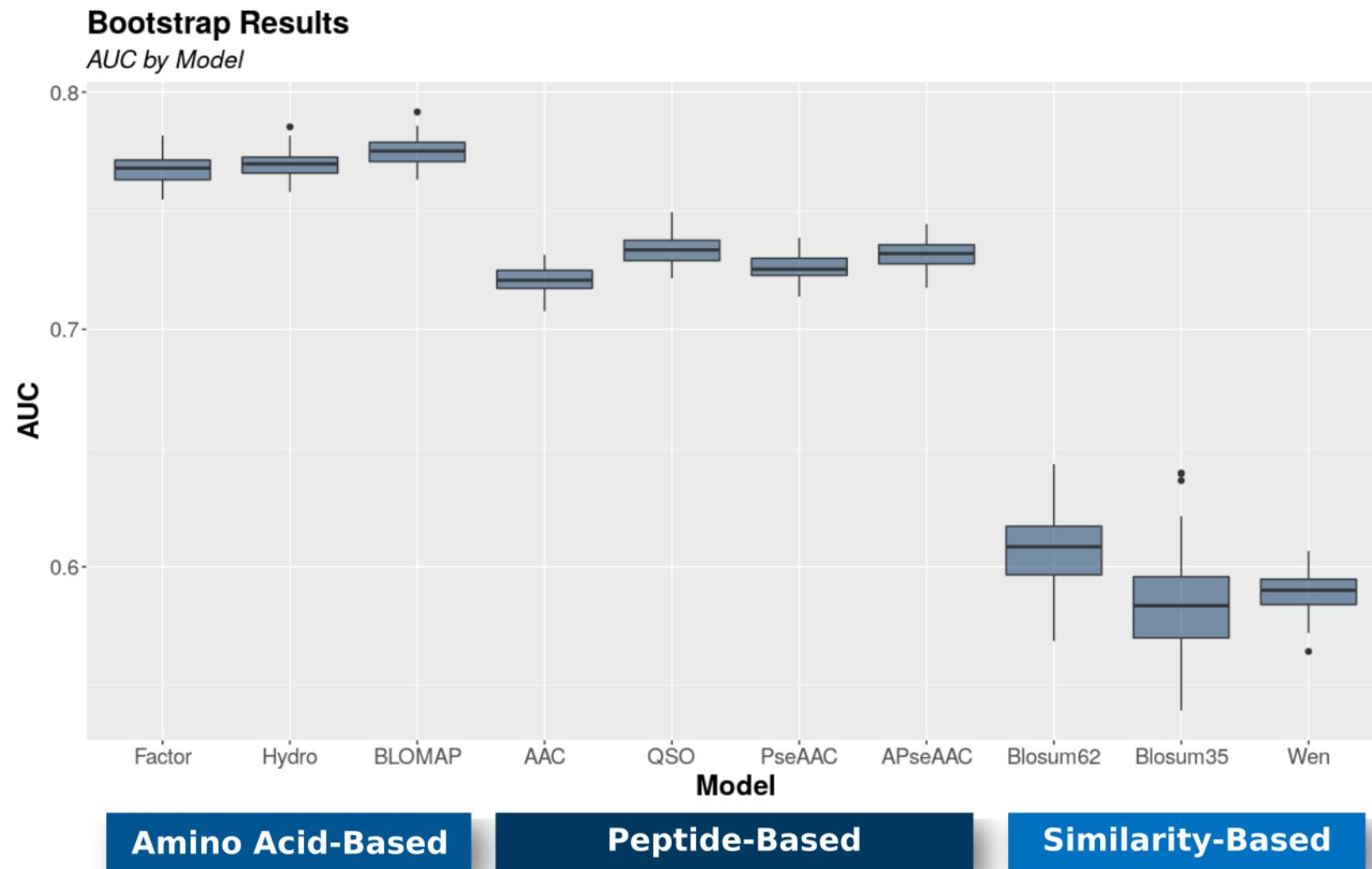
IMMUNOGENICITY PREDICTION

Experimental Set Up

- Epitope data collected from 19 sources
- **21,148** unique peptide-immunogenicity pairs
- **5,288** immunogenic peptides (~ 25%)
- **15,860** non-immunogenic peptides (~ 75%)
- Non-parametric bootstrap with 100 iterations

IMMUNOGENICITY PREDICTION

Results (1/2)

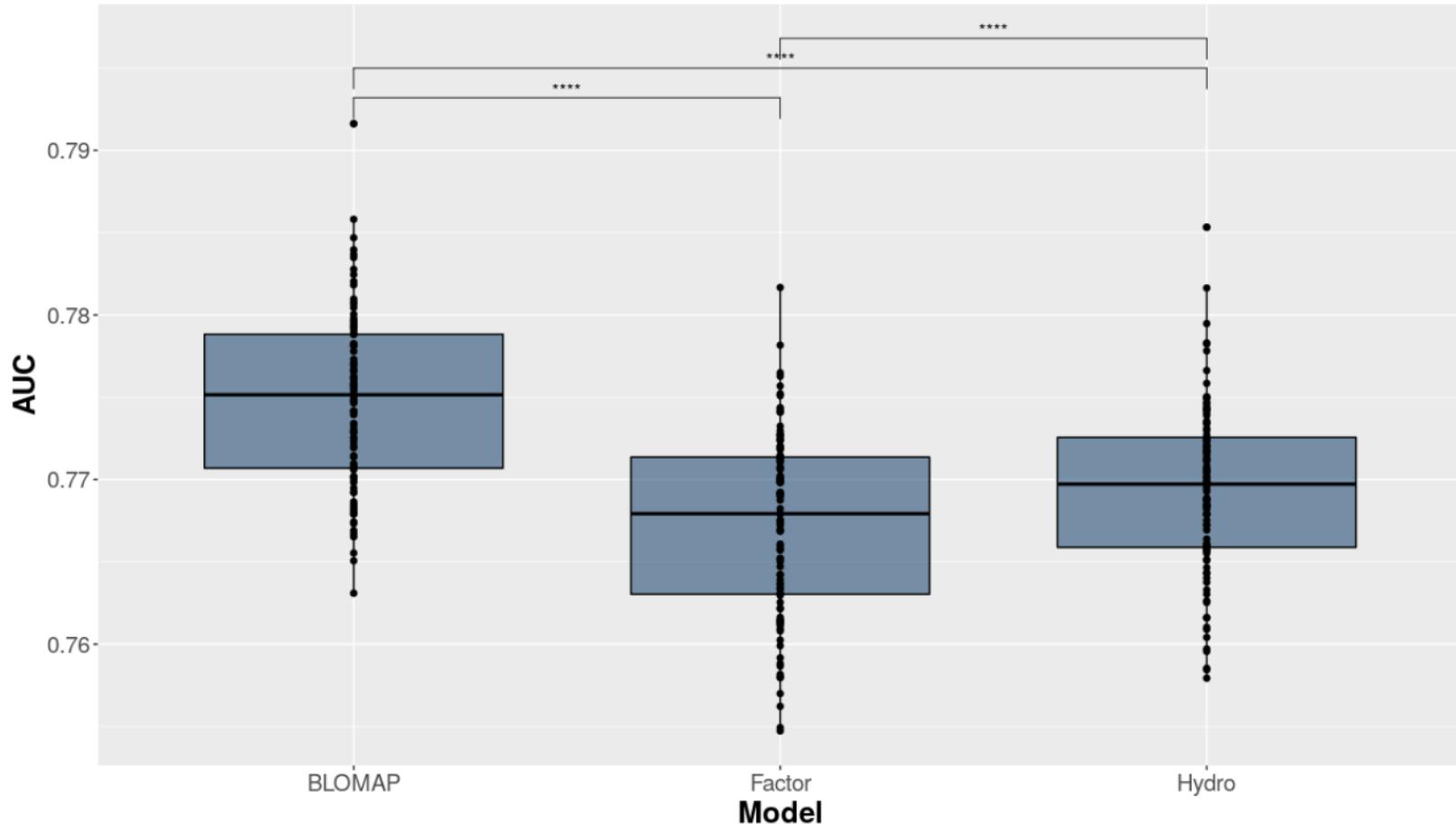


IMMUNOGENICITY PREDICTION

Results (2/2)

Repeated Measures ANOVA

Anova, $F(1.57,155.67) = 308.88, p = <0.0001, \eta_g^2 = 0.26$

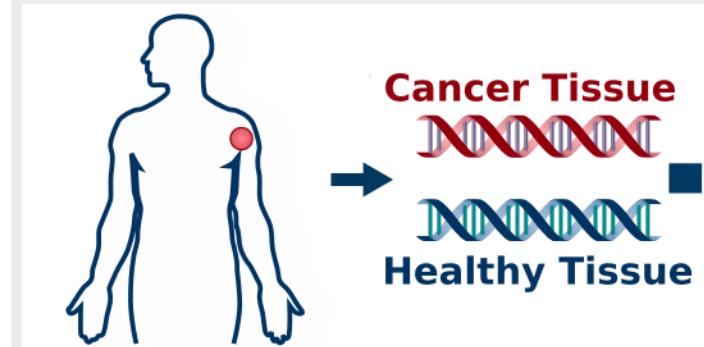


pwc: **T test**; p.adjust: **Bonferroni**

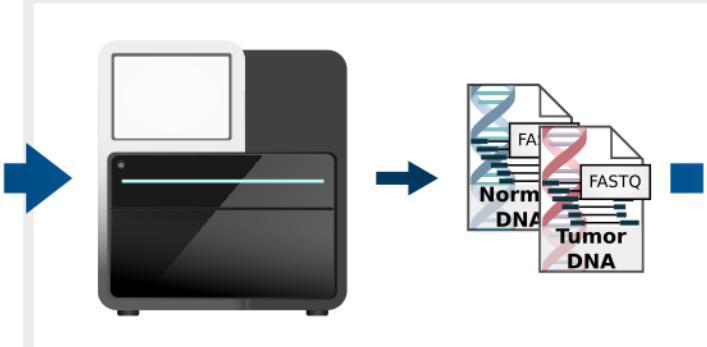
APPLICATION & SUMMARY

Pipeline Application

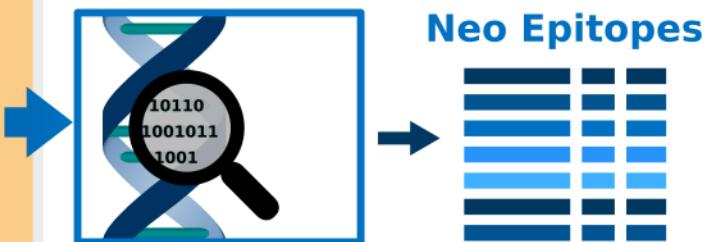
1 Sample Collection



2 Sequencing



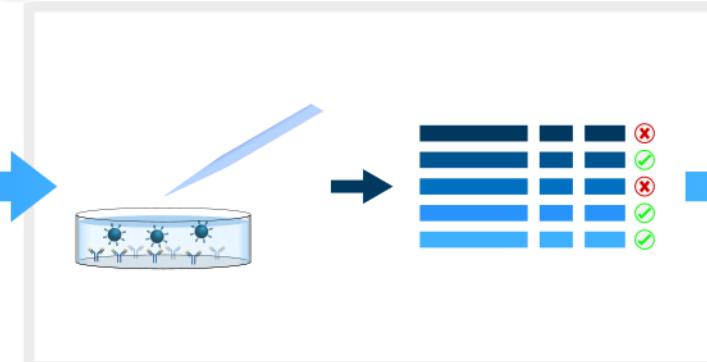
3 Epitope Prediction



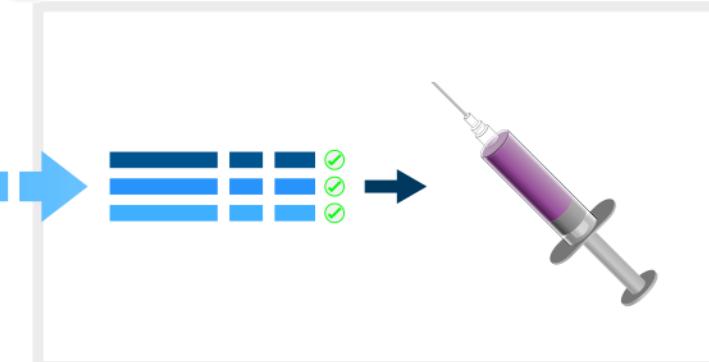
4 Epitope Selection



5 In Vitro Testing



6 Vaccine Synthesis



APPLICATION & SUMMARY

Output

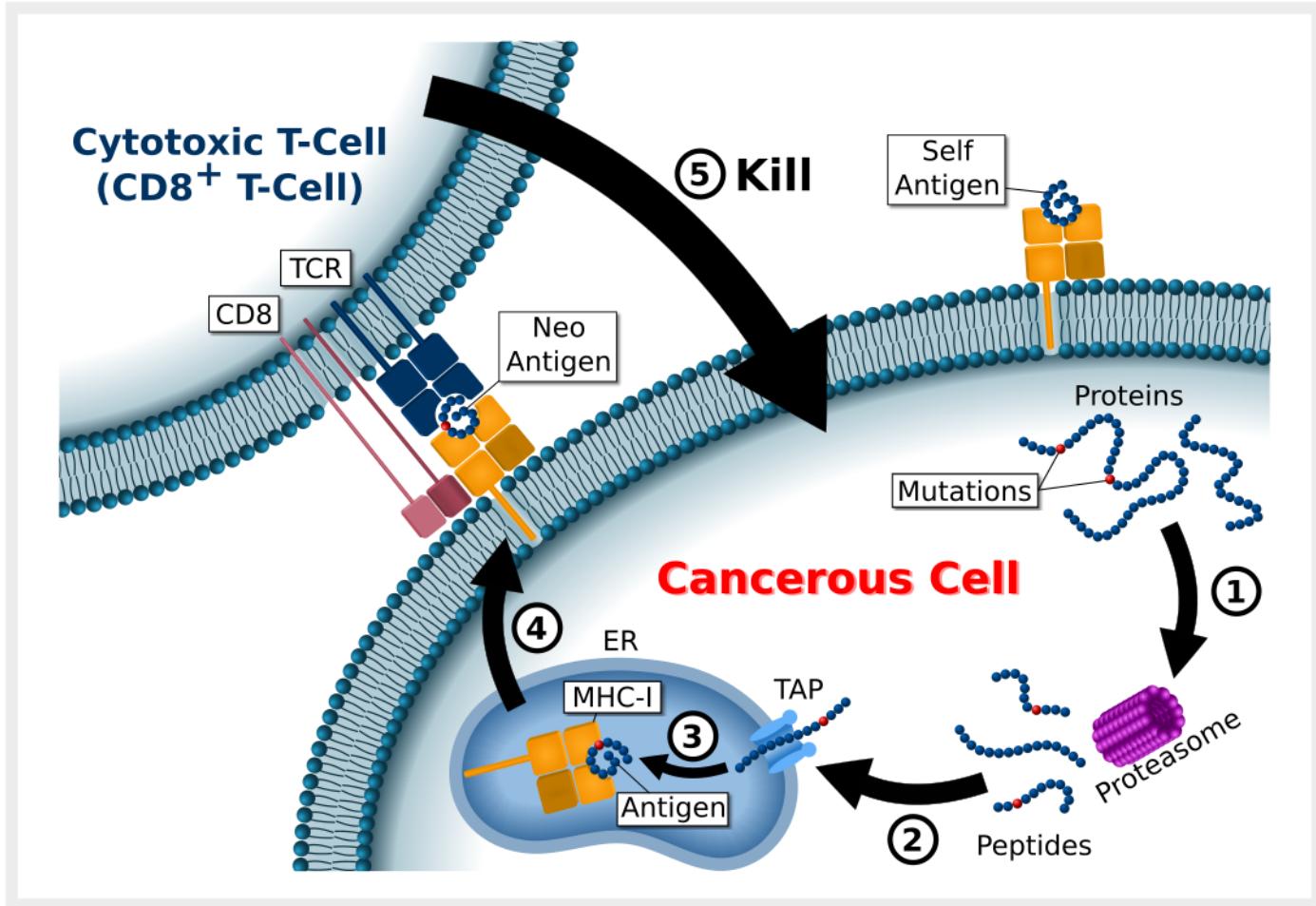
Peptide	Variant	Binding Affinity	Transcription
<ul style="list-style-type: none">• Peptide	<ul style="list-style-type: none">• Variants• VariantTypes• VariantCallers• DNA VAF Normal• DNA VAF Tumor• RNA VAF• Read Count DNA Normal• Read Count DNA Tumor• Read Count RNA Tumor	<ul style="list-style-type: none">• HLA• Binding Core• ICore• Raw Prediction Score• Affinity (nM)• %Rank• Exp• Strength	<ul style="list-style-type: none">• TPM• expressed
Protein			Immunogenicity
<ul style="list-style-type: none">• gene• Protein			<ul style="list-style-type: none">• Immunogenicity Score• Immunogenic

APPLICATION & SUMMARY

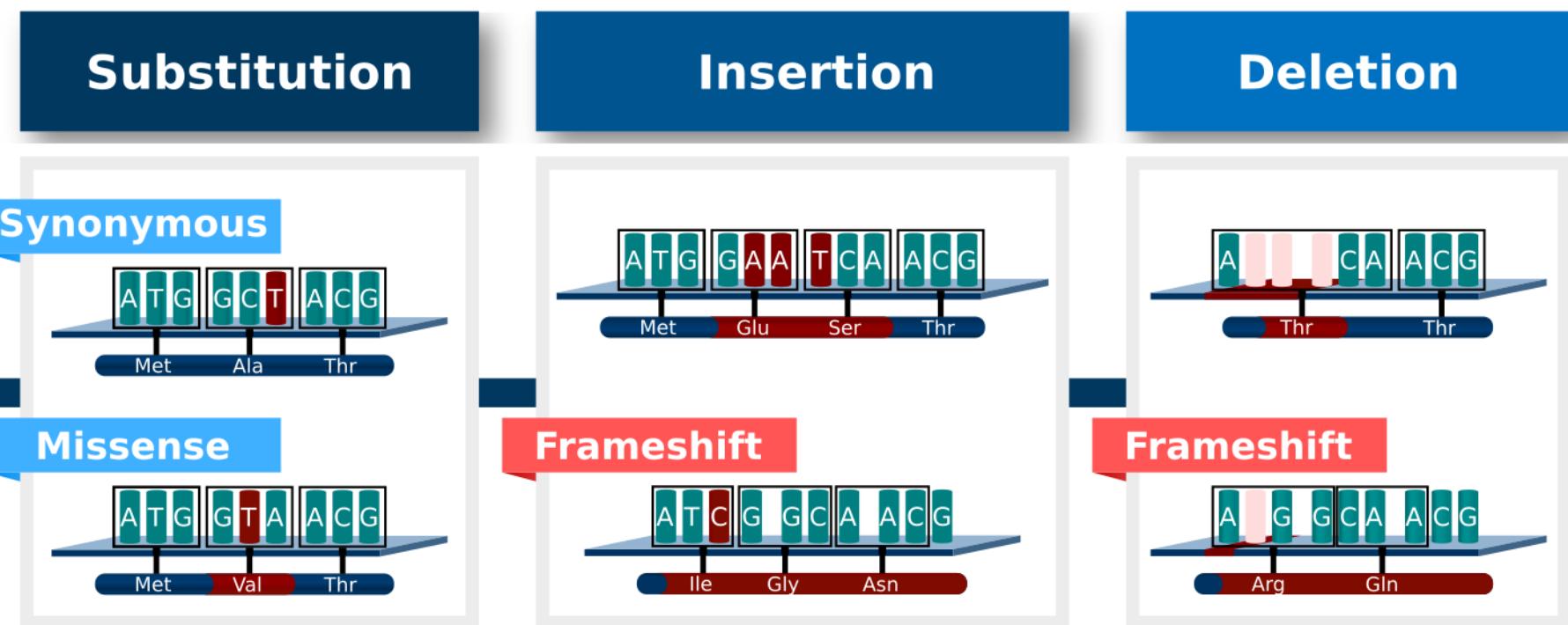
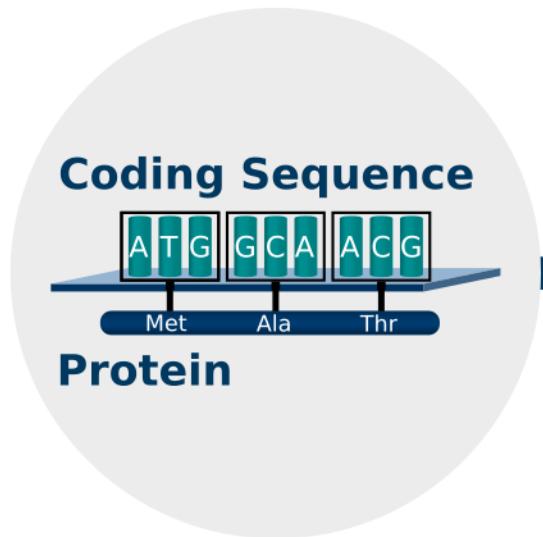
Summary

- Creation of a new analytical end-to-end pipeline for deriving cancer neo epitopes as a basis for vaccine generation.
- Inclusion of a variety of mutational types as well as phase information.
- Implementation of an Immunogenicity prediction model using a Random Forest based on BLOMAP encoding of amino acids.

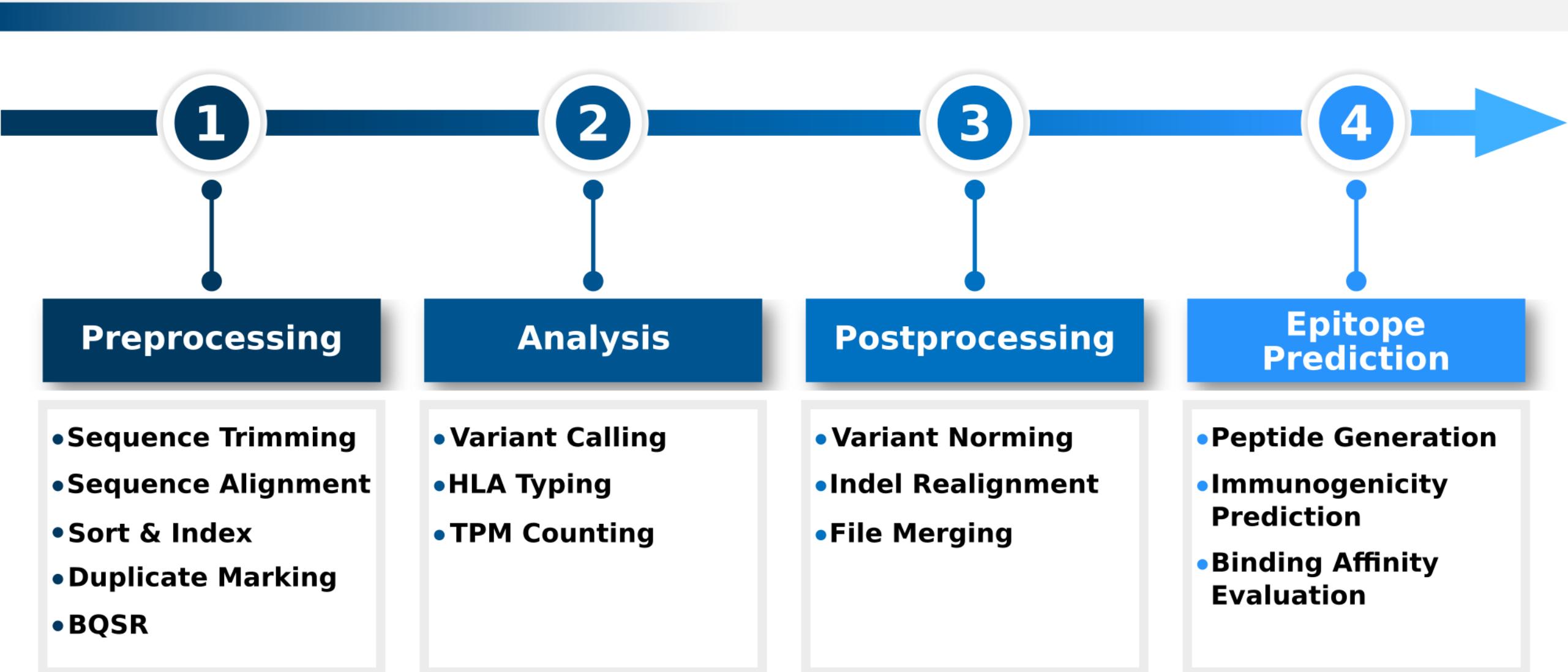
ANTIGEN PRESENTATION PATHWAY



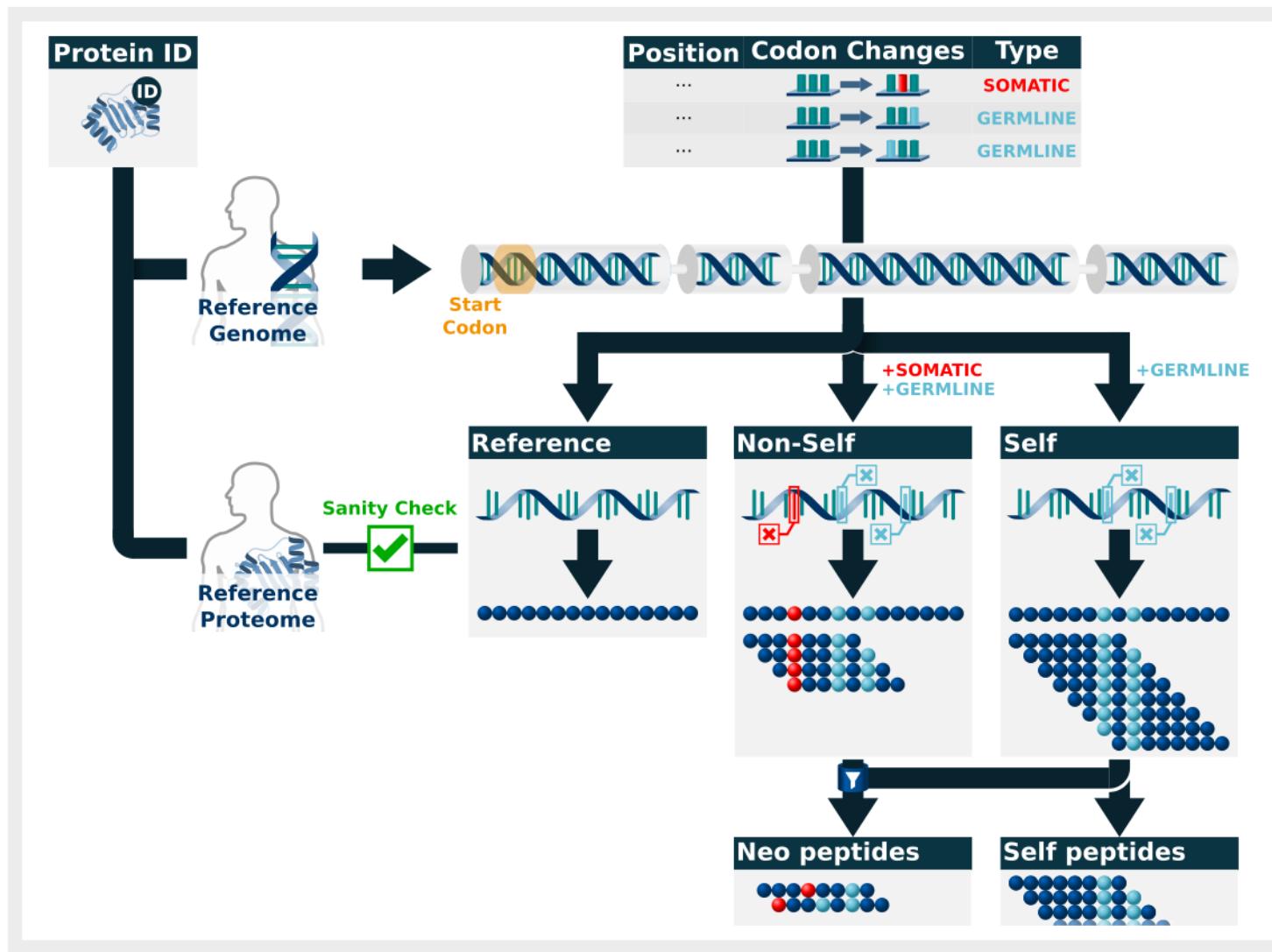
MUTATIONS



ANALYTICAL WORKFLOW

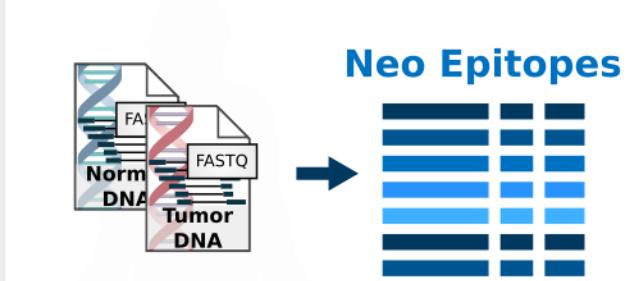


PEPTIDE GENERATION

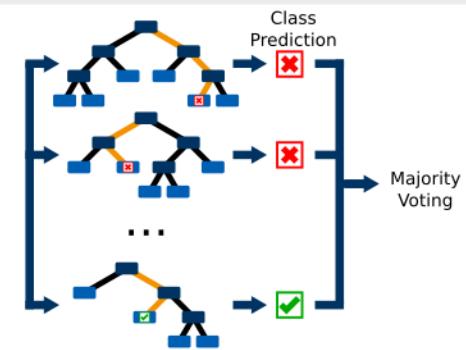


FEATURES

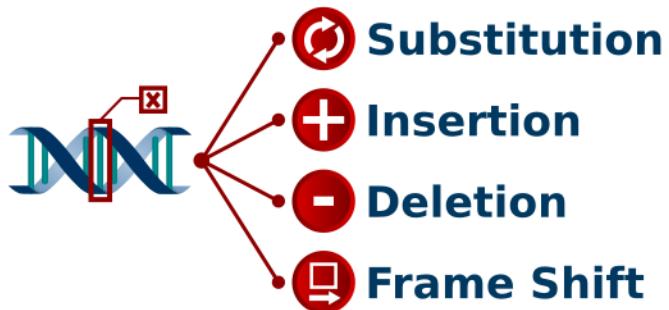
End-To-End



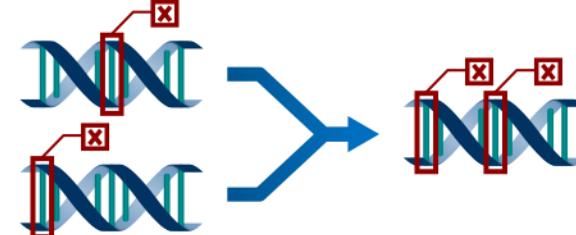
Immunogenicity



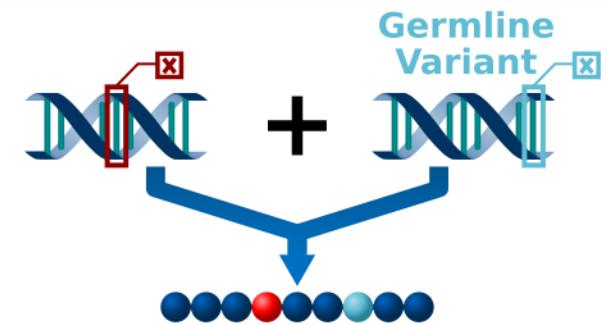
Mutation Types



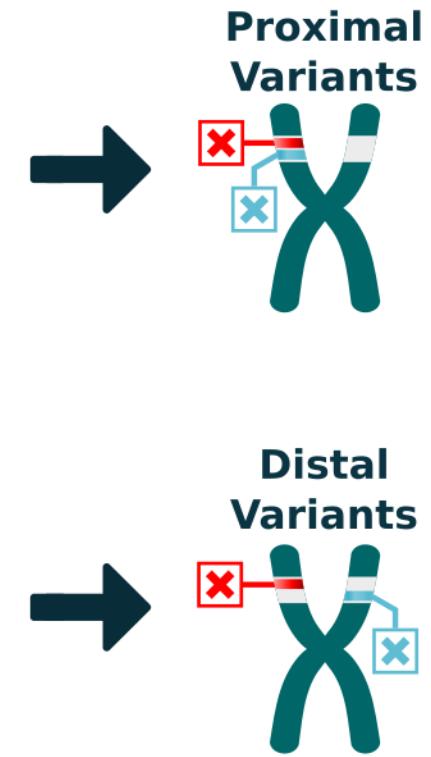
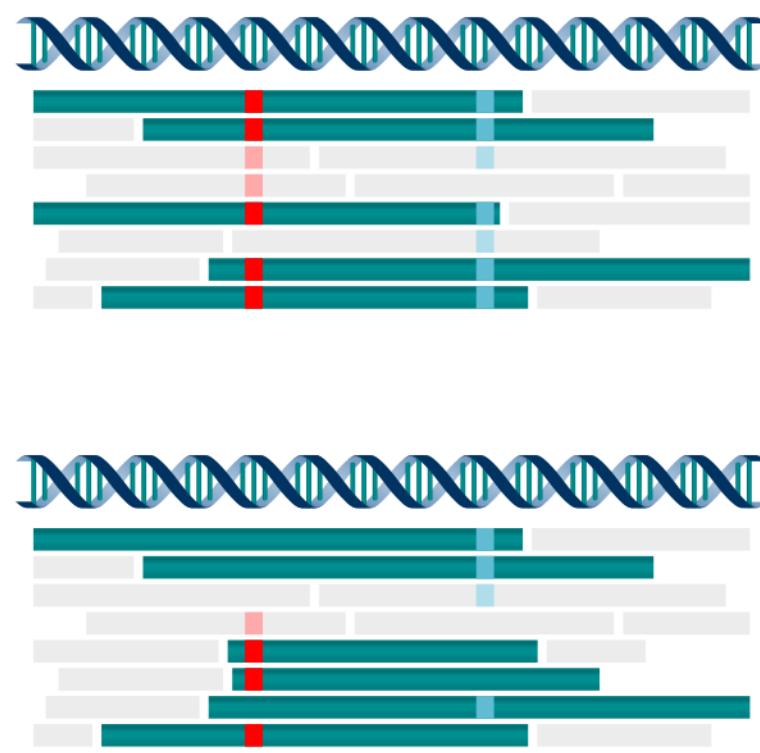
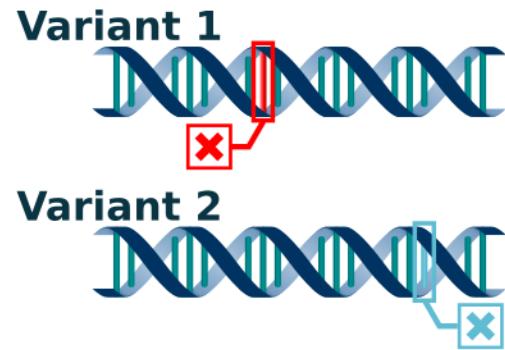
Variant Phasing



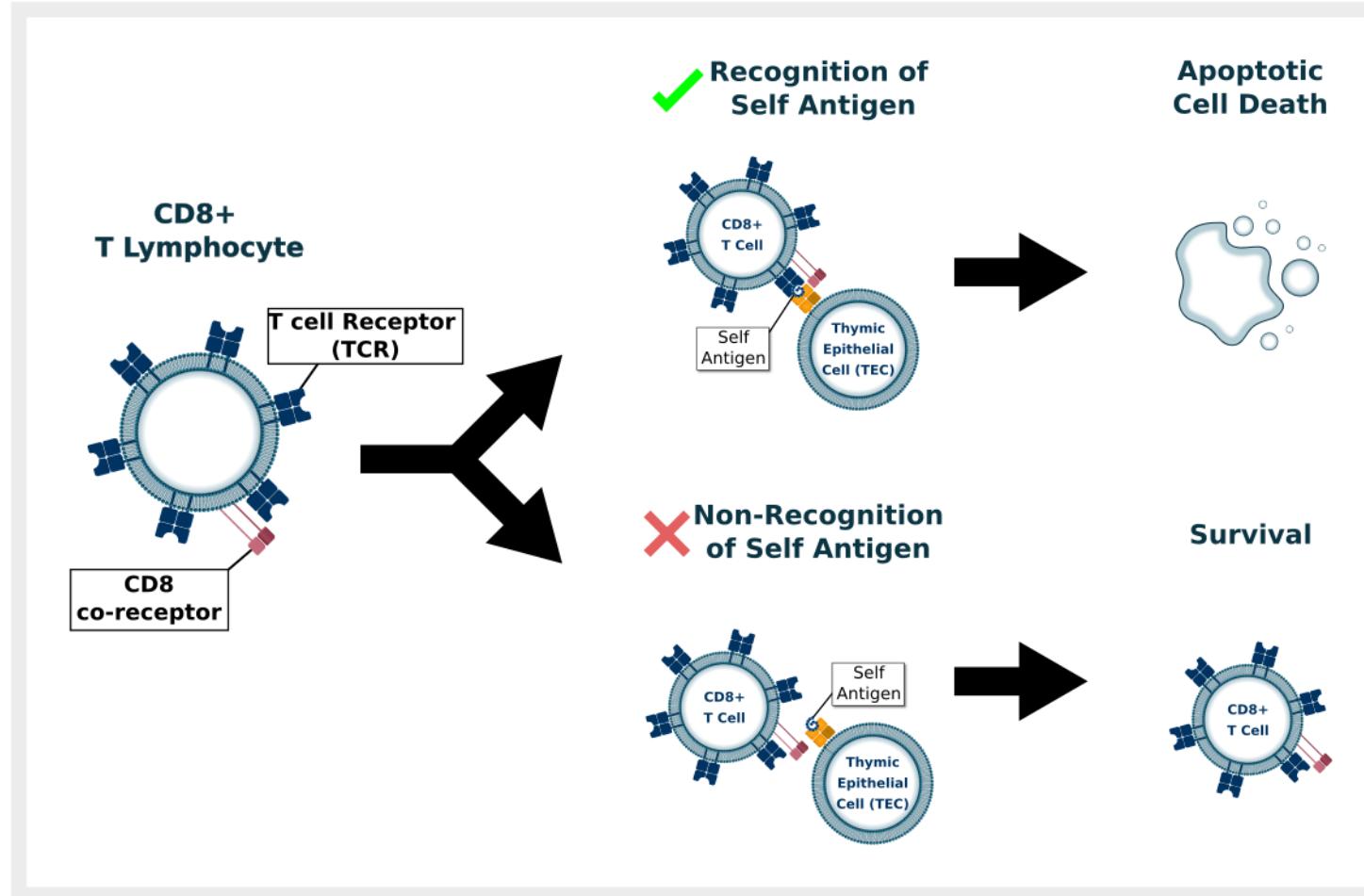
Germline Context



VARIANT PHASING



CENTRAL TOLERANCE

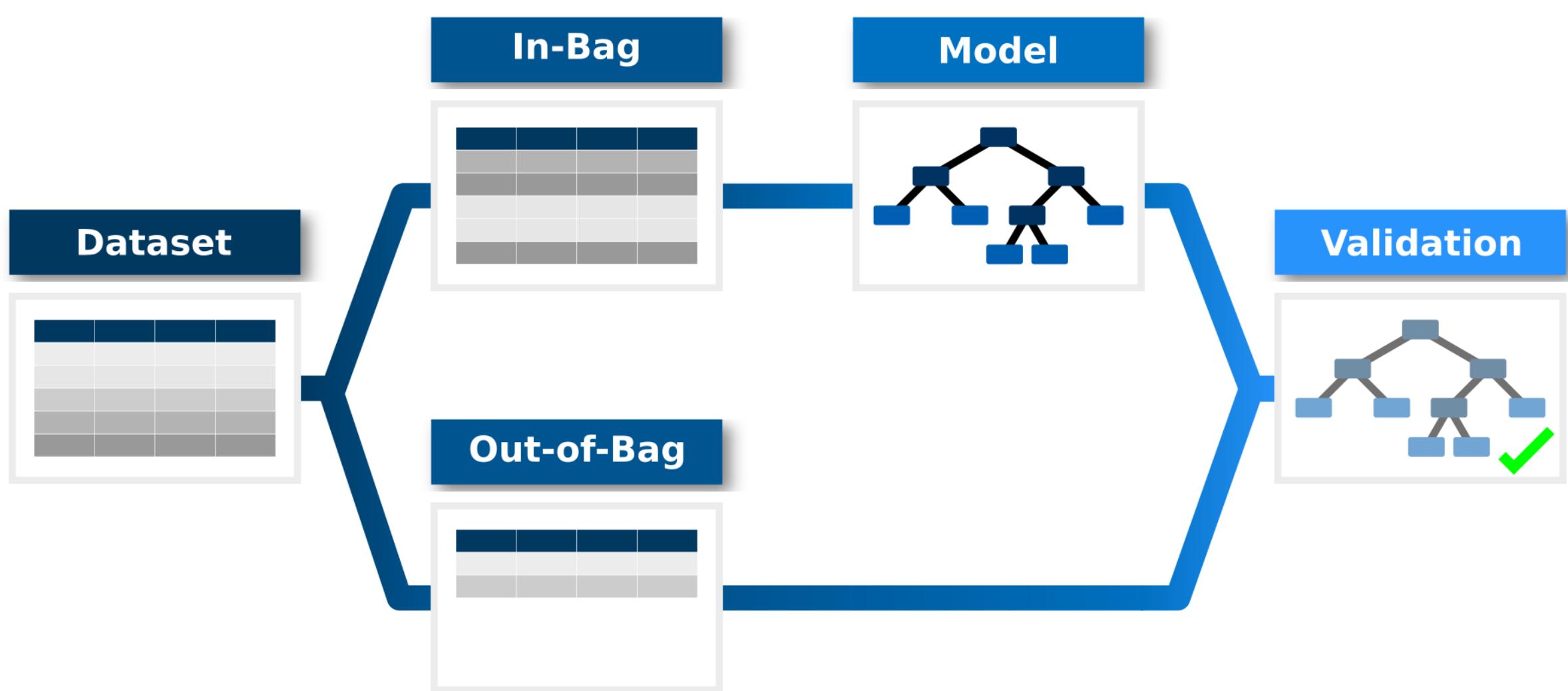


BLOMAP ENCODING

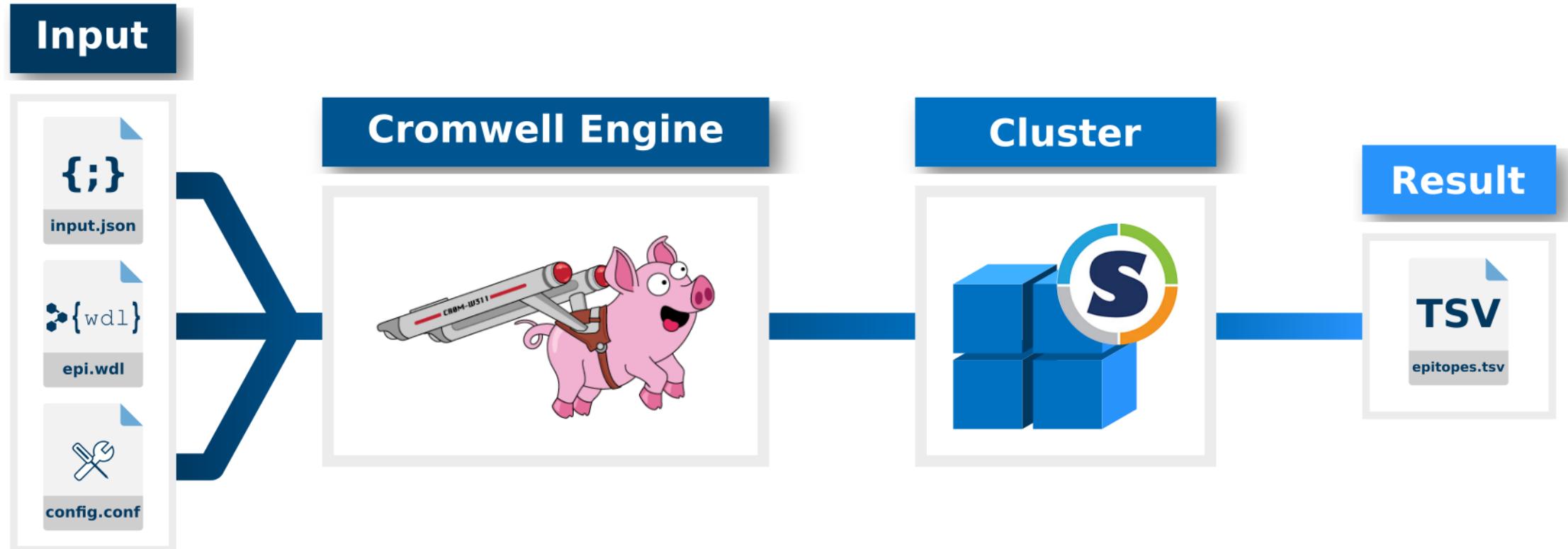
Amino Acids

	A	R	N	D	C	Q	E	G	H	I	L	K	M	F	P	S	T	W	Y	V
X1	-0.57	-0.4	-0.7	-1.62	0.07	-0.05	-0.64	-0.90	0.73	0.59	0.65	-0.64	0.76	1.87	-1.82	-0.39	-0.04	1.38	1.75	-0.02
X2	0.39	-0.83	-0.63	-0.52	2.04	-1.50	-1.59	0.87	-0.67	0.79	0.84	-1.19	0.05	1.04	-0.63	-0.27	-0.3	1.69	0.11	0.30
X3	-0.96	-0.61	-1.47	-0.67	0.65	-0.67	-0.39	-0.36	-0.42	1.44	1.25	-0.65	0.06	1.28	0.32	-1.51	-0.82	1.91	0.65	0.97
X4	-0.61	1.26	1.02	1.02	-1.13	0.49	0.69	1.08	1.13	-1.90	-0.99	0.68	-0.62	-0.61	0.03	-0.25	-1.02	1.07	0.21	-1.55
X5	-0.69	-0.28	1.06	1.47	-0.39	0.21	1.04	1.95	0.99	-0.93	-1.90	-0.13	-1.59	-0.16	0.68	0.31	-0.04	-0.05	-0.41	-1.16

BOOTSTRAP RESAMPLING



Pipeline Execution



```
> java -Dconfig.file=configuration.conf -jar cromwell.jar run epi.wdl --inputs input.json
```

FUTURE WORK

- Include more sources for neo antigens
(non-coding sequences, post translational modifications, ...)
- Allow for the analysis of more complex types of mutations
(e.g. genomic fusions)
- Evaluate other similarity measures in terms of immunogenicity prediction
- Improve predictive model
- Improve pipeline's concurrency for faster analysis