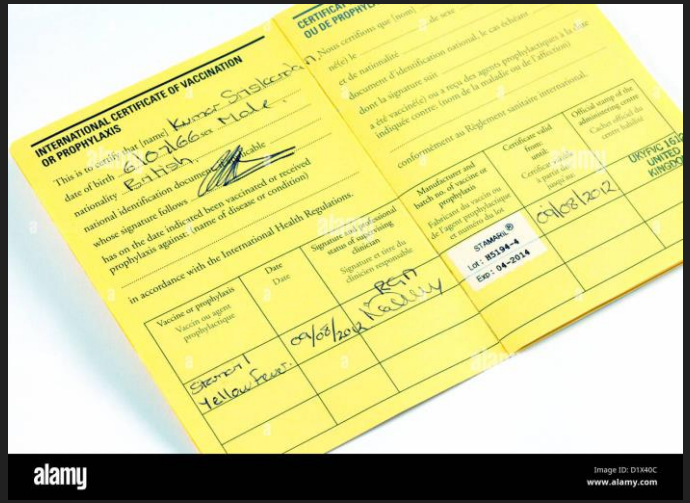
Systems Biology Approach Predicts Immunogenicity of the Yellow Fever Vaccine in Humans

Electra Scarpignato, Ethan Masters, Evan Horvath, Kara Walp

Querec, T., Akondy, R., Lee, E. et al. Systems biology approach predicts immunogenicity of the yellow fever vaccine in humans. Nat Immunol 10, 116–125 (2009). https://doi.org/10.1038/ni.1688

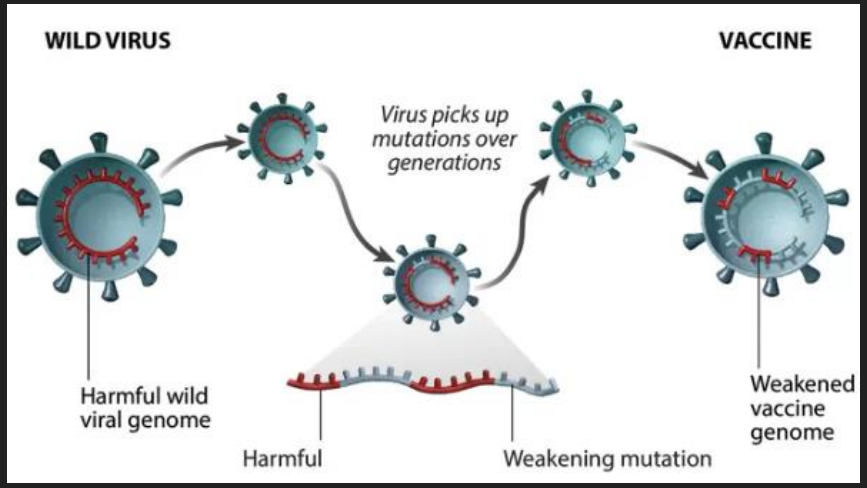
Yellow Fever Vaccine (YF-17D)

- Developed in 1930s by Max Theiler
- Administered to 600+ million people
- Attenuated Asibi strain of yellow fever virus
- One of the most effective vaccines ever made
  - Effective immunity within 30 days for 99%
  - Model vaccine
    - Innate immune response
    - Inform new vaccine design
    - Subsequent differences in adaptive immunity



https://www.alamy.com/stock-photo/yellow-fever-certificate.html?sortBy=relevant

Attenuated Virus Vaccines

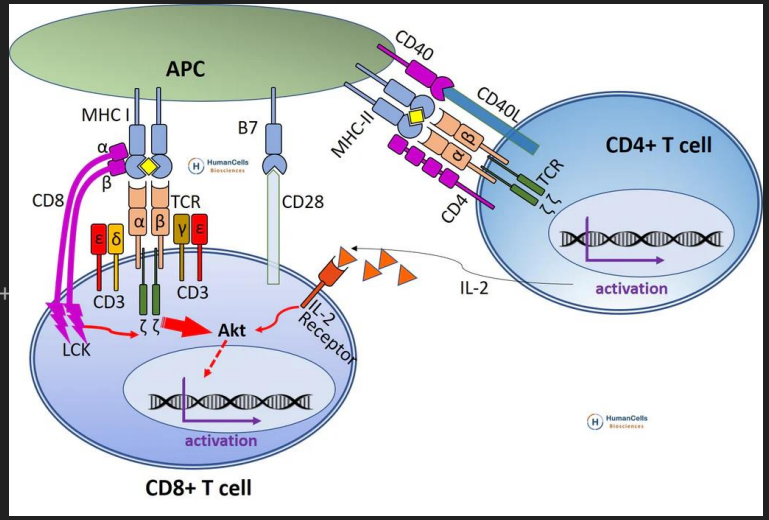


https://www.pfizer.com/news/articles/understanding_six_types_of_vaccine_technologies

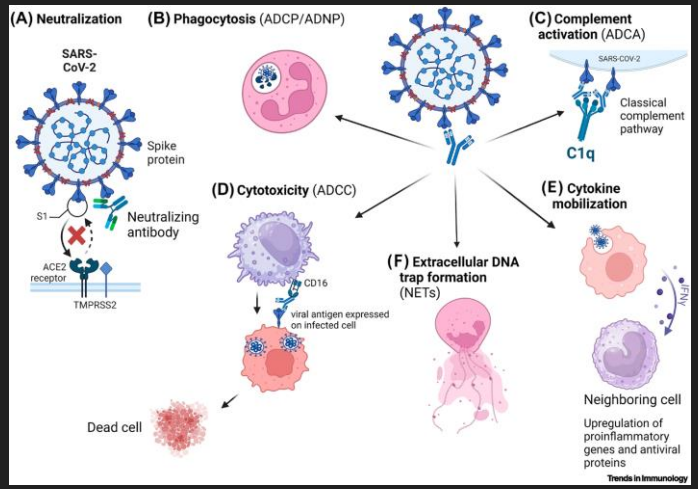
- Smallpox, polio, measles, mumps
- Live attenuated virus
- Control viral replication and virulence
  - Deletion or mutation of replication genes
- Safety concerns: potential to revert to original virulence, impaired immunity
- Greater efficacy than inactivated virus or subunit vaccines

Adaptive Response to YF-17D

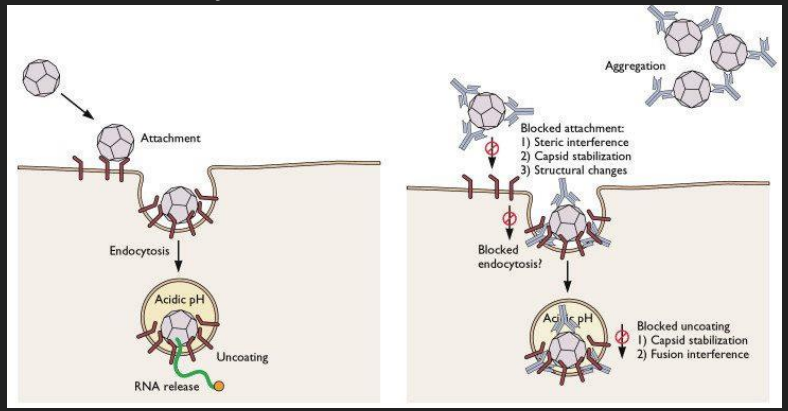
- Cytotoxic T lymphocytes (CTLs)
  - T helper type Ⅰ (T_H1, $CD8^+$) and type Ⅱ (T_H2, $CD4^+$
- Neutralizing antibodies up to 30 years
    - Complement pathway



https://humancellsbio.com/products/human-normal-peripheral-blood-cd8-cytotoxic-t-cells?variant=39874997715047



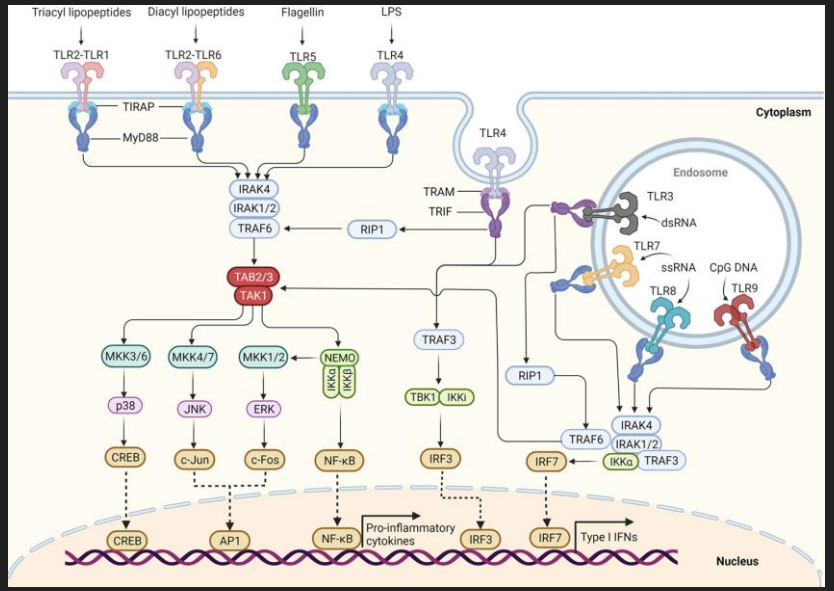
https://www.cell.com/trends/immunology/fulltext/S1471-4906%2824%2900152-2



https://virology.ws/2009/07/24/virus-neutralization-by-antibodies/

Innate Processes that Influence Adaptive Response

- Toll-like receptors (TLRs)
  - Innate immune cells
  - Release pro-inflammatory cytokines and/or interferons (IFNs) in response to a virus
    - Yellow fever has ssRNA: TLR7/TLR8 activation → Type Ⅰ IFNs released
- Antigen presenting cells (APCs, DCs)



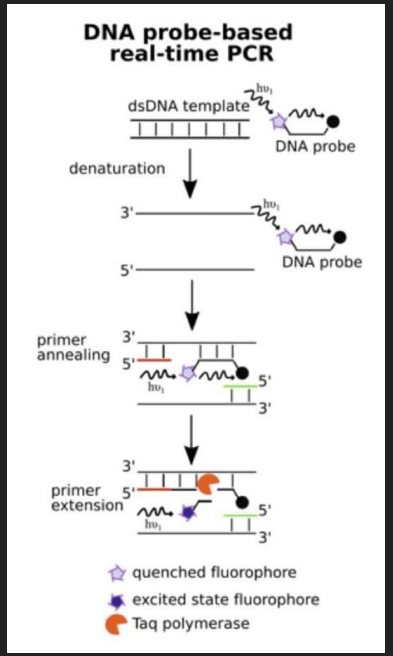
Janeway, Charles. Immunobiology: The Immune System in Health and Disease. Garland Science, 2005.

Goals of the Paper

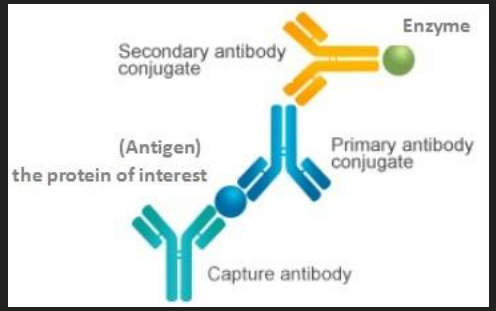
- Conduct multivariate analysis of innate response
  - Gene expression profiling
  - Multiplex analysis of cytokines and chemokines
  - Multiparameter flow cytometry
  - Computational modeling
- Previously utilized for oncology but not for vaccinology
- Identify innate immune signatures to predict adaptive immune response
  - Inform effects of differences in immune systems person to person
  - Contribution to systems immunology

Other Methodologies

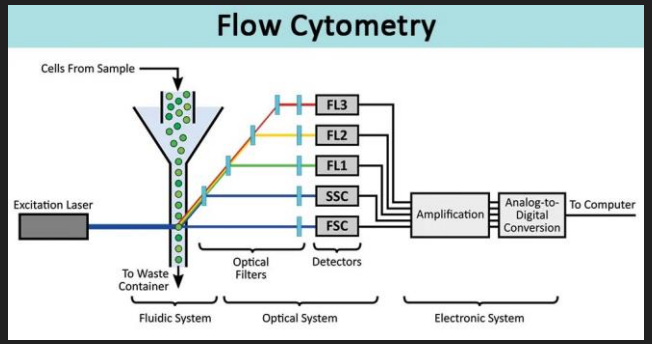
https://bitesizebio.com/29508/real-time-pcr-digest/



- Gene expression profiling (GEP, transcriptomics)
  - Microarray
  - RT-PCR
- Multiplex cytokine/chemokine analysis
  - Enzyme-linked immunosorbent assay (ELISA)
    - Antibodies
- Flow cytometry



https://www.cusabio.com/c-20931.html



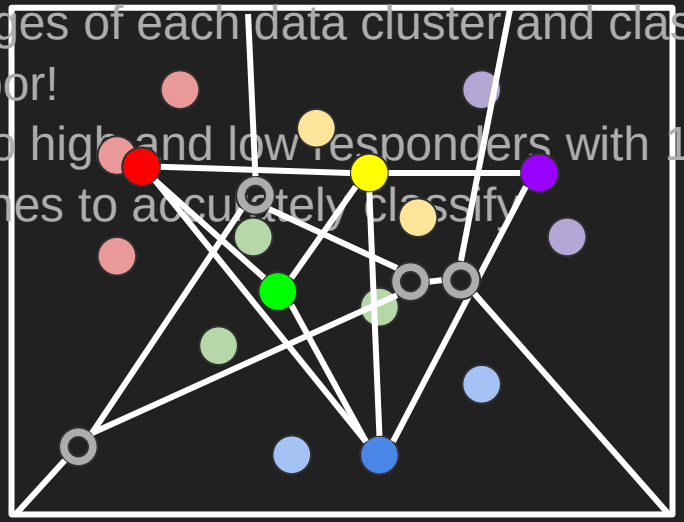
https://microbenotes.com/flow-cytometry/

Methodology Overview

- Two trials (n=15, n=10) examining the response to YF-17D vaccination
- Identified specific genes associated with immune response
  - Days 1, 3, 7, and 21
- Tested if genomic signatures identified in trial 1 predict response in trial 2
- Used primarily two methods for classification and prediction
  - ClaNC
  - DAMIP
- Sought to predict the high and low CD8+ T cell responders
- Verified predictions with RT-PCR

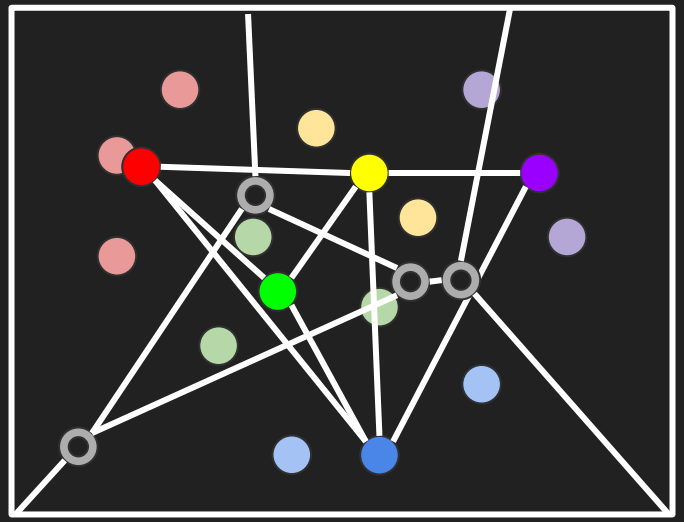
ClaNC Overview

- Centroid → average of the data points in a cluster
- Overall…
  - Find the averages of each data cluster and classify data points by the nearest neighbor!
- Used to classify into high and low responders with 15 genes
  - Needed 48 genes to accurately classify
  - Overtrained



ClaNC Overview

- Centroid → average of the data points in a cluster

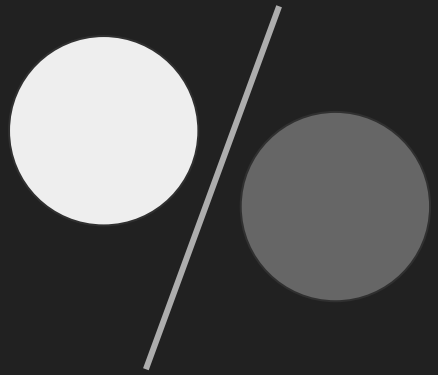


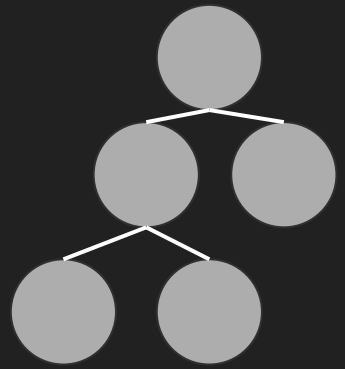
DAMIP

Discriminant Analysis via Mixed-Integer Programming

Discriminant Analysis

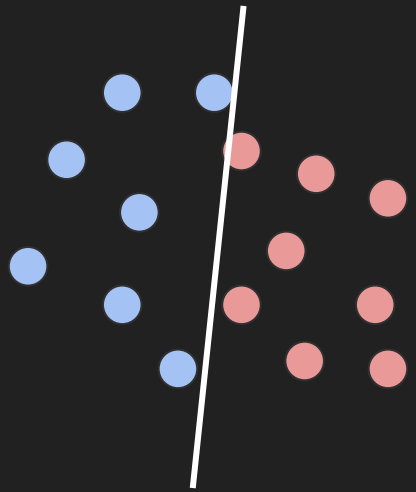
MIP





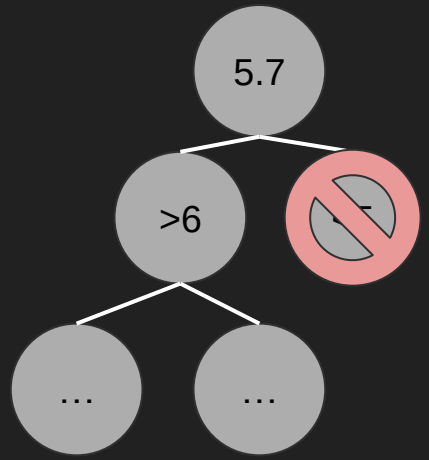
Discriminant Analysis Overview

- Goal is to determine a discriminant function to differentiate between multiple classes
- Low/High responders (T-Cells and antibodies)



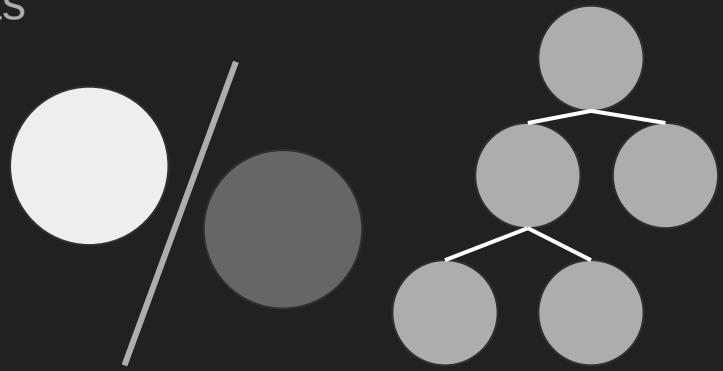
Mixed-Integer Programming Overview

- Common solving method → Branch-and-bound
- Begin with linear relaxation → remove integrality restrictions
- Solve constraint problem–max or min
- Select branching variable
    - Prune branch that does not maximize/minimize
- Repeat until optimality is reached



Putting it Together… DAMIP

- Assign binary variables to each observation
- MIP model shown earlier is used instead of the traditional linear/quadratic discriminant function
  - 2 group classification
  - Maximize the amount of correctly classified data points
- Provides more accurate classification results
- Determined 8-14 predictive signatures



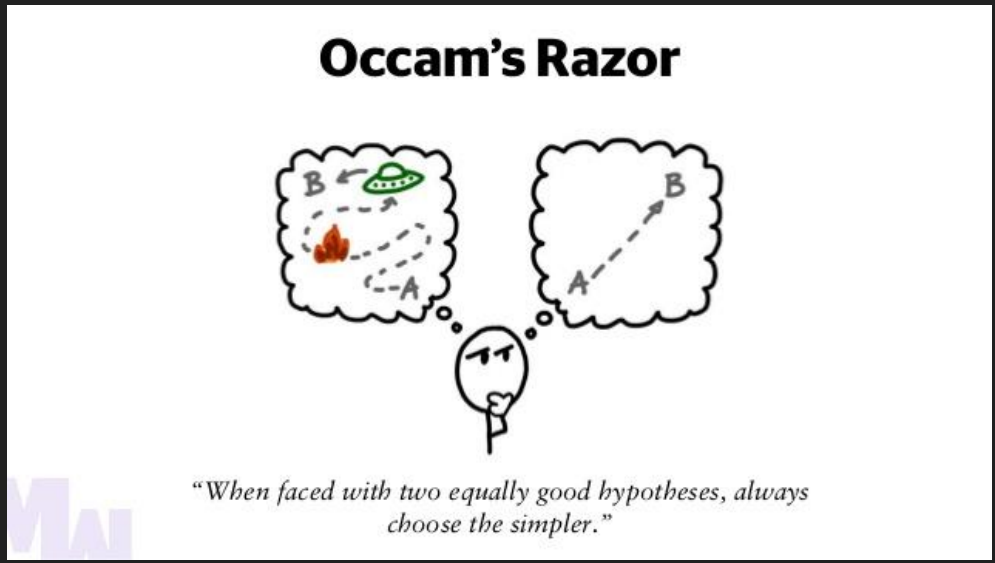
Advantages and Disadvantages:

ClaNC vs. (DA)MIP

Utility and Feasibility

ClaNC: Design & Implementation Advantages

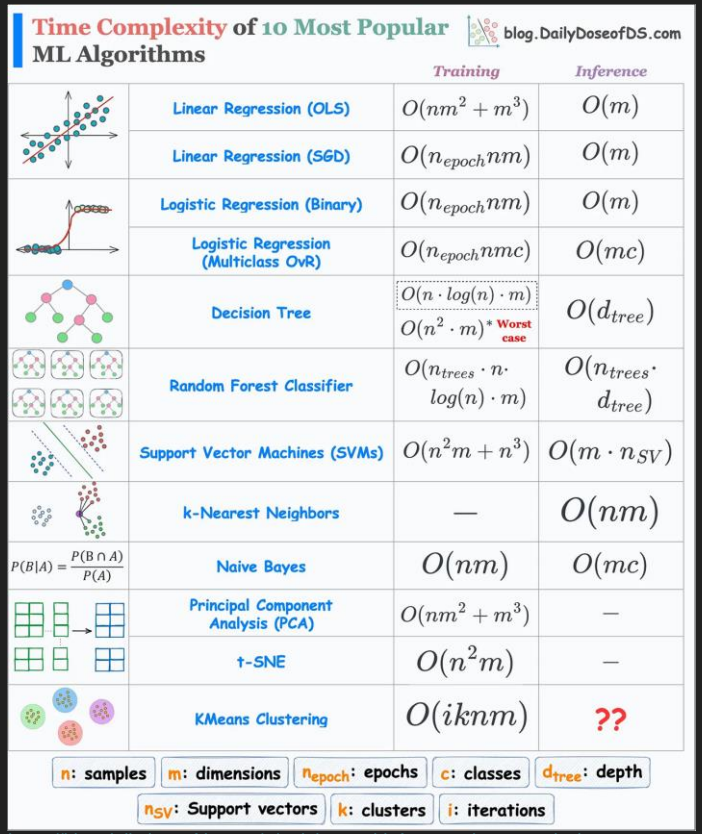
Algorithmic Simplicity



https://automaticaddison.com/occams-razor-and-machine-learning/

ClaNC: Design & Implementation Advantages

Algorithmic Simplicity



https://blog.dailydoseofds.com/p/training-and-inference-time-complexity

ClaNC: Design & Implementation Advantages

Algorithmic Simplicity

Less Data Points for Training

Generally, to prevent overfitting,

$$N \geq 10 * k$$

N = number of data points
k = number of dimensions/features per data point

ClaNC: Design & Implementation Advantages

Algorithmic Simplicity

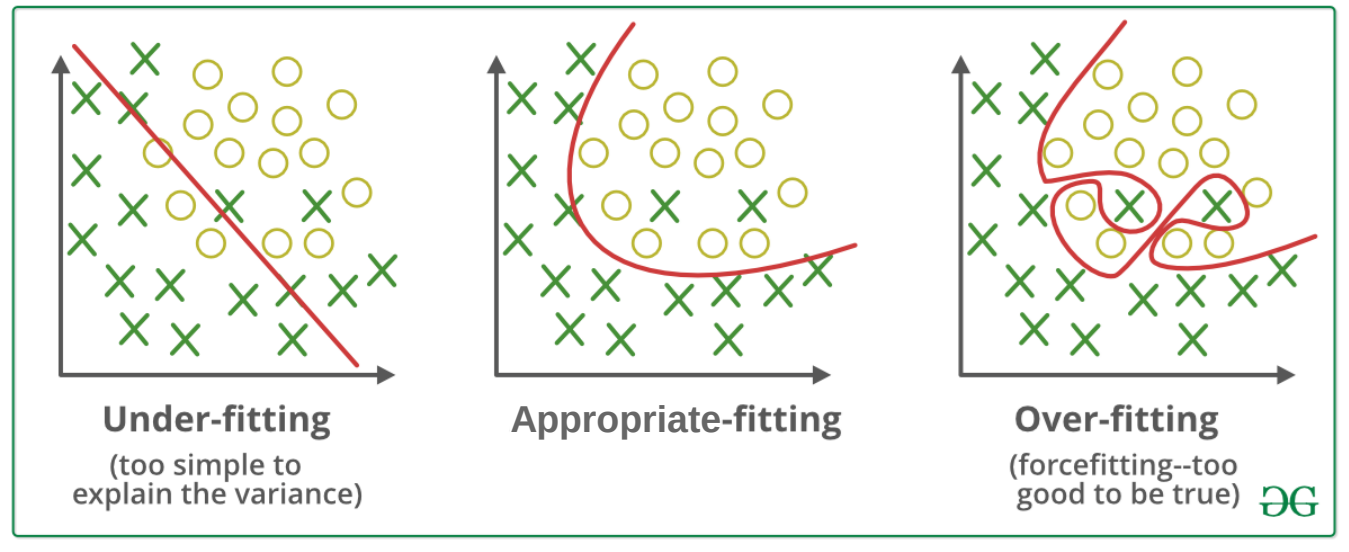
Less Data Points for Training

T-Cell Response Prediction and Identifying Immune Correlates

- can pinpoint factors of immune response (eg. related to complement activation)
- can capture complex metabolic influences on immunological response

ClaNC: Disadvantages

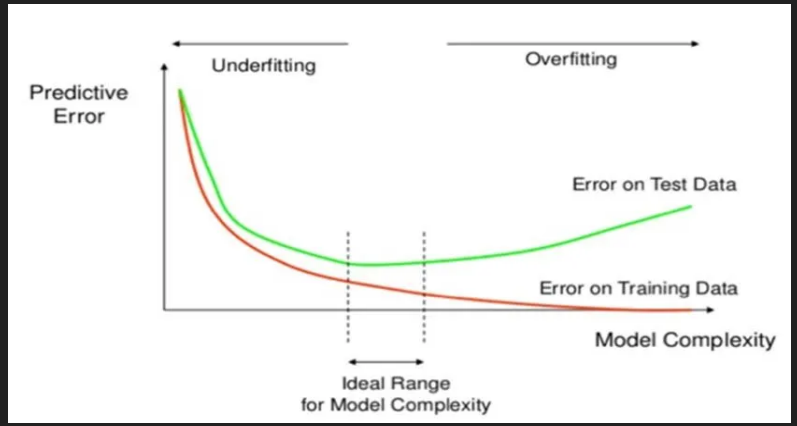
Overfitting



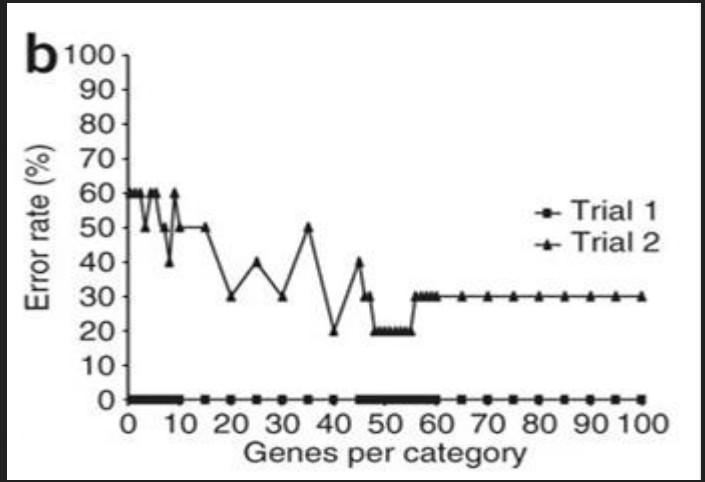
https://www.geeksforgeeks.org/underfitting-and-overfitting-in-machine-learning/

ClaNC: Disadvantages

Overfitting



https://aiml.com/what-is-overfitting/



ClaNC: Disadvantages

Overfitting

Underlying Mechanism Prediction

- Not good at gene role prediction
- Identifies correlation between genes and modulation outcome, but not of the innate mechanism

DAMIP: Design Advantages

Large Scale Discrete Optimization

Support Vector Machines

DAMIP

- Classification of any number of groups
- Can incorporate heterogeneous attribute types as input
- Pre-training Principal Component Analysis
- Reserved-Judgement Region

DAMIP: Implementation Advantages

Users can set:

- The number of discriminatory gene measurements per signature set

  →
  - This study set it to max of 5
  - Only 3 at most were needed before target classification rate was met :)

- A target value for the threshold classification rate (accuracy)

  →
  DAMIP will:
  - iterate through training cycles
  - terminate once threshold classification rate is achieved

DAMIP: Implementation Advantages

Versatility

- Predicted both T Cell and B Cell responses with high accuracy

Comprehensivity

- Uses multiple data types to identify a broad span of correlates

Utility

- Can identify early vaccine signatures, and is insensitive to prior probability specification

DAMIP: Disadvantages

Training Complexity

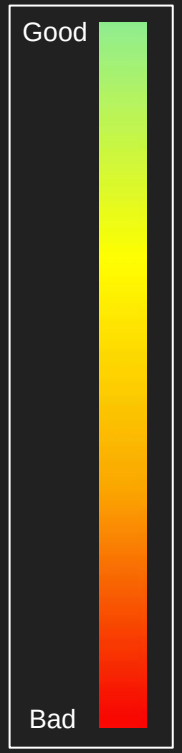
- High amount of computational work required to train DAMIP

Overfitting

- Small risk of overfitting, but this is true with almost any model
- DAMIP is still much better than ClaNC

Overall (Qualitative) Comparison of Model Metrics

Metrics	ClaNC	DAMIP
T Cell Response Prediction		
B Cell Response Prediction		
Specific Pathway Prediction		
Broad Analysis		
Prone to Overfitting	48 Features needed for any training accuracy	< 5 Features needed to meet threshold accuracy :)
Lack of Feasibility to Validate Findings		
Computational Intensity	~ O(*iknm*) training runtime	

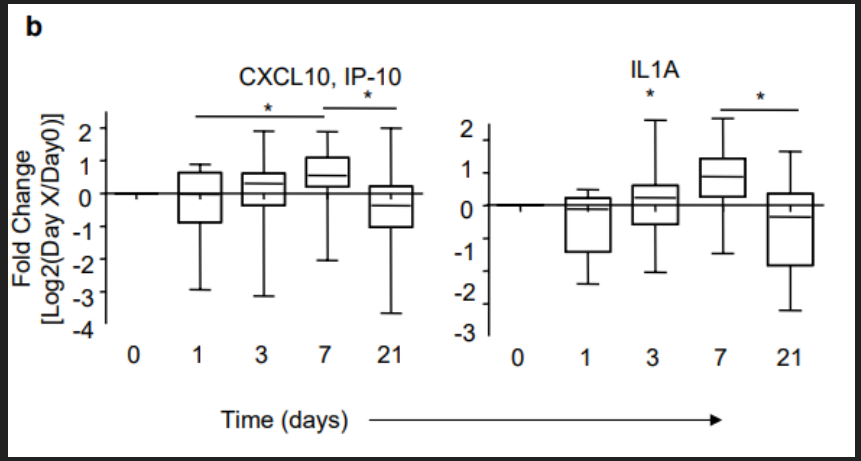


Results

Reiterating research question:

Can multivariate analysis of innate immune response in humans after vaccination be used to identify gene signatures that sufficiently predict adaptive immune response?

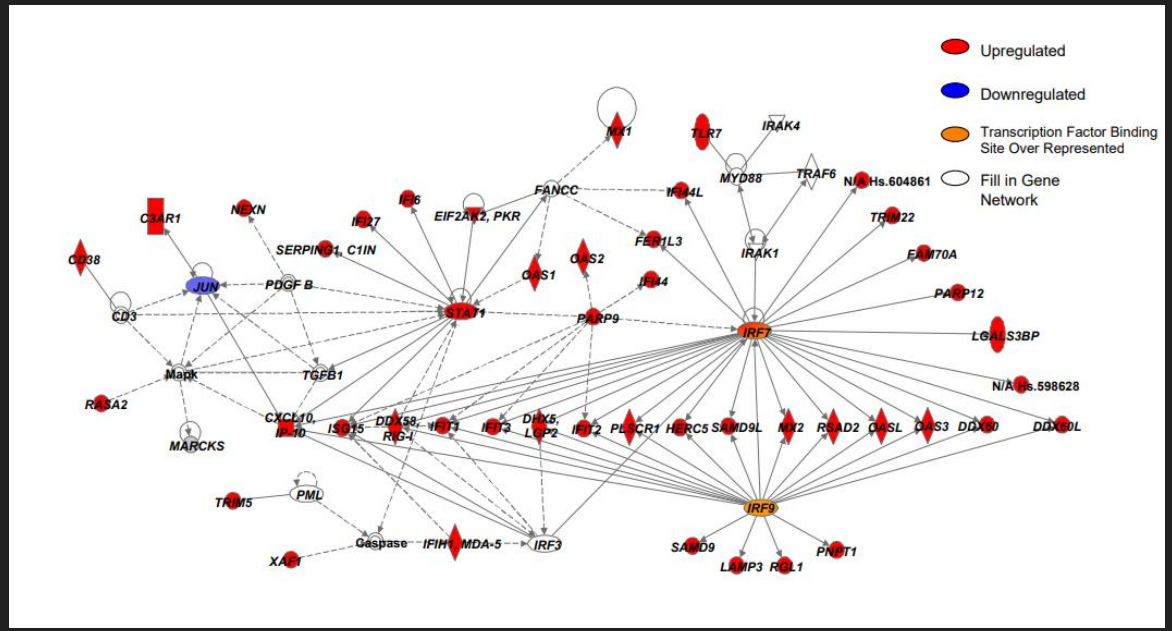
There is a unique innate immune response and gene expression in response to yellow fever vaccine and likely other vaccines



Innate Immune Response Characterization

- Induction of cytokines IP-10 and IL1A
- Upregulation of CD86 on DCs and monocytes

There is a unique innate immune response and gene expression in response to yellow fever vaccine and likely other vaccines



Gene Network

- Genes involved in viral recognition, & mediating antiviral immunity (including complement genes)

There is a unique innate immune response and gene expression in response to yellow fever vaccine and likely other vaccines

Given key finding #1: Yellow Fever vaccine induces characterizable innate immune response and network of antiviral genes

Can we predict the adaptive immunes response based on the induced gene expression or innate immune response from YF-17D?

Methods for Results in Key Finding #1

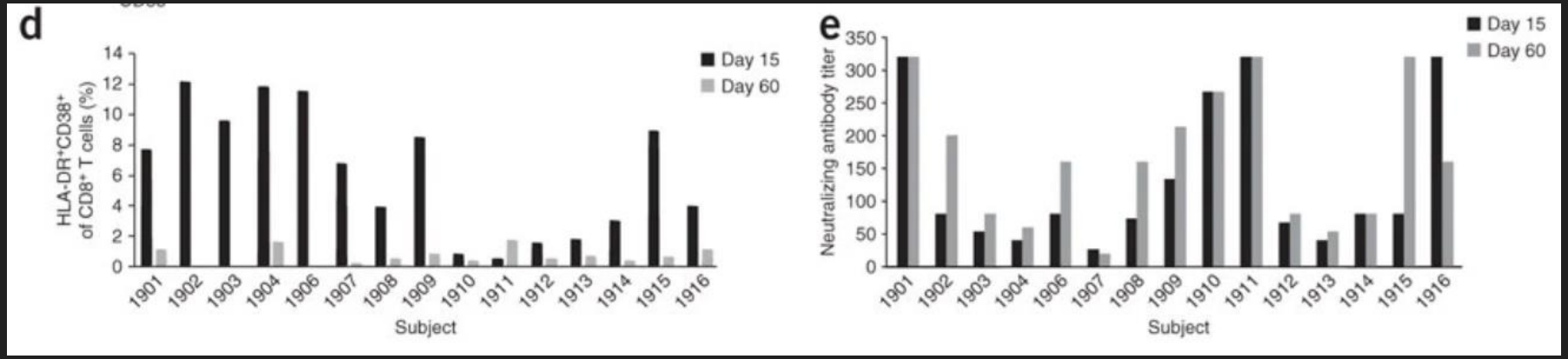
Characterizing Innate Immune Response

- Studied protein cytokine response with 24-plex Luminex assay
  - Time points: 0,1,3,7 and 21 days after vaccination
- Flow cytometry analysis to evaluate frequency and activation status of antigen-presenting cells

Obtaining Gene Network

- Transcriptional profiling of peripheral blood monocyte nuclear cells (PBMCs)
  - Trial 1: 15 subjects
- Verified with second trial: 10 subjects
- Imported genes into Toucan for transcription factor binding site analysis (TFBS)
- Visualized gene network with Ingenuity Pathway Analysis

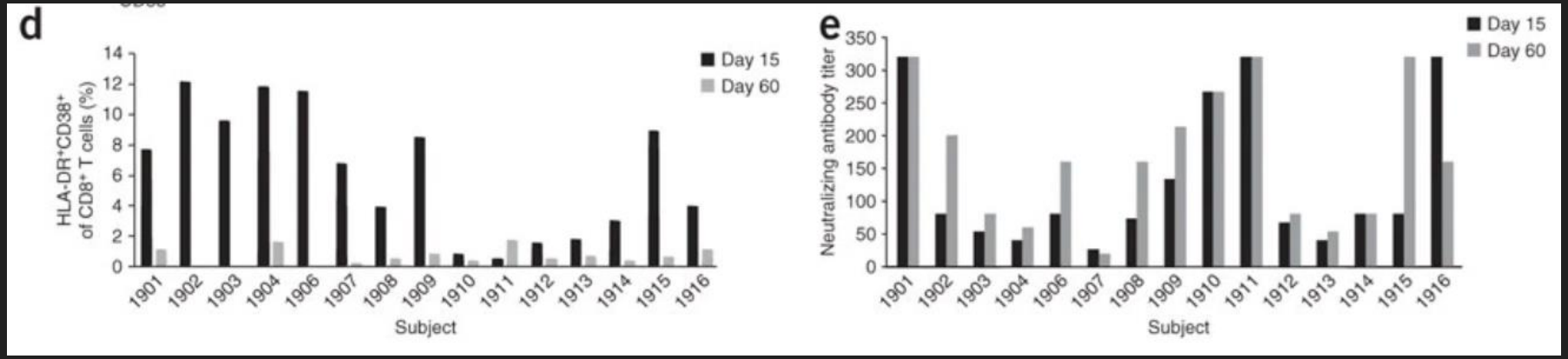
Unique innate immune response and gene expression as result of YF-I7D vaccine cannot predict adaptive immune response



- **Key finding #2:** Adaptive immune responses varied among individuals by more than tenfold
- Can potentially predict future response of CD8+ T-cells from early CD8+ T-cell response

Method: Flow cytometry → Magnitude of CD8+ T cell response proportional to HLA-DR+CD38+ population → analyzed CD8+ T cell activation by % of CD38+ HLA-DR+ cells at 15 and 60 days.

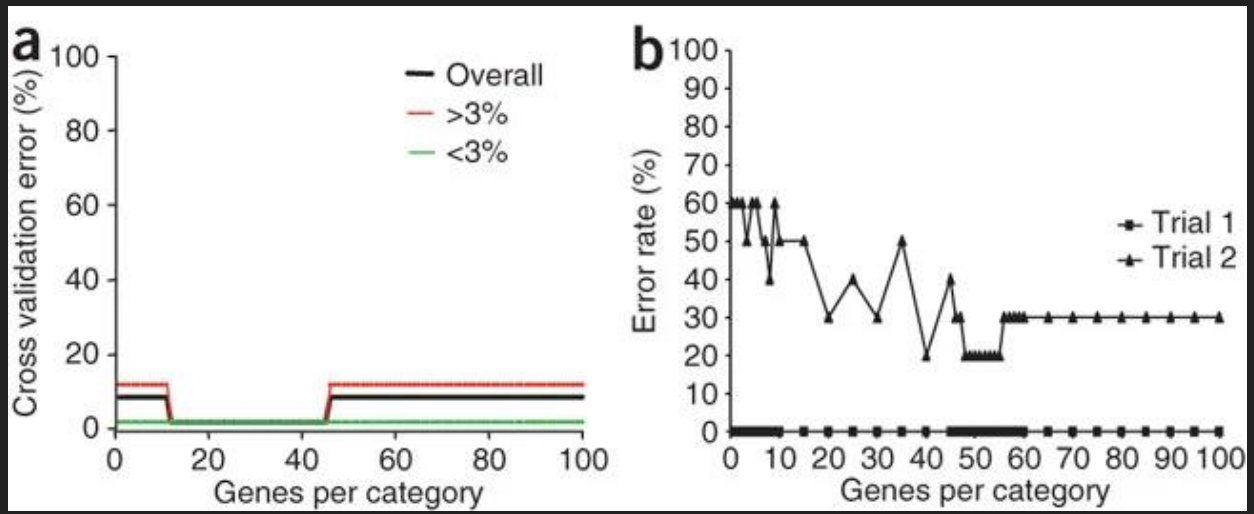
Unique innate immune response and gene expression as result of YF-I7D vaccine cannot predict adaptive immune response



Implications/Questions: Why is there such large variation among individuals even with a highly effective vaccine?

DAMIP is an effective multivariate analysis model in determining correlations between gene expression and adaptive immune response

ClaNC Model Predictions (Fig. 4a,b)



DAMIP is an effective multivariate analysis model in determining correlations between gene expression and adaptive immune response

Table 2 Genomic signatures that predict the magnitude of the CD8⁺ T cell responses using the DAMIP model

			DAMIP model predictive signatures																					
			Train on trial 1, test on trial 2								Train on trial 2, test on trial 1													
Gene name	Gene symbol	Gene ID	1	2	3	4	5	6	7	8	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Solute carrier family 2 (facilitated glucose transporter), member 6	*SLC2A6*	Hs.244378 Day 7	X		X	X	X	X	X	X	X	X	X		X		X			X	X	X	X	
Eukaryotic translation initiation factor 2 alpha kinase 4	*EIF2AK4*	Hs.412102 Day 7	X	X	X		X		X	X								X		X				X
Integrin, alpha L (antigen CD11A)	*ITGAL/LFA-1*	Hs.174103 Day 7			X									X		X	X	X	X					
C-terminal binding protein 1	*CTBP1*	Hs.208597 Day 7				X										X								
Tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein	*YWHAE*	Hs.513851 Day 3				X	X												X			X		
Transcribed locus		Hs.619443 Day 7						X			X	X	X	X	X	X	X	X	X	X	X	X	X	X
Protein phosphatase 1, regulatory (inhibitor) subunit 14A	*PPP1R14A*	Hs.631569 Day 3							X															X
Family with sequence similarity 62 member B	*FAM62B*	Hs.649908 Day 7		X						X			X										X	
Transcribed locus		Hs.42650 Day 7									X	X									X			
Accuracy of 10-fold cross-validation (%)			93	93	93	93	93	93	93	93	90	90	100	100	100	100	90	90	90	90	90	90	100	100
Accuracy of 1-fold blind prediction (%)			80	80	80	80	80	90	90	90	87	87	80	73	73	73	73	73	73	73	87	73	80	73
Accuracy of 10-fold blind prediction (%)			81	80	81	80	81	85	85	88	84	84	76	72	75	71	73	71	71	75	84	73	76	70

DAMIP is an effective multivariate analysis model in determining correlations between gene expression and adaptive immune response

Table 4 Genomic signatures that predict the magnitude of the neutralizing antibody responses using the DAMIP model

			DAMIP model predictive signatures														
			Train on trial 1, test on trial 2										Train on trial 2, test on trial 1				
Gene name	Gene symbol	Gene ID	1	2	3	4	5	6	7	8	9	10	1	2	3	4	5
BEN domain–containing 4	*BEND4*	Hs.120591				X	X	X	X	X		X	X			X	
Transcribed locus		Hs.139006	X					X			X	X					
6-Phosphofructo-2-kinase/fructose-2,6-biphosphatase 3	*PFKFB3*	Hs.195471		X													
Tumor necrosis factor receptor superfamily, member 17	*TNFRSF17*	Hs.2556	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Tumor protein D52	*TPD52*	Hs.368433				X			X		X			X		X	
Transcribed locus		Hs.481166					X	X		X		X					
Kelch repeat and BTB (POZ) domain containing 7	*KBTBD7*	Hs.63841	X	X	X	X	X	X	X	X	X	X			X		X
Transcribed locus		Hs.649726							X	X	X						X
Nucleosome assembly protein 1-like 2	*NAP1L2*	Hs.66180			X							X					
Accuracy of 10-fold cross-validation (%)			80	80	80	87	87	80	80	80	80	80	89	89	89	89	89
Accuracy of 1-fold blind prediction (%)			100	100	100	100	100	100	100	100	100	100	73	73	73	73	80
Accuracy of 10-fold blind prediction (%)			97	99	94	92	96	98	92	93	93	94	72	71	75	70	79

Analysis of signatures that predict the neutralizing antibody responses. Here all the discriminatory predictive signature sets turned out to consist of day 7 gene expression only. Further, training on trial 1 produces a high blind prediction accuracy on trial 2. *TNFRSF17* was present in all the predictive signature sets of the DAMIP model, and several genes, including *KBTBD7* and *BEND4* appeared in several signature sets.

DAMIP is an effective multivariate analysis model in determining correlations between gene expression and adaptive immune response

Key Finding #3: Successfully found early gene signatures of innate immune activation that predict T-Cell and antibody response using Independent Classification Methods (DAMIP and ClaNC)

Obtaining Results (Key Findings 3 & 4)

Identified genes correlated with magnitude of CD8+ T cell response and antibody response

Validated with Unsupervised Principal Component Analysis

- Visualize how well identified genes classified subjects into two groups as high or low CD8+ T cell or antibody responders (>3% CD8+ T cell activation was high performing, antibody cutoff was 170 titers)

Tested ability of gene signatures to predict immune response

- DAMIP
- ClaNC

Verified Genes with RT-PCR

Answer to Research Question

Can multivariate analysis of innate immune response in humans after vaccination be used to identify gene signatures that sufficiently predict adaptive immune response?

- The innate immune response to YF-17D cannot be used to predict magnitude of adaptive immune response
- Early gene signatures associated with CD8+ T-cells and antibodies correlate with the magnitude of immune response
- Multivariate analysis with DAMIP can identify gene signatures correlated with adaptive immune response

References

“Yellow Fever.” World Health Organization, World Health Organization, 31 May 2023, www.who.int/news-room/fact-sheets/detail/yellow-fever.

Lauring, A., Jones, J. & Andino, R. Rationalizing the development of live attenuated virus vaccines. Nat Biotechnol 28, 573–579 (2010). https://doi.org/10.1038/nbt.1635

Duan, Tianhao, et al. “Toll-like receptor signaling and its role in cell-mediated immunity.” Frontiers in Immunology, vol. 13, 3 Mar. 2022, https://doi.org/10.3389/fimmu.2022.812774.

“Yellow Fever Vaccine.” Centers for Disease Control and Prevention, Centers for Disease Control and Prevention, 15 May 2024, www.cdc.gov/yellow-fever/php/transmission/index.html.

“Microarray.” Nature News, Nature Publishing Group, 2014, www.nature.com/scitable/definition/microarray-202/.

Chiswick, Evan L., et al. “Detection and quantification of cytokines and other biomarkers.” Methods in Molecular Biology, vol. 844, 21 Dec. 2011, pp. 15–30, https://doi.org/10.1007/978-1-61779-527-5_2.

Janeway, Charles. Immunobiology: The Immune System in Health and Disease. Garland Science, 2005.

Ghosh N, Santoni D, Saha I, Felici G. Predicting Transcription Factor Binding Sites with Deep Learning. Int J Mol Sci. 2024 May 3;25(9):4990. doi: 10.3390/ijms25094990. PMID: 38732207; PMCID: PMC11084193.

Glen J.J. An iterative mixed integer programming method for classification accuracy maximizing discriminant analysis. Computers & Operations Research. 2003 February 30;2. https://doi.org/10.1016/S0305-0548(01)00088-0

Hassan M. “Discriminant Analysis - Methods, Types and Examples.” Research Method, August 15, 2023. https://researchmethod.net/discriminant-analysis/.

Kleiveland CR. Peripheral Blood Mononuclear Cells. In: Verhoeckx K, Cotter P, López-Expósito I, et al., editors. The Impact of Food Bioactives on Health: in vitro and ex vivo models [Internet]. Cham (CH): Springer; 2015. Chapter 15. Available from: https://www.ncbi.nlm.nih.gov/books/NBK500157/ doi: 10.1007/978-3-319-16104-4_15

“Mixed-Integer Programming (MIP) – A Primer on the Basics.” Gurobi Optimization, August 30, 2024. https://www.gurobi.com/resources/mixed-integer-programming-mip-a-primer-on-the-basics/.

“Nearest Centroid Classifier.” EvolkAI. Accessed September 24, 2024. https://evolkai.com/nearest-centroid.

https://automaticaddison.com/occams-razor-and-machine-learning/
https://blog.dailydoseofds.com/p/training-and-inference-time-complexity

Thank you!
Any Questions?