



# MACHINE LEARNING FOR ANTIBODY GLYCOSYLATION OPTIMISATION

**Dr. Charmaine Chia, Dr. Ian Walsh**

Analytical Science & Technologies (GlycoAnalytics)

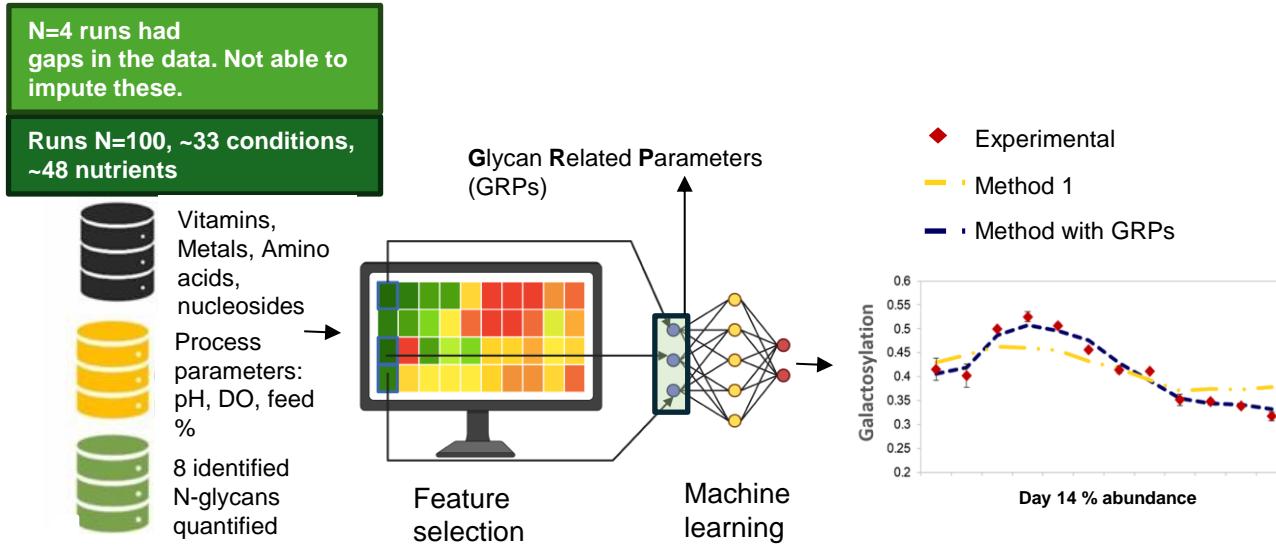
4<sup>th</sup> December 2024

Confidential



CREATING GROWTH, ENHANCING LIVES

# Prediction of glycan levels from process variables



## Aims:

- Discover glycan related parameters (GRPs)
- Discover how combinations of GRPs might influence glycan levels



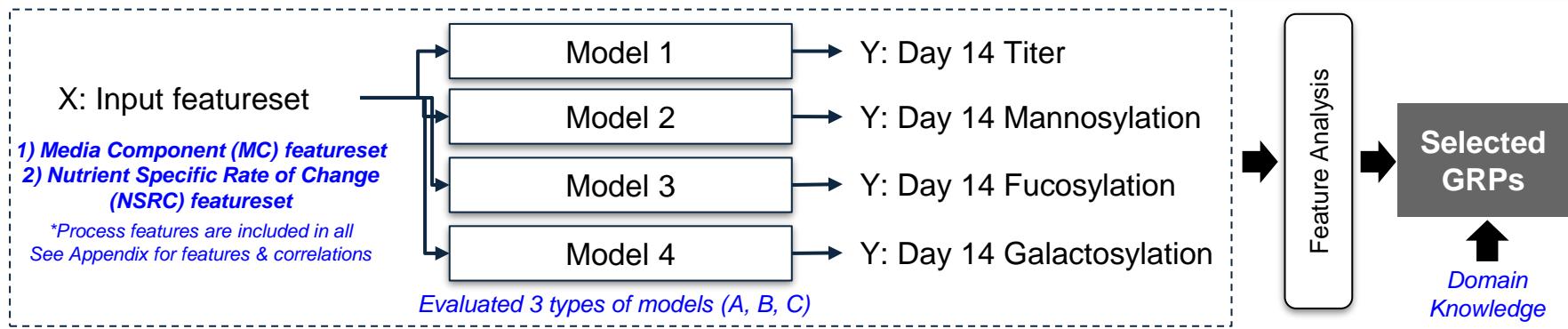
# Model Summary



**n=100, ~33 conditions, ~48 nutrients**

Label	File prefix	Basal media	Feed	Feed %	DO (%)	pH	n
dataset0	240402_20RP06-19	A	a, b, c, d	3, 6	40	7.0	4 x 2 (=8)
dataset1	240704_22DX05-12	A, B, C, D	e, f, d, a	6, 10	40	7.0	4 x 4 x 3 (=48)
dataset2	240710_22DX05-12	A	a	4, 6, 8	40, 60, 80	6.9, 7.0, 7.1	(=48)

**Y (Day 14 CQAs)**  
 0: Titer (mg/L)\_14  
 1: mannosylation\_14  
 2: fucosylation\_14  
     = sum(G0F, G1Fa, G1Fb, G2F)  
 3: galactosylation\_14  
     = sum(G1, G1Fa, G1Fb, G2, G2F)



## Observations:

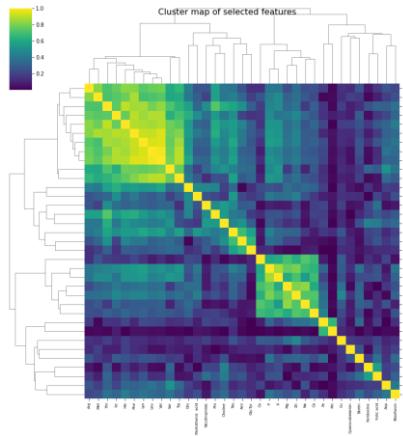
- High correlations between Media features ⇒ makes interpretability / causal discovery challenging
- Some missing data (e.g. nutrient concentrations not measured for certain feeds, etc.) → in generating the dataset, some features / samples had to be dropped to avoid gaps in the data array



# Summary of work done

- Machine learning models were trained on Ajinomoto data in which the Basal, Feed, and process parameters were varied. In addition, an automated feature selection module and feature optimization module were developed to simulate different combinations of features, given the model trained
- Models had good R<sup>2</sup> performance between 0.8 and 0.9 → this suggests that there *are* features in the dataset that have an impact on CQAs
- However, due to strong similarities (correlations) between the nutrient data, individual feature effects are challenging to isolate. This means that there are *many possible feature selection combinations that can fit the data well*.
  - >30% of basal and feed features have correlations above 0.7, respectively. Ideally, most correlations should be below 0.7.
- Using a combination of 1) feature selection methods, 2) domain knowledge, 3) nutrient consumption information, we propose a subset of features to investigate that balances plausibility with model performance
- This selection includes:
  - 17 nutrient features for individual testing
  - 3 process parameters for individual testing (pH, DO, feed %)
  - 2 combination conditions in which multiple nutrients and process parameters are varied at the same time
- Note:
  - Not all the selected features are expected to have an effect—some were added to help identify individual effects among highly correlated and plausible features
  - Direction of feature variation for testing is not always to have a ‘positive’ effect on CQAs (e.g. if Basal-A/Feed-a feature value is already close to optimal). In some cases, we vary the feature to verify if there’s an effect, or trend (e.g. by seeing if there’s negative effect)

# Model Summary



## NSRC FEATURE CORRELATIONS

### Highly correlated:

(Arg, Met), (Thr, Ile, His, Phe, Lys, Leu, Val), (Ser, Trp)  
Nicotinamide, Pro, Choline

Tau, Asn, Gly-Tyr

Co, P, K, Mg, Zn, Na, Ca,

Fe, Mn

Pyridoxine, Folic acid

### No strong correlations:

Glu

Pantothenic acid

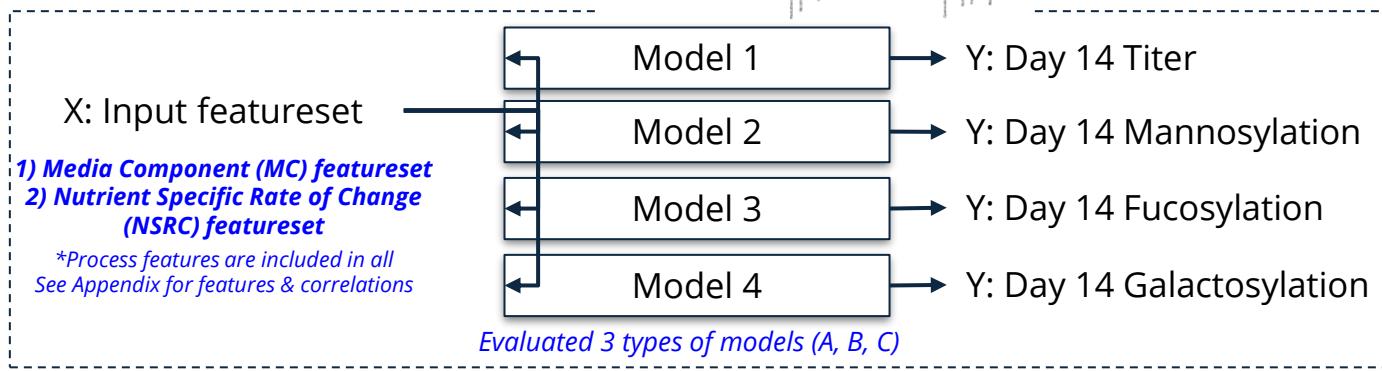
Cu

Cyanocobalamin

Biotin

Asp

Riboflavin



Selected GRPs

20-Feature Selection (not ranked)	
1. Arg	
2. Asn	
3. Asp	11. Ser
4. Folic acid	12. Thr
5. Co	13. Pro
6. Fe	14. Uridine
7. Ca	15. Riboflavin
8. Mg	16. Tyr
9. Mn	17. Glu
10. Zn	+ DO, pH, feed vol

Combination 1	
1. Asp	
2. Asn	
3. Ca	
4. Mg	
5. Uridine	
6. Riboflavin	
	+ DO, pH, feed vol

Combination 2	
1. Arg	
2. Folic acid	
3. Co	
4. Zn	
5. Thr	
6. Pro	
7. Uridine	
8. Riboflavin	
9. Tyr	
	+ DO, pH, feed vol

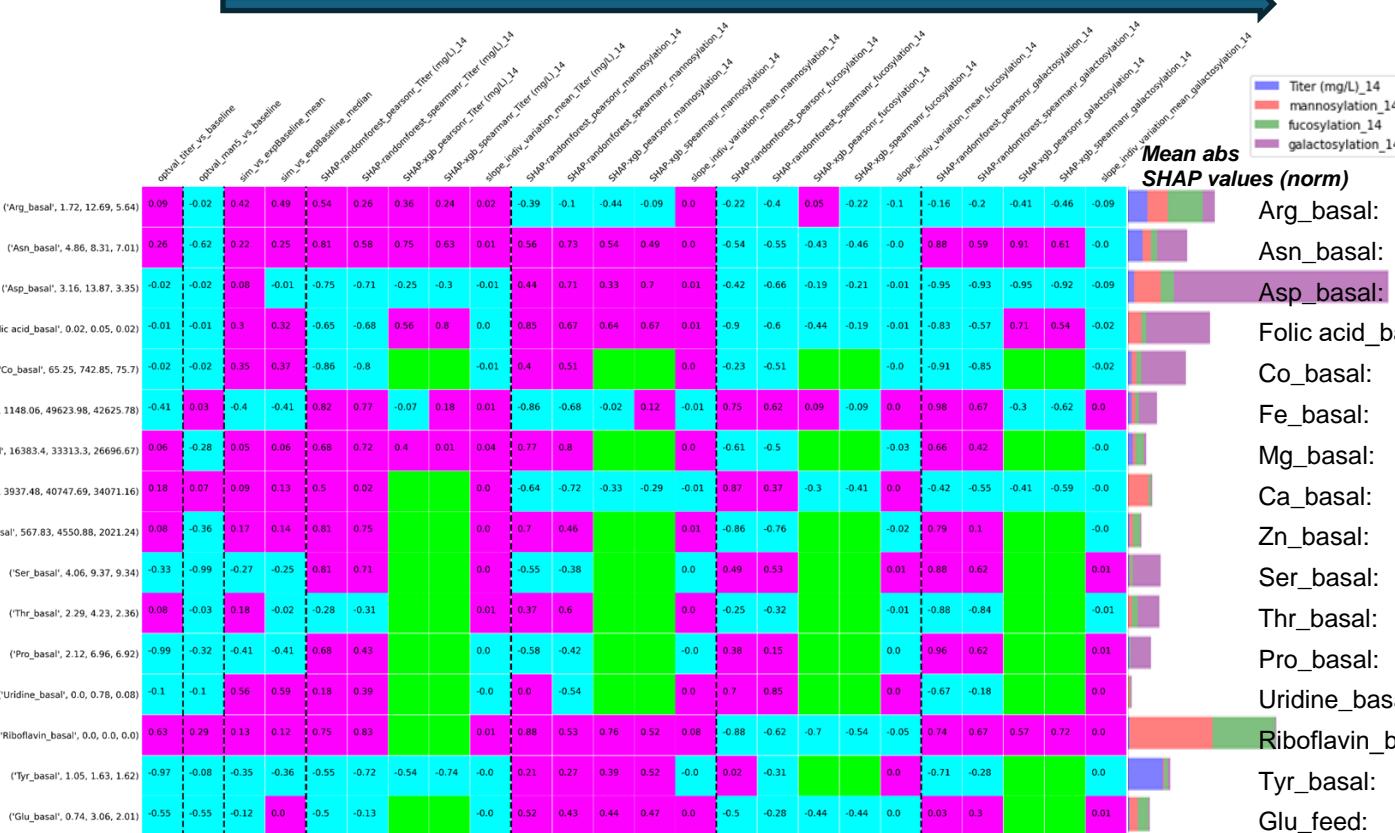
Direction of change (increase, decrease) given

# Individual feature effect direction summary (Basal)

a\*

Different metrics to describe the effect of varying the feature on each output CQA

Selected features for testing ↓



~ little change  
↓ decrease  
↑ increase

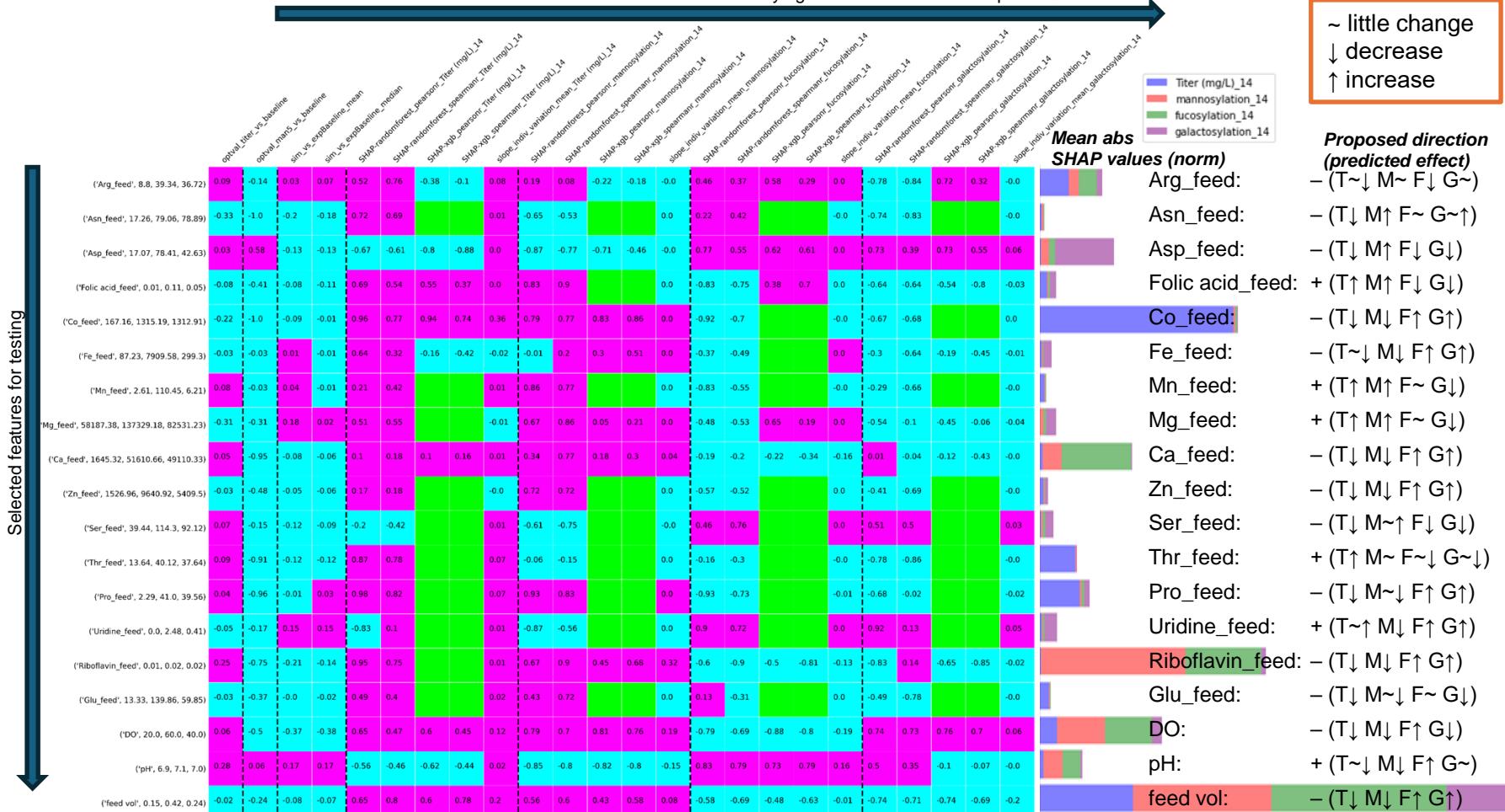
**Proposed direction (predicted effect)**

- Arg\_basal: + (T↑ M↓ F↓ G↓)
- Asn\_basal: - (T↓ M↓ F↑ G~↓)
- Asp\_basal: + (T↓ M↑ F↓ G↓)
- Folic acid\_basal: + (T~ M↑ F↓ G↓)
- Co\_basal: - (T↑ M↓ F↑ G↑)
- Fe\_basal: + (T↑ M↓ F~↑ G~↑)
- Mg\_basal: + (T↑ M↑ F↓ G~↑)
- Ca\_basal: + (T↑ M↓ F↑ G↓)
- Zn\_basal: - (T↓ M↓ F↑ G~↓)
- Ser\_basal: - (T↓ M↑ F↓ G↓)
- Thr\_basal: + (T~↑ M~↑ F~↓ G↓)
- Pro\_basal: - (T~↓ M~↑ F~↑ G↓)
- Uridine\_basal: + (T~↑ M~↑ F↑ G~↓)
- Riboflavin\_basal: + (T↑ M↑ F↓ G↑)
- Tyr\_basal: - (T↑ M~↓ F~ G~↑)
- Glu\_feed: - (T~↑ M↓ F↑ G↓)

## Individual feature effect direction summary (Feed + Process)



Different metrics to describe the effect of varying the feature on each output CQA





# Scoring of nutrients impact on Titer and Glycan (1 best, 0 worst)

a<sup>★</sup>

Nutrient	Titer		Glyco	
	Basal	Feed	Basal	Feed
Arg	0.50	0.52	0.16	0.10
Asn	0.44	0.26	0.32	0.23
Asp	0.35	0.29	0.85	0.30
Ca	0.14	0.17	0.22	0.45
Co	0.11	0.80	0.24	0.07
Fe	0.07	0.06	0.12	0.08
Folic acid	0.17	0.31	0.56	0.09
Glu	0.24	0.27	0.14	0.09
Mg	0.21	0.26	0.13	0.12
Mn	NA	0.01	NA	0.14
Pro	0.16	0.59	0.09	0.16
Riboflavin	0.03	0.31	0.78	1.00
Ser	0.25	0.26	0.30	0.15
Thr	0.20	0.41	0.49	0.24
Tyr	1.00	NA	0.05	0.00
Uridine	0.01	0.05	0.18	0.32
Zn	0.37	0.32	0.15	0.10

Importance class

1	High
0.9	High
0.8	High
0.7	High
0.6	Medium
0.5	Medium
0.4	Medium
0.3	Medium
0.2	Low
0.1	Low
0	Low

The values here are derived from slides Appendix 24-30 and will be explained in the final report.



# Individual Feature Selection (20) & Direction

NC: no change  
recommended



Feature	Basal					Feed				
	Lbnd	Ubnd	Baseline	Direction	Effect	Lbnd	Ubnd	Baseline	Direction	Effect
Arg	1.72	12.69	5.64	+	T↑ M↓ F↓ G↓	8.8	39.34	36.723	- (NC)	T~↓ M~ F↓ G~
Asn	4.86	8.31	7.01	-	T↓ M↓ F↑ G~↓	17.26	79.06	78.89	-	T↓ M↑ F~ G~↑
Asp	3.16	13.87	3.35	+	T↓ (high) M↑ F↓ G↓	17.066	78.409	42.63	- (NC)	T↓ (low) M↑ F↓ G↓
Folic acid	0.0185	0.054	0.0186	+	T~ M↑ F↓ G↓	0.0137	0.107	0.052	+	T↑ M↑ F↓ G↓
Co	65.25	742.85	75.70	-	T↑ M↓ F↑ G↑	167.16	1315.19	1312.90	-	T↓ M↓ F↑ G↑
Fe	1148	49624	42626	+	T↑ M↓ F~↑ G~↑	87.23	7909.58	299.30	- (NC)	T~↓ M↓ F↑ G↑
Mg	16383	33313	26697	+	T↑ M↑ F↓ G~↑	58187	137329	82531	+	T↑ M↑ F~ G↓
Mn						2.61	110.45	6.21	+	T↑ M↑ F~ G↓
Ca	3937	40748	34071	+ (NC)	T↑ M↓ F↑ G↓	1645	51611	49110	-	T↓ M↓ F↑ G↑
Zn	568	4551	2021	-	T↓ M↓ F↑ G~↓	1527	9641	5410	-	T↓ M↓ F↑ G↑
Ser	4.06	9.37	9.34	-	T↓ M↑ F↓ G↓	39.44	114.30	92.12	-	T↓ M~↑ F↓ G↓
Thr	2.29	4.23	2.36	+	T~↑ M~↑ F~↓ G↓	13.64	40.12	37.64	+	T↑ M~ F~↓ G~↓
Pro	2.12	6.96	6.92	-	T~↓ M~↑ F~↑ G↓	2.29	41.00	39.56	-	T↓ M~↓ F↑ G↑
Uridine	0.0	0.783	0.0807	+	T~↑ M~ F↑ G~↓	0.0	2.48	0.410	+	T~↑ M↓ F↑ G↑
Riboflavin	0.00252	0.00361	0.00257	+ (NC)	T↑ M↑ F↓ G↑	0.00837	0.0177	0.0153	-	T↓ M↓ F↑ G↑
Tyr	1.05	1.634	1.617	-	T↑ M~↓ F~ G~↑					
Glu	0.738	3.058	2.006	-	T~↑ M↓ F↑ G↓	13.33	139.85	59.85	-	T↓ M~↓ F~ G↓
DO						20	60	40	-	T↓ M↓ F↑ G↓
pH						6.9	7.1	7.0	+	T~↓ M↓ F↑ G~
Feed vol						0.149 (4%)	0.419 (10%)	0.244 (6%)	-	T↓ M↓ F↑ G↑

To quantify effect as high, medium or low see slide 8

e.g. Uridine has estimated low affect on titer yet has medium affect on lowering mannosylation if increased on the feed.

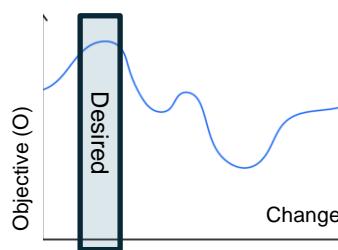
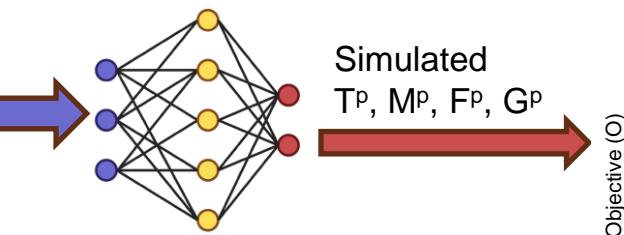
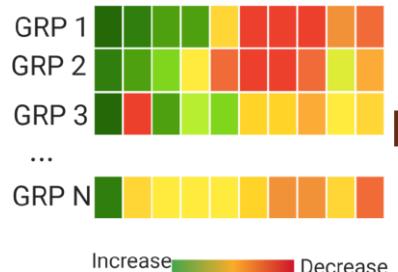
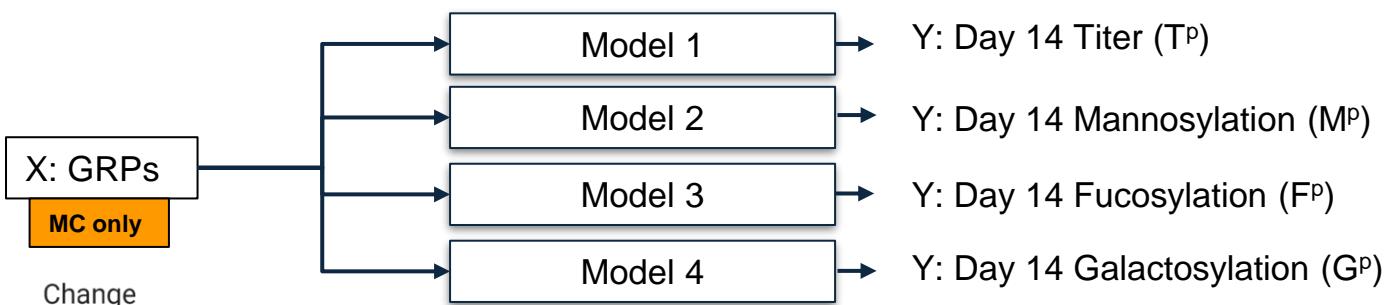
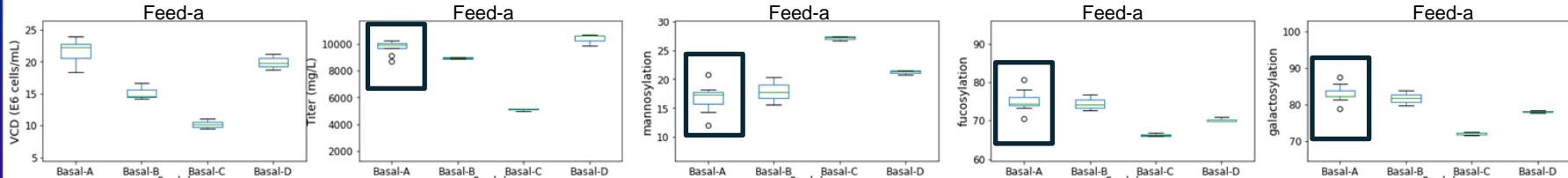
Therefore we should consider increasing uridine in the feed.

# Calculating combinatorial GRP effects



Reference (median):  $T^R$ ,  $M^R$ ,  $F^R$ ,  $G^R$  (e.g. )

Define **reference media** that has good titer and mannosylation that can “improved”: Feed-a, Basal-A (See appendix)

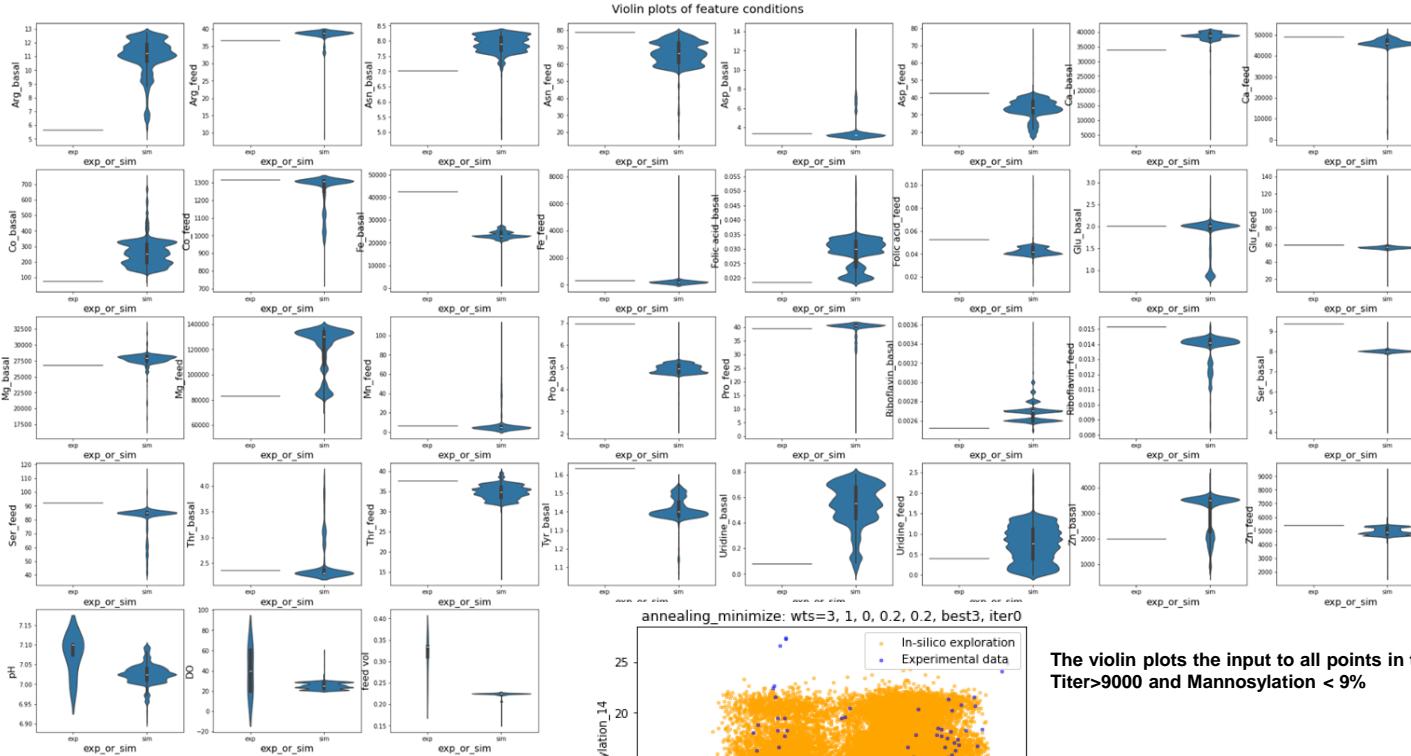


$$O = \alpha(T^p - T^R) + \beta(M^p - M^R) + \gamma(F^p - F^R)^2 + \delta(G^p - G^R)^2$$

# Comparison of optimized feature combinations vs baseline



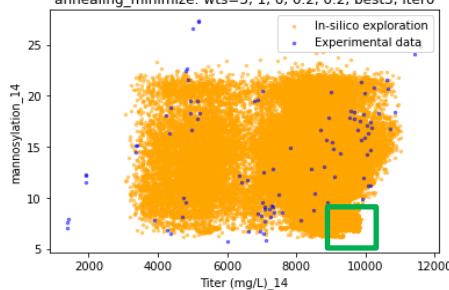
[BASELINE: MODAL CONDITION] Basal-A, Feed-a, pH7, DO40, feed 6% (feed vol 0.235)



Threshold condition for selecting in silico inputs =  
Titer > 9000, Mannosylation < 9% (see

Where the violin plots get “fatter” (median) the model suggests this most times as the point which improves our objective function O

Objective O is searched using simulated annealing



The violin plots the input to all points in the green box i.e.  
Titer > 9000 and Mannosylation < 9%

## SIMULATED ANNEALING

Basal-A, Feed-a, pH7, DO20, feed 8% (feed vol 0.335)  
Starting CQAs:  $T=9891$ ,  $M=9.89$ ,  $F=82.1$ ,  $G=12.6$   
Best CQAs:  $T=10252$ ,  $M=6.61$ ,  $F=85.4$ ,  $G=17.8$



# Combi 1: Feature proposal based on Simulated annealing

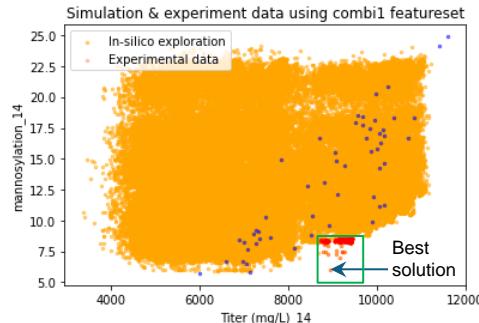


Best solutions input:

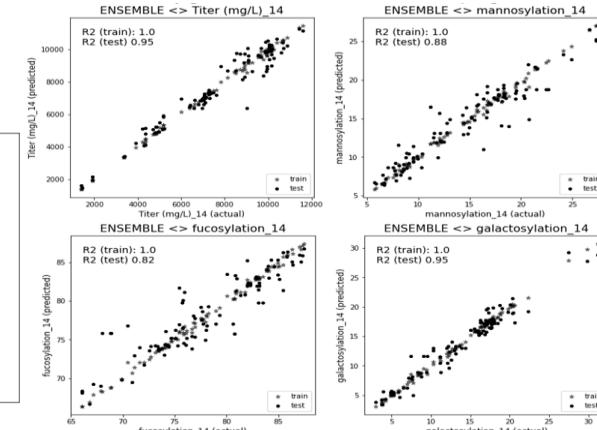
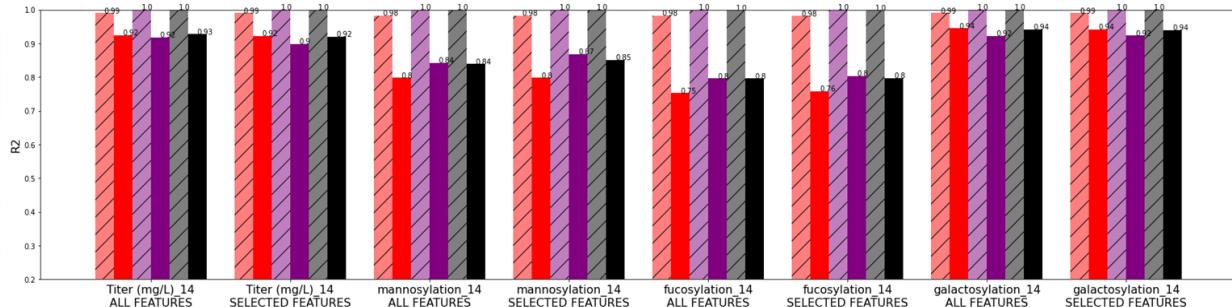
Feature	Basal					Feed				
	Lbnd	Ubnd	Baseline	Direction	Input to model	Lbnd	Ubnd	Baseline	Direction	Input to model
Asp	3.16	13.87	3.35	—	3.17	17.066	78.409	42.63	+	48.07
Asn	4.86	8.31	7.01	+	7.12	17.26	79.06	78.89	—	77.58
Mg	16383	33313	26697	+	28043	58187	137329	82531	—	62350
Ca	3937	40748	34071	+	40359	1645	51611	49110	—	46019
Uridine	0.0	0.783	0.0807	+	0.679	0.0	2.48	0.410	+	1.559
Riboflavin	0.00252	0.00361	0.00257	+	0.0027	0.00837	0.0177	0.0153	—	0.0091
DO						20.0	60.0	40.0	—	22
pH						6.9	7.1	7.0	~	7.0
Feed vol						0.149	0.419	0.244	—	0.154 (4%)

Best solutions output CQA:

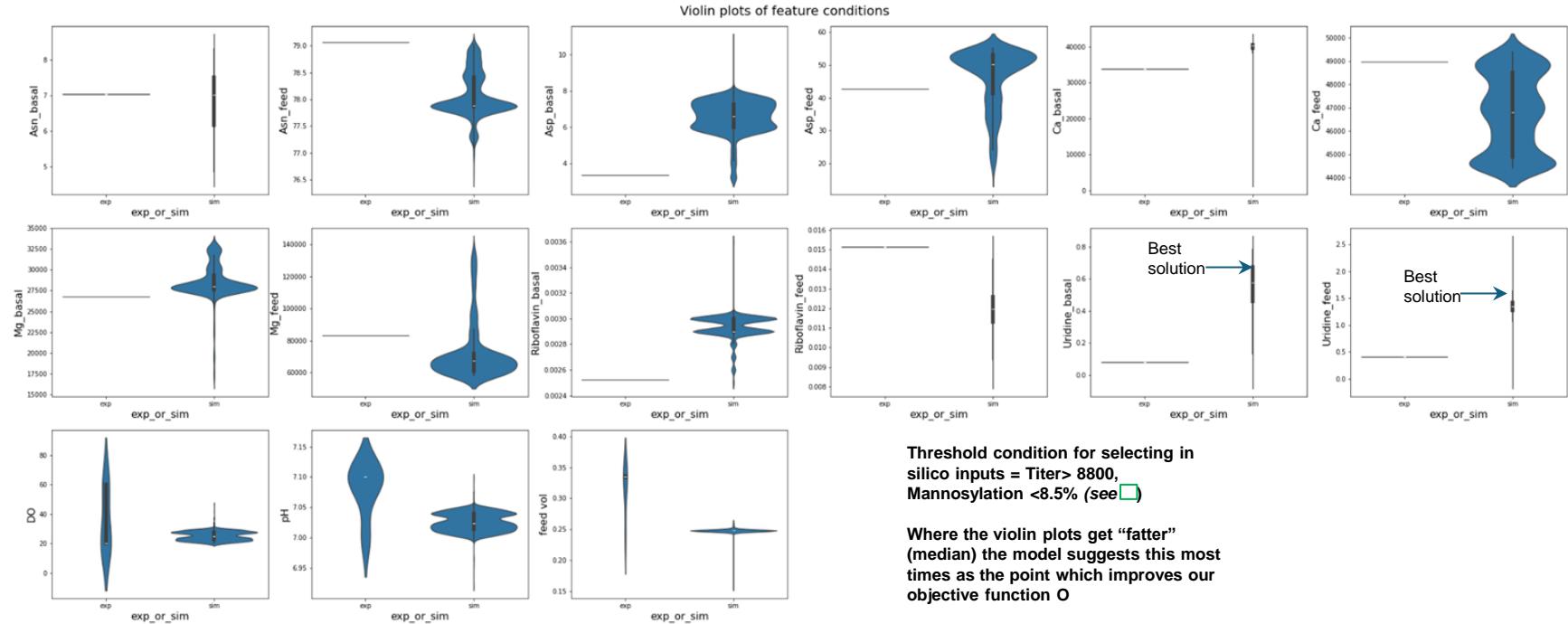
Titer (mg/L)\_14: 8936  
mannosylation\_14: 6.00  
fucosylation\_14: 86.4  
galactosylation\_14: 23.09



- randomforest\_train
- randomforest\_cv
- xgb\_train
- xgb\_cv
- ENSEMBLE\_train
- ENSEMBLE\_cv



# Combi 1: Feature proposal based on Simulated annealing



Threshold condition for selecting in silico inputs = Titer > 8800,  
Mannosylation < 8.5% (see

Where the violin plots get “fatter”  
(median) the model suggests this most times as the point which improves our objective function O

O is searched using simulated annealing

Uridine example shown.

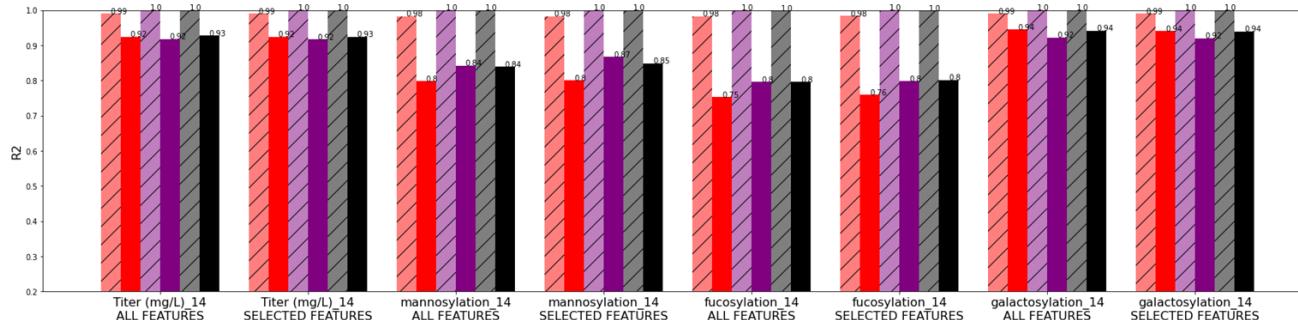
# Combi 2 Selection

Combi 1 and Combi 2 are not so correlated



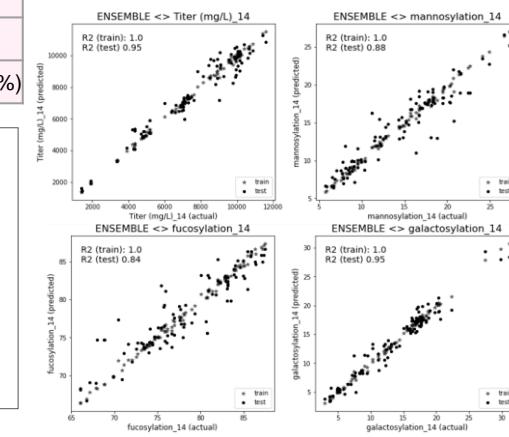
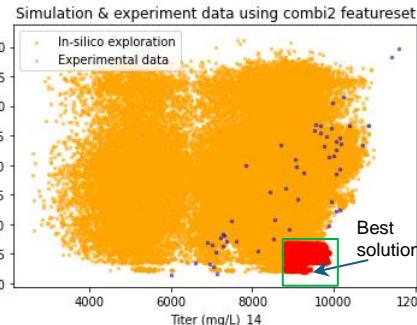
Best solutions input:

Feature	Basal					Feed				
	Lbnd	Ubnd	Baseline	Direction	Input to model	Lbnd	Ubnd	Baseline	Direction	Input to model
Arg	1.72	12.69	5.64	+	8.58	8.8	39.34	36.723	~	36.12
Folic acid	0.0185	0.054	0.0186	+	0.0197	0.0137	0.107	0.052	-	0.0455
Co	65.25	742.85	75.70	+	221	167.16	1315.19	1312.90	-	983
Zn	568	4551	2021	+	2661	1527	9641	5410	-	4551
Thr	2.29	4.23	2.36	-	2.31	13.64	40.12	37.64	-	32.07
Pro	2.12	6.96	6.92	+	4.87	2.29	41.00	39.56	-	30.42
Uridine	0.0	0.783	0.0807	+	0.407	0.0	2.48	0.410	+	0.583
Riboflavin	0.00252	0.00361	0.00257	+	0.0027	0.00837	0.0177	0.0153	-	0.0134
Tyr	1.05	1.634	1.617	-	1.52					
DO						20.0	60.0	40.0	-	28.6
pH						6.9	7.1	7.0	-	6.93
Feed vol						0.149	0.419	0.244	-	0.155 (4%)

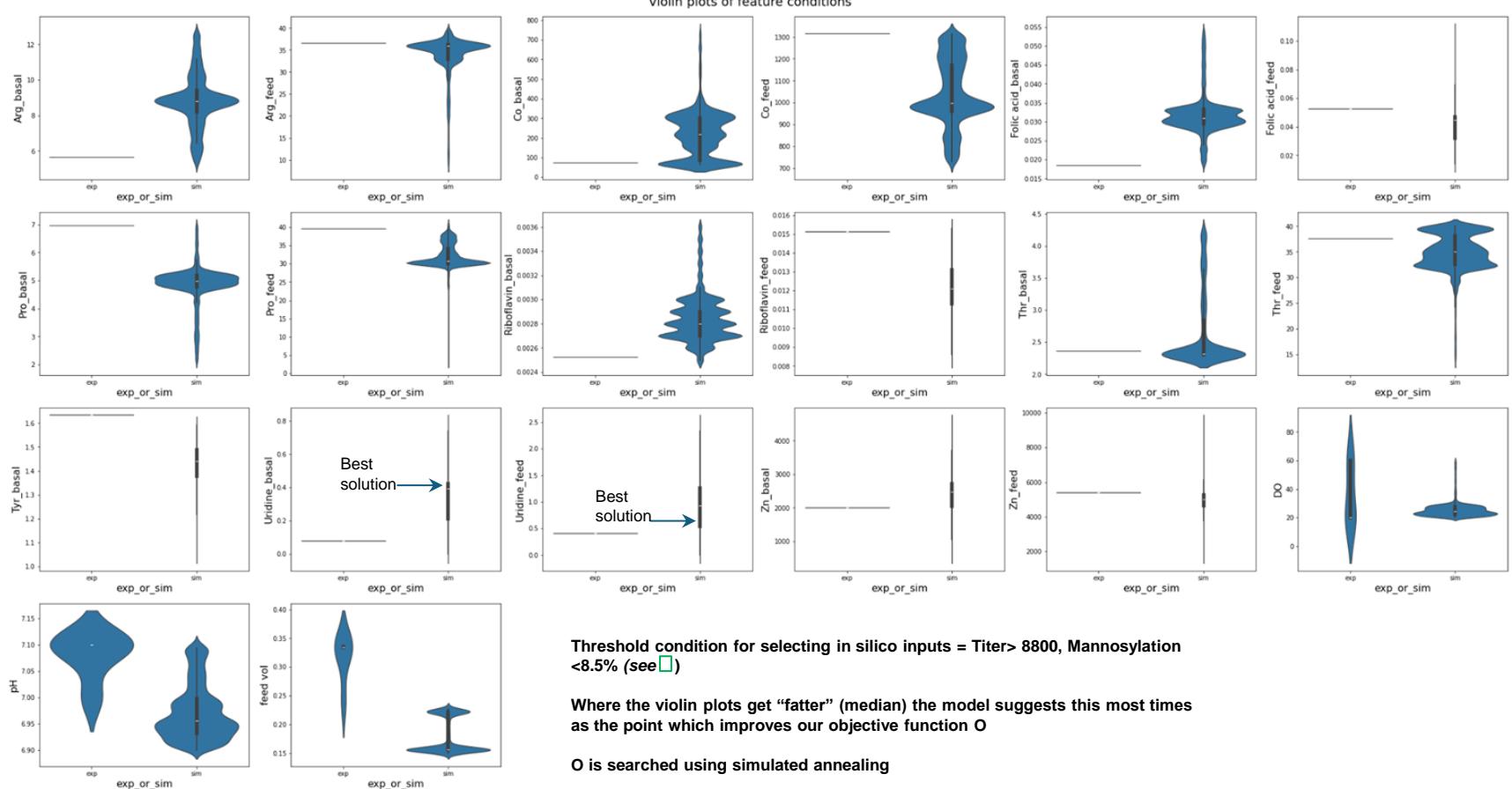


Best solutions output CQA:

Titer (mg/L)\_14: 9326  
Mannosylation\_14: 6.03  
fucosylation\_14: 86.8  
galactosylation\_14: 14.8



# Combi 2: Feature proposal based on Simulated annealing





# APPENDIX



# Responses to Questions

## Page 3 (Model Summary)

- I would like to know the details of missing data (evaluation label, vessel No., nutrient name, reason to drop).
  - From the dataset 240402\_20RP06-19\_data for Astar\_4conditions\_v2, experiments 3,4,5,6 (Basal-A, Feed-b/c) were dropped because Uridine\_feed was not measured in the Media composition data for Feed b/c (see screenshot)

## Page 4 (Summary of work done)

- Could you clarify the meaning of "3) nutrient consumption information" in line 10? Does it mean grouping by "NSRC feature correlations" in Page 5?
  - The nutrient consumption information used was the **Correlation between average Nutrient Specific Rate of Consumption (NSRC) and the final CQA (i.e. Titer, Man5, Fuc, Gal on Day 14), ( $\text{Corr}_{\text{NSRC,CQA14}}$ )**. (NSRC was calculated as outlined on [slide 18](#)). The general idea is that if there is a stronger relationship between NSRC and CQA, it might indicate that the nutrient is more important. However, this was not weighed too heavily as we had to interpolate a lot to get the NSRC data, which might have introduced some errors.
  - $\text{Corr}_{\text{NSRC,CQA14}}$  is not the same as the cross-correlations between the NSRC values. These are solely correlations between the NSRC features, and not the output CQA. So for example if NSRC for Arg and Met seem to be highly correlated, it means that for experimental samples/days when the consumption rate for Arg is high, it tends to be high for Met as well, and vice versa
- I would like to know contribution levels of "feature selection", "domain knowledge" and "nutrient consumption" on each CQAs to select each components. The table Ian suggested in the mtg would be really helpful.
  - [Slides 20-21](#) describe the algorithm for how we used these scores. Essentially, we constructed a composite score using the Model-based scores as a base (with a certain weightage), and boosting them by some percentage using the NSRC scores and Domain Knowledge scores (again with a certain weightages  $w_1, w_2, w_3, w_4$  as described in the slides). We then rank them and choose 6-10 top features per CQA with less than 2 features in the selected set having correlation  $>0.85$ . We ran this algorithm using many different weightages for the various components (i.e. a Gridsearch).
  - This resulted in ~5000 unique feature set combinations. We filtered for those that preserved model performance (< 0.2% decrease in R<sup>2</sup>), then averaged the resulting selections. This table of feature importances according to the model can be found on [Slide 22](#) (1st 4 rows), along with the domain knowledge rankings.
  - The final selection of 20 features relied on consideration of this table, together with the feature correlations (see [Slide 23](#)) – we want to select some highly correlated features (since it's hard for the model to tell them apart, but also a sufficiently diverse feature set according to domain knowledge
  - A new model was then trained on the selected subset. The column "contrib%" on Slide 24-25 indicates the average importance of each feature for the prediction
- I know that I do not fully understand the two points of Note. Could you elaborate them?
  - See Slide 19 for elaboration on the impact of severe multicollinearity in the dataset on the analysis. In summary, the models can still predict well, but the estimates of feature importance (e.g. the coefficients fitted) end up having high errors, making it hard to do causal discovery using just the dataset. This is because the models finds it hard to distinguish between highly correlated features; some types of models may even arbitrarily drop features if they seem 'redundant' to the prediction – we have tried to avoid models that tend to do this.
  - See Q1 response for the point on missing data resulting in samples / features being dropped

exp_label	lavin_feed	Biotin_feed	Ethanolamine_fe	Uridine_feed	Na_feed	Mg_fe
0.1	66737513	0.01473598035	5.151	<b>0.4095004095</b>	5059221.4	
0.2	66737513	0.01473598035	5.151	<b>0.4095004095</b>	5059221.4	
0.3	394792774	0.01289398281			2894069.46	
0.4	394792774	0.01289398281			2894069.46	
0.5	394792774	0.01289398281			4599954.816	
0.6	394792774	0.01289398281			4599954.816	
0.7	275374641	0.01187016495	1.043	0	6134567.554	111
0.8	275374641	0.01187016495	1.043	0	6134567.554	111
1.1	327792539	0.02926609635	0	0	6268071.72	1
1.2	327792539	0.02926609635	0	0	6268071.72	1

# Responses to Questions



## Page 9 (Feature effect direction summary (BASAL))

- Could you please clarify the meaning of “optval (does it mean the equation in Page6?)”, “sim\_vs”, “slope”?
  - Optval\_titer/man5\_vs\_baseline: Using the models trained on the feature subset, we can predict the effect of varying individual features (12 points across the full range). Optval refers to the % change in the *feature value* from baseline that gives maximum predicted titer and minimum mannosylation – we use it more as an indication of direction, rather than absolute magnitude (which would be more uncertain)
  - Slope indiv variation mean Titer/Man5: This is the mean slope of the CQA prediction as the feature is varied from the lower to upper bound value
  - Sim\_vs\_expbaseline\_mean/median: When we run simulated annealing (allowing multiple features to vary at the same time) to maximize Titer, minimize Man5, and keep Fuc and Gal the same, we generate thousands of simulated experiments'. We then look at the best results (i.e. those in bottom right corner of the ‘Titer vs Mannosylation’ scatter plot as highlighted in the green boxes on Slides 11 & 13), and compare the *mean value each feature for those optimized samples* VS the baseline Basal-A/Feed-a. That's what this value refers to.
  - **Note:** For determining feature direction, we would also want to consider SHAP\_randomforest/xgb\_pearsonr/spearmanr\_titer/man5. This is the correlation of the feature values to their respective SHAP values. If correlation is positive, a higher value of the feature tends to increase the output prediction. → **this might be a better metric to use** than Optval\_titer/man5\_vs\_baseline or sim\_vs\_expbaseline\_mean/median, if they don't all agree. The average value for this metric is shown in the summary tables on [Slides 26-27](#) under the column *shap\_corr*

## Page 10 (Individual feature selection and direction)

- To estimate best concentrations or setting points for each items, is it possible to indicate either estimation profile such as a figure below or each scatter plots of actual data between CQA (e.g mannosylation%) and individual data (e.g Arg).
  - Please see Appendix slides 26-27 for more detailed information in the table, and Slides 28-31 for scatter plots forecasting the effect of varying individual features, shown together with all the experimental data used. Note that these should be taken only as an approximate estimate of direction and effect size
- I would like to know the reason of “no change recommended”.
  - When the direction proposed for a certain nutrient in Basal or Feed are opposite, or when the predicted effects are opposite, we would recommend to change only 1 of Basal or Feed (not both), so that the effects don't cancel out and make interpretation more difficult. We would propose to use the feature that seems more likely to result in an overall positive change (e.g. predicted increase in titer along with decrease in mannosylation). If there is a tradeoff between titer and mannosylation, we would propose to drop the feature that seems less important to the models overall (e.g. looking at the magnitude of effect and the contribution% of that feature).

## Page 11 (combi 1)

- I am wondering that decreasing the amount of Feed vol would be the same as reducing the concentration of all components (including the components not selected in the model) at once. I would like to confirm if it reflects your consideration correctly.
  - This does seem intuitive from a reasoning point of view. However, this is not something we tested with the data/modeling, so we can't say for sure what the effect of x% decrease in feed vol will be vs. x% decrease in concentration of components.

# Calculating Nutrient Specific Rate of Change (NSRC)

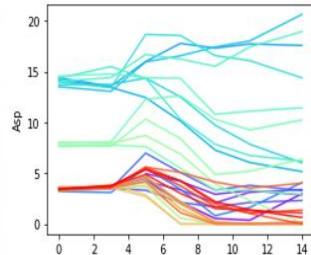
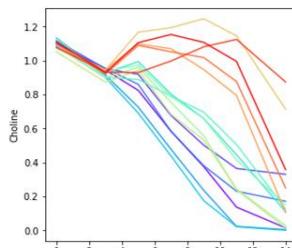
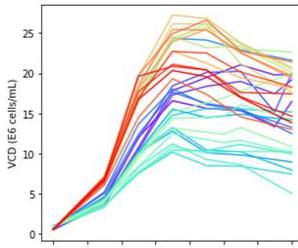
- We only have data *before* each feed, every 2 days → doesn't capture the change in concentration after feeds ('undersampled')
- Accurate data *before* and *after* each feed is needed to calculate NSRC
- ⇒ Estimate nutrient concentrations after each feed using sampling, media composition and volume data

## Post-feed Nutrient Concentration: $I(t,x)$

Nutrient Mass (Post-Feed) = Nutrient Mass (Pre-Feed) + Nutrient Mass (Feed)

⇒ Nutrient Mass = Nutrient Conc \* VesselVol

$$I(t,x) * \text{VesselVol}(t-1) = I(t-1,x) * \text{VesselVol}(t-1) + I(\text{feed}) * \text{FeedVol}(t-1,x)$$



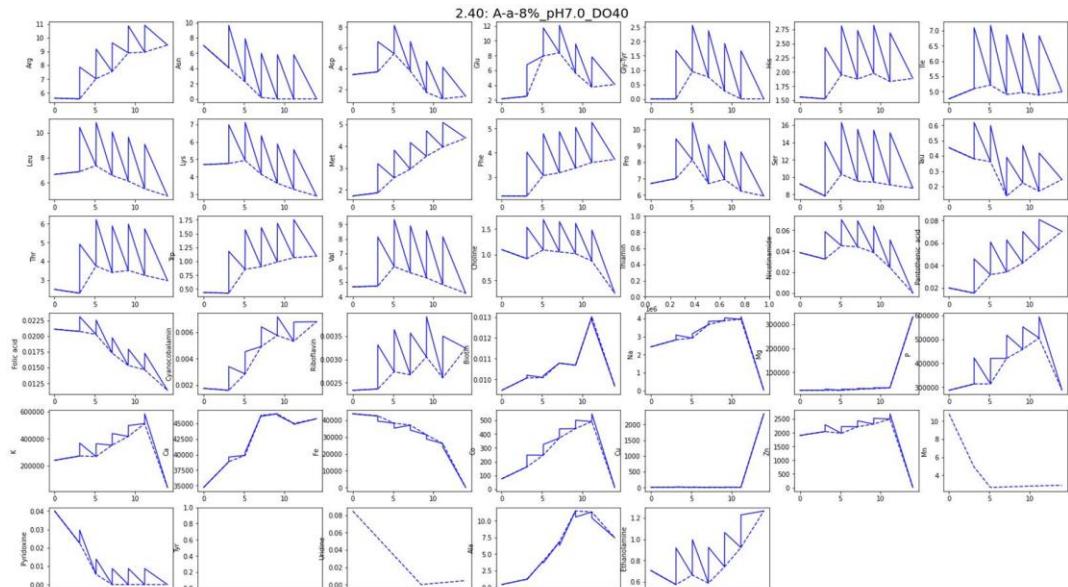
**Calculate the average NSRC from Days 5-12**

## Nutrient Specific Rate of Change: NSRC

**NSRC** =  $\Delta$  in Nutrient Conc. /  $\Delta$  in Integrated VCD (IVCD)

$\Delta$  in Nutrient Conc. =  $I(t,x) - I(t-1,x)$

$\Delta$  in IVCD = IVCD(t) - IVCD(t-1)





# A note on multicollinearity



**Perfect multicollinearity** refers to a situation where the **predictive variables** have an *exact* linear relationship. When there is perfect collinearity, the **design matrix**  $X$  has less than full **rank**, and therefore the **moment matrix**  $X^T X$  cannot be **inverted**. In this situation, the **parameter estimates** of the regression are not well-defined, as the system of equations has **infinitely** many solutions.

**Imperfect multicollinearity** refers to a situation where the **predictive variables** have a *nearly* exact linear relationship.

There is no justification for the practice of removing collinear variables as part of regression analysis,<sup>[1][4][5][6][7]</sup> and doing so may constitute **scientific misconduct**. Including collinear variables does not reduce the predictive power or **reliability** of the model as a whole,<sup>[6]</sup> and does not reduce the accuracy of coefficient estimates.<sup>[1]</sup>

High collinearity indicates that it is exceptionally important to include all collinear variables, as excluding any will cause worse coefficient estimates, strong **confounding**, and downward-biased estimates of **standard errors**.<sup>[2]</sup>

To address the high collinearity of a dataset, **variance inflation factor** can be used to identify the collinearity of the predictor variables.

## Effects on coefficient estimates

In addition to causing numerical problems, **imperfect multicollinearity makes precise estimation of variables difficult**. In other words, **highly correlated variables lead to poor estimates and large standard errors**.

As an example, say that we notice Alice wears her boots whenever it is raining and that there are only puddles when it rains. Then, we cannot tell whether she wears boots to keep the rain from landing on her feet, or to keep her feet dry if she steps in a puddle.

The problem with trying to identify how much each of the two variables matters is that they are **confounded** with each other: our observations are explained equally well by either variable, so we do not know which one of them causes the observed correlations.

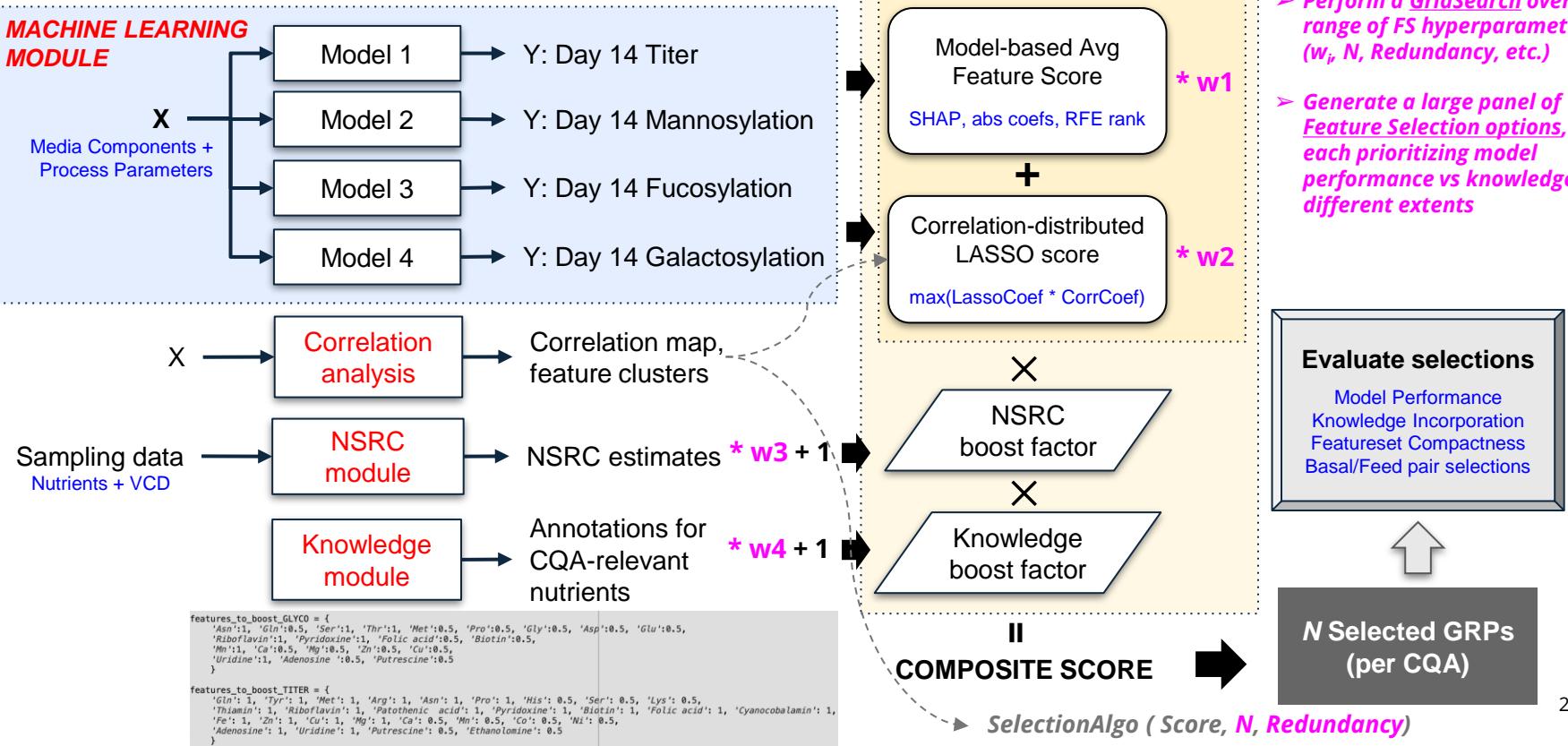
There are two ways to discover this information:

1. **Using prior information or theory.** For example, if we notice Alice never steps in puddles, we can reasonably argue puddles are not why she wears boots, as she does not need the boots to avoid puddles.
2. Collecting more data. If we observe Alice enough times, we will eventually see her on days where there are puddles but not rain (e.g. because the rain stops before she leaves home).

This confounding becomes substantially worse when researchers attempt to ignore or suppress it by excluding these variables from the regression (see **#Misuse**). Excluding multicollinear variables from regressions will invalidate **causal inference** and produce worse estimates by removing important confounders.

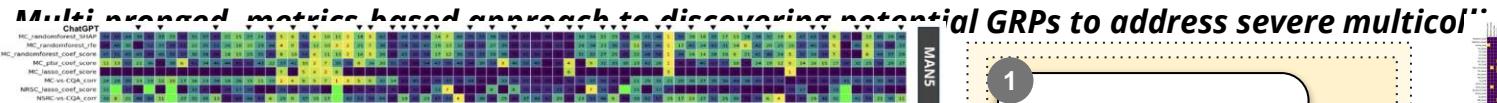
# Modeling + Feature Selection (FS) Modules

*Multi-pronged, metrics-based approach to discovering potential GRPs to address severe multicollinearity in the data*

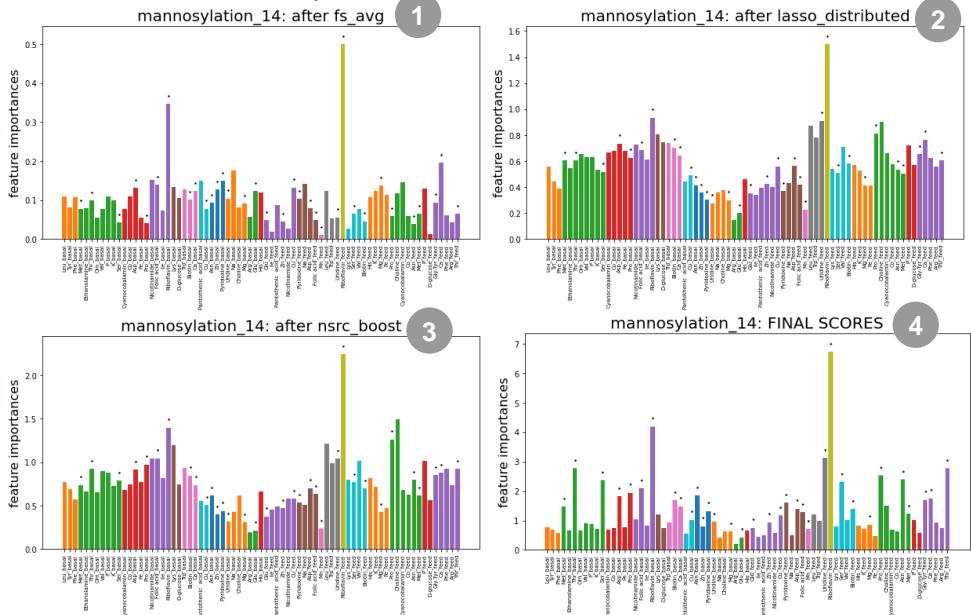




# Modeling + Feature Selection (FS) Modules

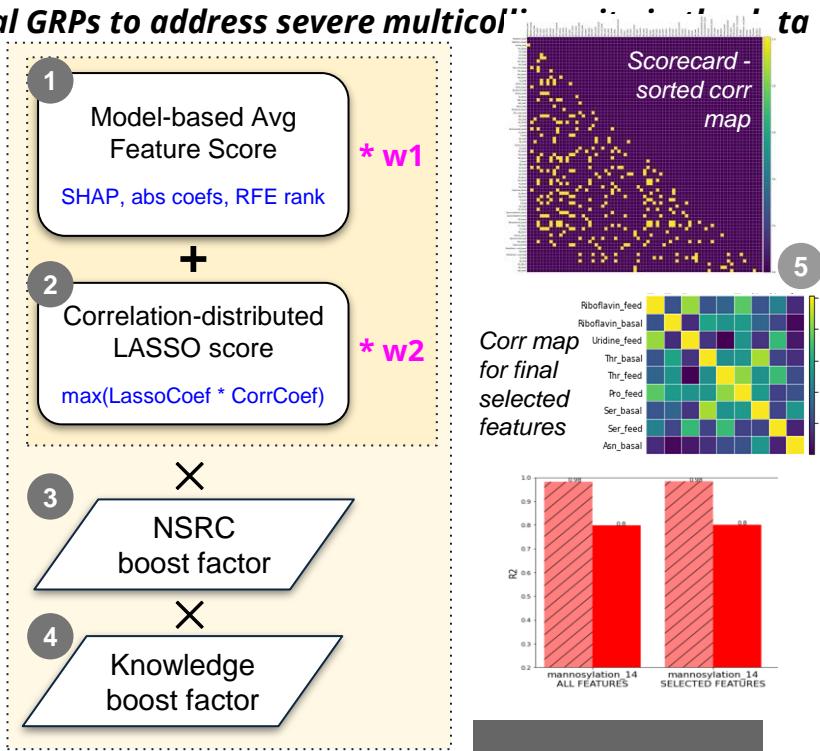


Scorecard after each step



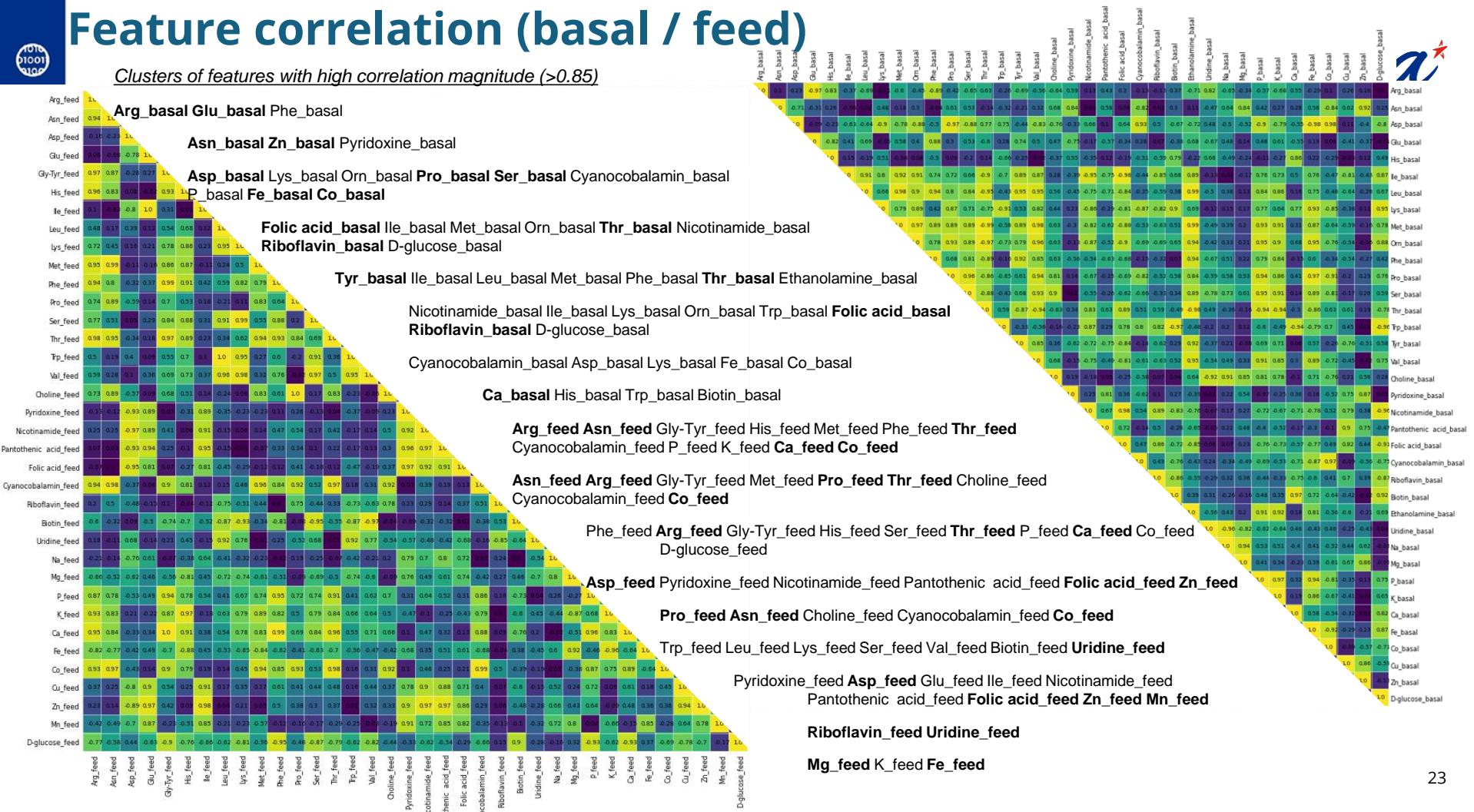
E.g.  
FS settings

```
'fs_avg_weight': 0.5,
'lasso_corr_scaling_power': 1,
'lasso_distributed_scores_weight': 1,
'nsr_boosting_factor': 0.66,
'knowledge_scaling_factor': 2,
'corr_thres': 0.8,
'num_features_to_select': 9,
'max_num_highcorr_features': 2
```



# Feature correlation (basal / feed)

Clusters of features with high correlation magnitude (>0.85)



To be explained in final report

# Scoring of features for Titer and Glycan (justifications)



model_feature_importance_titer	73	46	51	39	41	42	33	38	31	40	18	21	43	44	45	4	47	48	49	50	13	53	19	64	72	71	70	69	68	67	66	65	64	63	62	61	60	59	58	57	56	55	54	53	52																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																	
model_feature_importance_man5	73	18	6	43	41	38	36	11	39	40	44	22	23	4	29	35	46	37	18	7	20	4	41	1	15	50	51	27	25	24	23	22	21	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1	0	-1	-2	-3	-4	-5	-6	-7	-8	-9	-10	-11	-12	-13	-14	-15	-16	-17	-18	-19	-20	-21	-22	-23	-24	-25	-26	-27	-28	-29	-30	-31	-32	-33	-34	-35	-36	-37	-38	-39	-40	-41	-42	-43	-44	-45	-46	-47	-48	-49	-50	-51	-52	-53	-54	-55	-56	-57	-58	-59	-60	-61	-62	-63	-64	-65	-66	-67	-68	-69	-70	-71	-72	-73	-74	-75	-76	-77	-78	-79	-80	-81	-82	-83	-84	-85	-86	-87	-88	-89	-90	-91	-92	-93	-94	-95	-96	-97	-98	-99	-100	-101	-102	-103	-104	-105	-106	-107	-108	-109	-110	-111	-112	-113	-114	-115	-116	-117	-118	-119	-120	-121	-122	-123	-124	-125	-126	-127	-128	-129	-130	-131	-132	-133	-134	-135	-136	-137	-138	-139	-140	-141	-142	-143	-144	-145	-146	-147	-148	-149	-150	-151	-152	-153	-154	-155	-156	-157	-158	-159	-160	-161	-162	-163	-164	-165	-166	-167	-168	-169	-170	-171	-172	-173	-174	-175	-176	-177	-178	-179	-180	-181	-182	-183	-184	-185	-186	-187	-188	-189	-190	-191	-192	-193	-194	-195	-196	-197	-198	-199	-200	-201	-202	-203	-204	-205	-206	-207	-208	-209	-210	-211	-212	-213	-214	-215	-216	-217	-218	-219	-220	-221	-222	-223	-224	-225	-226	-227	-228	-229	-230	-231	-232	-233	-234	-235	-236	-237	-238	-239	-240	-241	-242	-243	-244	-245	-246	-247	-248	-249	-250	-251	-252	-253	-254	-255	-256	-257	-258	-259	-260	-261	-262	-263	-264	-265	-266	-267	-268	-269	-270	-271	-272	-273	-274	-275	-276	-277	-278	-279	-280	-281	-282	-283	-284	-285	-286	-287	-288	-289	-290	-291	-292	-293	-294	-295	-296	-297	-298	-299	-300	-301	-302	-303	-304	-305	-306	-307	-308	-309	-310	-311	-312	-313	-314	-315	-316	-317	-318	-319	-320	-321	-322	-323	-324	-325	-326	-327	-328	-329	-330	-331	-332	-333	-334	-335	-336	-337	-338	-339	-340	-341	-342	-343	-344	-345	-346	-347	-348	-349	-350	-351	-352	-353	-354	-355	-356	-357	-358	-359	-360	-361	-362	-363	-364	-365	-366	-367	-368	-369	-370	-371	-372	-373	-374	-375	-376	-377	-378	-379	-380	-381	-382	-383	-384	-385	-386	-387	-388	-389	-390	-391	-392	-393	-394	-395	-396	-397	-398	-399	-400	-401	-402	-403	-404	-405	-406	-407	-408	-409	-410	-411	-412	-413	-414	-415	-416	-417	-418	-419	-420	-421	-422	-423	-424	-425	-426	-427	-428	-429	-430	-431	-432	-433	-434	-435	-436	-437	-438	-439	-440	-441	-442	-443	-444	-445	-446	-447	-448	-449	-450	-451	-452	-453	-454	-455	-456	-457	-458	-459	-460	-461	-462	-463	-464	-465	-466	-467	-468	-469	-470	-471	-472	-473	-474	-475	-476	-477	-478	-479	-480	-481	-482	-483	-484	-485	-486	-487	-488	-489	-490	-491	-492	-493	-494	-495	-496	-497	-498	-499	-500	-501	-502	-503	-504	-505	-506	-507	-508	-509	-510	-511	-512	-513	-514	-515	-516	-517	-518	-519	-520	-521	-522	-523	-524	-525	-526	-527	-528	-529	-530	-531	-532	-533	-534	-535	-536	-537	-538	-539	-540	-541	-542	-543	-544	-545	-546	-547	-548	-549	-550	-551	-552	-553	-554	-555	-556	-557	-558	-559	-560	-561	-562	-563	-564	-565	-566	-567	-568	-569	-570	-571	-572	-573	-574	-575	-576	-577	-578	-579	-580	-581	-582	-583	-584	-585	-586	-587	-588	-589	-590	-591	-592	-593	-594	-595	-596	-597	-598	-599	-600	-601	-602	-603	-604	-605	-606	-607	-608	-609	-610	-611	-612	-613	-614	-615	-616	-617	-618	-619	-620	-621	-622	-623	-624	-625	-626	-627	-628	-629	-630	-631	-632	-633	-634	-635	-636	-637	-638	-639	-640	-641	-642	-643	-644	-645	-646	-647	-648	-649	-650	-651	-652	-653	-654	-655	-656	-657	-658	-659	-660	-661	-662	-663	-664	-665	-666	-667	-668	-669	-670	-671	-672	-673	-674	-675	-676	-677	-678	-679	-680	-681	-682	-683	-684	-685	-686	-687	-688	-689	-690	-691	-692	-693	-694	-695	-696	-697	-698	-699	-700	-701	-702	-703	-704	-705	-706	-707	-708	-709	-7010	-7011	-7012	-7013	-7014	-7015	-7016	-7017	-7018	-7019	-7020	-7021	-7022	-7023	-7024	-7025	-7026	-7027	-7028	-7029	-7030	-7031	-7032	-7033	-7034	-7035	-7036	-7037	-7038	-7039	-7040	-7041	-7042	-7043	-7044	-7045	-7046	-7047	-7048	-7049	-7050	-7051	-7052	-7053	-7054	-7055	-7056	-7057	-7058	-7059	-7060	-7061	-7062	-7063	-7064	-7065	-7066	-7067	-7068	-7069	-7070	-7071	-7072	-7073	-7074	-7075	-7076	-7077	-7078	-7079	-7080	-7081	-7082	-7083	-7084	-7085	-7086	-7087	-7088	-7089	-7090	-7091	-7092	-7093	-7094	-7095	-7096	-7097	-7098	-7099	-70100	-70101	-70102	-70103	-70104	-70105	-70106	-70107	-70108	-70109	-70110	-70111	-70112	-70113	-70114	-70115	-70116	-70117	-70118	-70119	-70120	-70121	-70122	-70123	-70124	-70125	-70126	-70127	-70128	-70129	-70130	-70131	-70132	-70133	-70134	-70135	-70136	-70137	-70138	-70139	-70140	-70141	-70142	-70143	-70144	-70145	-70146	-70147	-70148	-70149	-70150	-70151	-70152	-70153	-70154	-70155	-70156	-70157	-70158	-70159	-70160	-70161	-70162	-70163	-70164	-70165	-70166	-70167	-70168	-70169	-70170	-70171	-70172	-70173	-70174	-70175	-70176	-70177	-70178	-70179	-70180	-70181	-70182	-70183	-70184	-70185	-70186	-70187	-70188	-70189	-70190	-70191	-70192	-70193	-70194	-70195	-70196	-70197	-70198	-70199	-70200	-70201	-70202	-70203	-70204	-70205	-70206	-70207	-70208	-70209	-70210	-70211	-70212	-70213	-70214	-70215	-70216	-70217	-70218	-70219	-70220	-70221	-70222	-70223	-70224	-70225	-70226	-70227	-70228	-70229	-70230	-70231	-70232	-70233	-70234	-70235	-70236	-70237	-70238	-70239	-70240	-70241	-70242	-70243	-70244	-70245	-70246	-70247	-70248	-70249	-70250	-70251	-70252	-70253	-70254	-70255	-70256	-70257	-70258	-70259	-70260	-70261	-70262	-70263	-70264	-70265	-70266	-70267	-70268	-70269	-70270	-70271	-70272	-70273	-70274	-70275	-70276	-70277	-70278	-70279	-70280	-70281	-70282	-70283	-70284	-70285	-70286	-70287	-70288	-70289	-70290	-70291	-70292	-70293	-70294	-70295	-70296	-70297	-70298	-70299	-70300	-70301	-70302	-70303	-70304	-70305	-70306	-70307	-70308	-70309	-70310	-70311	-70312	-70313	-70314	-70315	-70316	-70317	-70318	-70319	-70320	-70321	-70322	-70323	-70324	-70325	-70326	-70327	-70328	-70329	-70330	-70331	-70332	-70333	-70334	-70335	-70336	-70337	-70338	-70339	-70340	-70341	-70342	-70343	-70344	-70345	-70346	-70347	-70348	-70349	-70350	-70351	-70352	-70353	-70354	-70355	-70356	-70357	-70358	-70359	-70360	-70361	-70362	-70363	-70364	-70365	-70366	-70367	-70368	-70369	-70370	-70371	-70372	-70373	-70374	-70375	-70376	-70377	-70378	-70379	-70380	-70381	-70382	-70383	-70384	-70385	-70386	-70387	-70388	-70389	-70390	-70391	-70392	-70393	-70394	-70395	-70396	-70397	-70398	-70399	-70400	-70401	-70402	-70403	-70404	-70405	-70406	-70407	-70408	-70409	-70410	-70411	-70412	-70413	-70414	-70415	-70416	-70417	-70418	-70419	-70420	-70421	-70422	-70423	-70424	-70425	-70426	-70427	-70428	-70429	-70430	-70431	-70432	-70433	-70434	-70435	-70436	-70437	-70438	-70439	-70440	-70441	-70442	-70443	-70444	-70445	-70446	-70447	-70448	-70449	-70450	-70451	-70452	-70453	-70454	-70455	-70456	-70457	-70458	-70459	-70460	-70461	-70462	-70463	-70464	-70465	-70466	-70467	-70468	-70469	-70470	-70471	-70472	-70473	-70474	-70475	-70476	-70477	-70478	-70479	-70480	-70481	-70482	-70483	-70484	-70485	-70486	-70487	-70488	-70489	-70490	-70491	-70492	-70493	-70494	-70495	-70496	-70497	-70498	-70499	-70500	-70501	-70502	-70503	-70504	-70505	-70506	-70507	-70508	-70509	-70510	-70511	-70512	-70513	-70514	-70515	-70516	-70517	-70518	-70519	-70520	-70521	-70522	-70523	-70524	-70525	-70526	-70527	-70528	-70529	-70530	-70531	-70532	-70533	-70534	-70535	-70536	-70537	-70538	-70539	-70540	-70541	-70542	-70543	-70544	-70545	-70546	-70547	-70548	-70549	-70550	-70551	-70552	-70553	-70554	-70555	-70556

# Individual feature effects and contributions



CREATING GROWTH, ENHANCING LIVES

base	yvar	BASAL							FEED / PROCESS							
		x_min	x_baseline	x_max	dy% (lbd)	dy% (ubd)	shap_corr	contrib%	x_min	x_baseline	x_max	dy% (lbd)	dy% (ubd)	shap_corr	contrib%	
Arg	Titer	1.72	5.54	12.69	-2.20	1.10	0.35	2.94	8.80	36.40	39.34	-12.59	0.80	0.20	5.73	
		4.86	6.97	8.31	-0.56	0.38	0.69	2.29	17.26	78.19	79.06	-0.76	0.19	0.35	0.53	
		3.16	3.30	13.87	0.19	0.09	-0.50	1.11	17.07	42.20	78.41	-0.36	0.07	-0.74	0.37	
		3,937.48	33,694.93	40,747.69	0.39	-0.08	0.13	0.18	1,645.32	49,215.80	51,610.66	-0.79	0.37	0.13	0.66	
		65.25	72.26	742.85	0.14	-0.51	-0.42	0.67	167.16	1,304.74	1,315.19	-40.44	0.19	0.85	36.65	
		1,148.06	41,916.73	49,623.98	-0.59	0.15	0.43	0.70	87.23	270.14	7,909.58	0.10	-1.27	0.09	0.58	
		0.02	0.02	0.05	0.19	0.35	0.01	0.19	0.01	0.05	0.11	-1.01	-0.37	0.54	1.32	
		0.74	2.01	3.06	0.03	-0.27	-0.16	0.28	13.33	58.69	139.86	-2.01	0.15	0.22	1.69	
		16,383.40	26,154.50	33,313.30	0.04	3.37	0.45	0.94	58,187.38	83,555.87	137,329.18	0.92	-0.61	0.27	0.27	
		Mn							2.61	5.73	110.45	-0.99	-0.15	0.16	0.92	
		Pro	2.12	6.69	6.96	-0.03	0.14	0.28	0.29	2.29	38.84	41.00	-5.89	0.39	0.45	6.60
		Riboflavin	0.00	0.00	0.00	0.19	1.24	0.40	0.29	0.01	0.02	0.02	-0.45	0.50	0.43	0.40
		Ser	4.06	9.21	9.37	-0.04	0.19	0.38	0.36	39.44	91.37	114.30	0.12	0.62	-0.16	0.38
		Thr	2.29	2.34	4.23	0.09	0.82	-0.15	0.17	13.64	37.29	40.12	-6.26	0.40	0.41	5.88
		Tyr	1.05	1.59	1.63	0.21	0.00	-0.64	4.99							
		Uridine	0.00	0.08	0.78	0.49	0.15	0.14	0.07	0.00	0.41	2.48	-0.47	0.19	-0.18	0.53
		Zn	567.83	1,990.24	4,550.88	0.02	0.87	0.39	0.31	1,526.96	5,395.97	9,640.92	-0.62	-1.12	0.09	0.79
		DO							20.00	40.00	60.00	-4.19	-0.63	0.54	3.21	
		feed vol							0.15	0.24	0.42	-12.36	6.35	0.71	16.78	
		pH							6.90	7.00	7.10	-0.04	-1.43	-0.52	0.94	
Arg	Man5	1.72	5.54	12.69	-1.46	2.31	-0.26	2.93	8.80	36.40	39.34	-3.00	0.69	-0.03	2.05	
		Asn	4.86	6.97	8.31	-2.78	0.69	0.58	1.10	17.26	78.19	79.06	0.52	0.69	-0.29	0.21
		Asp	3.16	3.30	13.87	-14.00	0.15	0.55	3.62	17.07	42.20	78.41	0.61	0.15	-0.70	1.55
		Ca	3,937.48	33,694.93	40,747.69	0.60	0.28	-0.50	2.75	1,645.32	49,215.80	51,610.66	-0.76	1.12	0.40	3.34
		Co	65.25	72.26	742.85	-0.86	0.71	0.23	0.55	167.16	1,304.74	1,315.19	0.59	0.69	0.81	0.30
		Fe	1,148.06	41,916.73	49,623.98	0.46	0.71	-0.36	0.40	87.23	270.14	7,909.58	0.68	0.35	0.25	0.12
		Folic acid	0.02	0.02	0.05	0.95	1.56	0.71	1.80	0.01	0.05	0.11	-0.09	0.68	0.43	0.27
		Glu	0.74	2.01	3.06	0.64	0.74	0.47	1.14	13.33	58.69	139.86	0.45	0.23	0.29	0.10
		Mg	16,383.40	26,154.50	33,313.30	0.44	1.07	0.39	0.31	58,187.38	83,555.87	137,329.18	0.36	0.70	0.45	0.52
		Mn							2.61	5.73	110.45	0.48	0.69	0.41	0.17	
		Pro	2.12	6.69	6.96	0.83	0.69	-0.25	0.12	2.29	38.84	41.00	-0.30	0.69	0.44	0.36
		Riboflavin	0.00	0.00	0.00	0.69	28.83	0.67	10.92	0.01	0.02	0.02	2.03	2.01	0.68	26.22
		Ser	4.06	9.21	9.37	0.45	0.69	-0.23	0.19	39.44	91.37	114.30	0.41	0.69	-0.34	0.21
		Thr	2.29	2.34	4.23	-0.06	2.59	0.24	0.56	13.64	37.29	40.12	0.29	0.69	-0.05	0.15
		Tyr	1.05	1.59	1.63	1.06	0.69	0.35	0.25							
		Uridine	0.00	0.08	0.78	1.07	1.25	-0.13	0.26	0.00	0.41	2.48	0.71	0.55	-0.36	0.18
		Zn	567.83	1,990.24	4,550.88	0.51	1.04	0.29	0.44	1,526.96	5,395.97	9,640.92	0.37	0.50	0.36	0.18
		DO							20.00	40.00	60.00	-8.70	0.91	0.76	8.59	
		feed vol							0.15	0.24	0.42	4.88	-3.85	0.54	24.79	
		pH							6.90	7.00	7.10	1.81	-1.18	-0.82	3.34	

To be explained in final report

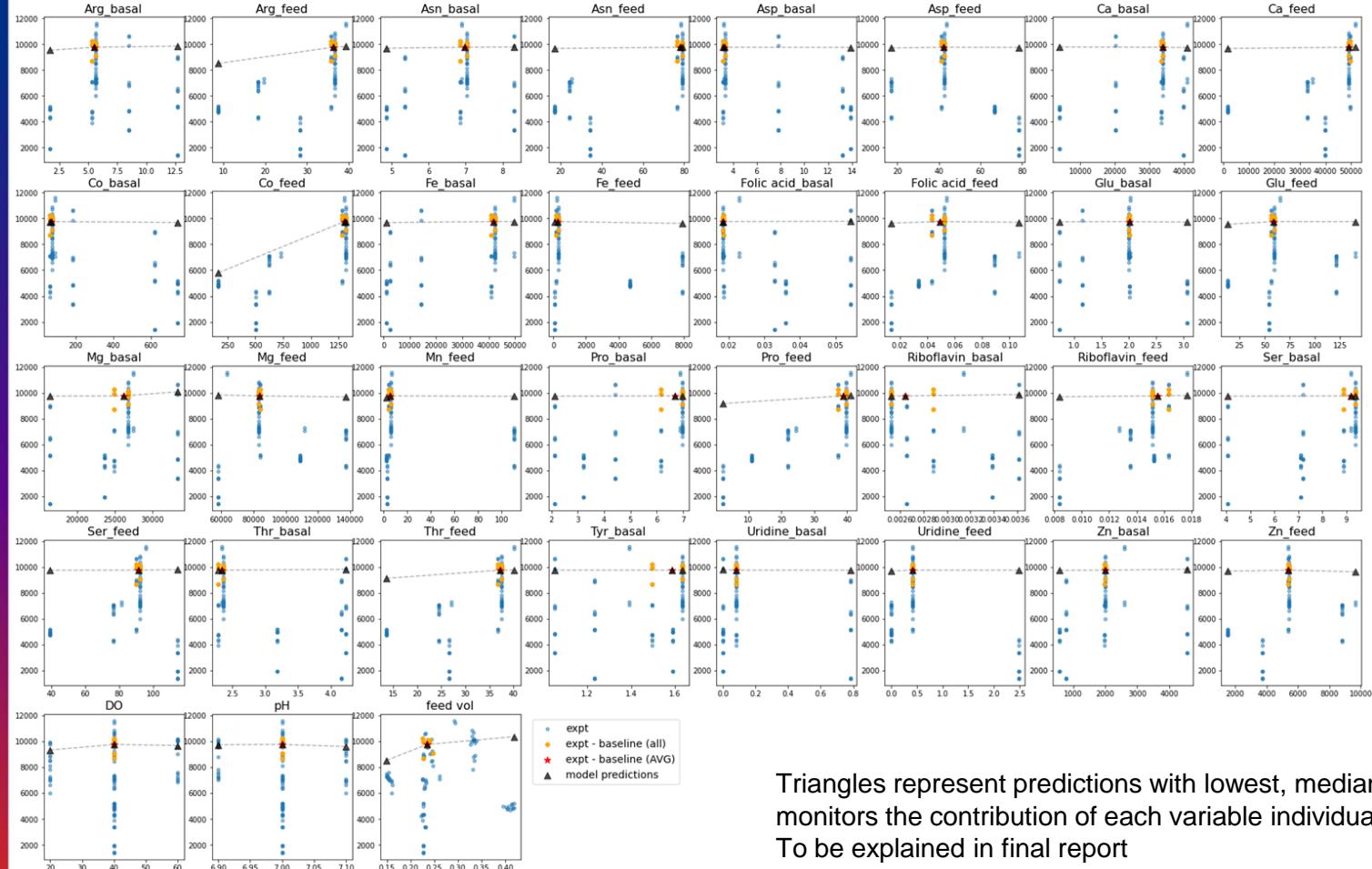
# Individual feature effects and contributions

base	yvar	BASAL							FEED / PROCESS						
		x_min	x_baseline	x_max	dy% (lbd)	dy% (ubd)	shap_corr	contrib%	x_min	x_baseline	x_max	dy% (lbd)	dy% (ubd)	shap_corr	contrib%
Arg	Fuc	1.72	5.54	12.69	0.27	0.73	-0.20	4.93	8.80	36.40	39.34	1.55	0.15	0.43	3.41
Asn		4.86	6.97	8.31	0.35	0.00	-0.50	0.70	17.26	78.19	79.06	0.23	0.15	0.16	0.19
Asp		3.16	3.30	13.87	2.47	0.16	-0.37	1.87	17.07	42.20	78.41	0.25	0.65	0.64	1.05
Ca		3,937.48	33,694.93	40,747.69	0.10	0.11	0.13	0.45	1,645.32	49,215.80	51,610.66	0.20	0.15	-0.24	12.81
Co		65.25	72.26	742.85	1.02	0.12	-0.18	0.72	167.16	1,304.74	1,315.19	0.34	0.15	-0.40	0.31
Fe		1,148.06	41,916.73	49,623.98	0.12	0.11	0.34	0.39	87.23	270.14	7,909.58	0.24	0.27	-0.22	0.14
Folic acid		0.02	0.02	0.05	0.15	-0.02	-0.54	0.66	0.01	0.05	0.11	0.23	0.15	-0.12	0.19
Glu		0.74	2.01	3.06	0.13	0.09	-0.41	1.48	13.33	58.69	139.86	0.15	0.31	-0.04	0.12
Mg		16,383.40	26,154.50	33,313.30	0.17	-0.31	-0.28	1.00	58,187.38	83,555.87	137,329.18	0.19	0.47	-0.04	0.48
Mn									2.61	5.73	110.45	0.19	0.17	-0.35	0.10
Pro		2.12	6.69	6.96	0.14	0.15	0.13	0.11	2.29	38.84	41.00	0.36	0.11	-0.42	0.46
Riboflavin		0.00	0.00	0.00	0.15	-3.39	-0.69	8.33	0.01	0.02	0.02	-0.93	0.36	-0.70	13.31
Ser		4.06	9.21	9.37	0.09	0.15	0.26	0.33	39.44	91.37	114.30	0.34	0.26	0.30	0.51
Thr		2.29	2.34	4.23	1.17	0.19	-0.14	0.71	13.64	37.29	40.12	0.24	0.12	-0.11	0.11
Tyr		1.05	1.59	1.63	0.10	0.12	-0.07	0.36							
Uridine		0.00	0.08	0.78	0.14	0.26	0.39	0.23	0.00	0.41	2.48	0.15	0.24	0.40	0.27
Zn		567.83	1,990.24	4,550.88	0.23	0.02		0.98	1,526.96	5,395.97	9,640.92	0.17	0.30	-0.27	0.10
DO									20.00	40.00	60.00	4.72	0.08	-0.79	8.42
feed vol									0.15	0.24	0.42	5.03	3.39	-0.59	31.58
pH									6.90	7.00	7.10	-0.24	1.11	0.79	3.18
Arg	Gal	1.72	5.54	12.69	1.13	0.25	-0.31	1.66	8.80	36.40	39.34	0.99	1.13	-0.14	0.86
Asn		4.86	6.97	8.31	1.52	-0.29	0.75	4.19	17.26	78.19	79.06	1.17	0.81	-0.39	0.10
Asp		3.16	3.30	13.87	3.97	-29.33	-0.94	28.82	17.07	42.20	78.41	0.26	14.92	0.60	10.56
Ca		3,937.48	33,694.93	40,747.69	1.06	0.96	-0.49	0.14	1,645.32	49,215.80	51,610.66	0.96	2.90	-0.14	0.18
Co		65.25	72.26	742.85	1.09	-4.62	-0.44	5.86	167.16	1,304.74	1,315.19	1.76	1.06	-0.34	0.16
Fe		1,148.06	41,916.73	49,623.98	-1.06	1.01	0.18	2.48	87.23	270.14	7,909.58	5.08	-2.42	-0.40	1.47
Folic acid		0.02	0.02	0.05	1.06	-7.15	-0.04	8.61	0.01	0.05	0.11	4.95	1.09	-0.65	1.14
Glu		0.74	2.01	3.06	1.06	1.40	0.08	0.18	13.33	58.69	139.86	1.66	1.06	-0.32	0.13
Mg		16,383.40	26,154.50	33,313.30	1.00	0.96	0.27	0.32	58,187.38	83,555.87	137,329.18	7.88	1.13	-0.29	1.77
Mn									2.61	5.73	110.45	1.07	1.13	-0.24	0.07
Pro		2.12	6.69	6.96	-1.72	0.92	0.40	2.77	2.29	38.84	41.00	2.93	1.24	-0.17	1.00
Riboflavin		0.00	0.00	0.00	1.06	1.20	0.67	0.20	0.01	0.02	0.02	3.04	2.56	-0.55	0.83
Ser		4.06	9.21	9.37	-2.57	0.89	0.38	3.68	39.44	91.37	114.30	0.71	6.44	0.25	1.46
Thr		2.29	2.34	4.23	1.09	-1.80	-0.43	2.92	13.64	37.29	40.12	1.14	1.13	-0.41	0.07
Tyr		1.05	1.59	1.63	1.37	1.06	-0.25	0.35							
Uridine		0.00	0.08	0.78	1.06	1.06	-0.21	0.07	0.00	0.41	2.48	1.06	9.56	0.26	2.16
Zn		567.83	1,990.24	4,550.88	1.06	0.96	0.22	0.21	1,526.96	5,395.97	9,640.92	1.16	-1.46	-0.28	0.45
DO									20.00	40.00	60.00	1.64	1.08	0.73	1.79
feed vol									0.15	0.24	0.42	18.52	-4.52	-0.72	12.90
pH									6.90	7.00	7.10	-0.51	1.87	0.17	0.44

To be explained in final report

# Effect of varving single feature vs. baseline (TITER)

Titer (mg/L)\_14



Triangles represent predictions with lowest, median and highest values. This monitors the contribution of each variable individually on titer.  
To be explained in final report

# Effect of varying single feature vs. baseline (MANNOSYLATION)

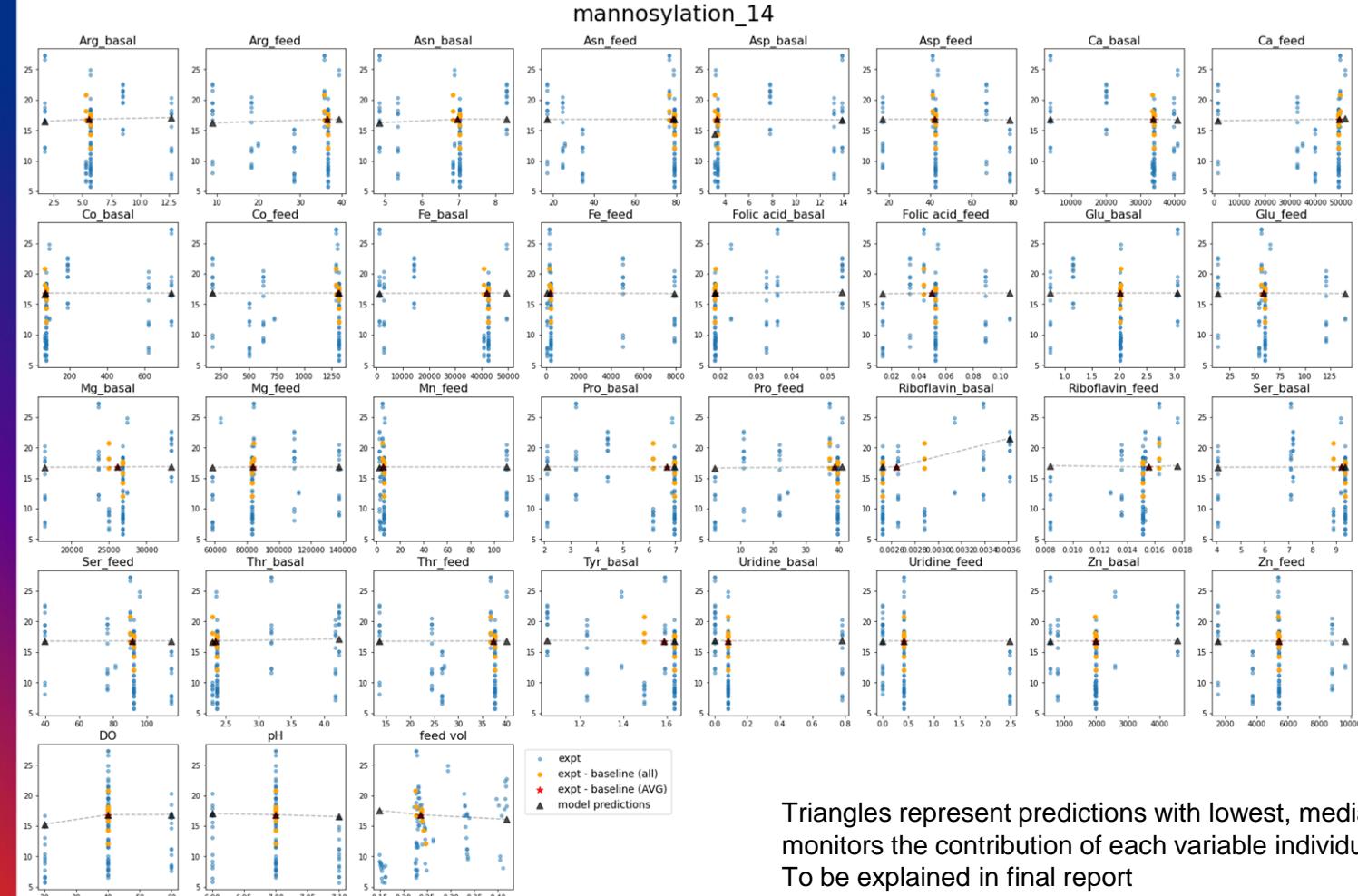
a<sup>★</sup>

10101  
10101

\*

Y

CREATING GROWTH, ENHANCING LIVES



Triangles represent predictions with lowest, median and highest values. This monitors the contribution of each variable individually on titer.  
To be explained in final report

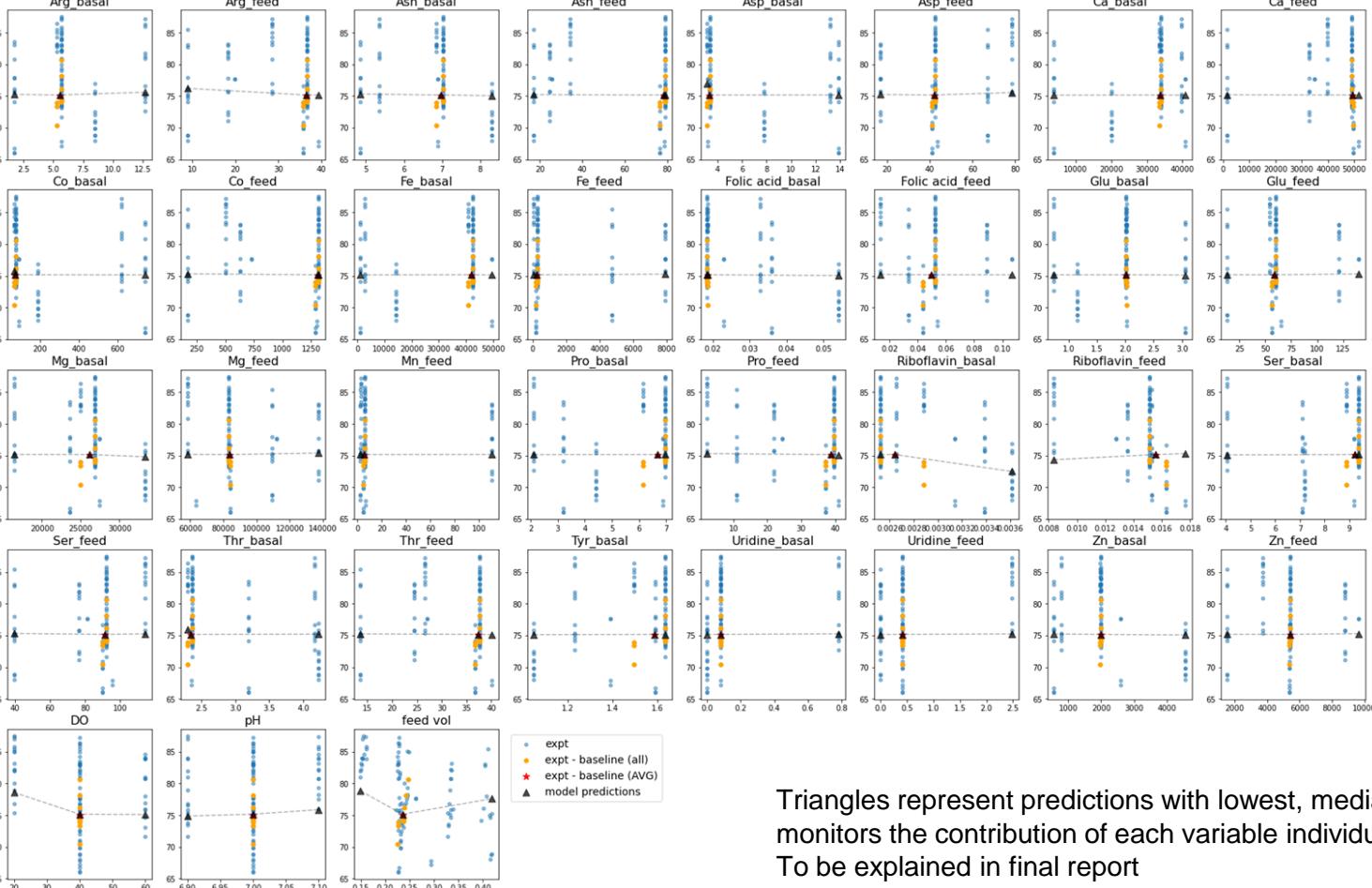
# Effect of varying single feature vs. baseline (FUCOSYLATION)

a



CREATING GROWTH, ENHANCING LIVES

fucosylation\_14

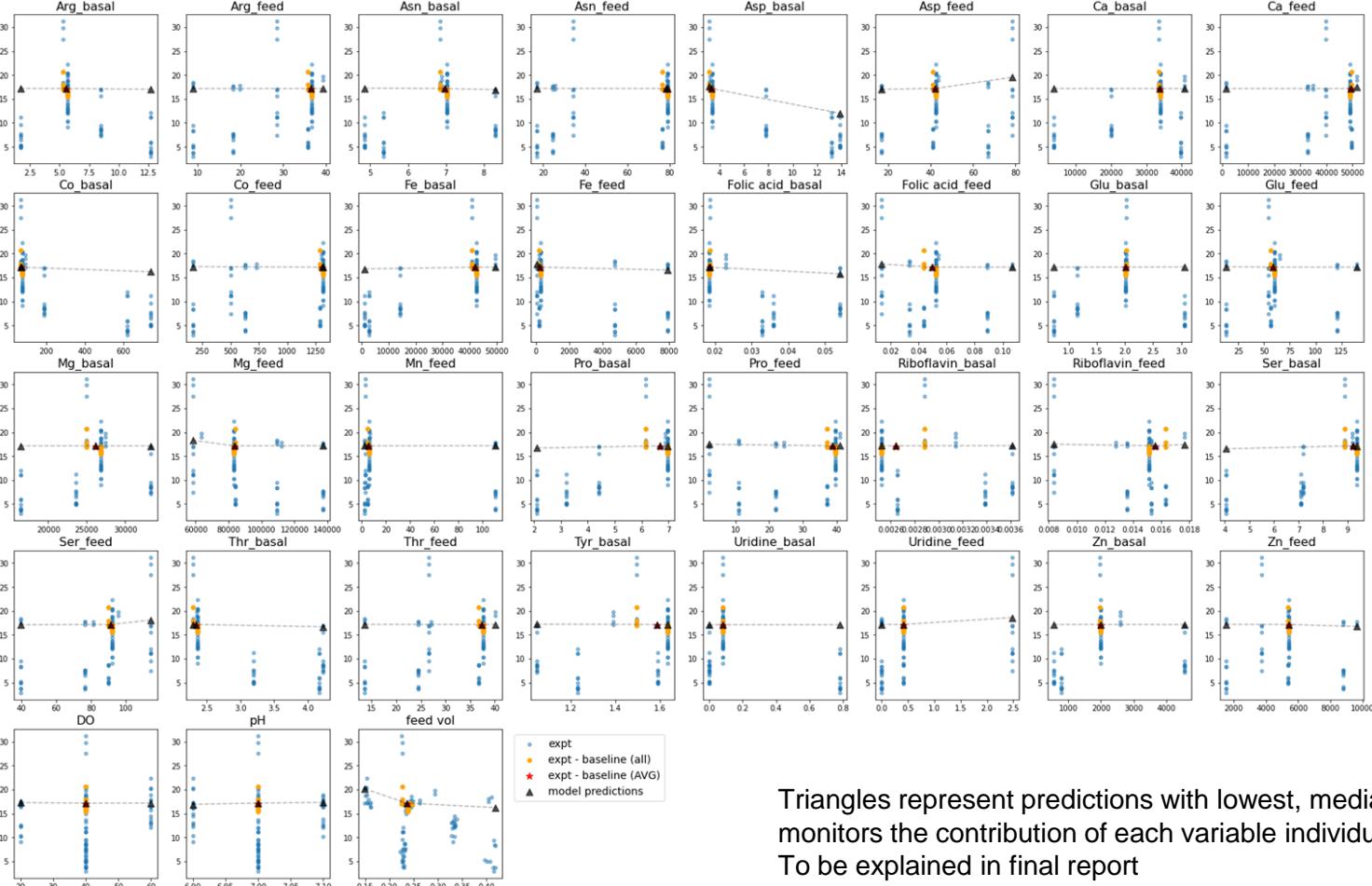


Triangles represent predictions with lowest, median and highest values. This monitors the contribution of each variable individually on titer.  
To be explained in final report

# Effect of varving single feature vs. baseline (GALACTOSYLATION)

galactosylation\_14

a\*



Triangles represent predictions with lowest, median and highest values. This monitors the contribution of each variable individually on titer.  
To be explained in final report