Apertura
GENE THERAPY

Machine Learning-Enhanced Design of Experiments (DoE) for Optimizing AAV Plasmid Ratios in Gene Therapy Manufacturing

Von Wiltman; Program Manager, CMC

Jorge Santiago-Ortiz, Andrew Steinsapir, Blake Albright, Max Ezekiel

ASGCT Annual Meeting 2025; Upstream Manufacturing for AAV Vectors I

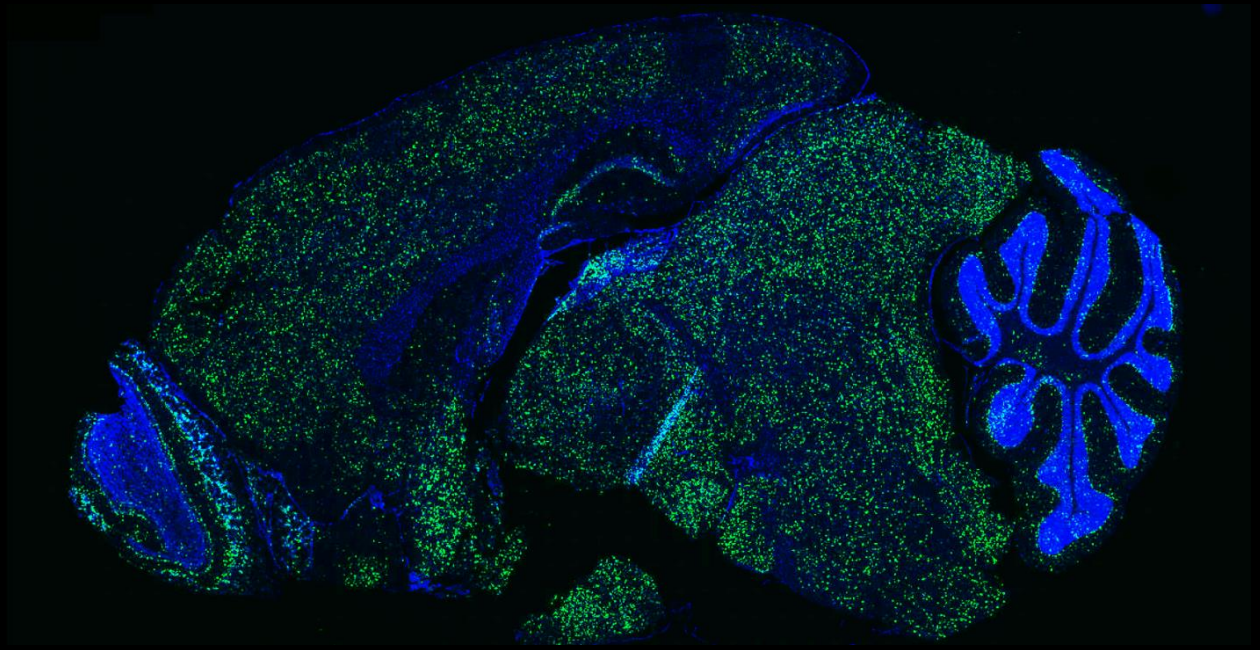
Widening the aperture of gene therapy through receptor-targeted engineering.

Novel, first-in-class family of BBB crossing capsids

- BBB-crossing targeting human-specific receptor
- FDA support for preclinical path

Modular NAb evasion capsid engineering platform

- Up to 50% reduction in NAb binding relative to natural serotypes
- Capsid tropism maintained after evasion modification

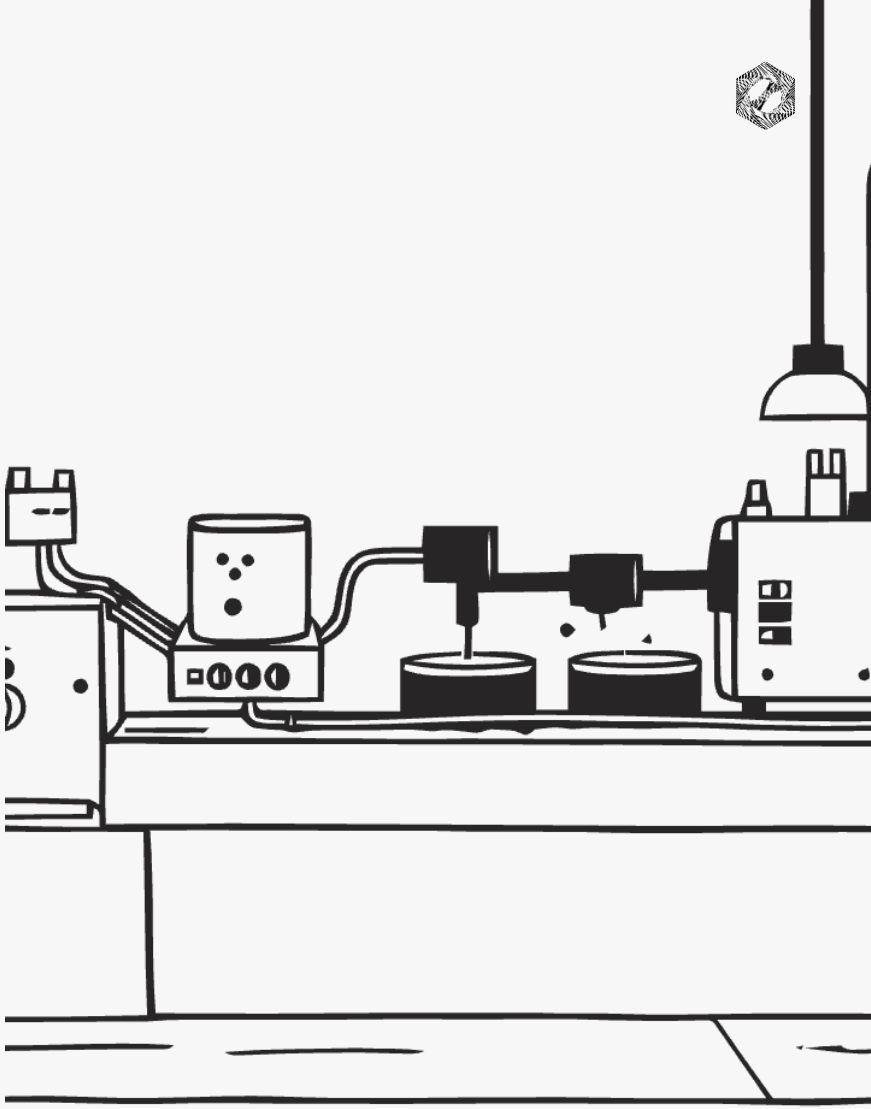


Broad CNS biodistribution via IV administration shows successful cell transduction of **payload** using our hTfR1 capsids.

Payload	CEG-hGBA-HA-WPRE
Mouse model	hTfR1 knock-in (Biocytogen)
Age of mice	14 weeks
Route of Administration	Intravenous (Tail vein)

Benefit of Fast & Data-Lean Upstream Optimization at Apertura

- Wide variety of capsids and payload combinations being produced at Apertura
  - Utilizing transient triple transfection (TTT) suspension HEK293 for flexibility
- Every new capsid and payload combination can exhibit unique production dynamics
- Trade-off between time and expense of production optimization and benefit from improved productivity
- Opportunity to explore DoE approaches for AAV production that can provide more process insight with less data



Machine Learning Design of Experiment (DoE) Approach Applied to AAV Upstream Optimization

Typical Approaches

Common uses of DoE for USP AAV Optimization

- Brief overview on DoE concepts
- Applications of DoE in transient triple transfection (TTT) AAV production
  - Factors
  - Responses
  - Design Strategies
  - Modeling
- Shortcomings of these design and modeling approaches when applied to TTT AAV production

Alternative Approaches

Self-Validating Ensemble Modeling (SVEM)

- Neural network (NN) ML models applied to plasmid ratio
- Self-Validating ensemble modeling utilized to compensate for smaller data sets relative to design space

Experiment Design, Methods, and Results

Plasmid Ratio Space-Filling Design: Scheffé Cubic vs SVEM

- Overview of plasmid ratio design and methods
- Modeling and results using Scheffé cubic regression model
- Predicted solution verification experiment
- Model comparison to SVEM NN

Implications and Future Work

Validation of Approach

- SVEM NN DoE modeling approach validated for AAV production applications
- Larger designs with more factors, not typically handled well by traditional models, can be confidently explored

Design of Experiment (DoE): Understanding Interactions Between Process Inputs and Outputs

Structured approach to study factor–response relationships efficiently

Define Factors

Inputs, or **factors**, for a process with their respective boundaries are defined. This is the **design space**

Generate Design

Use an algorithm that takes the defined **design space** and the number of tests, or runs, able to be performed and provides a set of process conditions to test

Collect Responses

Generated process condition combinations are performed and the process output(s), or **response(s)**, of interest are measured and recorded

Generate Models

These sets of process conditions and responses are then input into a model-generating algorithm that creates a model that estimates responses for any given set of inputs

Key advantage: Maximizes insight from minimal experimental runs

Montgomery, D.C. (2020) Design and Analysis of Experiments (10th ed.). Wiley.

Design of Experiment (DoE): Understanding Interactions Between Process Inputs and Outputs

Structured approach to study factor–response relationships efficiently

Define Factors

Inputs, or **factors**, for a process with their respective boundaries are defined. This is the **design space**

Generate Design

Use an algorithm that takes the defined **design space** and the number of tests, or runs, able to be performed and provides a set of process conditions to test

Collect Responses

Generated process condition combinations are performed and the process output(s), or **response(s)**, of interest are measured and recorded

Generate Models

These sets of process conditions and responses are then input into a model-generating algorithm that creates a model that estimates responses for any given set of inputs

Key advantage: Maximizes insight from minimal experimental runs

Montgomery, D.C. (2020) Design and Analysis of Experiments (10th ed.). Wiley.

Established DoE Approaches for TTT AAV Production Optimization

Factors[1]

- Plasmid ratios (pRC, pTG, pHLP)
- Total pDNA
- TFXN reagent; amounts
- Cell lines, TFXN cell density
- Media composition

Responses[1]

- AAV genomic titer
- % Full/empty capsids
- Impurity profile
- *In-vitro* Potency

Designs[1]

- D or I optimal designs
- Response surface design (Box-Behnken)
- Mixture design
- Space filling design
- Hybrids of several

Modeling[1]

- Regression models
  - Linear
  - Interaction
  - Quadratic
  - Scheffé cubic

[1]Illustrative, not a comprehensive list

Cornell, J. A. (2011).; Tran, N., et al. (2021).

Established DoE Approaches for TTT AAV Production Optimization

Factors[1]

- Plasmid ratios (pRC, pTG, pHLP)
- Total pDNA
- TFXN reagent; amounts
- Cell lines, TFXN cell density
- Media composition

Responses[1]

- AAV genomic titer
- % Full/empty capsids
- Impurity profile
- *In-vitro* Potency

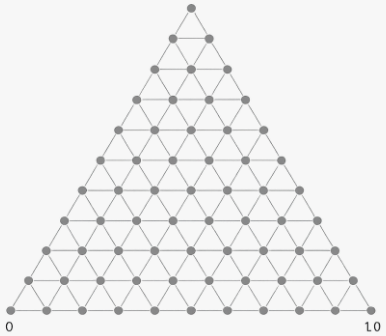
Designs[1]

- D or I optimal designs
- Response surface design (Box-Behnken)
- Mixture design
- Space filling design
- Hybrids of several

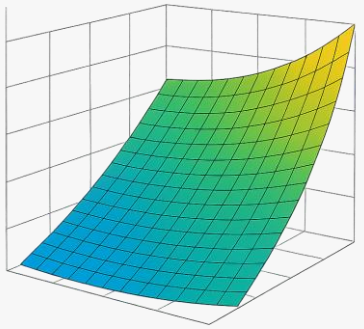
Modeling[1]

- Regression models
  - Linear
  - Interaction
  - Quadratic
  - Scheffé cubic

Plasmid Ratio Designs: Preferred Regression Approach



Space-Filling Mixture Design



Scheffé Cubic Regression Model

[1]Illustrative, not a comprehensive list

Cornell, J. A. (2011).; Tran, N., et al. (2021).

Established DoE Approaches for TTT AAV Production Optimization

Factors[1]

- Plasmid ratios (pRC, pTG, pHLP)
- Total pDNA
- TFXN reagent; amounts
- Cell lines, TFXN cell density
- Media composition

Responses[1]

- AAV genomic titer
- % Full/empty capsids
- Impurity profile
- *In-vitro* Potency

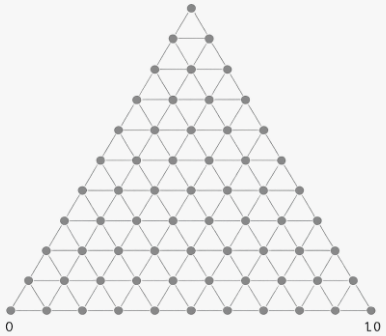
Designs[1]

- D or I optimal designs
- Response surface design (Box-Behnken)
- Mixture design
- Space filling design
- Hybrids of several

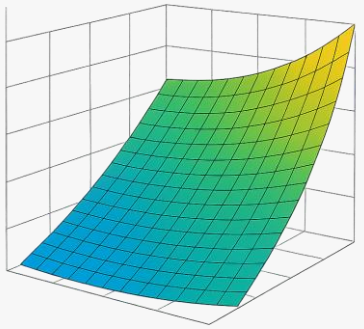
Modeling[1]

- Regression models
  - Linear
  - Interaction
  - Quadratic
  - Scheffé cubic

Plasmid Ratio Designs: Preferred Regression Approach



Space-Filling Mixture Design



Scheffé Cubic Regression Model

[1]Illustrative, not a comprehensive list

Cornell, J. A. (2011).; Tran, N., et al. (2021).

Neural Network Models and ML in DoE: Not Typically Used

Characteristic	Regression (Scheffé Cubic)	Neural Network
Form of Output Function	Polynomial in component proportions	Flexible, nonlinear mapping via layers and activation functions
Handling non-Linearities	Limited to what cubic terms can capture	Excellent at modeling complex nonlinearities and interaction effects
Assumptions	Assumes response can be captured via polynomial terms; model terms pre-specified	Fewer assumptions about underlying function form
Handling of High-Dimensional Inputs	Poor – Terms increase exponentially	Strong
Impact of Outliers	High – Least squares fit	**Lower – If ensembled**
Data Requirements	Moderate – Requires balanced	**High – Requires partitioning of data into training and validating sets**

Bishop, C. M. (2006).; Lemkus, Trent, et al. (2021); Rasmussen, C. E., & Williams, C. K. I. (2006).; Tibshirani, R. (1996).

Neural Network Models and ML in DoE: Not Typically Used

Characteristic	Regression (Scheffé Cubic)	Neural Network
Form of Output Function	Polynomial in component proportions	Flexible, nonlinear mapping via layers and activation functions
Handling non-Linearities	Limited to what cubic terms can capture	Excellent at modeling complex nonlinearities and interaction effects
Assumptions	Assumes response can be captured via polynomial terms; model terms pre-specified	Fewer assumptions about underlying function form
Handling of High-Dimensional Inputs	Poor – Terms increase exponentially	Strong
Impact of Outliers	High – Least squares fit	**Lower – If ensembled**
Data Requirements	Moderate – Requires balanced	**High – Requires partitioning of data into training and validating sets**

Bishop, C. M. (2006).; Lemkus, Trent, et al. (2021); Rasmussen, C. E., & Williams, C. K. I. (2006).; Tibshirani, R. (1996).

Self-Validated Ensemble Modeling Approach (SVEM): Potential to Compensate for High Data Requirements

A New Approach to Predictive Modeling in DoE

What is SVEM?

- Recently developed approach that has demonstrated utility in mixture DoEs and combines:
  - Auto-validation of data sets
  - Fractionally weighted bootstrapping
  - Ensemble Modeling

How Does it Work?

- Auto-validation of data sets
  - Anti-correlated weighting to duplicate data set (gamma distribution)
- Fractionally weighted bootstrapping
  - Bootstrapped models with responses weighted
- Ensemble Modeling
  - Averages models generated

Key Benefits

- Ability to tune with small data set allowing for theoretical utility in TTT AAV Production optimization with:
  - Reduced risk of overfitting
  - Reduced instability

Lemkus, T., Gotwalt, C., Ramsey, P., & Weese, M. L. (2021).

Self-Validated Ensemble Modeling Approach (SVEM): Potential to Compensate for High Data Requirements

A New Approach to Predictive Modeling in DoE

What is SVEM?

- Recently developed approach that has demonstrated utility in mixture DoEs and combines:
  - Auto-validation of data sets
  - Fractionally weighted bootstrapping
  - Ensemble Modeling

How Does it Work?

- Auto-validation of data sets
  - Anti-correlated weighting to duplicate data set (gamma distribution)
- Fractionally weighted bootstrapping
  - Bootstrapped models with responses weighted
- Ensemble Modeling
  - Averages models generated

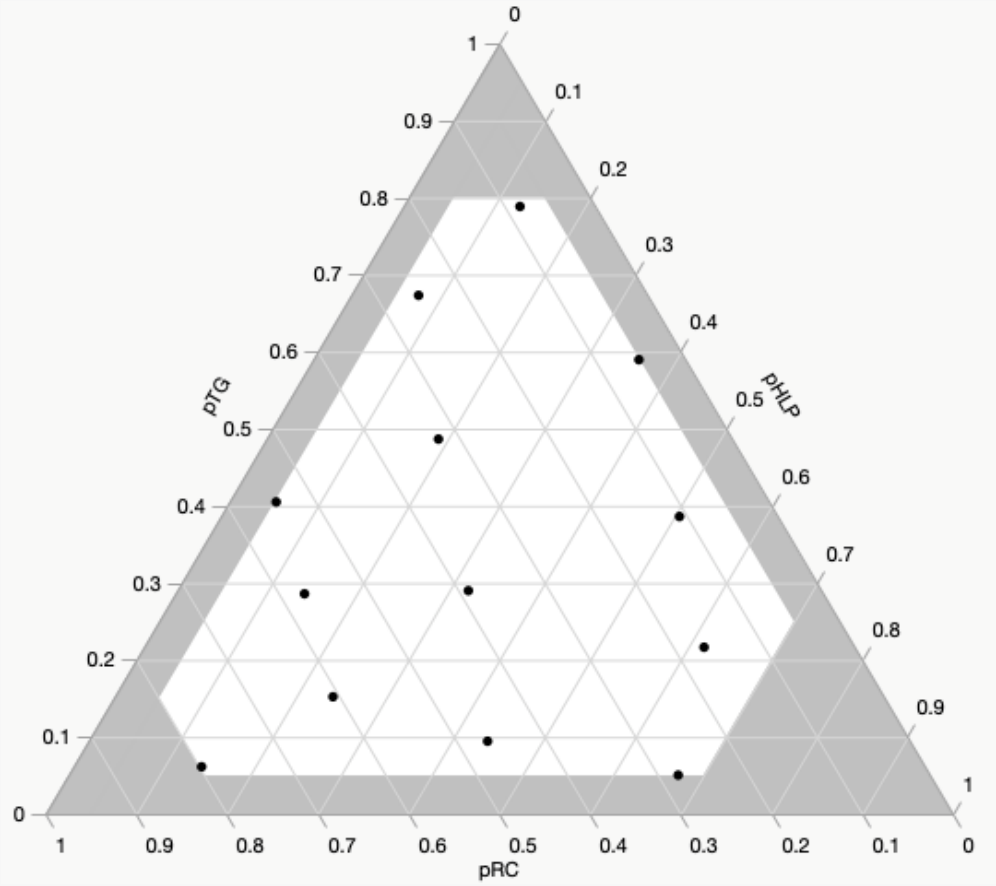
Key Benefits

- Ability to tune with small data set allowing for theoretical utility in TTT AAV Production optimization with:
  - Reduced risk of overfitting
  - Reduced instability

Lemkus, T., Gotwalt, C., Ramsey, P., & Weese, M. L. (2021).

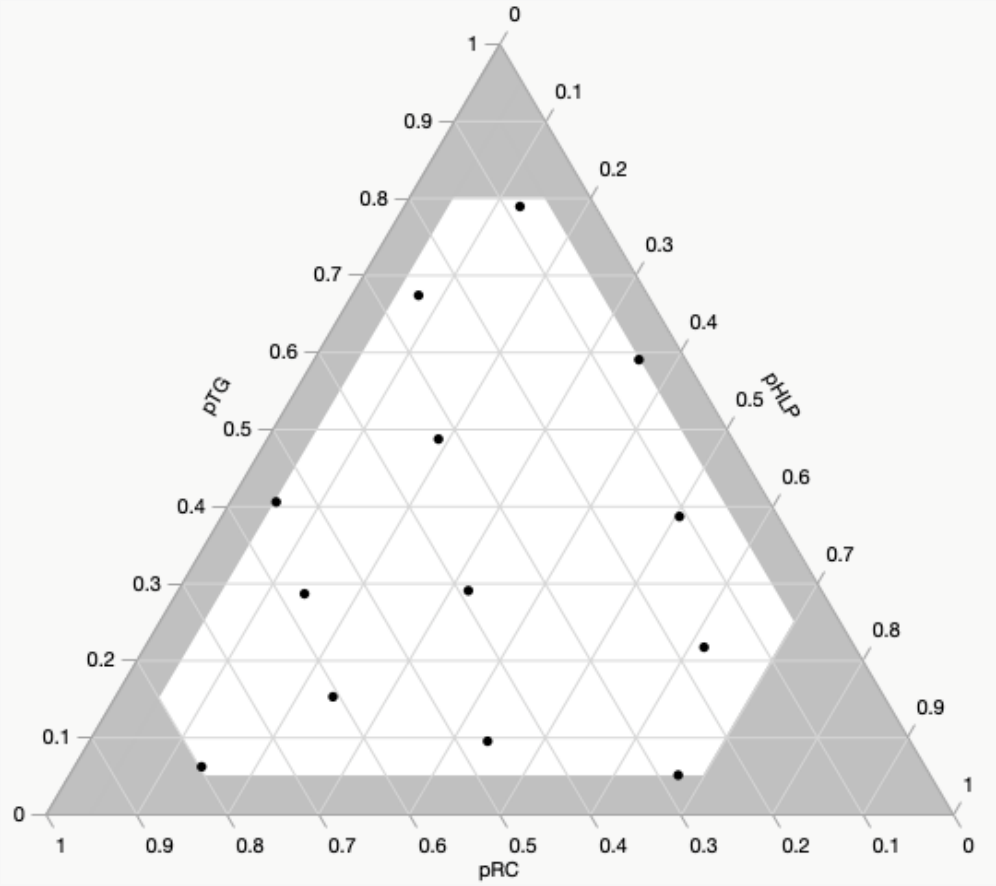
Plasmid Ratio DoE Experiment Design and Methods

Factors (ranges)	Plasmid Molar Ratios: pRepCap (0.05-.08) pTG (0.05-0.8) pHLP (0.05-0.7)
Responses	Genomic titer by ddPCR Percent empty by mass photometry
Design	Space filling, Mixture
Conditions /Replicates	13/2
System	125mL SF - 30mL WV
Analyte	Batch bound lysate eluate
Acceptance Criteria	Order independence ddPCR plate independence
Model Generation and Analysis	Scheffé cubic NN SVEM: 100 BS, Excl. 50



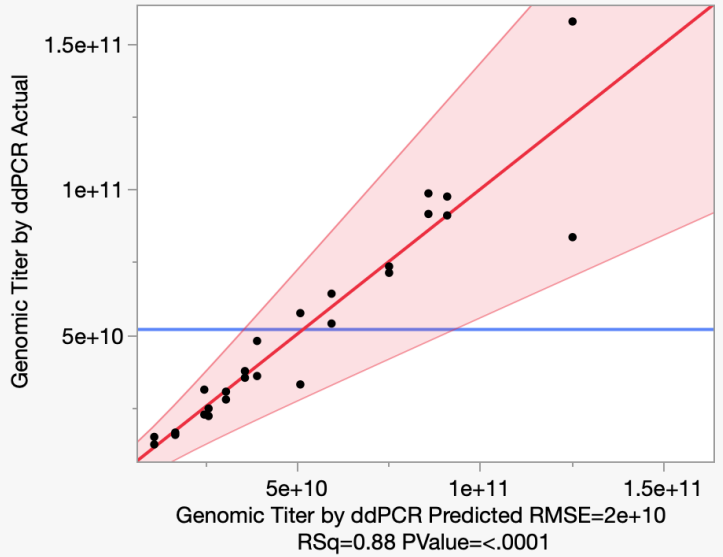
Plasmid Ratio DoE Experiment Design and Methods

Factors (ranges)	Plasmid Molar Ratios: pRepCap (0.05-.08) pTG (0.05-0.8) pHLP (0.05-0.7)
Responses	Genomic titer by ddPCR Percent empty by mass photometry
Design	Space filling, Mixture
Conditions /Replicates	13/2
System	125mL SF - 30mL WV
Analyte	Batch bound lysate eluate
Acceptance Criteria	Order independence ddPCR plate independence
Model Generation and Analysis	Scheffé cubic NN SVEM: 100 BS, Excl. 50



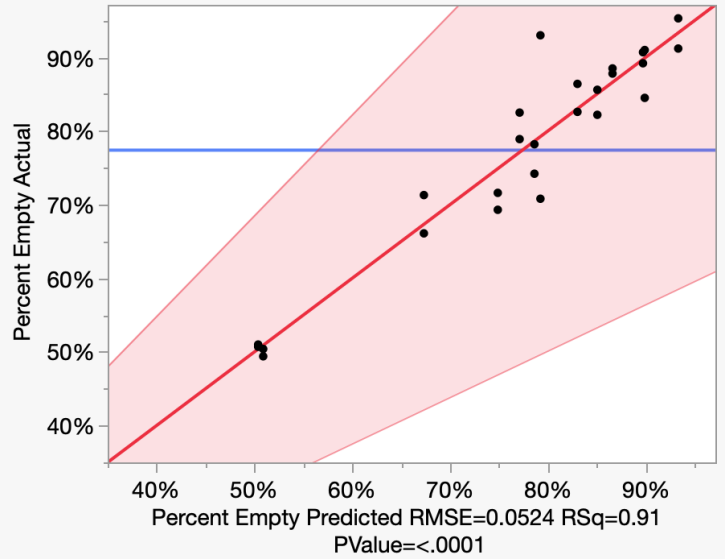
Scheffé Cubic Models Yield Robust Models with Strong Predictive Accuracy

Genomic Titer by ddPCR



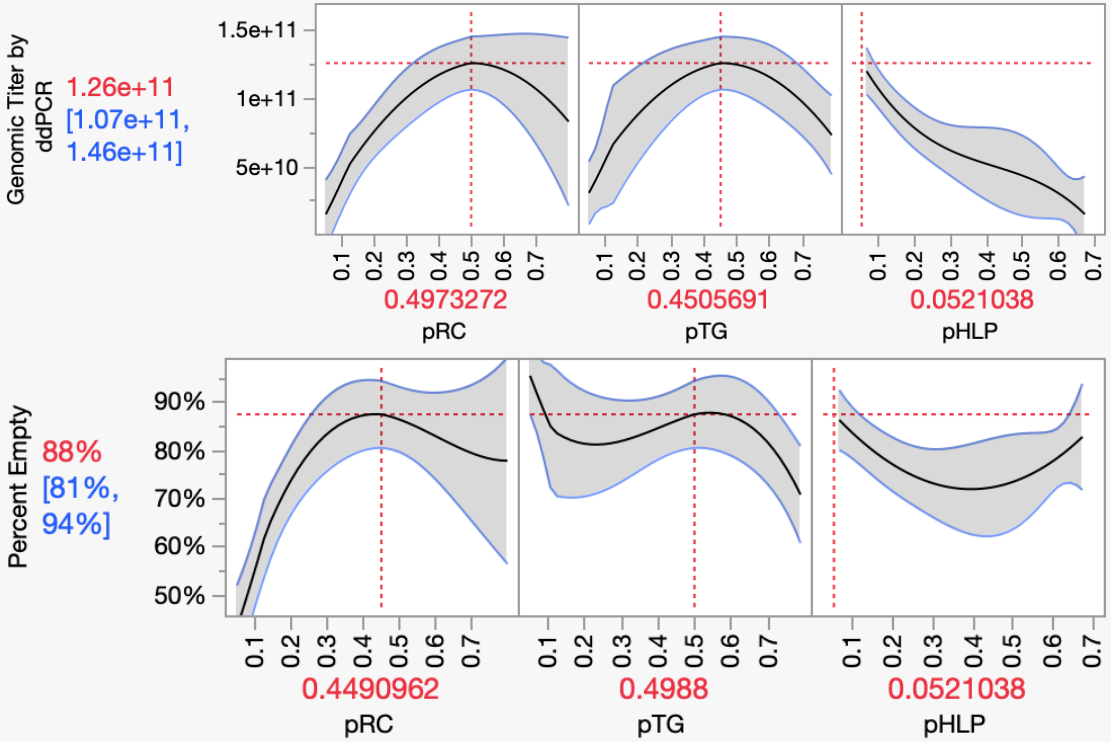
Source	Logworth	PValue
pRC*pTG	1.859	0.01383
(pTG-0.0504)/0.84649	0.630	0.23439
pTG*pHLP*(pTG-pHLP)	0.560	0.27563
(pRC-0.05101)/0.84649	0.555	0.27869
pRC*pTG*pHLP	0.517	0.30405
pRC*pHLP	0.449	0.35541
(pHLP-0.0521)/0.84649	0.313	0.48648
pTG*pHLP	0.312	0.48697
pRC*pHLP*(pRC-pHLP)	0.271	0.53569
pRC*pTG*(pRC-pTG)	0.013	0.97070

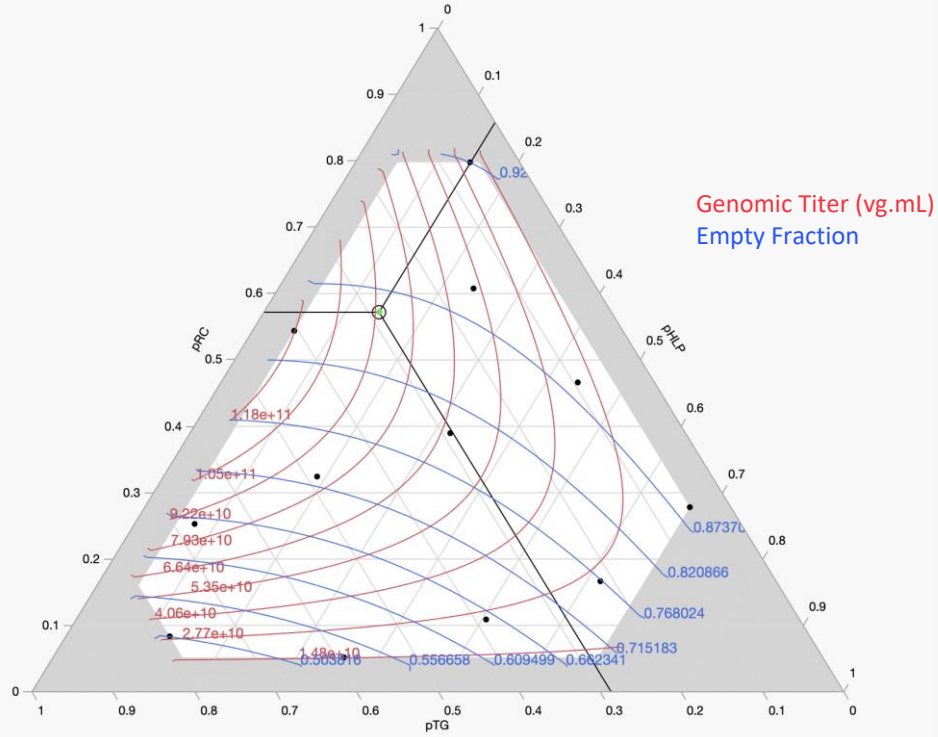
Percent Empty



Source	Logworth	PValue
(pRC-0.05101)/0.84649	4.095	0.00008
(pTG-0.0504)/0.84649	3.899	0.00013
(pHLP-0.0521)/0.84649	1.631	0.02340
pRC*pTG*(pRC-pTG)	1.442	0.03611
pRC*pTG	1.341	0.04562
pRC*pTG*pHLP	0.603	0.24967
pRC*pHLP*(pRC-pHLP)	0.428	0.37341
pTG*pHLP	0.396	0.40224
pTG*pHLP*(pTG-pHLP)	0.319	0.47939
pRC*pHLP	0.003	0.99344

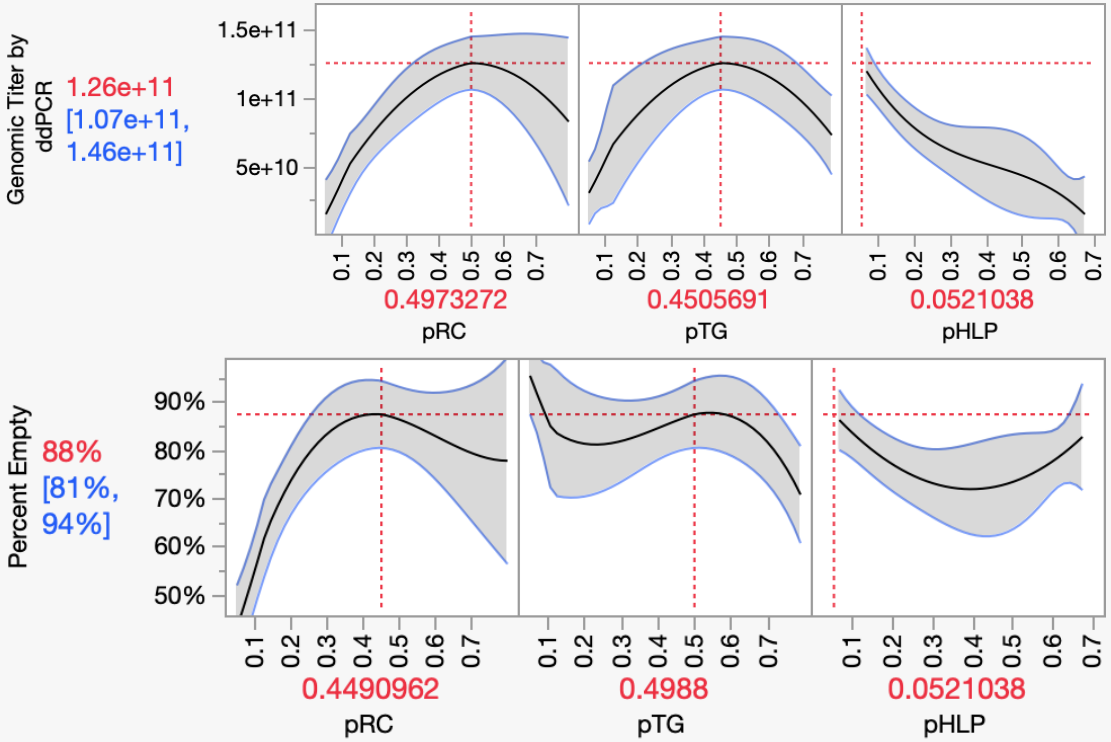
Maximized Genomic Titer Solution Predicts 47% Increase From Baseline

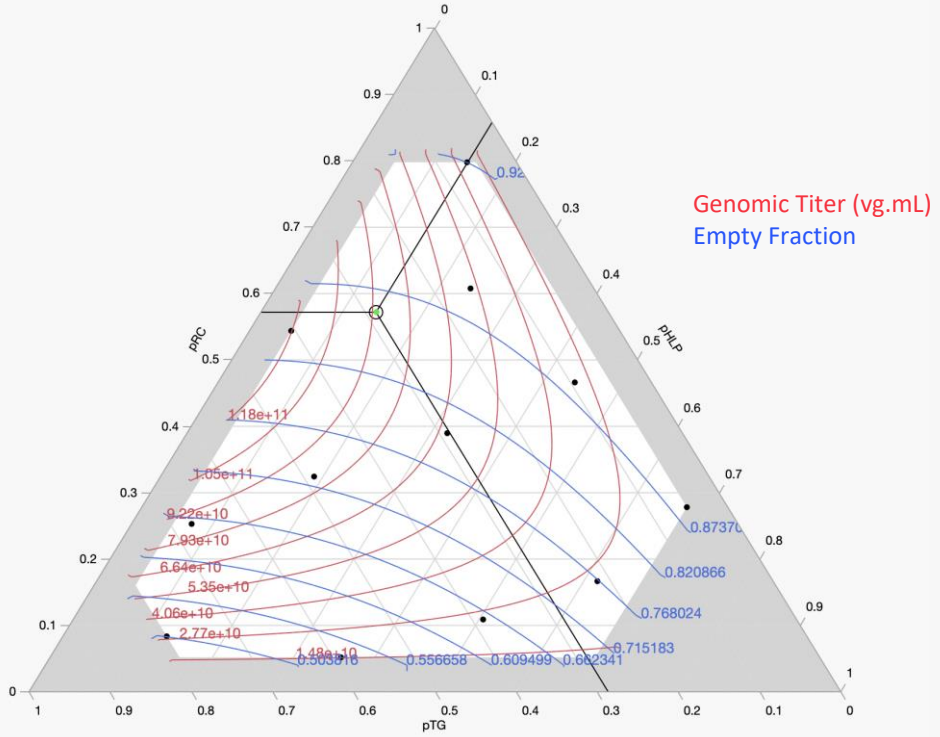




Predicted Plasmid Molar Ratios	Predicted GT	Predicted Baseline GT	Predicted GT Increase from Baseline	Predicted % Empty at Maximized GT
pRepCap: ~0.50 pTG: ~0.45 pHLP: ~0.05	1.26e11 vg/mL	8.60e10 vg/mL	47%	88%

Maximized Genomic Titer Solution Predicts 47% Increase From Baseline

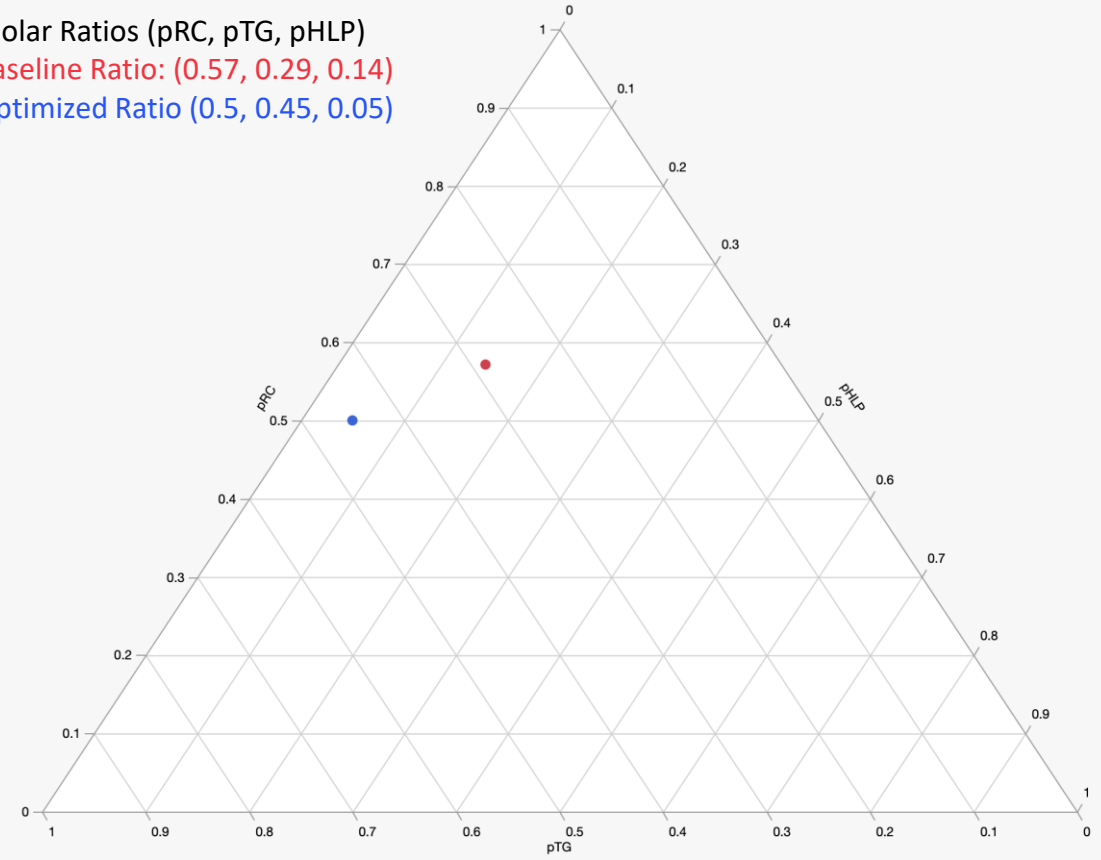


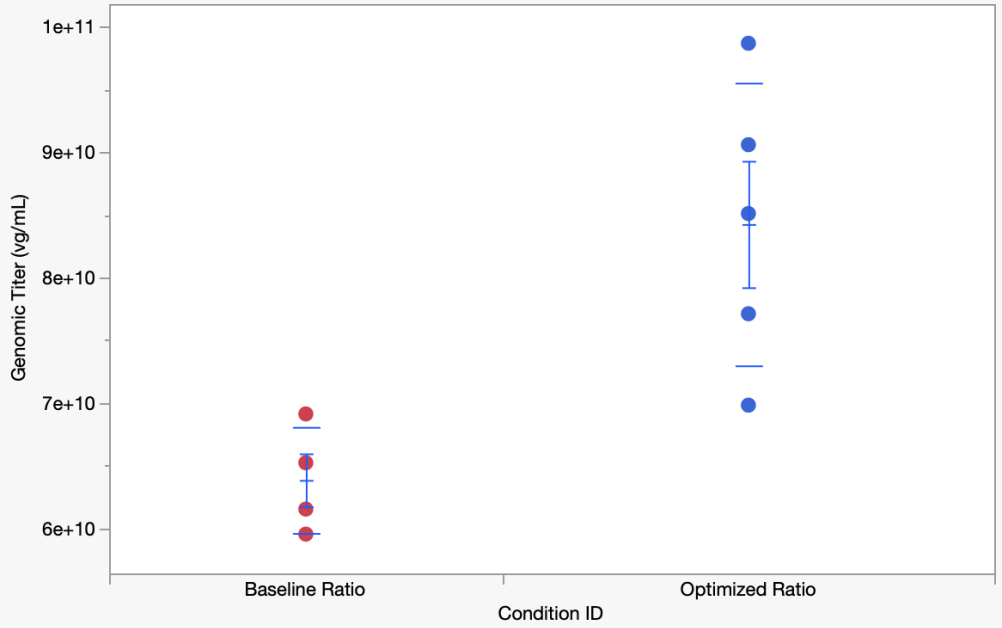


Predicted Plasmid Molar Ratios	Predicted GT	Predicted Baseline GT	Predicted GT Increase from Baseline	Predicted % Empty at Maximized GT
pRepCap: ~0.50 pTG: ~0.45 pHLP: ~0.05	1.26e11 vg/mL	8.60e10 vg/mL	47%	88%

Genomic Titer Solution Confirmation Study Recapitulates Predicted Response

Molar Ratios (pRC, pTG, pHLP)
Baseline Ratio: (0.57, 0.29, 0.14)
Optimized Ratio (0.5, 0.45, 0.05)



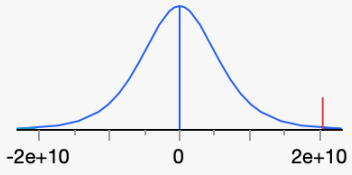


Means and Std Deviations

Level	Number	Mean	Std Dev
Baseline Ratio	4	6.382e+10	4.2359e+9
Optimized Ratio	5	8.426e+10	1.129e+10

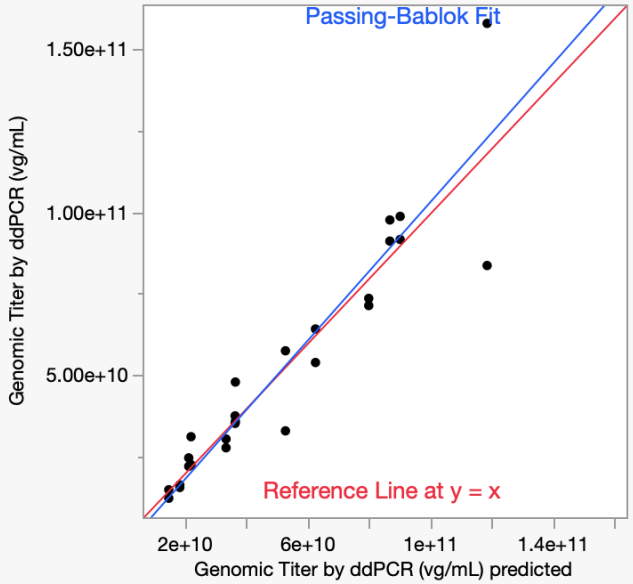
Assuming unequal variances

t Ratio	3.732507
DF	5.312395
Prob > |t|	0.0121*
Prob > t	0.0061*
Prob < t	0.9939



SVEM NN Models Offer Comparable Robustness and Enhanced Predictive Compared to Scheffé Cubic

Genomic Titer by ddPCR

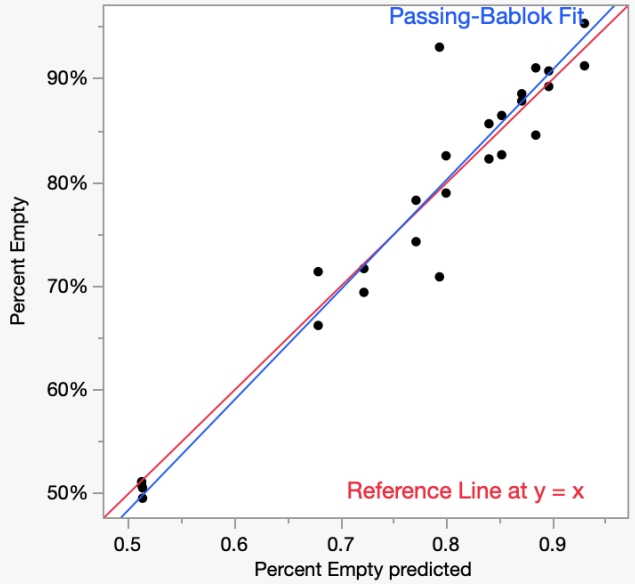


Model Summary

Platform	Neural
Response	Genomic Titer by ddPCR
Layers	NTanH(3)
Validation Method	Auto Validation
Number of Bootstrapped Models	100
Random Seed	74643
Proportion of Excluded Runs	0.5
Number of Rows	26

Measure	Training
Final Model RASE	1.2e+10

Percent Empty



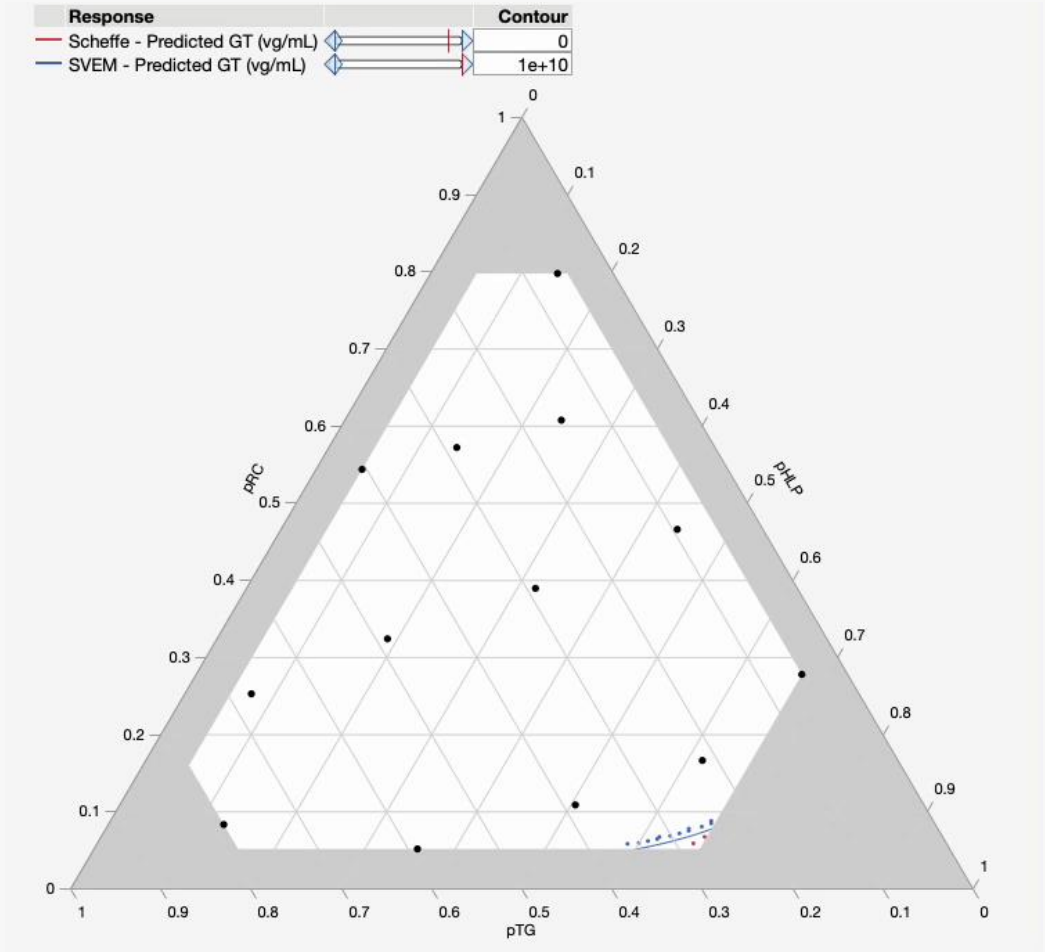
Model Summary

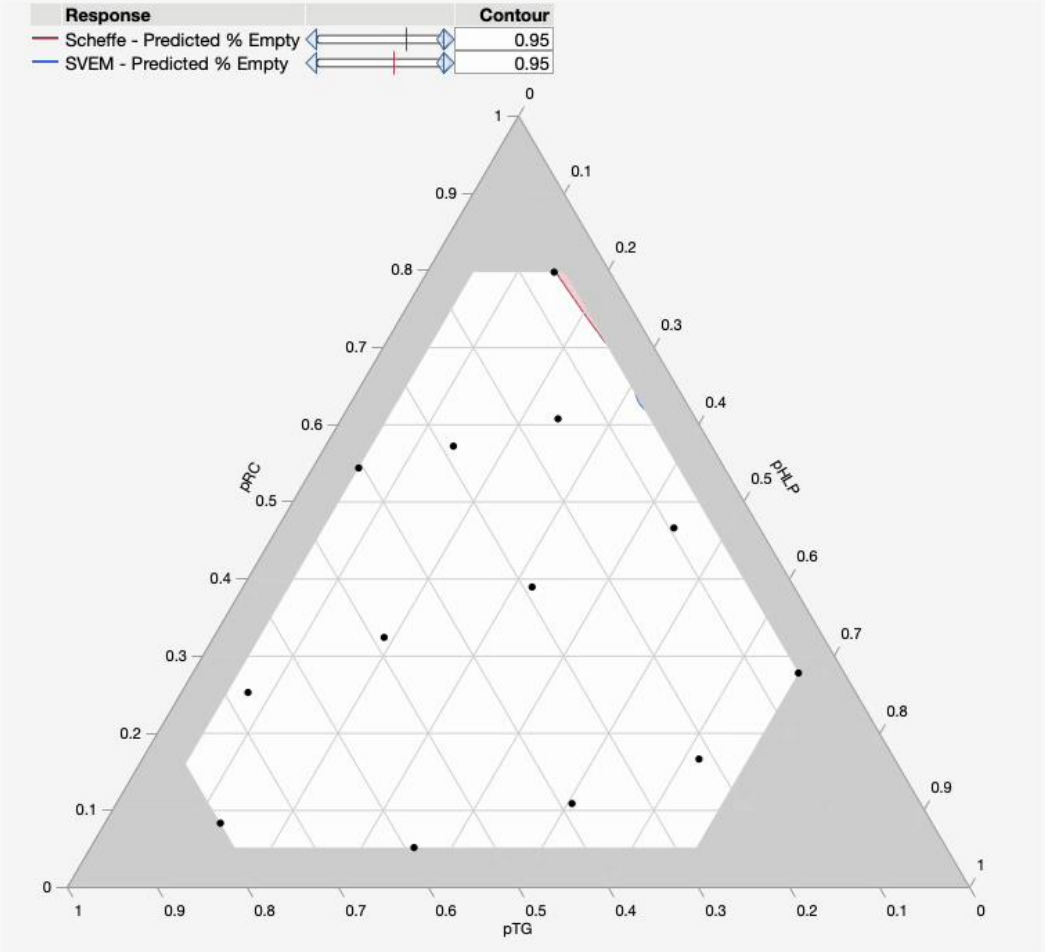
Platform	Neural
Response	Percent Empty
Layers	NTanH(3)
Validation Method	Auto Validation
Number of Bootstrapped Models	100
Random Seed	74643
Proportion of Excluded Runs	0.5
Number of Rows	26

Measure	Training
Final Model RASE	0.03687

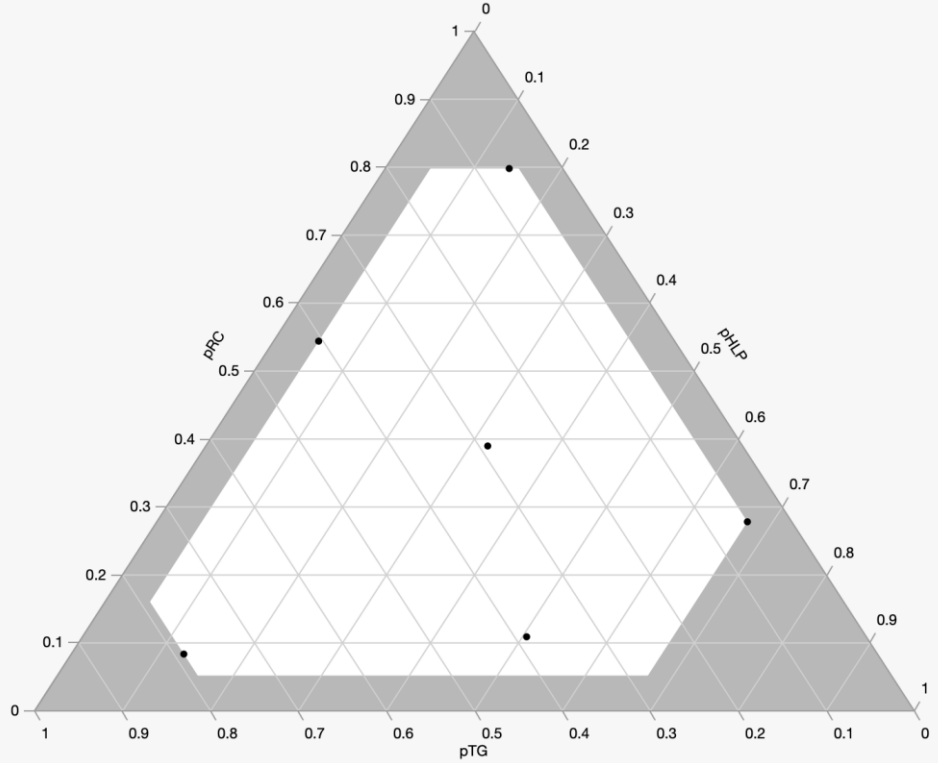
Genomic Titer		Percent Empty	
Scheffé Cubic RMSE	SVEM RASE	Scheffé Cubic RMSE	SVEM RASE
1.6e10	1.2e10	5.2%	3.7%

SVEM vs Scheffé Cubic – Highly Similar Response Profile Patterns



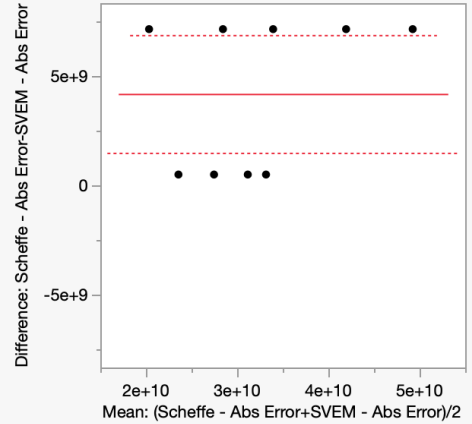


Run Count Sensitivity Test: SVEM Demonstrates Improved Predictability When Runs are Constrained



Genomic Titer		Percent Empty	
Scheffé Cubic RMSE	SVEM RASE	Scheffé Cubic RMSE	SVEM RASE
2.2e10	1.5e10	2.5%	1.8%

Confirmation Data as Validation Set



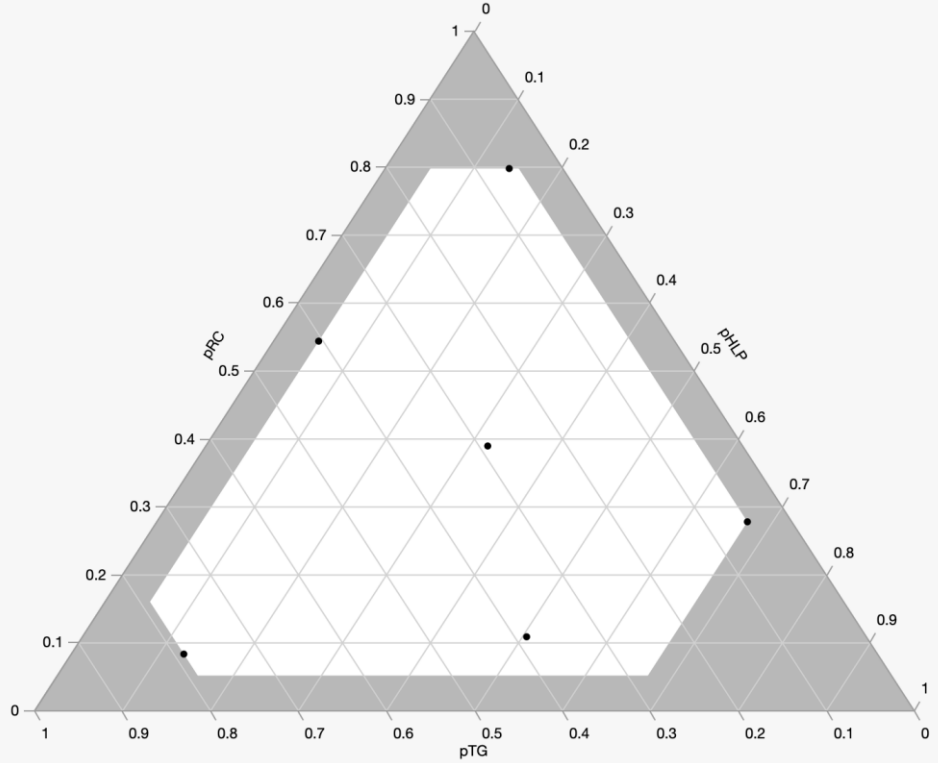
Scheffe - Abs Error	3.4e+10	t-Ratio	3.584256
SVEM - Abs Error	3e+10	DF	8
Mean Difference	4.18e+9	Prob > |t|	0.0071*
Std Error	1.17e+9	Prob > t	0.0036*
Upper 95%	6.87e+9	Prob < t	0.9964
Lower 95%	1.49e+9		
N	9		
Correlation	0.93457		

Wilcoxon Signed Rank

	Scheffe - Abs Error-SVEM - Abs Error
Test Statistic S	22.500
Prob>|S|	0.0039*
Prob>S	0.0020*
Prob<S	0.9980

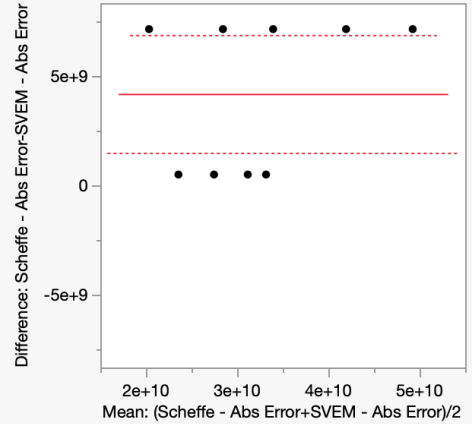
SVEM illustrates enhanced predictability with reduced training data

Run Count Sensitivity Test: SVEM Demonstrates Improved Predictability When Runs are Constrained



Genomic Titer		Percent Empty	
Scheffé Cubic RMSE	SVEM RASE	Scheffé Cubic RMSE	SVEM RASE
2.2e10	1.5e10	2.5%	1.8%

Confirmation Data as Validation Set



Scheffe - Abs Error	3.4e+10	t-Ratio	3.584256
SVEM - Abs Error	3e+10	DF	8
Mean Difference	4.18e+9	Prob > |t|	0.0071*
Std Error	1.17e+9	Prob > t	0.0036*
Upper 95%	6.87e+9	Prob < t	0.9964
Lower 95%	1.49e+9		
N	9		
Correlation	0.93457		

Wilcoxon Signed Rank

	Scheffe - Abs Error-SVEM - Abs Error
Test Statistic S	22.500
Prob>|S|	0.0039*
Prob>S	0.0020*
Prob<S	0.9980

SVEM illustrates enhanced predictability with reduced training data

Conclusions: Broader Applications in AAV Optimization

- Space-filling mixture designs for plasmid ratio yield robust models in both Scheffé cubic and self-validating ensemble modeling approaches
- The self-validating ensemble approach offers internal robustness checks, allowing for NN ML methods without the need for large verification datasets
- SVEM demonstrated enhanced predictive accuracy compared to Scheffé cubic regression for plasmid ratio optimization in AAV production under limited data sets
- Future work will extend SVEM to broader upstream design spaces, including media composition and other transfection conditions

Acknowledgements

- **Apertura Gene Therapy Team**
  - Jorge Santiago-Ortiz, PhD
  - Christopher Davis, PhD
  - Sisi Chen, PhD
  - Ellen Horste, PhD
  - Tina Hu-Seliger, PhD
  - Hikari Sorensen, BS
  - Devanshi Shanghavi, PSM
  - Von Wiltman, BS
  - Sampada Kalan, PhD
  - Ada Zhang, MS
  - Jade Ganter, MS
  - Anthony Stabile, BS
  - Jessica Luo, BS
  - Max Seaman, MS
  - Megan Powell, BS
  - Blake Albright, PhD
  - Max Ezekiel, MS
  - Paul Puglisi, MS
  - Diego Garzón, PhD, MBA
  - Andrew Steinsapir, MBA
- **Ben Deverman, PhD**
- **Dave Greenwald, PhD**

Apertura
GENE THERAPY

Contact: info@aperturagtx.com