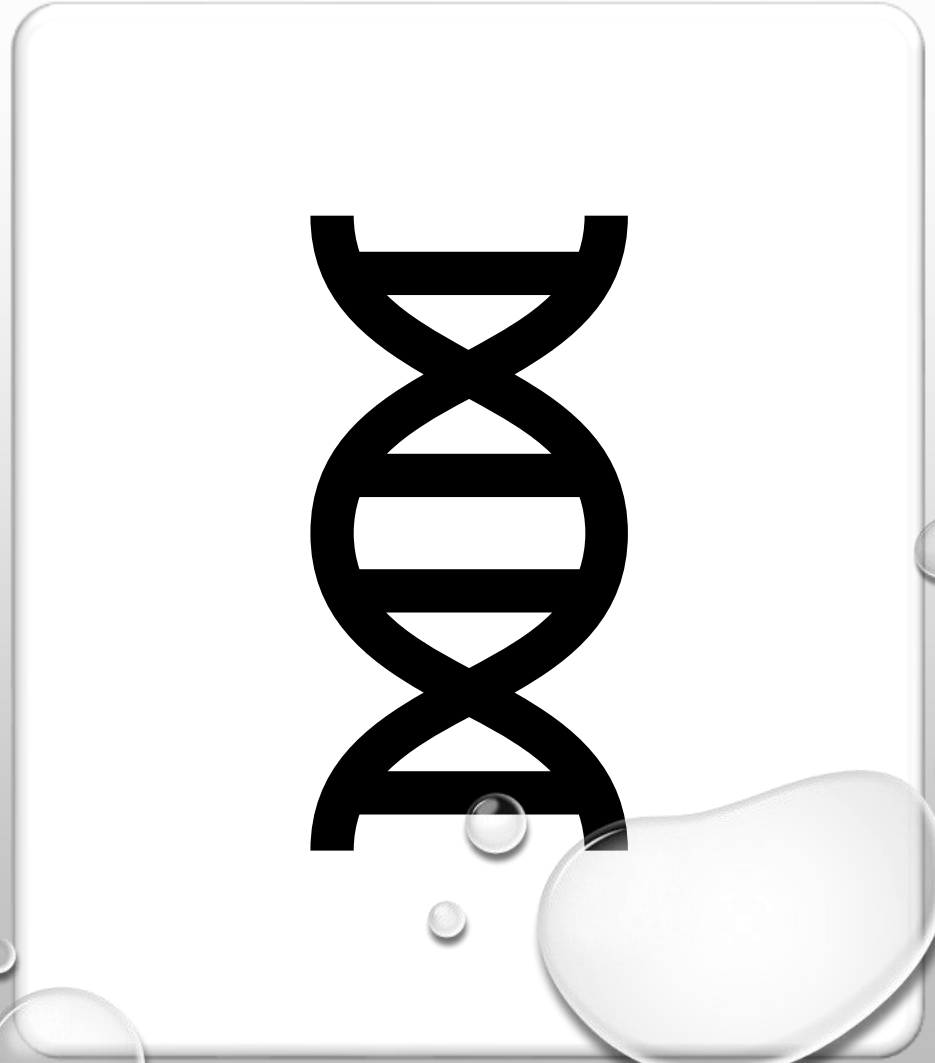
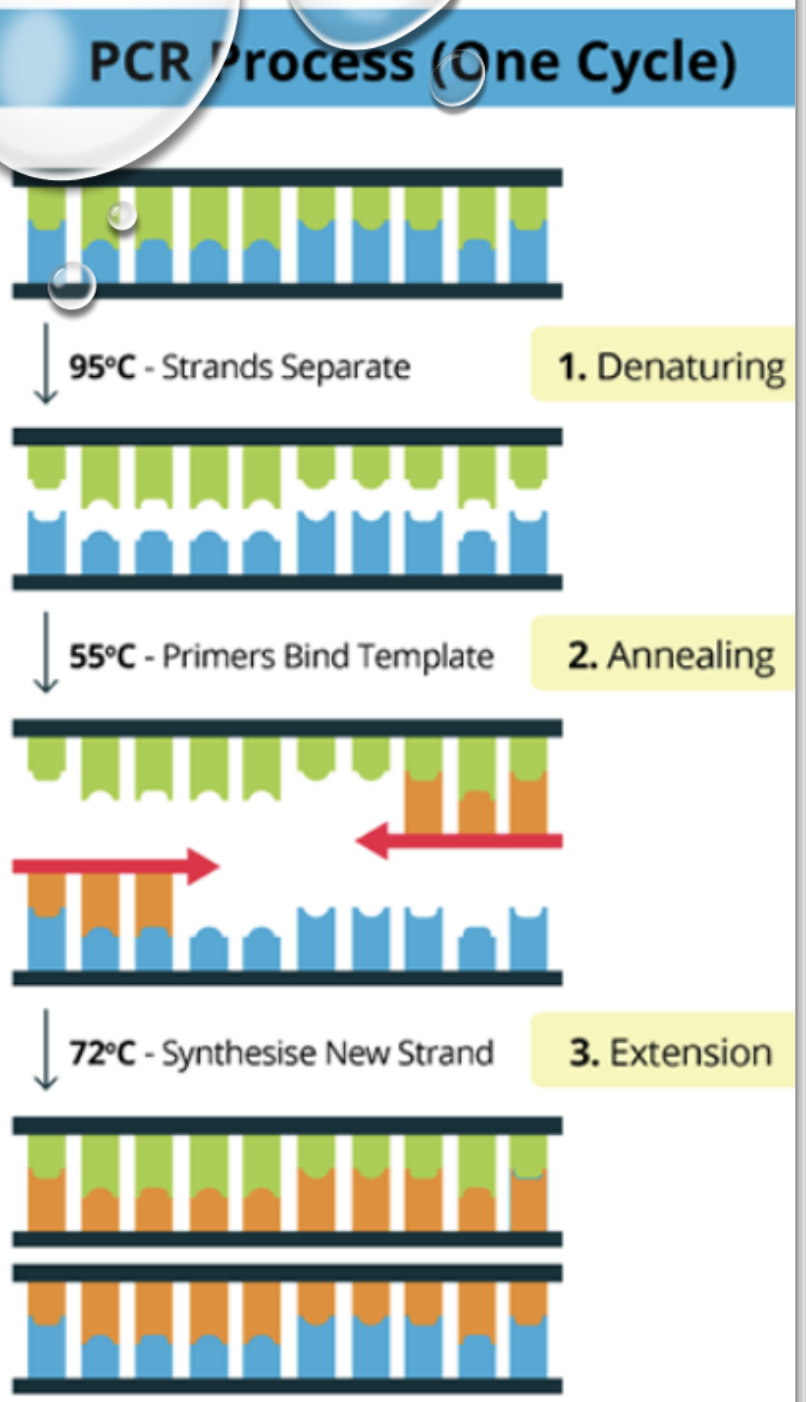
OPTIMIZING A PCR PROCESS

Erica Stein, Fatemeh Esmaeilbeiki,
Alyssa Rogers-Armstrong

WHAT IS A PCR?

INTRODUCTION





POLYMERASE CHAIN REACTION (PCR)

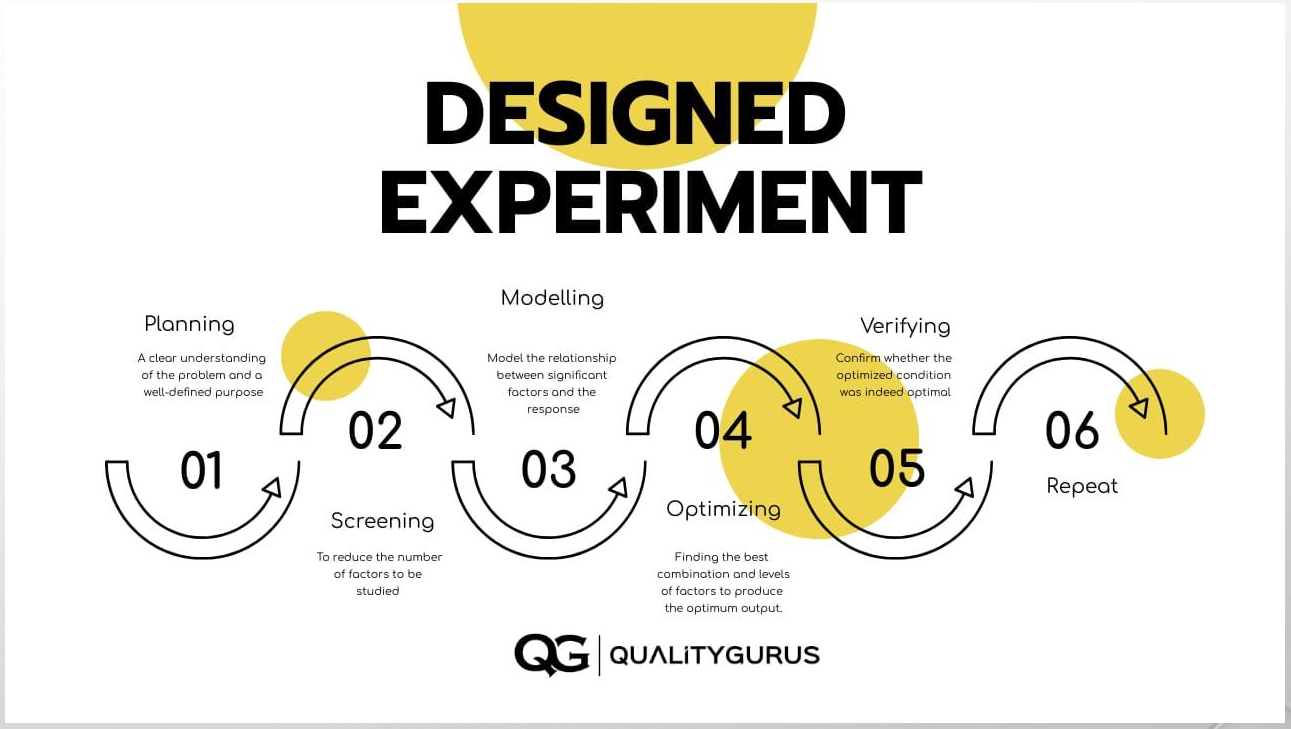
- PCR is a method for DNA replication, developed in the 1980s.
- Applications:
  - Identifying genetic defects.
  - DNA fingerprinting in forensics
  - Studying evolutionary relationships among species.

RESEARCH QUESTION

What are the optimum settings to maximize yield and purity while minimizing time?

GOALS OF THE PROJECT

- Primary goal
  - Optimize PCR by maximizing yield and purity while minimizing cycle time.
- Strategy
  - Scouting Experiment: Initial exploration of variables
  - Screening Design: Identify critical factors using Fractional Factorial Resolution III.
  - Space Filling Design: Inform final recommendations
  - Future Steps: Explore interactions, non-linearity and refine optimization with factorial or response surface designs.



OPERATIONAL CONSTRAINTS

- Reactor equation governs feasibility:

$$5 \cdot dNTP + 2.5 \cdot primer + 0.0373 \cdot plasmid + 0.25 \cdot polymerase < 50$$

- Ensures experimental runs remain within operational boundaries.
- Key implications:
  - Helps narrow feasible variable ranges for testing.
  - Invalid combinations result in failed experiments.

DATA DESCRIPTION: VARIABLES

Response Variables

Variable	Type	Description
Yield	Continuous	The amount of DNA material produced
Purity	Continuous	Purity of DNA material, values range from 0 to 100
Time	Continuous	The amount of time to complete the process

Variable	Type	Description
Number of Cycles	Continuous	The number of cycles determines how many times the DNA is amplified.
Denaturation Temperature	Continuous	Denaturation is the step where the double-stranded DNA template is separated into two single strands.
Denaturation Time	Continuous	This is the amount of time the reaction mixture is held at the denaturation temperature.
Annealing Temperature	Continuous	During the annealing step, the primers bind (anneal) to the complementary sequences on the single-stranded DNA template.
Annealing Time	Continuous	This is the amount of time the reaction mixture is held at the annealing temperature.
Extension Temperature	Continuous	The extension step is when the DNA polymerase synthesizes a new strand of DNA complementary to the template strand.
Extension Time	Continuous	The length of time the reaction is held at the extension temperature. This depends on the length of the target sequence being amplified.
DNTP Concentration	Continuous	Deoxynucleotide triphosphates (dNTPs) are the building blocks for DNA synthesis.
Primer Concentration	Continuous	Primers are short sequences of nucleotides that bind to the template DNA to initiate replication.
Plasmid Mass	Continuous	In cloning experiments, the plasmid mass refers to the amount of plasmid DNA (vector) used as a template.
Polymerase Concentration	Continuous	DNA polymerase is responsible for synthesizing the new DNA strand. The concentration of polymerase is important for controlling the rate of amplification.
Polymerase Type	Categorical	Different polymerases have different properties.

SCOUTING PHASE

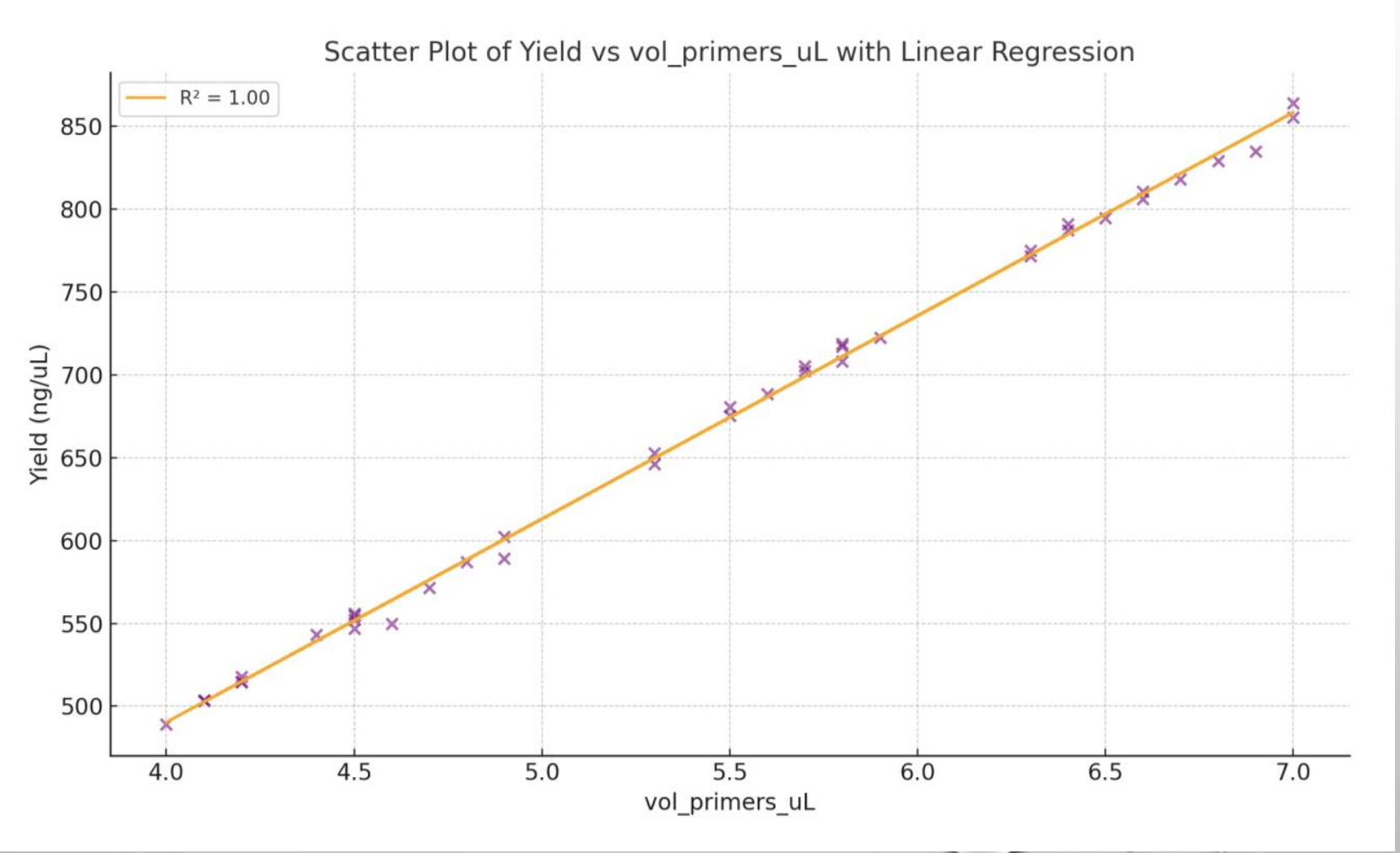
SCOUTING EXPERIMENT SUMMARY: PREDICTORS

Variable	Min	Median	Mean	Max	Correlation with Yield	Correlation with Purity
Cycles	30.0	35.00	34.88	40	0.04	0.19
Temp denat	94.0	96.15	95.97	98.0	0.03	0.05
Time denat	20.1	22.95	22.72	25.0	-0.04	0.14
Temp anneal	50.0	57.55	57.84	64.9	-0.23	-0.27
Time anneal	20.0	30.25	30.75	39.80	-0.13	0.18
Temp extend	68.2	74.95	74.44	80.0	0.01	-0.20
Time extend	8.0	30.20	32.78	59.2	0.19	0.04
DNTP	4.0	5.7	5.49	7.0	-0.08	-0.05
Primer	4.0	5.5	5.42	7.0	1.00	0.01
Plasmid Mass	2.2	11.00	11.14	19.9	-0.01	0.19
Polymerase Concentration	1.1	6	5.45	10.0	0.04	-0.26

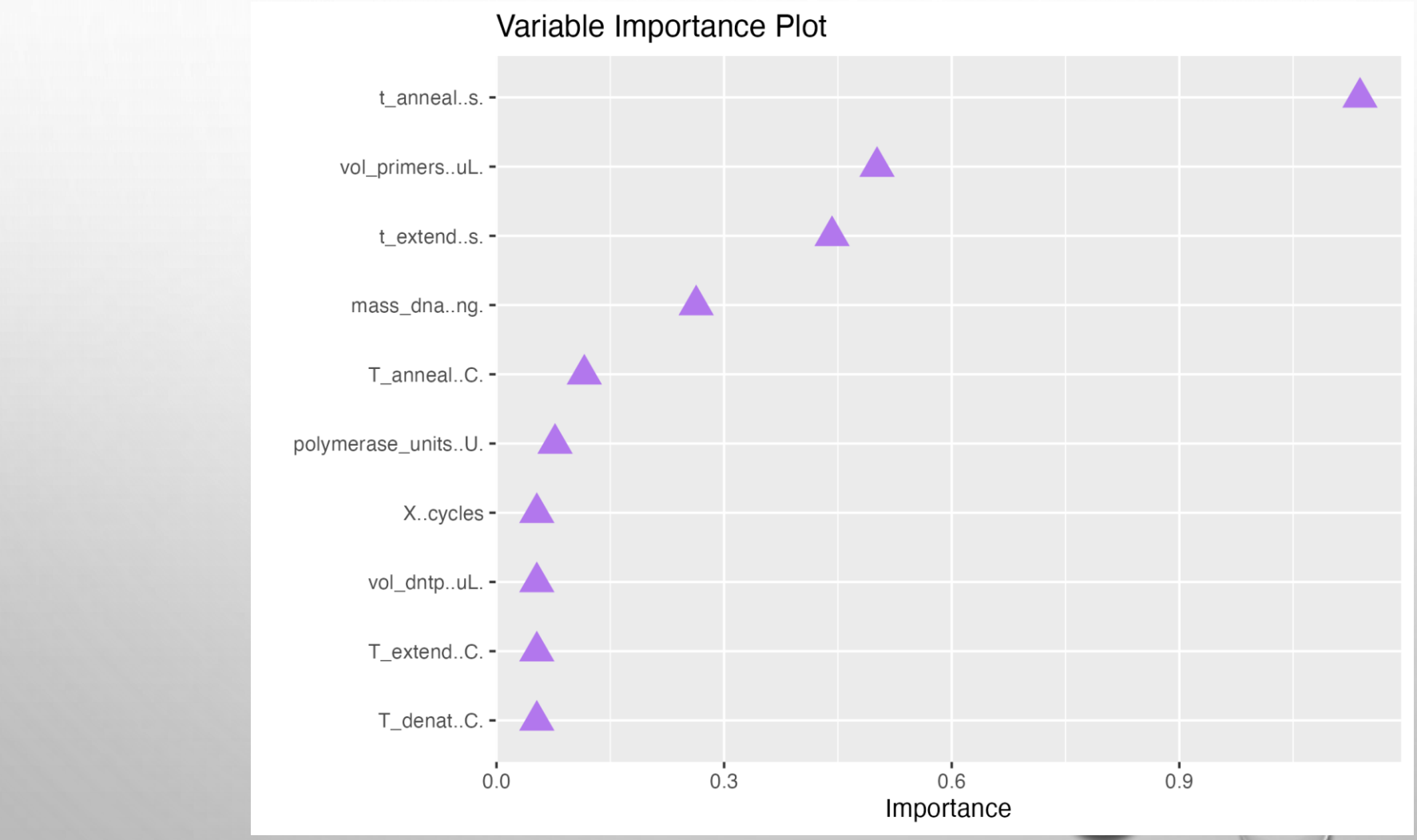
SCOUTING EXPERIMENT SUMMARY: RESPONSE

Variable	Min	Median	Mean	Max
Yield	488.9	678.10	663.92	863.9
Purity	98.2	98.9	99.2	99.5
Time	39.8	57.20	60.22	90.9

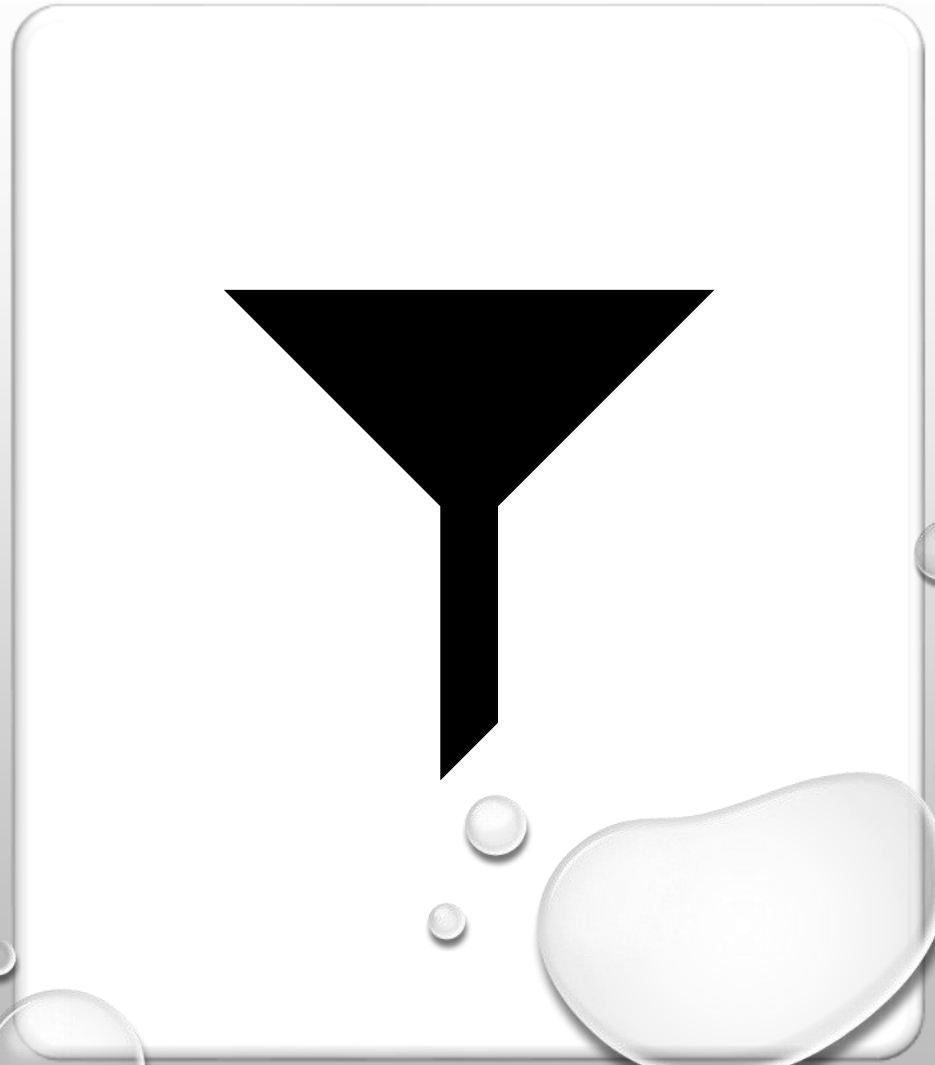
EXPLORATORY DATA ANALYSIS: YIELD



EXPLORATORY DATA ANALYSIS: PURITY



SCREENING PHASE: 16 RUNS + 6 CENTER POINTS



SCREENING EXPERIMENT DESIGN

- Experiment designed to gather information to determine the most important variables in meeting experimental goals
- 2^{12-8} Resolution III Design – Allows for analysis of main effects
  - This design was chosen to efficiently screen 12 factors in just 16 runs + 6 center point runs.
- Analysis methods
  - Combined screening data with scouting data to develop models
  - Tree regression
  - Multiple linear regression

SCREENING EXPERIMENT DESIGN CONTINUED

- Low and High Factor Levels selected in consideration of scouting data observations and patterns, operational constraints, and ensuring adequate distance between the high and low values to detect a significant effect
- Center points included to aid in identifying any non-linear patterns for a total of 22 runs

Variable	Low	Center	High
Number of Cycles	34	37	40
Temp_denat	95	96.5	98
time_denat	20	24	28
Temp_anneal	50	57.5	65
time_anneal	24	32	40
Temp_extend	70	75	80
time_extend	30	45	60
plasmid	5	12.5	20
Vol_Primers	6	6.85	7.7
Vol_Dntp	3.5	4.5	5.5
Polymerase Units	5	7.5	10
Polymerase	Phusion	3 Phusion, 3 Taq	Taq

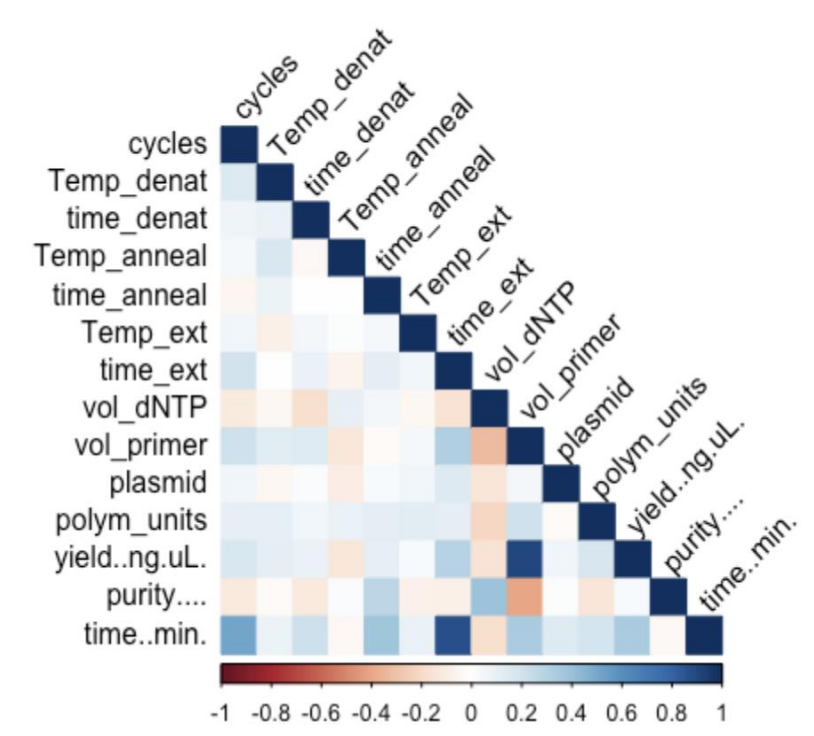
SCREENING EXPERIMENT DESIGN: SUMMARY STATISTICS

Response Variable	Min	Median	Mean	Max	SD
Yield	652.5	745.9	797.0	952.5	89.7
Purity	73.2	99.1	95.6	99.6	7.8
Time	51.9	72.3	73.3	95.3	10.4

- Moving forward in the screening phase, time became less of a focus as the ranges of observed times were deemed acceptable. However, all other considerations being equal, the final recommendations included reduced settings for predictor time variables and cycles were selected as these were each highly correlated with run time.

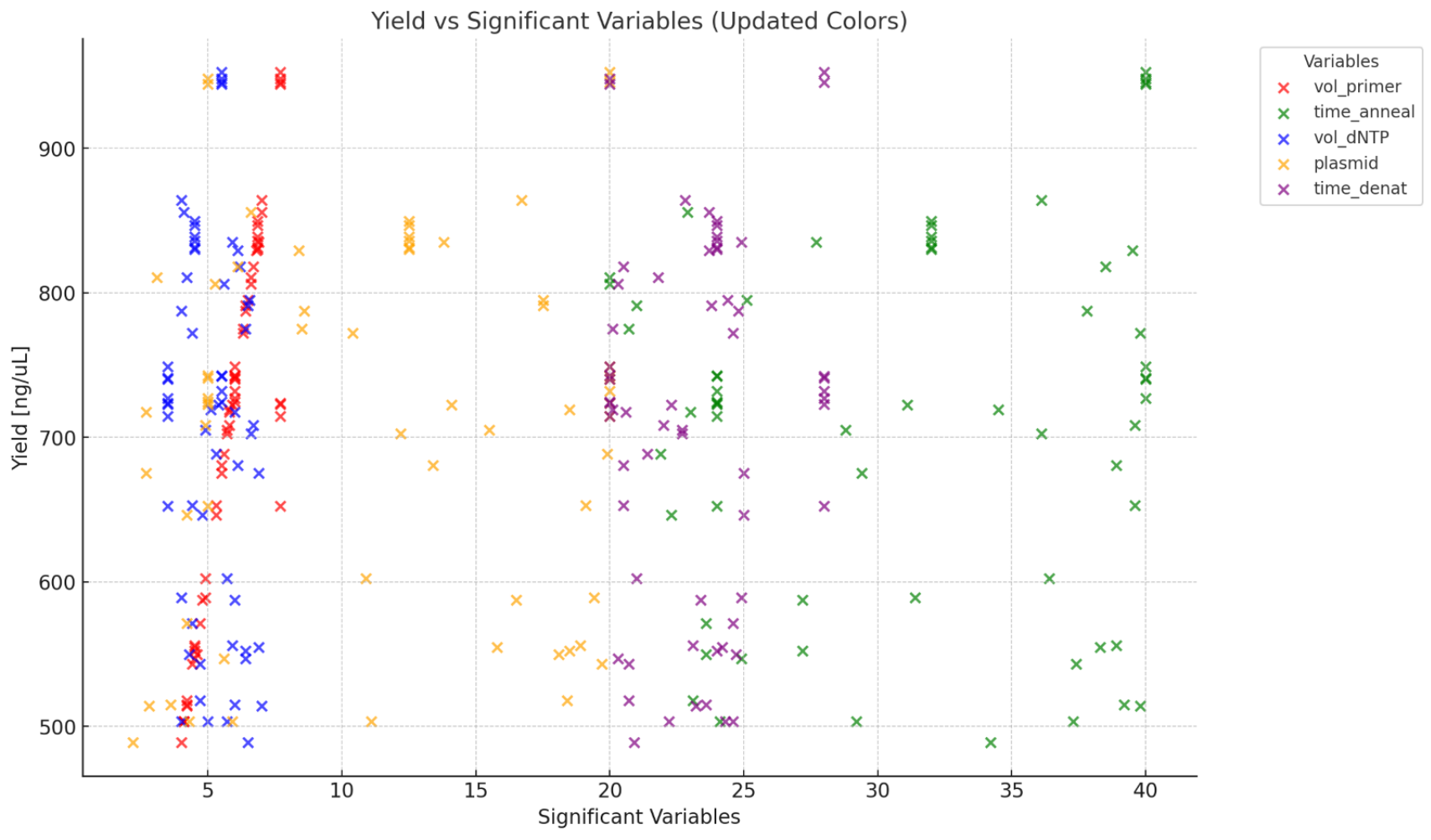
APPROACH TO COMBINING DATA

- The screening data (16 runs + 6 center points) provided critical insights into significant variables, while the scouting data expanded the design space.
- Combining these datasets allowed us to refine our analysis by leveraging both exploratory and designed experimental data.
  - Screening: Focused on identifying vital factors and detecting curvature.
  - Scouting: Offered broader coverage of the design space.

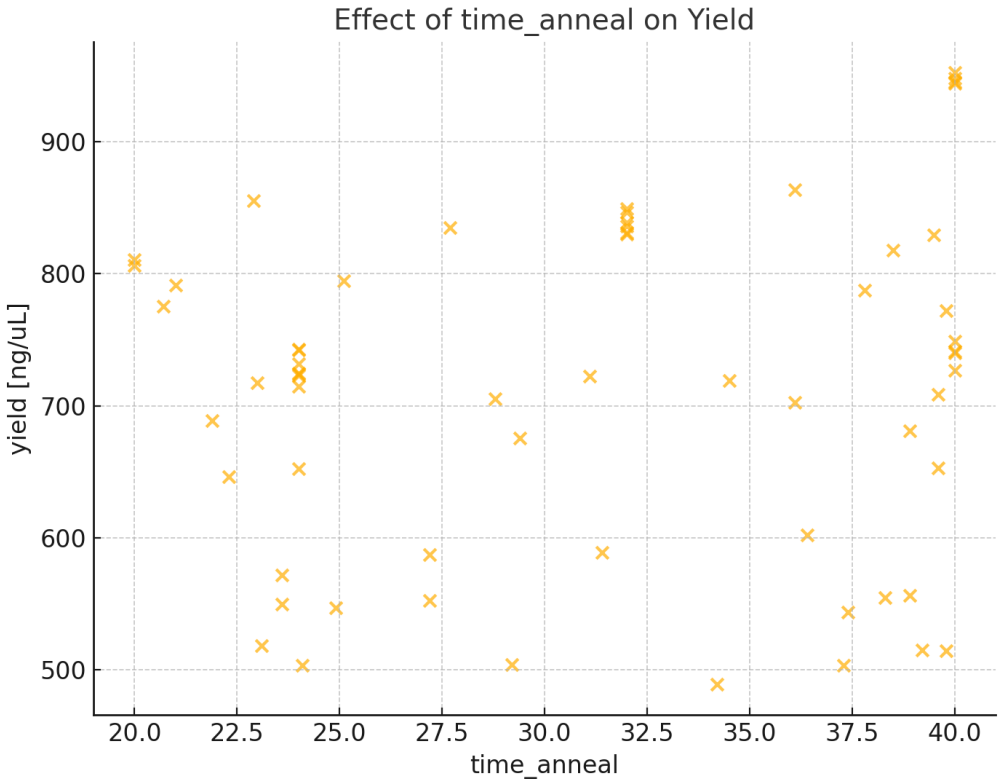


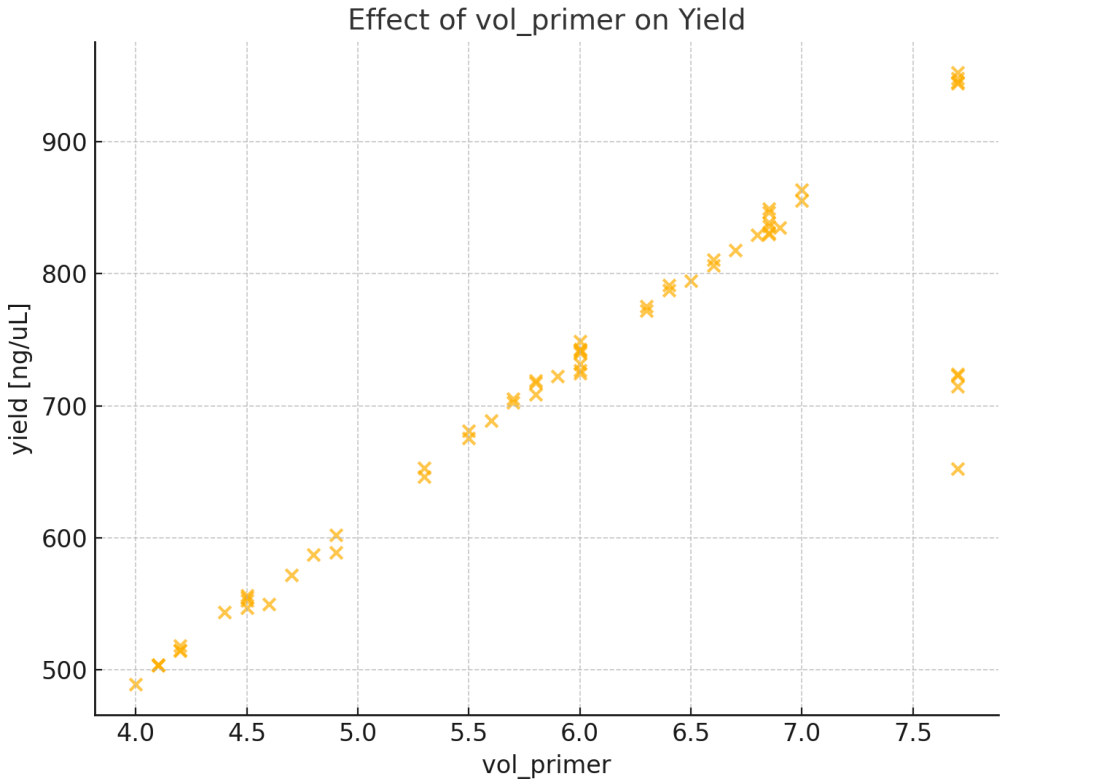
Response Variable	Mean	SD	Min	Max
Yield	711.15	126.92	488.9	952.5
Purity	97.86	4.87	73.2	99.6

SCREENING DESIGN: TREE MODELING FOR YIELD

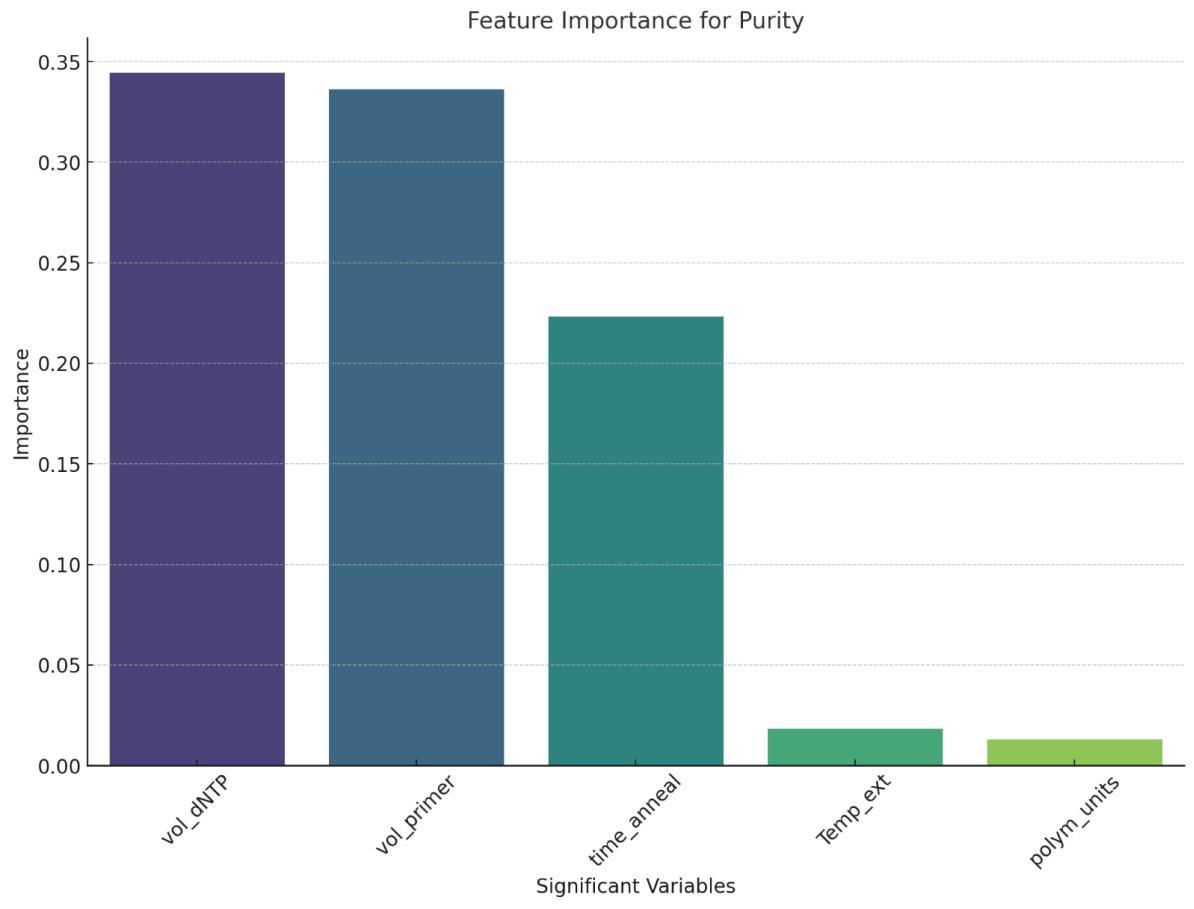


SCREENING DESIGN: TREE MODELING YIELD CONT.





SCREENING DESIGN: TREE MODELING FOR PURITY



SCREENING DESIGN: MULTIPLE LINEAR REGRESSION

MODEL FIT: PART 1

- Stepwise Regression
  - Included all predictors and response variables
    - **Purity and Yield**: Annealing Time + Volume dNTP + Volume Primer were significant.
    - **Time**: Cycles + Denaturation Time + Annealing Time + Extention Time were significant.

MODEL FIT: PART 2

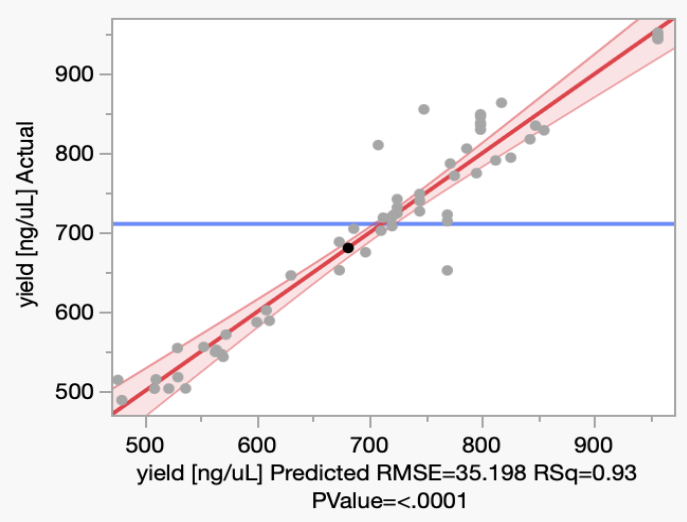
- Stepwise Regression
  - Included Annealing Time + Volume dNTP + Volume Primer + All two-way interactions
    - **Purity + Yield:** All variables were significant to these models.
    - **Time**: All variables except
      - Annealing Time + Volume Primer
      - Volume dNTP + Volume Primer

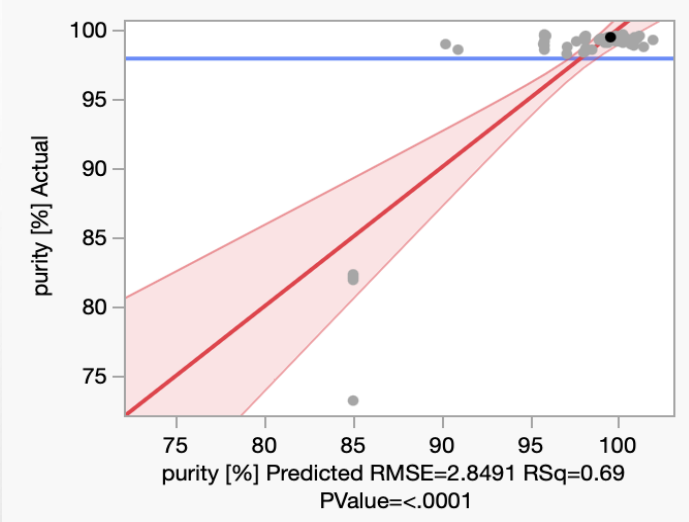
SCREENING DESIGN: MULTIPLE LINEAR REGRESSION

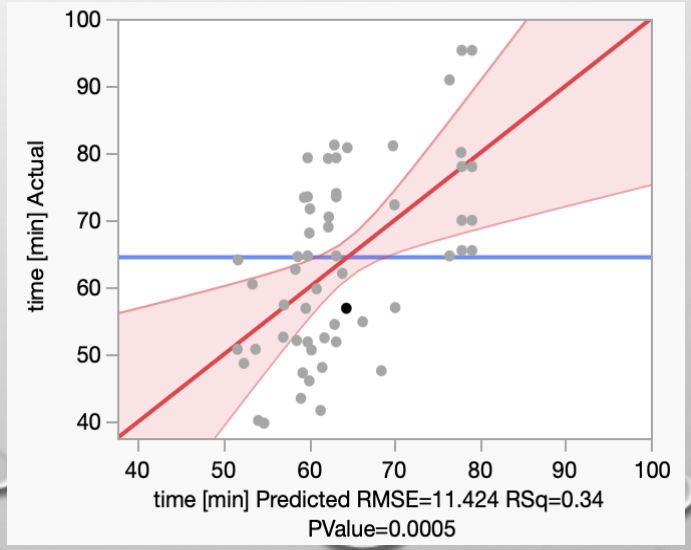
FINAL MODEL

- Response Variables
  - Yield
  - Purity
  - Time
- Predictor Variables
  - Volume Primer
  - Volume dNTP
  - Annealing Time
  - Volume Primer + Annealing Time
  - Volume dNTP + Annealing Time
  - Volume Primer + Volume dNTP

ACTUAL VS. PREDICTED PLOTS





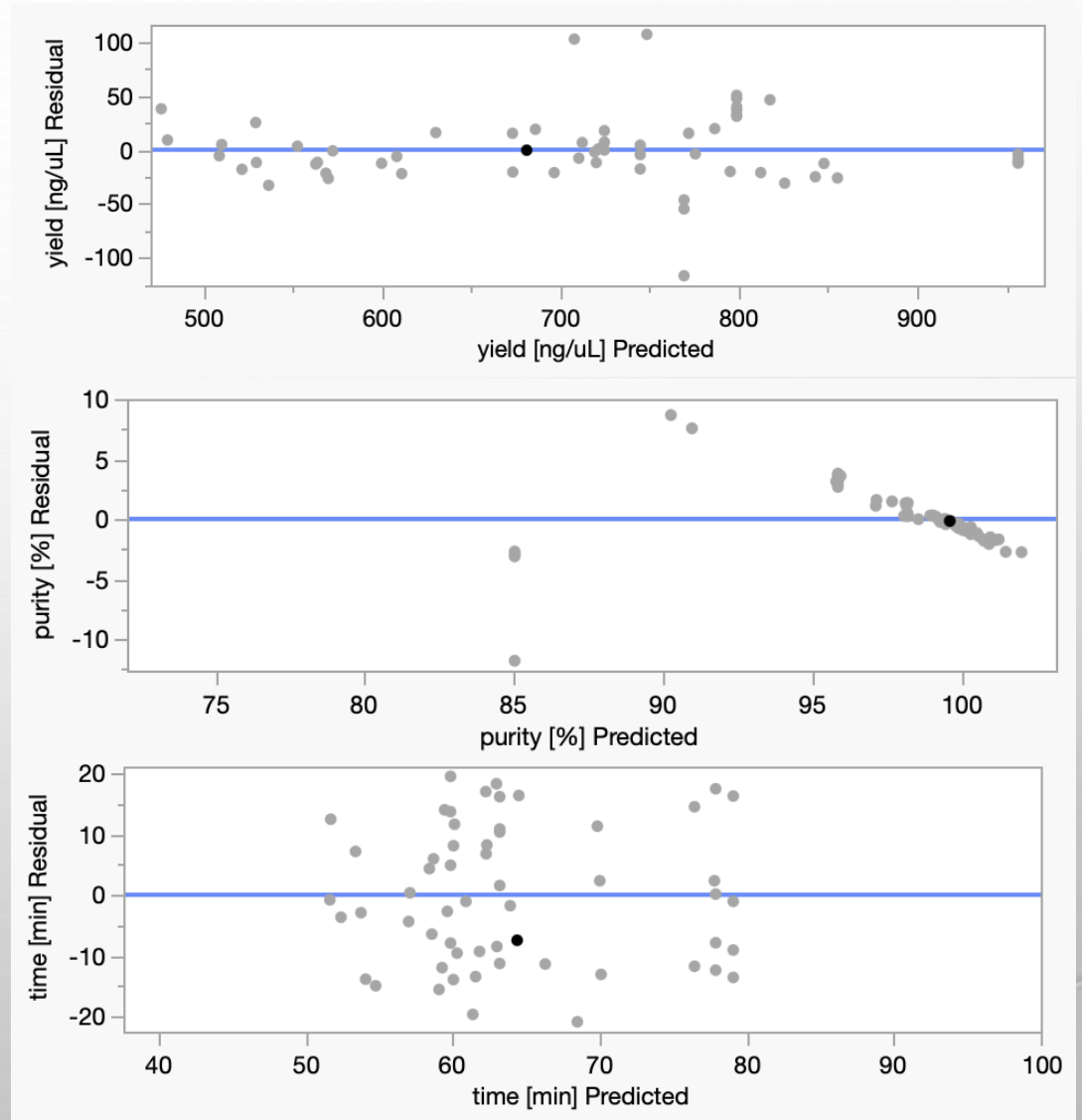


SCREENING DESIGN: MLR RESIDUALS

FINAL MODEL

- Yield: Residuals are evenly distributed with no visible curvature, indicating a good model fit.
- Purity: Minor deviations in residuals at higher predicted values; overall fit remains acceptable
- Time: Residuals are well-centered, suggesting no major bias or pattern.

ACTUAL VS. PREDICTED PLOTS



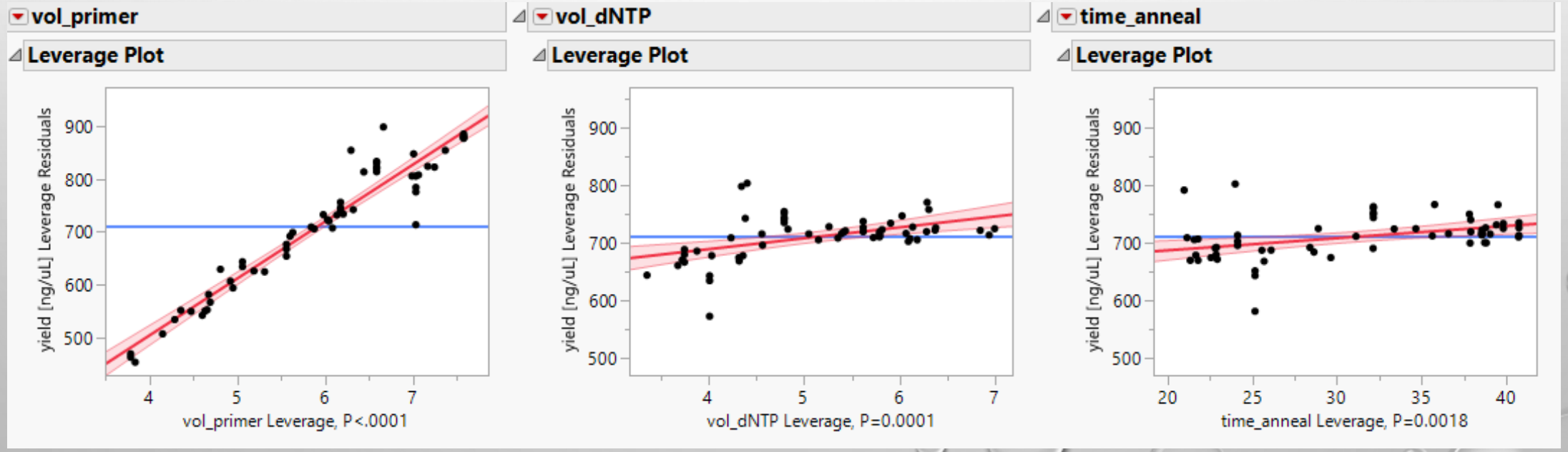
SCREENING DESIGN: ANALYSIS AND RESULTS

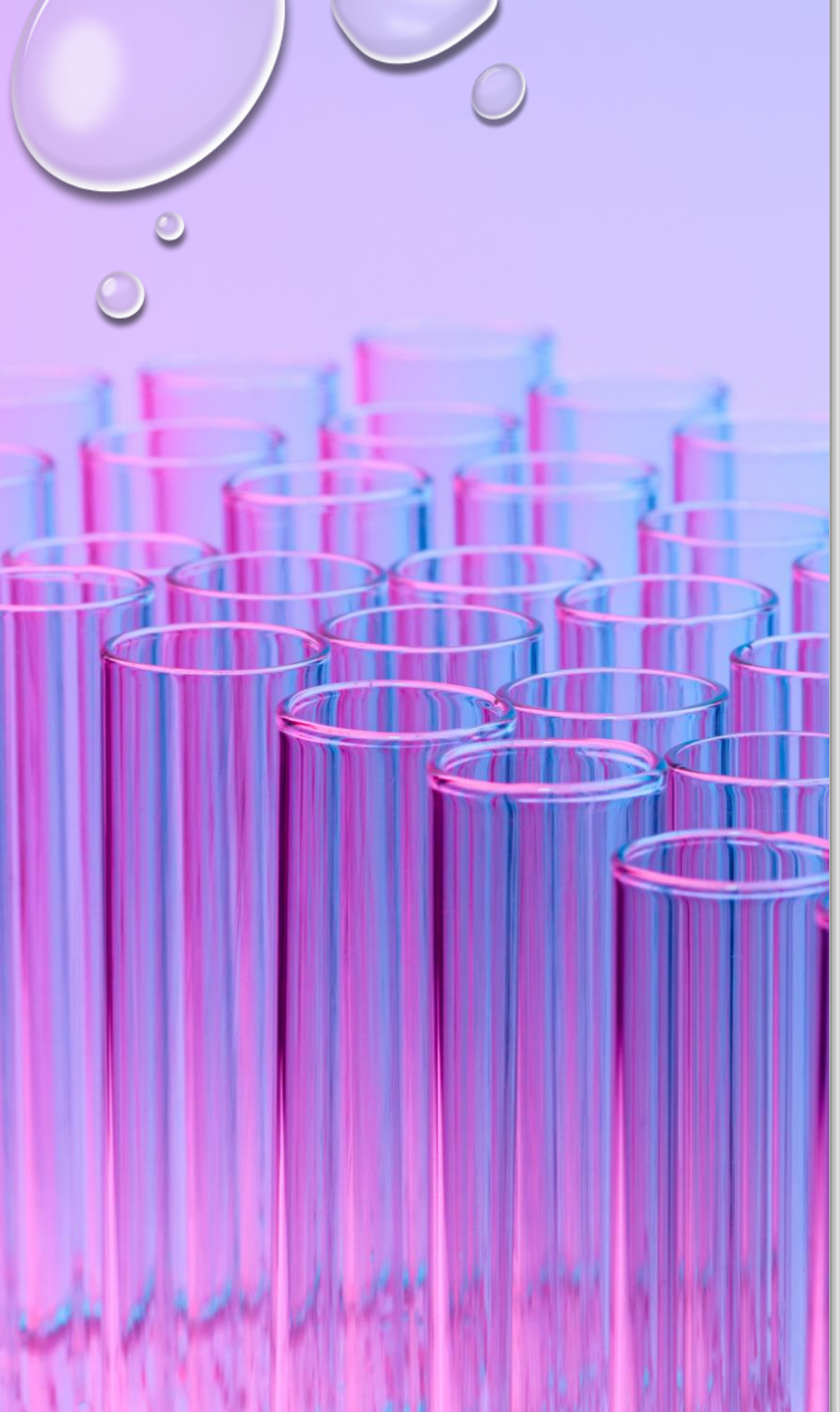
- Parameter Estimates:
  - Yield
    - Volume Primer and Volume dNTP significantly increase Yield.
    - Interaction between Volume Primer + Volume dNTP positively impacts Yield.
  - Purity
    - Increasing the Volume dNTP increases the Purity.
    - Increasing the Volume Primer decreases the purity.
  - Time
    - Volume Primer and Annealing Time increase Time
    - The interaction between Volume dNTP + Annealing Time reduces Time.

SPACE-FILLING PHASE

MOTIVATION FOR FINAL EXPERIMENT

- Patterns found in leverage plots for the final Screening Phase model led to questions about whether further optimization was possible in space beyond ranges previously tested
- Relationship between yield and primer volume of primary interest
- Leverage plots with Yield as the response:





SPACE FILLING DESIGN

- Explore results at various settings of the most important variables
  - Primer Volume Range: 7 – 8 µL
  - dNTP Volume Range: 4.8 - 6 µL
  - Annealing Time Range: 35 – 55 seconds
- Allowed for exploration of variable in areas of interest, beyond the constraints of the fractional factorial design
- Conducted 40 runs

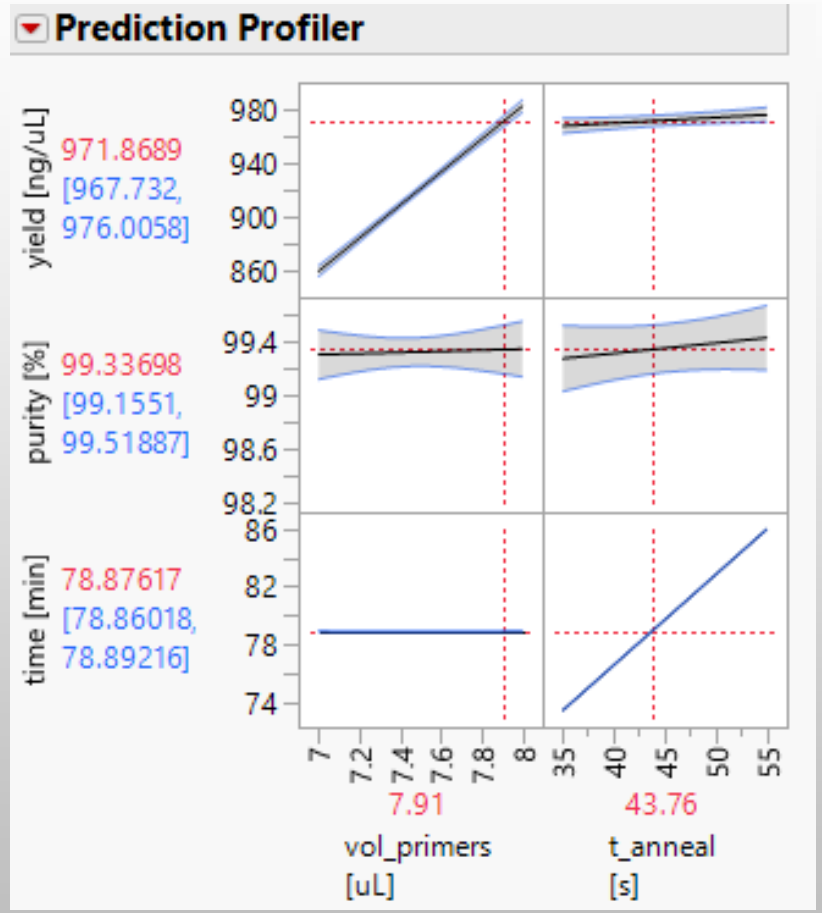
SPACE FILLING EXPERIMENT RESULTS

- Multiple Linear Regression Results: Only primer volume and annealing time found to be significant in prediction of any of the response variables in this space filling experiment
- Neither dNTP volume nor any interactions were found to be significant

Variable	Yield	Purity	Time
Volume Primers	122.6 (p=<0.0001)	0.0424 (p=0.794)	-0.0012 (p=0.928)
Annealing Time	0.419 (p=0.028)	0.00077 (p=0.348)	0.633 (p=<0.0001)

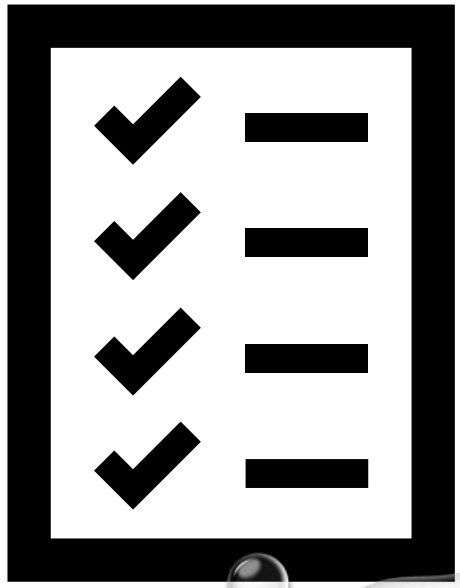
SPACE FILLING EXPERIMENT RECOMMENDATIONS

- Setting Recommendations
  - 7.9 µL Prime Volume
  - 43.8 seconds Annealing Time
  - 5.3 µL dNTP Volume



Model Predictions			
Response	Estimate	95% CI Lower Bound	**95% CI Upper Bound**
Yield	971.87	967.73	976.00
Purity	99.33	99.16	99.52
Time	78.88	78.86	78.89

FINAL RECOMMENDATIONS + CONCLUSIONS



FINAL RECOMMENDED CONFIGURATION

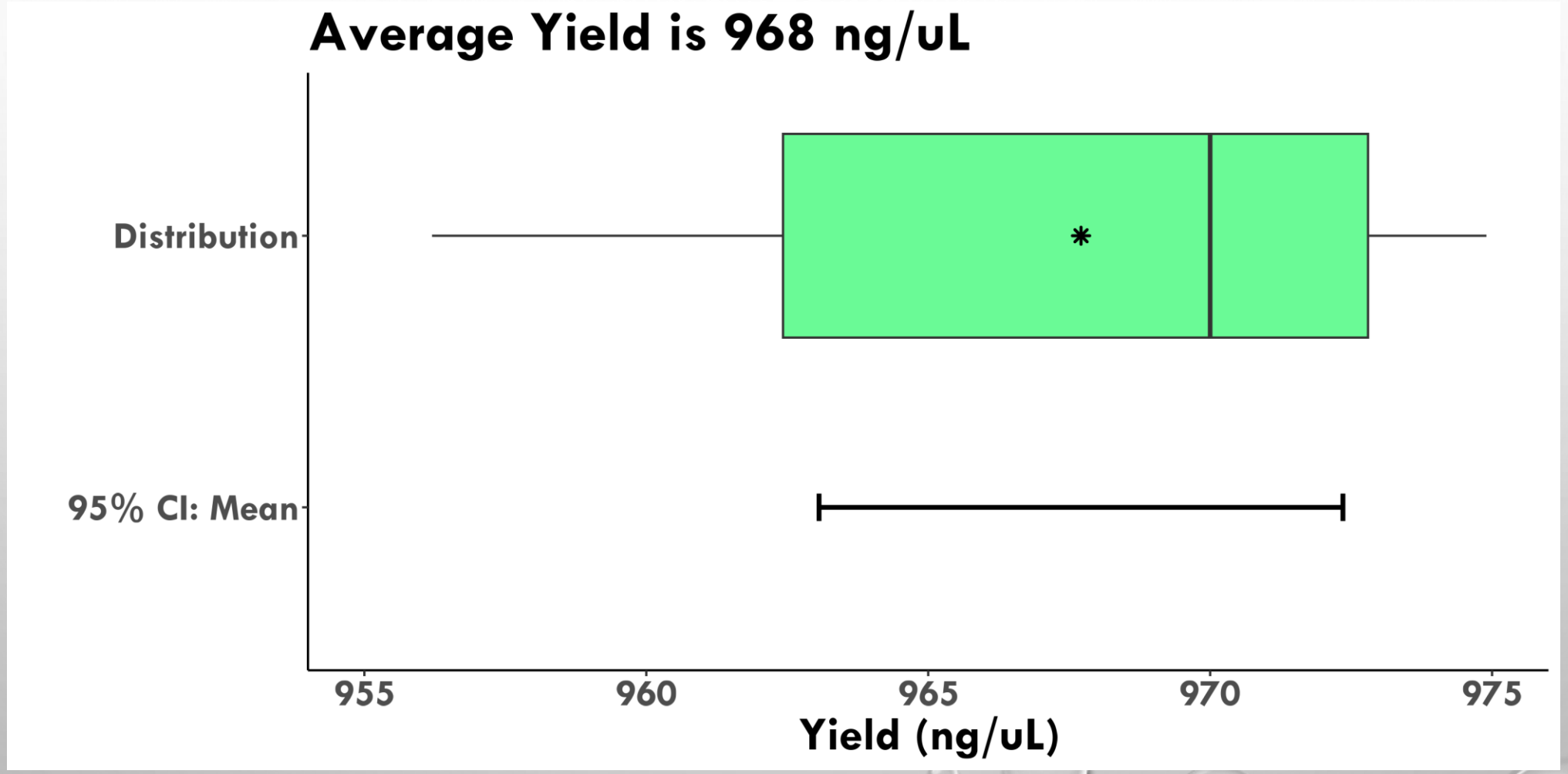
Variable	Setting		Variable	Setting
Denaturation Temp (C)	96.5		Cycles	38
Denaturation Time (s)	20		Volume Primer (µL)	7.9
Annealing Temp (C)	57.5		Volume dNTP (µL)	5.3
Annealing Time (s)	43.8		Plasmid	16
Extension Temp (C)	72.5		Polymerase Units	7.5
Extension Time (s)	45		Polymerase Type	Phusion

RESULTS

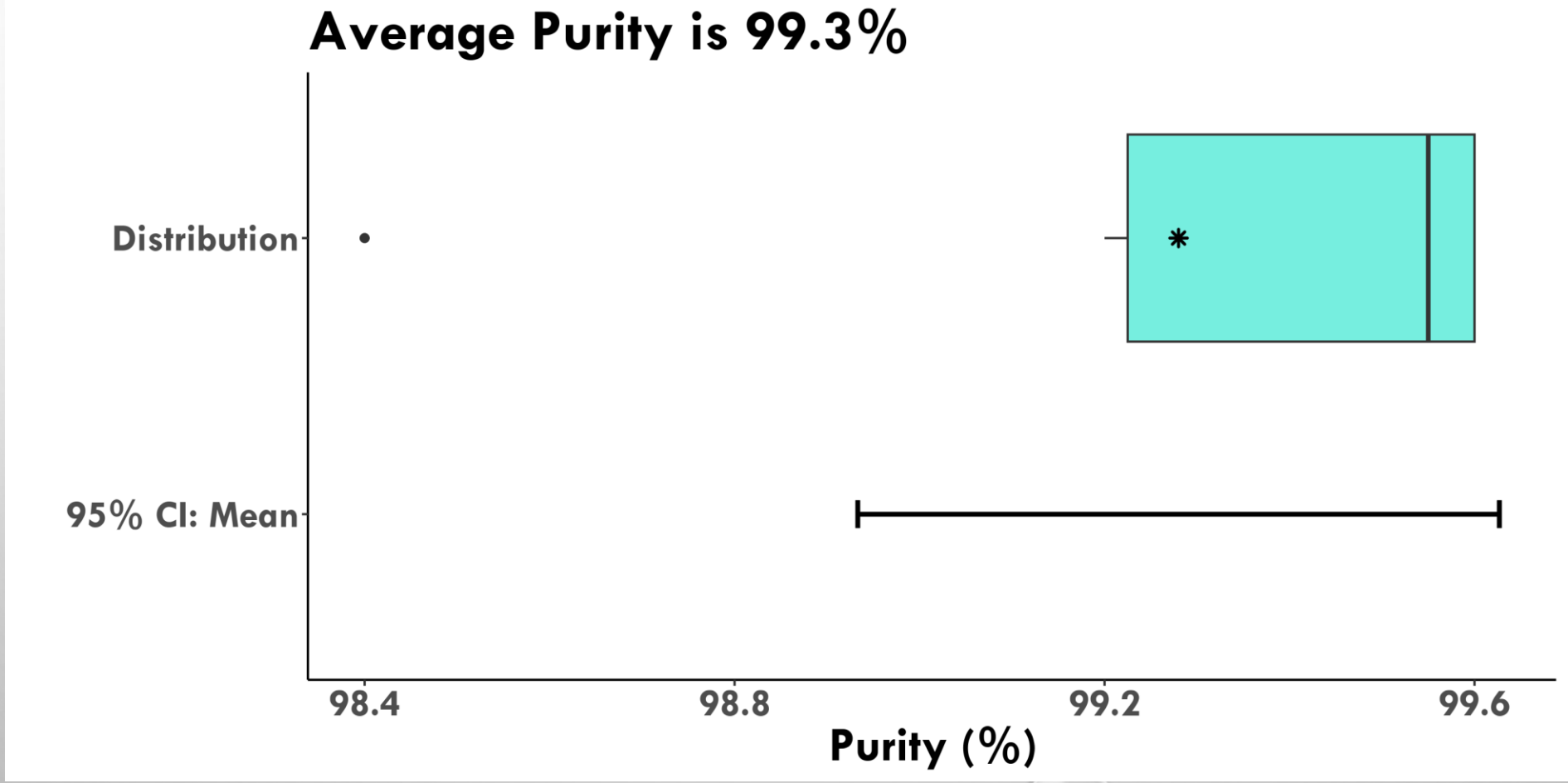
- Executed 10 confirmation runs

Response	Mean	Standard Error	Lower Bound 95% CI	Upper Bound 95% CI
Yield (ng/uL)	967.71	2.054	963.064	972.356
Purity (%)	99.28	0.153	98.933	99.627
Time (min)	78.90	0.000	78.900	78.900

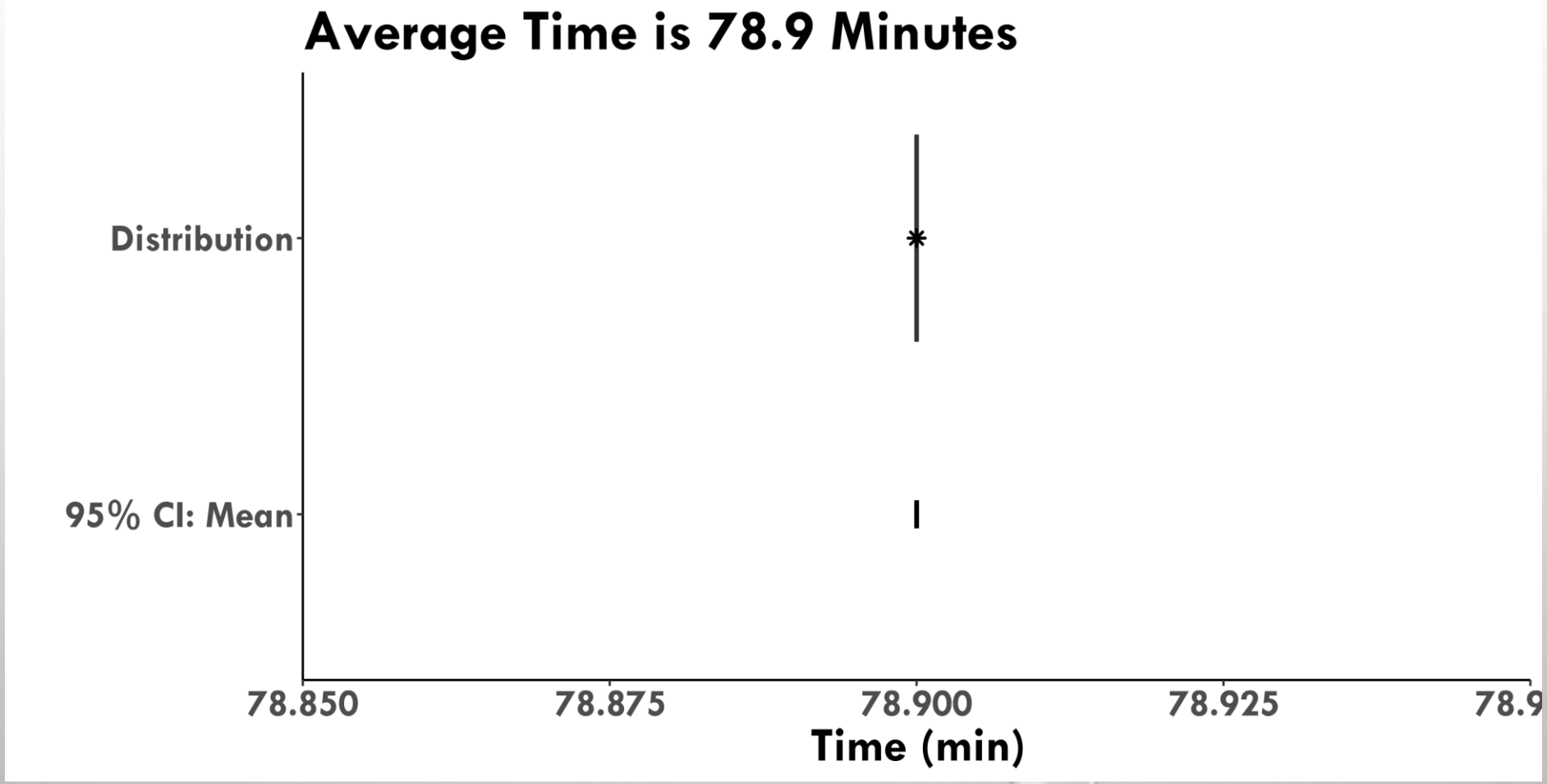
CONFIRMATION RESULTS: YIELD



CONFIRMATION RESULTS: PURITY

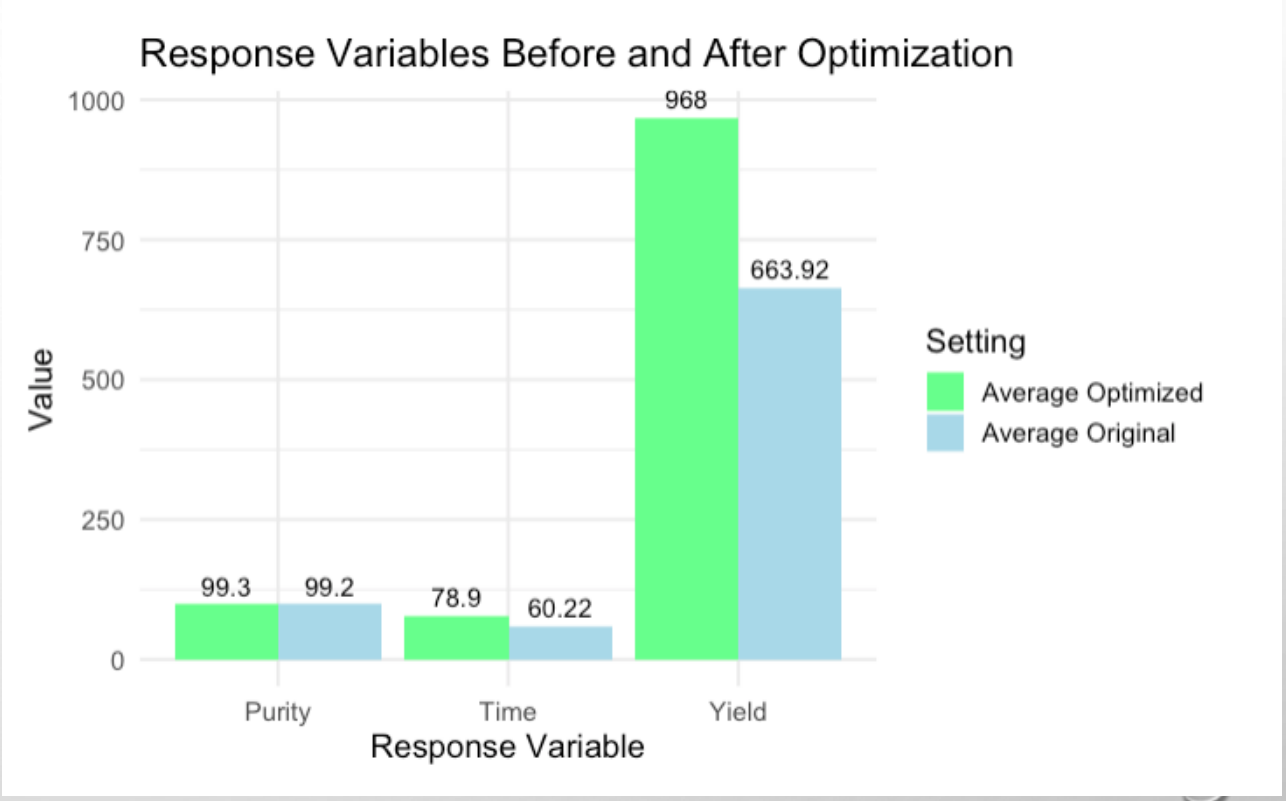


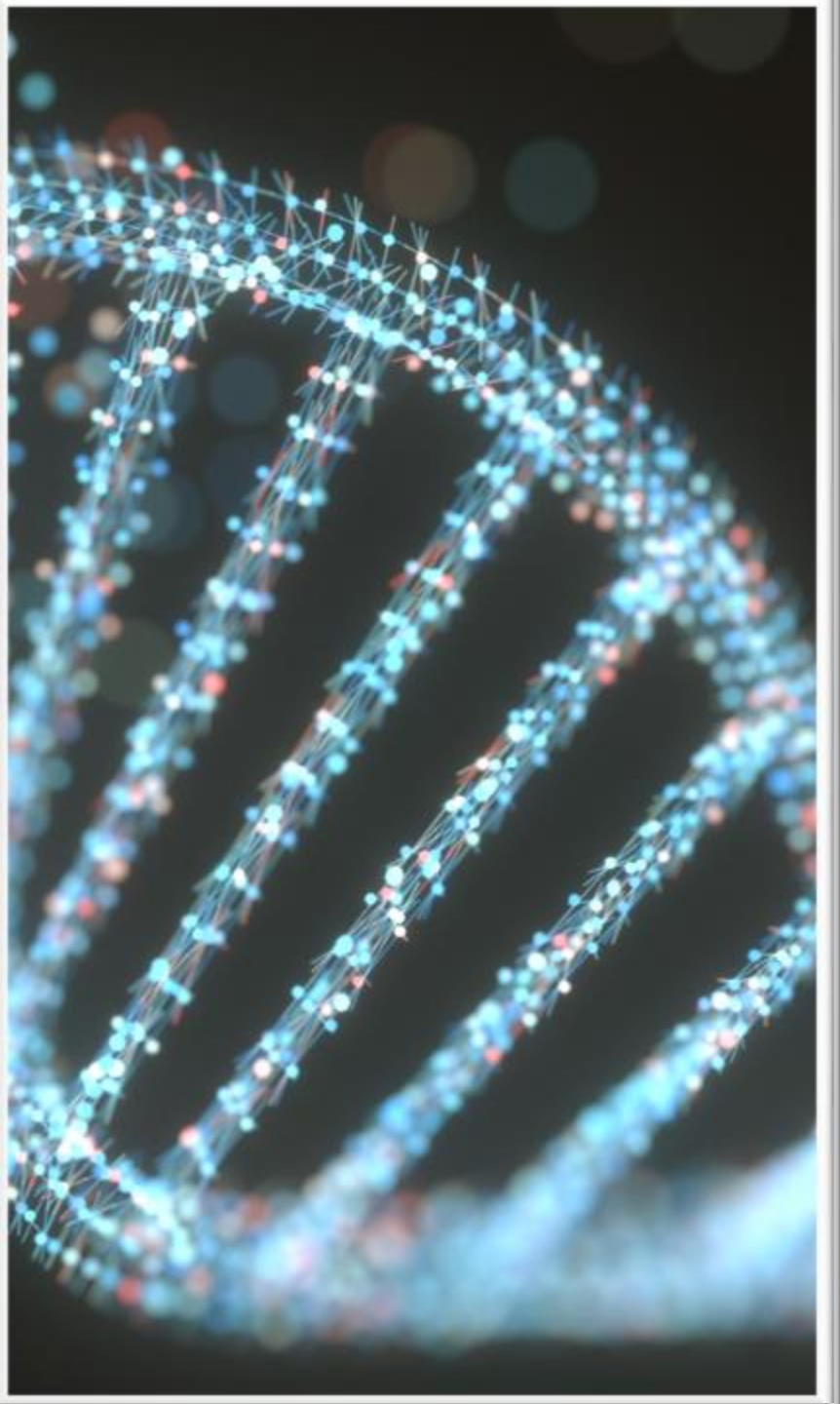
CONFIRMATION RESULTS: TIME



OPTIMIZING THE PCR PROCESS

Variable	Average Original Settings (scouting data)	Final Optimized Settings
Volume Primer	5.42	7.9
Volume dNTP	5.49	5.3
Annealing Time	30.75	43.8
Annealing Temp	57.84	72.5
Extension Time	32.78	45
Extension Temp	74.44	72.5
Number of Cycles	34.88	38
Polymerase Concentration	5.45	7.5
Denaturation Time	22.72	20
Denaturation Temp	96.97	96.5
Plasmid Mass	11.14	16
Polymerase Type	Phusion + Taq	Phusion





CONCLUSIONS

- Screening design identified primer volume, dNTP concentration, and annealing time as critical variables in influencing yield and purity.
- Space-filling design fine-tuned settings, achieving maximum yield and purity within operational constraints.

FURTHER STUDIES

- Further exploration of interactions using a Resolution IV Fractional Factorial Design
- Explore non-linearity further within different predictor variables.
- Perform a Response Surface Design to refine optimization further.
- Pilot the recommended configuration in an operational setting to validate results.