Modeling Heterogeneity in phrA-Mediated Signaling

Jacqueline Deng, Betty Hou, Calum Zhang

May 2025

1. Introduction

This report investigates cell–cell communication in *Streptococcus pneumoniae* (SPN) mediated by the signaling peptide phrA, with an emphasis on quantifying spatial heterogeneity in gene induction. We analyze how variability in phrA-driven fluorescence differs within and across microfluidic droplets containing bacterial cells. Each droplet functions as a semi-isolated environment with stochastic variation in cell count and local signaling conditions.

phrA participates in a peptide-based quorum sensing system: cells secrete phrA peptides, which accumulate extracellularly and activate downstream transcriptional pathways upon internalization. A fluorescent reporter gene under phrA-responsive regulation enables quantification of this signaling activity at the single-cell level. Fluorescence intensity serves as a proxy for the degree of phrA induction.

Our analysis focuses on three experimental conditions:

- **Wild-type (WT)**: cells with intact endogenous *phrA* signaling.
- **Knockout (KO)**: peptide-generating machinery is disrupted; cells cannot initiate signaling but retain the capacity to respond.
- **WT + Peptide**: wild-type cells exposed to exogenously added synthetic *phrA* peptide.

The primary analytical objective is to quantify heterogeneity in fluorescence expression attributable to droplet-level effects. Specifically, we ask:

> *Does phrA induction show greater heterogeneity within droplets or across droplets? How does this differ across experimental conditions?*

To answer this, we apply mixed-effects models to single-cell fluorescence data, estimating the proportion of total variance explained by droplet-level clustering (intraclass correlation coefficient, ICC). While full time-course data are

available, the 4-hour (4h) time point—where fluorescence intensity peaks in WT—is used as the focal point for most analyses. Additional models evaluate group and interaction effects over time.

2. Methods

Exploratory Data Analysis (EDA)

Initial visualization of single-cell fluorescence data was performed to assess distributional properties and potential sources of heterogeneity. Each observation corresponds to the fluorescence intensity of an individual cell within a microfluidic droplet. To address overplotting and facilitate interpretability, boxplots were constructed for a randomly selected subset of 10 droplets per experimental condition.

Raw fluorescence values exhibited substantial right skew, consistent with heterogeneous transcriptional activation and saturation effects at high expression levels. A log-transformation was applied to stabilize variance, suppress outliers, and better satisfy assumptions of normality required for downstream Gaussian-based modeling approaches.

Variance Decomposition via Mixed-Effects Modeling

To quantify the relative contribution of within-droplet versus between-droplet variability in *phrA*-induced fluorescence, we employed a linear mixed-effects model of the form:

$$\log(\text{intensity}_{ij}) = \mu + u_j + \varepsilon_{ij}$$

where:

- i indexes cells within droplet j,
- μ is the overall log-scale mean fluorescence intensity,
- $u_j \sim \mathcal{N}(0, \sigma_u^2)$ represents the random intercept for droplet j (i.e., between-droplet variation),
- $\varepsilon_{ij} \sim \mathcal{N}(0, \sigma^2)$ captures within-droplet residual variation.

This model allows decomposition of total fluorescence variance into between-droplet and within-droplet components. The proportion of total variance attributable to droplet-level clustering is summarized by the intraclass correlation coefficient (ICC):

$$\text{ICC} = \frac{\sigma_u^2}{\sigma_u^2 + \sigma^2}$$

An ICC close to zero suggests minimal inter-droplet heterogeneity, whereas a higher ICC indicates substantial droplet-level effects—potentially due to localized peptide signaling environments or microfluidic variability.

Separate models were fit for the wild-type (WT) and knockout (KO) conditions at the 4-hour time point to isolate genotype-specific patterns in spatial signaling heterogeneity. Additional models incorporate time as a covariate and evaluate interaction terms (e.g., Group $\times$ Time) where applicable.

Implementation and Validation

All models were implemented using the `lme4` package in R. Log-transformed fluorescence intensity served as the response variable. Model diagnostics, including residual plots and convergence criteria, confirmed adequate model fit. ICC estimates were also computed across multiple time points to characterize temporal trends in signaling variability.

3. Results

3.1 WT and KO at 4 Hours (Analyzed Separately)

3.1.1 Basic Exploratory Data Analysis

We begin with an exploratory analysis of the 4-hour datasets comparing wild-type (WT) and knockout (KO) strains. Each experiment (230111, 230113, 230611) includes multiple droplets, each containing a variable number of cells.

Figure 1 shows raw fluorescence intensity for 10 randomly sampled droplets from the WT group in experiment 230111. Substantial heterogeneity is observed both within and between droplets. Some droplets exhibit wide interquartile ranges and outliers exceeding 15,000 units (e.g., droplet 4h_3_10), while others are tightly clustered, suggesting reduced intra-droplet variability.

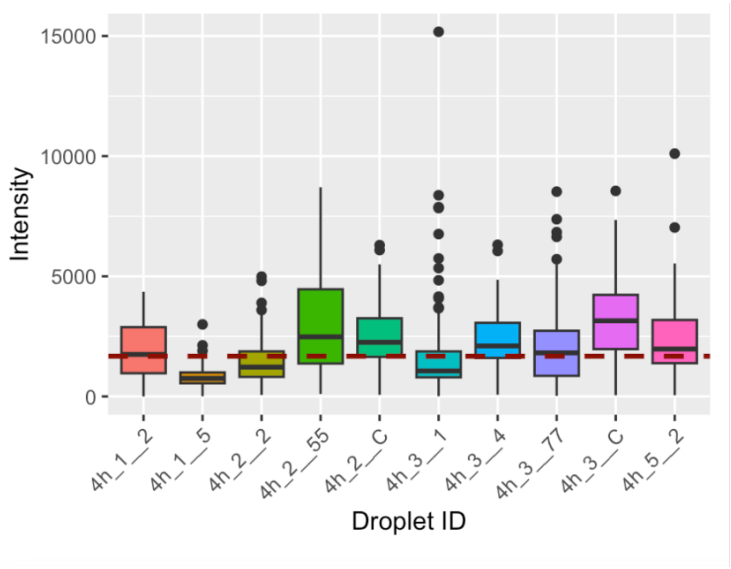

Figure 1: Raw Fluorescence Intensity for 10 Randomly Selected Droplets (WT_230111, 4h)

Due to extreme skew in the raw data, a log transformation was applied to compress the dynamic range and improve comparability across droplets. The resulting distribution (Figure 2) is approximately symmetric, facilitating model-based inference.

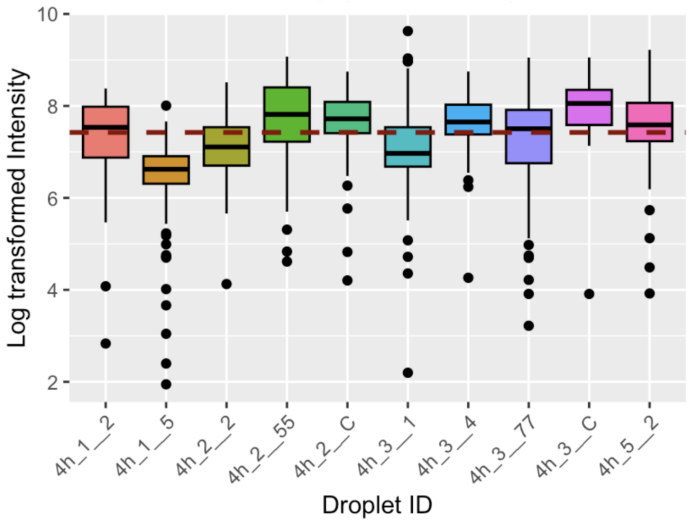

Figure 2: Log-Transformed Fluorescence Intensity for 10 Random Droplets (WT_230111, 4h)

We then compared WT and KO samples side-by-side (Figure 3). WT droplets show greater variance and more extreme upper outliers than KO droplets. Horizontal lines mark the mean of droplet-level means within each group. WT consistently exhibits higher average log intensity and greater dispersion across droplets, suggesting functional peptide signaling and inter-droplet heterogeneity.

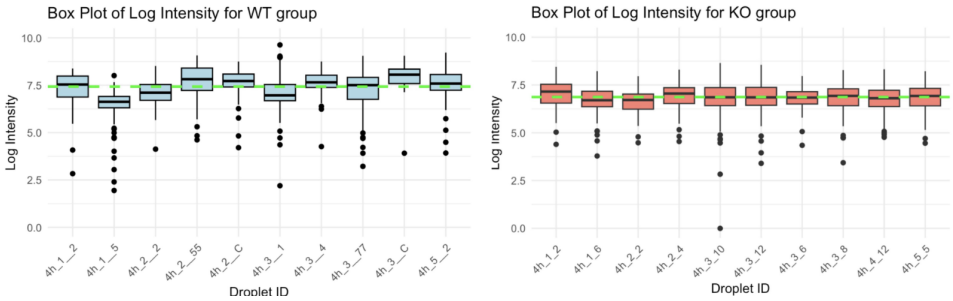

Figure 3: Log Fluorescence Intensity by Droplet for WT and KO (230111, 4h)

Trends in experiments 230113 and 230611 are similar and included in the appendix.

3.1.2 Random-Effect Model of WT and KO

To quantify the degree of fluorescence heterogeneity across spatial compartments, we fit a linear mixed-effects model separately for WT and KO groups, respectively. This model includes a single fixed intercept term and a random intercept for each droplet, allowing us to decompose total variance into within-droplet and between-droplet components. The random effect captures spatial clustering by estimating the extent to which cells within the same droplet exhibit correlated fluorescence levels.

The model specification is: in the computer package R is

$$\texttt{lmer(log.intensity} \sim 1 + (1 \mid \texttt{droplet\_ID}))$$

Table 1: Log-Intensity @ 4 hrs (WT and KO dataset)

Experiment	KO	WT
Exp #1		
Mean	6.753	7.335
ICC	4.90%	13.57%
Exp #2		
Mean	7.1048	7.6331
ICC	7.27%	18.49%
Exp #3		
Mean	6.7116	7.237
ICC	8.79%	18.64%

The comparison between the wild-type (WT) and knockout (KO) groups provides foundational insight into the biological significance of phrA-induced communication in Streptococcus pneumoniae. Across all three experiments, WT cells consistently demonstrated higher mean log-fluorescence intensity values at the 4-hour timepoint relative to KO cells (Table 1). Specifically, the mean intensity in WT groups exceeded KO groups by approximately 0.5 to 0.6 log units across all experiments, which corresponds to a substantial increase in signal magnitude when interpreted on the original fluorescence scale.

This observed elevation in mean intensity among WT cells supports our hypothesis that endogenous production of the phrA peptide actively contributes to cell-cell communication. In contrast, the KO groups, which lack the peptide generator but retain the receptor, show substantially lower fluorescence, implying that without internal phrA signaling, cells remain relatively inactive in the signaling pathway.

However, the group means alone do not fully address our central research question, which concerns the nature of heterogeneity in signal expression — specifically, whether signaling variance is greater within a droplet (i.e., among

cells sharing the same microenvironment) or between droplets (i.e., across spatially separated communities). To rigorously assess this, we turn to the intraclass correlation coefficient (ICC), which quantifies the proportion of total variance in fluorescence attributable to between-droplet variation.

As shown in Table 1, ICC values for WT groups ranged from 13.57% to 18.64% across the three experiments, whereas KO groups displayed markedly lower ICCs, ranging from only 4.90% to 8.79%. This suggests that WT cells not only express higher average fluorescence but do so with greater heterogeneity across droplets. In other words, there is considerable variation in signaling intensity from droplet to droplet in the WT condition, whereas KO droplets are more homogeneous.

Biologically, this differential pattern of heterogeneity aligns well with our understanding peptide signaling systems in bacterial communities. In WT groups, endogenous production of phrA may vary across droplets due to difference levels of induction and localized differences in cell density. As a result, signaling activation is not uniformly distributed. In contrast, the KO condition effectively removes this endogenous layer of variability. Since these cells cannot produce phrA, their activation levels depend purely on environmental noise or experimental artifact, which leads to relatively consistent (and low) fluorescence across all droplets.

Droplet-Level Variance Confirmed by ANOVA

To independently validate the presence and significance of droplet-level heterogeneity, we conducted a one-way ANOVA where droplet ID was treated as a categorical predictor of fluorescence intensity. This analysis was performed separately for WT and KO datasets. The ANOVA tests whether the variance across different droplets exceeds what we would expect due to random noise.

Table 2: One-way ANOVA: Droplet-level clustering effect

Group	Df (Droplets)	F-value	P-value	Mean Sq (Droplets)
WT	35	15.86	$< 2 \times 10^{-16}$	12.172
KO	33	7.081	$< 2 \times 10^{-16}$	3.842

As shown in Table 2, the effect of droplet ID is highly statistically significant in both groups ($p < 2 \times 10^{-16}$). However, the effect size, as measured by the F-statistic and the mean square between groups, is substantially larger in the WT group (F = 15.86) than in the KO group (F = 7.081), though both are statistically significant. This result suggests that droplet-level variability is more dominant in WT populations, which aligns with our ICC results where WT groups consistently showed ICCs exceeding 10%, compared to KO groups' lower ICCs (e.g., 4.90%–8.79%).

In conclusion, the WT vs KO comparison highlights two key insights: (1) phrA is essential for initiating effective cell-cell signaling in SPN, as evidenced by

the consistent increase in fluorescence in WT cells; and (2) this signaling is not uniformly expressed, but instead shows significant between-droplet variation, which we can statistically confirm through ICC analysis and ANOVA tests.

3.2 WT and KO at 4 Hours (Group Effect)

3.2.1 Mixed-Effect Model of WT and KO

In this section, we directly compare wild-type (WT) and knockout (KO) groups by modeling the group effect explicitly. We used two modeling frameworks: a mixed-effects model that includes droplet-level random intercepts, and a fixed-effects model that treats all observations as independent. The purpose of this comparison is to evaluate both the consistency of group effect estimation across methods and the influence of droplet-level clustering on statistical inference.

The mixed-effects model is specified as follows: using the R programming language

$$\texttt{lmer(log.intensity} \sim 1 + \texttt{factor(group)} + (1 \mid \texttt{droplet\_ID}))$$

The fixed-effects model is specified as follows:

$$\texttt{lm(log.intensity} \sim 1 + \texttt{factor(group)})$$

Table 3: **Comparison of KO vs WT @ 4 hrs**

Exp	Measure	Mixed Model		Fixed Effects	
		KO	WT	KO	WT
Exp #1	Mean (KO)	6.752		6.764	
	Group Effect (WT)	**0.582**		**0.552**	
	t-value (group)	8.554		29.61	
	ICC	10.35%		NA	
Exp #2	Mean (KO)	7.10398		6.82107	
	Group Effect (WT)	**0.43544**		**0.52545**	
	t-value (group)	5.401		32.27	
	ICC	9.55%		NA	
Exp #3	Mean (KO)	6.722		6.873	
	Group Effect (WT)	**0.498**		**0.546**	
	t-value (group)	22.04		32.74	
	ICC	12.06%		NA	

Across all three experiments, the estimated group effects—defined as the difference in mean log fluorescence intensity between WT and KO cells—were consistently positive and statistically significant. This confirms that WT cells exhibit stronger signal activation than KO cells, reinforcing our earlier observation that phrA-mediated signaling is functionally active and contributes meaningfully to

the observed fluorescence intensities. In the mixed-effects model, the WT group effect estimates ranged from approximately 0.435 to 0.582 log-units. These values indicate a biologically meaningful increase in signal activation that cannot be explained by random variation.

Importantly, the fixed-effects and mixed-effects models yielded similar point estimates for the group effect, suggesting that the WT-KO group difference is robust and not highly sensitive to model specification. This consistency validates the use of simpler linear models for estimating central tendencies. However, there is a marked divergence in the associated t-statistics.

Significance Testing and the Role of Standard Errors

To further contextualize the group effect comparison between mixed-effects and fixed-effects models, it is crucial to understand how statistical significance is derived. The t-statistic is calculated as the ratio of the estimated effect to its standard error:

$$t = \frac{\hat{\beta}}{\mathrm{SE}(\hat{\beta})}$$

This formula implies that underestimating the standard error (SE) will inflate the resulting t-value, potentially leading to an overstatement of statistical significance. This is precisely what occurs in the fixed-effects model as it does not account for between-droplet variability, leading to artificially small SEs and, in turn, substantially larger t-values. This inflation can be observed in Table 3, where fixed-effects t-values reach above 30, compared to mixed-effects t-values between 5 and 22 for the same experiments.

In contrast, the mixed-effects model adjusts for the between-droplet variability by incorporating droplet-level random intercepts. This adjustment results in a more accurate (and larger) estimate of the standard error, thus yielding more conservative but statistically valid t-values. Therefore, the mixed-effects framework should be preferred in settings where natural clustering, such as droplets or batches, is expected to induce correlation among units.

Together, we can conclude that:

- Fluorescence variability across droplets is a statistically significant component of the total variance.
- WT groups exhibit greater droplet-level heterogeneity than KO groups, consistent with functional peptide signaling.
- Fixed-effects models fail to capture this structure, leading to underestimated standard errors and exaggerated t-values.

Accordingly, the mixed-effects framework not only provides more accurate inference but also helps uncover the ecological and spatial dynamics underpinning microbial communication systems.

3.3 WT and WT + peptide at 4 Hours (Analyzed Separately)

3.3.1 Basic Exploratory Data Analysis

This section focuses on the 4-hour data from experiment 230807. All intensity values are analyzed on the log scale to account for skewness and improve interpretability.

Figure 4 displays the log fluorescence distributions for 10 randomly selected droplets from the WT condition. While most droplets exhibit moderate dispersion and consistent medians, a few show elevated variance and skewed distributions. This suggests heterogeneous phrA induction even under endogenous conditions, potentially driven by differences in local cell density or peptide diffusion across droplets.

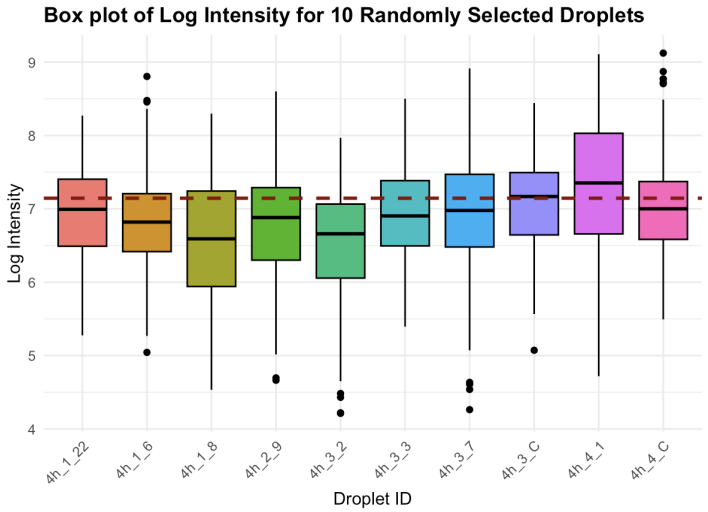

Figure 4: Boxplot of Log Intensity for 10 Randomly Selected Droplets (WT_230807, 4h)

Figure 5 compares WT and WT+peptide groups side-by-side. Each panel shows the log intensity distributions for 10 randomly selected droplets per condition. WT+peptide droplets display more prominent low outliers and heavier lower tails, suggesting that exogenous phrA addition may produce more variable individual responses. However, the differences in median fluorescence levels between droplets appear less visually distinct compared to the WT group. This reduced between-droplet spread hints at more uniform peptide availability when supplied externally.

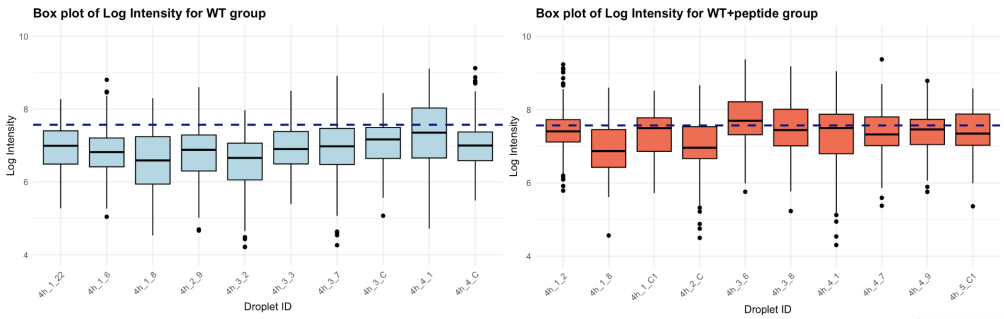

Figure 5: Boxplot of Log Intensity for 10 Randomly Selected Droplets (WT_230807 and WT+peptide_230807, 4h)

These qualitative patterns raise two contrasting hypotheses: (1) peptide-treated cells may exhibit stronger overall induction (higher mean), and (2) exogenous phrA may reduce spatial heterogeneity by homogenizing the local signaling environment. We evaluate these hypotheses in detail through formal mixed-effects modeling.

Consistent trends were observed in experiments 230804 and 240322, and the corresponding EDA plots are included in the appendix.

Taken together, EDA reveals heterogeneity in phrA induction at three levels: within droplets (cell-to-cell variability), across droplets (spatial clustering), and between experimental conditions (genotypic or treatment differences). These findings motivate the use of hierarchical models to rigorously decompose and quantify variance components.

3.3.2 Mixed-Effect Model of WT and WT + Peptide (Analyzed Separately)

To assess the effects of exogenous phrA peptide on both signal strength and spatial heterogeneity, we modeled WT and WT+peptide conditions separately using mixed-effects models with random droplet intercepts. Table 4 summarizes log-transformed mean fluorescence and intraclass correlation coefficients (ICCs) for each condition across three experiments.

The model specification is:

$$\texttt{lmer(log.intensity} \sim 1 + (1 \mid \texttt{droplet\_ID}))$$

Across all experiments, WT+peptide samples consistently exhibit higher average fluorescence intensity than WT, confirming that exogenous phrA enhances signal activation. However, the impact on spatial heterogeneity, as measured by ICC, varies across experiments.

In experiment 1, ICC is noticeably lower in WT+peptide (9.63%) compared to WT (15.24%), suggesting that exogenous phrA partially homogenizes inter-droplet variability. This may reflect saturation of signaling responses or more uniform peptide diffusion.

These differences may only be due to chance variation. I might suggest deleting this.

Table 4: WT and WT+Peptide @ 4 Hours (Separate Analysis)

Experiment	Group	Mean	ICC
Exp #1	WT	7.014	15.24%
	WT+Peptide	7.398	9.63%
Exp #2	WT	6.853	6.90%
	WT+Peptide	7.358	8.54%
Exp #3	WT	7.109	10.84%
	WT+Peptide	7.518	11.83%

Conversely, in experiments 2 and 3, ICC is actually higher in WT+peptide than WT (8.54% vs. 6.90%, and 11.83% vs. 10.84%, respectively). This implies that peptide addition does not universally suppress heterogeneity. In these contexts, exogenous phrA may interact with local factors—such as cell density or prior activation state—to create more diverse droplet-level responses.

Taken together, these results suggest:

- WT+peptide treatment consistently increases mean phrA-driven fluorescence;
- The impact on droplet-level heterogeneity is variable and context-dependent;
- Peptide addition does not eliminate clustering effects and may, in some cases, amplify them.

These findings reinforce the need for mixed-effects modeling when analyzing spatially clustered biological systems and underscore that external signal supplementation can modulate—but not fully control—heterogeneity in multicellular microbial environments.

Table 5: One-way ANOVA: Droplet-Level Clustering in WT+Peptide (Exp #3)

Source	Df	Mean Sq	F-value	P-value
Droplet ID	258	12.64	23.86	$< 2 \times 10^{-16}$
Residuals	21750	0.53		

To independently verify the presence of droplet-level heterogeneity in the WT+peptide condition, we conducted a one-way ANOVA using droplet ID as a categorical predictor of log-transformed fluorescence. As shown in Table 5, the analysis reveals a highly significant clustering effect ($p < 2 \times 10^{-16}$). The F-statistic of 23.86 indicates that average fluorescence levels vary substantially across droplets, while the mean square for droplet-level differences (12.64) greatly exceeds that of the residual variance (0.53). These findings validate the ICC estimates and confirm that spatial structure persists in the presence

of exogenous phrA, underscoring the importance of accounting for droplet-level variation in models of peptide-mediated signaling.

3.4 WT and WT + Peptide at 4 Hours (Group Effect)

3.4.1 Mixed-Effect Model of WT and WT + peptide

To assess the effect of exogenous phrA peptide on signal activation, we modeled the WT and WT+peptide conditions jointly at the 4-hour time point. As before, we used both mixed-effects models (accounting for droplet-level clustering) and fixed-effects models (assuming independent observations). The estimated group effects, standard errors, t-values, and intraclass correlation coefficients (ICCs) are summarized in Table 6.

The mixed-effects model is specified as follows:

$$\texttt{lmer(log.intensity} \sim 1 + \texttt{factor(peptide)} + (1 \mid \texttt{droplet\_ID}))$$

The fixed-effects model is specified as follows:

$$\texttt{lm(log.intensity} \sim 1 + \texttt{factor(peptide)})$$

Table 6: Log-Intensity @ 4 hrs (WT vs WT+Peptide)

Exp	Measure	Mixed Model	Fixed Effects
Exp #1	Intercept (WT)	7.006	6.964
	Group Effect (+ peptide)	**0.391**	**0.426**
	t-value (group)	14.30	22.86
	ICC	10.90%	NA
Exp #2	Intercept (WT)	6.900	6.821
	Group Effect (+ peptide)	**0.423**	**0.525**
	t-value (group)	16.99	32.27
	ICC	7.05%	NA
Exp #3	Intercept (WT)	7.191	7.092
	Group Effect (+ peptide)	**0.244**	**0.383**
	t-value (group)	8.41	18.31
	ICC	12.38%	NA

Across all three experiments, WT+peptide cells consistently display higher log-transformed fluorescence intensities than WT, with group effect estimates ranging from 0.244 to 0.423 in the mixed-effects model. These values confirm that exogenous peptide enhances phrA-mediated signaling, though the magnitude of enhancement varies across experiments.

As with the WT vs KO comparison, both modeling approaches yield similar point estimates for the group effect, indicating robustness to model specification. However, the fixed-effects models systematically produce larger t-values—sometimes nearly double those from the mixed-effects model—due to underestimation of standard errors. For example, in Exp #2, the t-value increases from 16.99 (mixed) to 32.27 (fixed), despite nearly identical group effect estimates.

Statistical Significance and the Role of Droplet Clustering

The elevated t-values in the fixed-effects model are a consequence of ignoring droplet-level dependencies. By assuming independence, the model inflates statistical certainty by reducing the standard error represented by the t-value.

When droplet-level variation is present, this leads to artificially high t-statistics and an overstatement of significance. In contrast, the mixed-effects model adjusts for this structure by introducing random intercepts for droplets. This correction produces more conservative, but statistically valid, inference. In all three experiments, the mixed model still detects strong peptide effects (all $t > 8$), but with appropriately adjusted error terms.

Summary and Interpretation

From this comparison, we conclude:

- WT+peptide treatment significantly increases phrA-driven fluorescence compared to WT alone, across all experiments.
- Fixed-effects models overestimate significance by ignoring droplet-level clustering.
- Mixed-effects models provide more reliable estimates by accounting for spatial dependencies in the data.

These findings reinforce our earlier observations of peptide-driven signal amplification and underscore the importance of hierarchical modeling when analyzing spatially structured biological systems such as droplet-encapsulated bacterial populations.

3.5 WT vs KO Over Time

3.5.1 Basic Exploratory Data Analysis

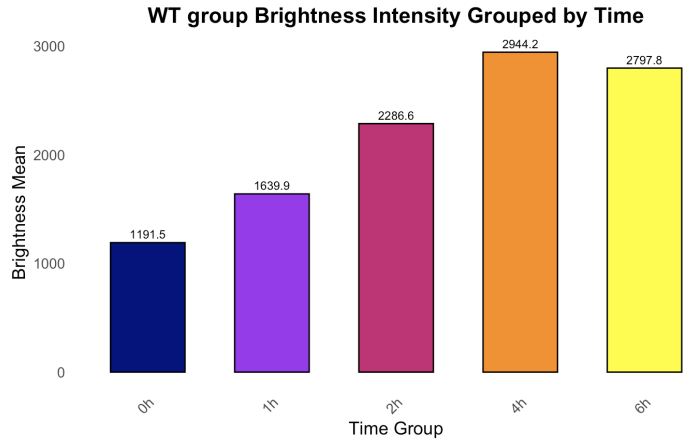

Figure 6: Brightness Intensity Grouped by Time, WT_230113

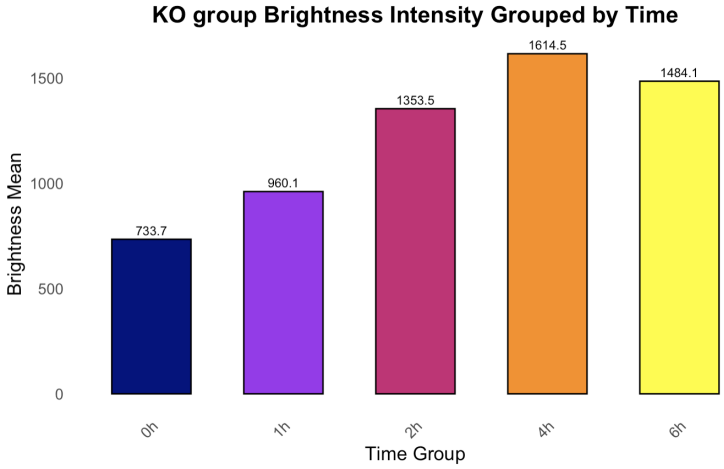

Figure 7: Brightness Intensity Grouped by Time, KO_230113

To investigate how phrA-mediated signaling evolves over time, we conducted an exploratory analysis using fluorescence brightness values grouped by time points. This helps assess whether signaling is static or dynamic, and whether WT and KO groups exhibit different temporal trends.

The bar plot of brightness intensity for the WT group reveals the trend of a clear increase from 1191.5 at 0h to a peak of 2944.2 at 4h, with a slight decline

by 6h (2797.8). This upward trajectory suggests that phrA signaling in WT cells is not constant but rather intensifies over time—indicative of a time-dependent enhancement in signal activation.

In contrast, KO cells also show a general upward trend, but the increase is both less pronounced and more gradual. Starting from 733.7 at 0h, KO brightness peaks at 1614.5 by 4h, with a slight decline to 1484.1 at 6h. This muted trend aligns with our expectations, as KO cells lack endogenous phrA peptide production and therefore rely solely on external cues or background noise.

These descriptive patterns motivate the need for a statistical model that accounts for fluorescence as a dynamic quantity that changes with time. In particular, we aim to test whether:

- Fluorescence follows a nonlinear trajectory over time (e.g., rises then plateaus),
- The rate of change in fluorescence differs between WT and KO groups,
- The WT-KO difference expands over time, suggesting interaction between group and time.

To address these questions, we introduce a mixed-effects model with time, time^2, and a group $\times$ time interaction term, which allows us to capture both linear trends and nonlinear saturation effects while accounting for droplet-level variability.

3.5.2 WT vs KO (Group + Time Interaction Model)

As we mentioned in the last section, we fit a mixed-effect model that includes fixed effects for group, time, time squared, and their interaction, along with a droplet-level random intercept to account for within-droplet clustering.

The inclusion of both linear and quadratic time terms enables us to model potential nonlinear trajectories, such as decline in the intensity growth potentially due to saturation, which are common in peptide-based signaling systems. Our model takes the form:

$$\texttt{lmer(log.intensity} \sim 1 + \texttt{factor(group)} * \texttt{time} + \texttt{time}^2 + (1 \mid \texttt{droplet\_ID}))$$

The term $\texttt{time}^2$ allows for curvature in the time trajectory, and the random intercept term (1 | droplet_ID) accounts for droplet-level variability in baseline intensity.

Interpreting the Group $\times$ Time Interaction

The interaction term between `group` and `time` captures whether the rate of fluorescence change over time differs between WT and KO cells. In other words, this term evaluates whether the phrA signaling trajectory over time has a different slope in WT versus KO groups.

A positive and statistically significant interaction coefficient implies that the WT group experiences a steeper increase in fluorescence over time relative to KO. Rather than a fixed difference between WT and KO (i.e., WT always being higher by the same amount), the gap between the two groups widens as time progresses. This finding is biologically meaningful: WT cells, capable of producing the phrA peptide, likely engage in time-dependent amplification of signaling due to feedback mechanisms and increasing local cell density within droplets. In contrast, KO cells—lacking endogenous phrA production—do not show the same dynamic amplification and therefore remain relatively flat in their signaling response.

From a modeling perspective, the group × time interaction allows the slope of the time trajectory to vary by group. Without this term, the model would assume that both WT and KO follow the same time trend, differing only by a constant offset. The significance of the interaction in Exp #1 ($t = 2.066$) and Exp #2 ($t = 5.069$), and its marginal presence in Exp #3 ($t = 1.612$), provides strong evidence that phrA signaling is temporally dynamic and more responsive in WT populations.

Table 7: Comparison of WT vs KO across Experiments (Interaction Model)

Exp	Measure	Estimate	t-value
Exp #1	Intercept (KO)	6.170	141.707
	Group Effect (WT)	0.509	9.561
	Time Effect	0.278	9.326
	Time^2	-0.032	-7.290
	Group : Time	**0.030**	**2.066**
	ICC	10.25%	
Exp #2	Intercept (KO)	6.078	195.926
	Group Effect (WT)	0.299	12.896
	Time Effect	0.291	11.531
	Time^2	-0.030	-7.624
	Group : Time	**0.029**	**5.069**
	ICC	8.14%	
Exp #3	Intercept (KO)	6.316	130.310
	Group Effect (WT)	0.426	7.553
	Time Effect	0.372	10.918
	Time^2	-0.042	-8.194
	Group : Time	**0.026**	**1.612**
	ICC	11.04%	

The coefficients for both `time` and `time`^{2} are statistically significant across all experiments, with consistent signs: a positive time coefficient and a negative quadratic term. This implies that fluorescence intensity initially rises but then

slows down over time, forming a concave trend. This decelerating pattern is consistent with what we observe from the EDA, where from 0h to 4h, the average intensity is gradually increasing, reaching the peak at 4 hours, and become to decrease at 6 hours.

The interaction term between group and time is positive in all three experiments, indicating that WT cells tend to increase fluorescence more rapidly than KO cells. Notably, the interaction reaches statistical significance in Exp #1 ($t = 2.066$) and Exp #2 ($t = 5.069$), while remaining marginal in Exp #3 ($t = 1.612$). This suggests that time-dependent amplification of phrA signaling is strongest in the first two replicates, possibly due to experimental variation or different levels of baseline activation.

The ICC values in this model, ranging from 8.14% to 11.04%, reflect modest but consistent droplet-level heterogeneity, reinforcing the earlier conclusion that droplet identity contributes meaningfully to variation in signal intensity. Importantly, the inclusion of time terms did not eliminate or overwhelm the between-droplet variance, confirming that both temporal and spatial structure must be accounted for in the modeling framework.

In summary, these results demonstrate that:

- WT and KO differ ~~not only~~ in average fluorescence levels ~~but also in how those levels evolve over time.~~
- WT cells exhibit a ~~faster~~ rise in signal followed by deceleration, suggesting dynamic regulation of phrA signaling.
- ~~The Group × Time interaction confirms that the WT-KO gap grows over time, particularly in early experiments.~~

Together, these findings support a model in which phrA signaling in WT is both stronger and more dynamically responsive to time. The use of nonlinear temporal modeling and interaction terms provides deeper insight into how microbial communication evolves in complex microenvironments.

3.6 WT vs WT + Peptide Over Time

3.6.1 Basic Exploratory Data Analysis

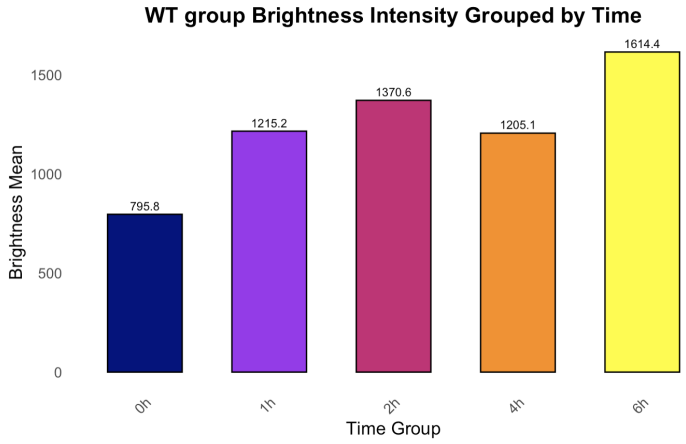

Figure 8: Brightness Intensity Grouped by Time, WT_230807

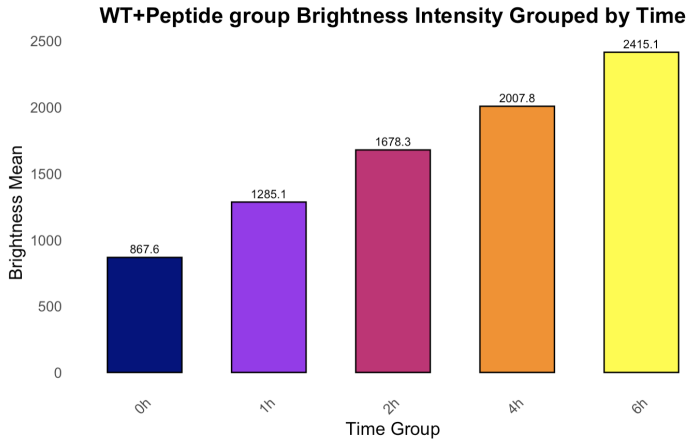

Figure 9: Brightness Intensity Grouped by Time, WT+Peptide_230807

Both groups exhibit a general increase in brightness from 0 to 6 hours, indicating time-dependent activation. However, the WT+Peptide group consistently shows higher fluorescence intensity at every time point. Notably, while the WT group peaks at 2 hours (1370.6), it experiences a slight dip at 4 hours (1205.1) before increasing again at 6 hours (1614.4). In contrast, the WT+Peptide group shows a steady rise throughout the entire time course, reaching a much higher peak of

2415.1 at 6 hours. This suggests that peptide addition enhances the strength and consistency of phrA activation, leading to a more robust and sustained fluorescence response compared to the WT condition alone. The trajectory of brightness intensity for the two groups is unknown now, and we are going to test it in the following section.

3.6.2 WT vs WT+Peptide (Group + Time Interaction Model)

Similarly, to investigate how peptide treatment modulates the temporal dynamics of phrA signaling, we fit a mixed-effects model that includes fixed effects for group (WT vs WT+peptide), time, time squared, and the group × time interaction. A droplet-level random intercept accounts for baseline intensity differences across droplets.

The model structure is as follows:

$$\texttt{lmer(log.intensity} \sim 1+\texttt{factor(peptide)}*\texttt{time}+\texttt{time}^2+(1 \mid \texttt{droplet\_ID}))$$

This formulation allows us to test for nonlinear signal trajectories (via the $\texttt{time}^2$ term) and whether the rate of fluorescence change differs between treatment groups (via the interaction term). The results are summarized in Table 8.

Table 8: Comparison of WT vs WT+Peptide across Experiments

Exp	Measure	Estimate	t-value
Exp #1	Intercept (WT)	5.997775	168.925
	Group Effect (WT + peptide)	0.184849	6.869
	Time Effect	0.479576	17.996
	Time^2	-0.0543261	-13.367
	Group : Time	**0.043925**	**6.374**
	ICC	0.07405908	
Exp #2	Intercept (WT)	6.648662	154.125
	Group Effect (WT + peptide)	0.085142	2.675
	Time Effect	0.280384	8.224
	Time^2	-0.03105	-5.853
	Group : Time	**0.046486**	**6.208**
	ICC	0.1349974	
Exp #3	Intercept (WT)	6.524299	196.884
	Group Effect (WT + peptide)	0.05487	2.335
	Time Effect	0.219826	8.19
	Time^2	-0.021383	-5.099
	Group : Time	**0.073259**	**13.818**
	ICC	0.08162916	

Interpreting the Group × Time Interaction

In all three experiments, the coefficients for the linear (`time`) and quadratic ($\texttt{time}^2$) terms are statistically significant, consistently exhibiting a positive linear effect and a negative quadratic effect. Specifically, linear term estimates range from 0.220 (Exp #3) to 0.480 (Exp #1), and quadratic term estimates range from −0.021 (Exp #3) to −0.054 (Exp #1). These results confirm that fluorescence intensity driven by phrA signaling increases over time but at a diminishing rate, reflecting a nonlinear, concave trajectory. Such patterns are typical of quorum-sensing systems, likely due to mechanisms like receptor saturation or negative feedback regulation that limit further increases at later time points.

The addition of exogenous peptide markedly influences both the baseline fluorescence levels and their temporal trajectories. The WT+peptide group exhibits significantly elevated baseline fluorescence compared to untreated WT across all replicates, with group effect estimates ranging from 0.055 (Exp #3, $t = 2.335$) to 0.185 (Exp #1, $t = 6.869$). More notably, the interaction between group and time is consistently positive and strongly significant in all experiments (Exp #1: estimate = 0.044, $t = 6.374$; Exp #2: estimate = 0.046, $t = 6.208$; Exp #3: estimate = 0.073, $t = 13.818$). This indicates that the presence of peptide not only raises the initial signaling level but also substantially accelerates the subsequent increase in fluorescence over time. This acceleration likely arises from the peptide's role in initiating positive feedback mechanisms earlier, amplifying phrA signaling dynamics more robustly in treated droplets.

Despite accounting for group and time effects, droplet-level heterogeneity remains a meaningful contributor to variability in fluorescence intensities. The intraclass correlation coefficients (ICC) across the three experiments range from approximately 7.4% (Exp #1) to 13.5% (Exp #2), emphasizing persistent droplet-level clustering. This reinforces the importance of explicitly modeling spatial variability through random intercept terms, as local conditions within droplets continue to significantly shape signaling responses even after accounting for temporal trends and peptide treatment.

In summary, these findings collectively demonstrate that peptide treatment significantly elevates both baseline phrA signaling and its temporal responsiveness, resulting in nonlinear, rapidly accelerating fluorescence trajectories. Moreover, the persistent droplet-level variability highlights the necessity of mixed-effects modeling approaches to comprehensively capture the complexity of microbial signaling dynamics in heterogeneous environments.

4 Conclusions

Biological Insights

Our analyses reveal robust and biologically meaningful differences in phrA-mediated signaling between WT and KO strains of Streptococcus pneumoniae. Across all experiments and models:

- WT cells consistently exhibit higher fluorescence intensities than KO cells, confirming stronger communication of phrA signaling pathways.
- Group effects are statistically significant across both fixed-effect and mixed-effect models, reinforcing the reproducibility of the observed WT-KO difference.
- Intraclass correlation coefficients (ICCs) and ANOVA results show significant between-droplet variability, especially in WT groups, highlighting between-group heterogeneity in signal induction.
- WT cells, which can produce their own phrA peptides, show a clear pattern of increasing fluorescence over time. This means their signaling system becomes more active as time passes, likely because the cells are communicating with each other and amplifying the signal through positive feedback.
- In contrast, KO cells cannot produce the phrA peptide, so they have no way to trigger or boost this signaling on their own. As a result, their fluorescence levels increase more slowly over time, since they lack the ability to activate the signaling pathway internally.
- In additional conditions with exogenous phrA peptide added, WT +peptide cells display both higher initial signal and steeper increases over time, demonstrating that phrA enhances both baseline activation and dynamic responsiveness.

Modeling Insights

From a statistical modeling standpoint:

- Mixed-effects models provide a more accurate and conservative framework for inference by accounting for clustering at the droplet level.
- These models capture unobserved heterogeneity that would otherwise inflate type I error rates in simpler fixed-effects models.
- In contrast, fixed-effects models underestimate standard errors, leading to exaggerated t-values and potentially overstated significance.

Together, these findings support the utility of mixed-effects modeling in analyzing droplet-based single-cell data and underscore the biological importance of both baseline and dynamic phrA signaling. Our results confirm that phrA-mediated communication is not only active in WT populations but also evolves meaningfully over time.

Exhibit 1

Boxplot of Log Intensity for 10 Randomly Selected Droplets

WT_230113

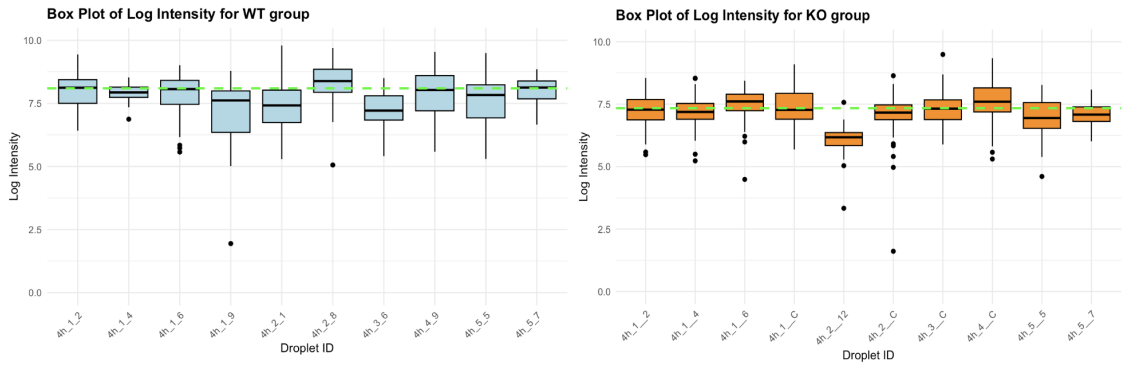

Figure 10: Boxplot of Intensity for 10 Randomly Selected Droplets After Log Transformation (WT_230113, 4h)

Exhibit 2

Boxplot of Log Intensity for 10 Randomly Selected Droplets

WT_230111

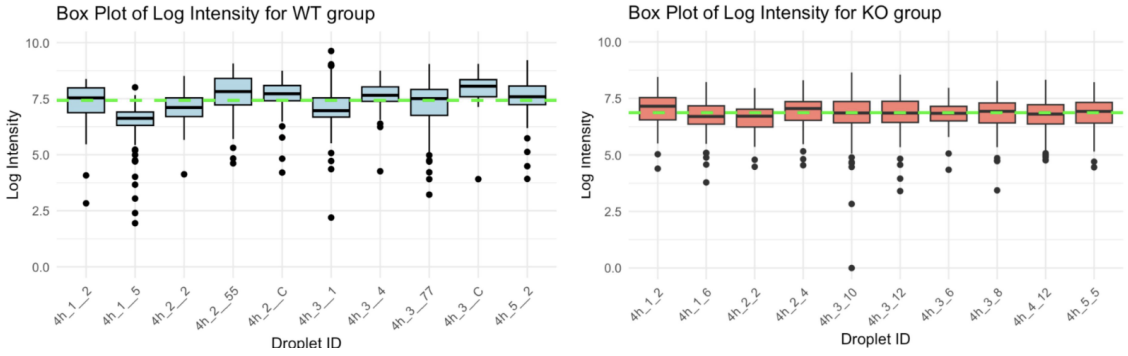

Figure 11: Boxplot of Intensity for 10 Randomly Selected Droplets After Log Transformation (WT_230111, 4h)

Exhibit 3

Boxplot of Log Intensity for 10 Randomly Selected Droplets

WT_230611

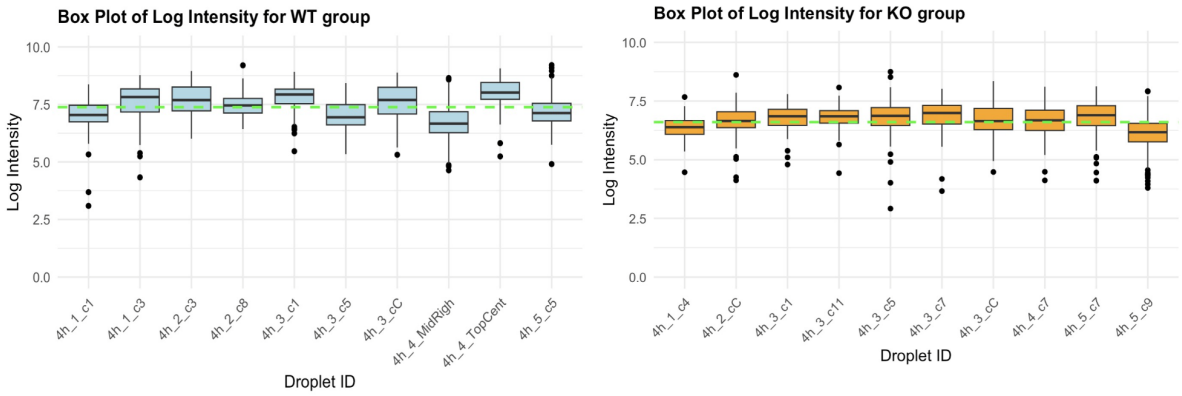

Figure 12: Boxplot of Intensity for 10 Randomly Selected Droplets After Log Transformation (WT_230611, 4h)

Exhibit 4

Result table for WT vs KO, model both time and group as factor

The model here is quite similar to the models we discussed in sections 3.5.2 and 3.6.2, the only difference is how we set the time variable. For the models in 3.5.2 and 3.6.2, the time variable is a continuous quantitative variable (regression modeling as numeric predictors), but for the model here, the time variable is a categorical variable (treating time points as distinct groups, not numbers).

The model structure is as follows:

$$\texttt{lmer(log.intensity} \sim 1+\texttt{factor(group)}+\texttt{factor(time)}+(1 \mid \texttt{droplet\_ID}))$$

Table 9: Comparison of WT vs KO across Experiments (model with time being a categorical variable)

Exp	**Measure**	**Estimate**	**t-value**
Exp #1	Intercept (KO, time = 0)	6.00969	179.812
	Group Effect (WT)	0.40254	36.127
	Time effect, t = 1	0.35374	7.645
	Time effect, t = 2	0.45357	9.371
	Time effect, t = 4	0.76306	17.210
	Time effect, t = 6	0.76706	17.515
	ICC	7.92%	
Exp #2	Intercept (KO, time = 0)	6.27747	128.497
	Group Effect (WT)	0.49734	14.259
	Time effect, t = 1	0.32698	5.086
	Time effect, t = 2	0.64949	10.404
	Time effect, t = 4	0.84416	14.008
	Time effect, t = 6	0.80997	13.792
	ICC	11.09%	
Exp #3	Intercept (KO, time = 0)	6.08324	139.011
	Group Effect (WT)	0.59946	18.997
	Time effect, t = 1	0.35453	6.384
	Time effect, t = 2	0.50899	9.502
	Time effect, t = 4	0.66026	12.403
	Time effect, t = 6	0.64613	12.506
	ICC	10.27%	

Table 10: Comparison of WT vs WT+Peptide across Experiments (model with time being a categorical variable)

Exp	Measure	Estimate	t-value
Exp #1	Intercept (WT, time = 0)	6.55121	142.177
	Group Effect (WT+peptide)	0.26008	17.393
	Time effect, t = 1	0.2157	3.489
	Time effect, t = 2	0.63941	10.373
	Time effect, t = 4	0.63054	10.456
	Time effect, t = 6	0.75007	12.383
	ICC	12.96%	
Exp #2	Intercept (WT, time = 0)	6.31915	179.753
	Group Effect (WT+peptide)	0.3383	29.439
	Time effect, t = 1	0.3715	7.614
	Time effect, t = 2	0.58248	12.220
	Time effect, t = 4	0.62358	13.192
	Time effect, t = 6	0.8636	18.400
	ICC	7.67%	
Exp #3	Intercept (WT, time = 0)	5.93154	156.714
	Group Effect (WT+peptide)	0.32913	22.334
	Time effect, t = 1	0.37824	7.419
	Time effect, t = 2	0.83197	17.017
	Time effect, t = 4	1.10879	22.917
	Time effect, t = 6	1.06209	22.334
	ICC	7.42%	

We can interpret the table as follow (take the EXP #2 from Table 10 as an example):

Term	Estimate	t-value	Interpretation
Intercept	6.31915	179.75	Mean log-intensity for **WT at time = 0h**.
group	0.33830	29.44	WT+peptide is **higher** than WT by 0.34 log-units (significant).
time effect, time = 1	0.37150	7.61	At 1h, log-intensity is **0.37 higher than at 0h**.
time effect, time = 2	0.58248	12.22	At 2h, intensity is even higher than at 0h (0.58 log-units higher than 0h).
time effect, time = 4	0.62358	13.19	At 4h, continued increase from baseline (0.62358 log-units higher than 0h).
time effect, time = 6	0.86360	18.40	At 6h, log-intensity is highest (0.86360 log-units higher than 0h).

All tables in the report can be interpreted as this, and all coefficients are relative to the baseline point which is the intercept.