Supporting information

Combinatorial Polycation Synthesis and Causal Machine Learning Reveal Divergent Polymer Design Rules for Effective pDNA and Ribonucleoprotein Delivery

Ramya Kumar,[†] Ngoc Le,[†] Felipe Oviedo,[‡] Mary E. Brown,[¶] and Theresa M. Reineke[*,†]

[†]*Department of Chemistry, University of Minnesota, Minneapolis, MN 55414*

[‡]*Nanite Inc., 6 Liberty Square #6128, Boston, MA 02109*

[¶]*University Imaging Centers, University of Minnesota, Minneapolis, MN 55414*

E-mail: treineke@umn.edu

Contents

1 Polyplex size distributions

The size distributions of 129 complexes formed between pDNA and 43 polymers in the library (at N/P ratios of 5, 10, and 20)are reported below. For some formulations, we were unable to acquire auto-correlation functions due to severe aggregation. These results are reported as "nan" or "0 nm"

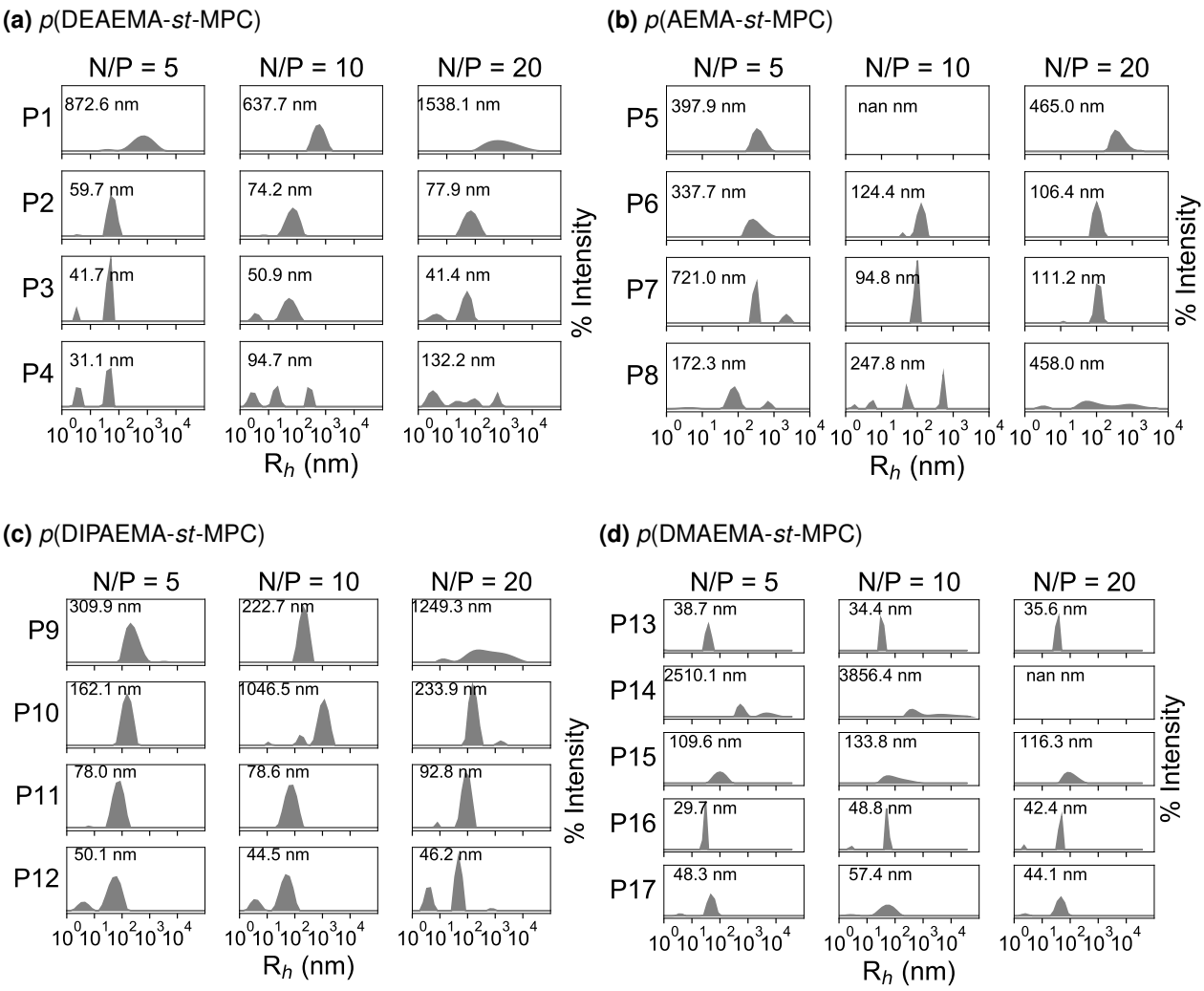

Figure S1. DLS measurements of polyplexes formulated using MPC-based copolymers (P1 through P17). We were unable to acquire auto-correlation functions for P5 polyplexes (N/P=10) and P14 polyplexes (N/P=20) due to severe aggregation.

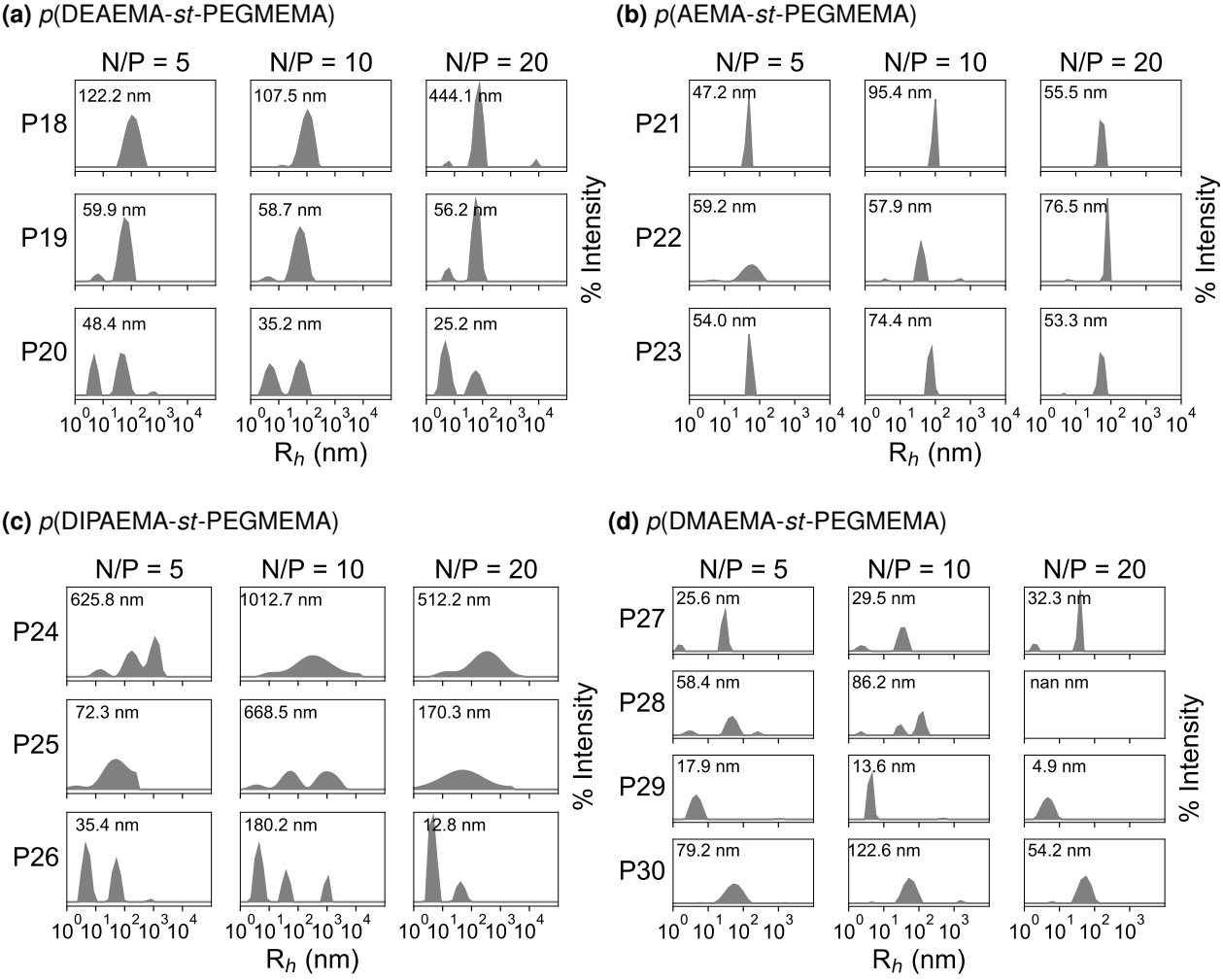

Figure S2. DLS measurements of polyplexes formulated using PEGMEMA-based copolymers (P18 through P30). We were unable to acquire auto-correlation functions for P28polyplexes (N/P=20) due to severe aggregation.

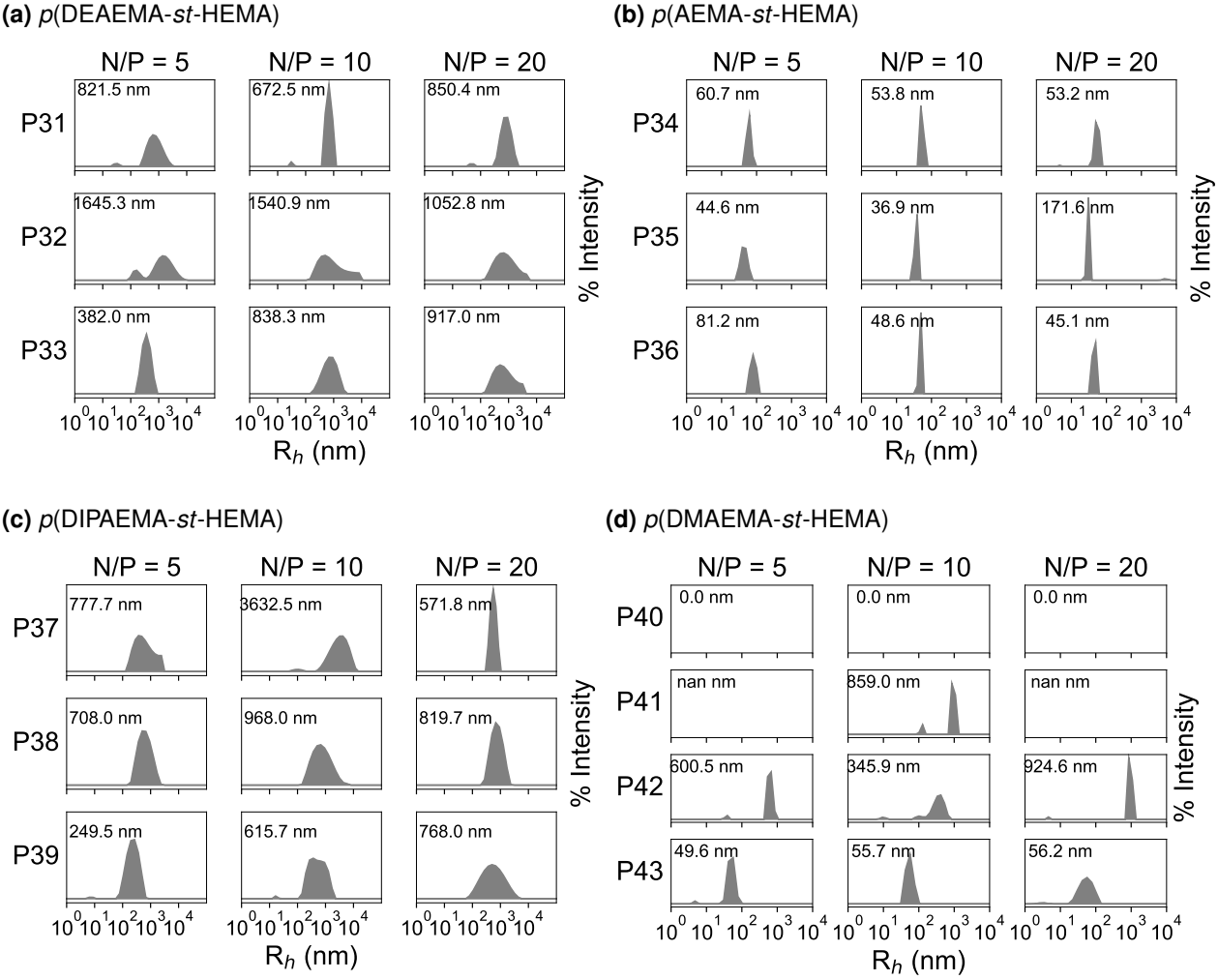

Figure S3. DLS measurements of polyplexes formulated using HEMA-based copolymers (P31 through P43). We were unable to acquire auto-correlation functions for P40 polyplexes and P41polyplexes (N/P=5) due to severe aggregation.

2 Gel migration assays of polyplexes formed with pDNA payloads

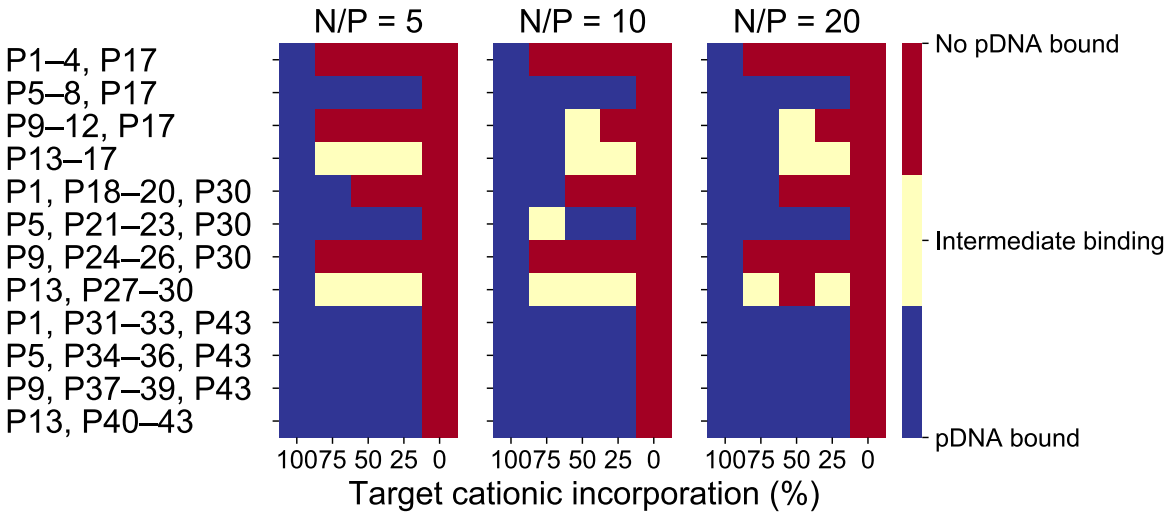

Figure S4. Summary of gel electrophoresis assays where N/P ratios of 5, 10 and 20 were evaluated across the entire polymer library. Polyplex formulations where pDNA migration was observed are depicted in red while those that retained pDNA are shown in blue. Intermediate binding behavior where both retention and migration were observed, are shown in yellow.

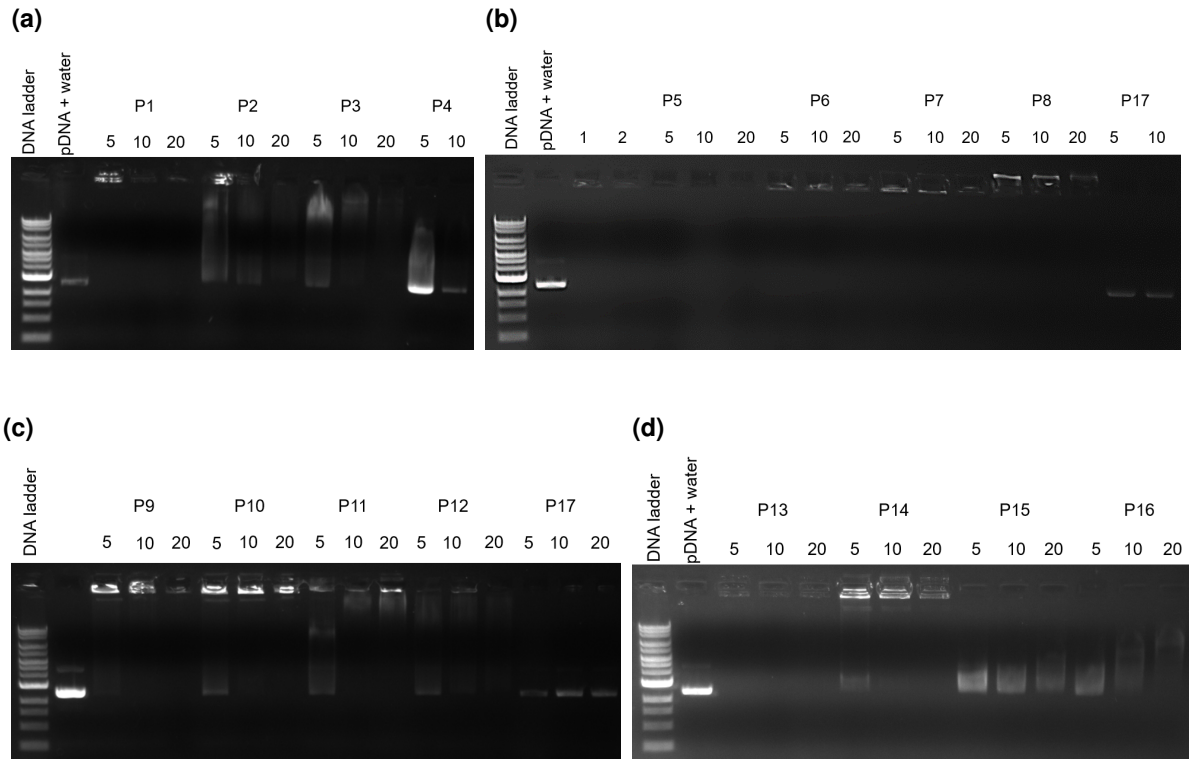

Figure S5. Gel migration assays for polyplexes formed from RNPs and MPC-based polymers (P1 through P17). Numbers denote N/P ratios.

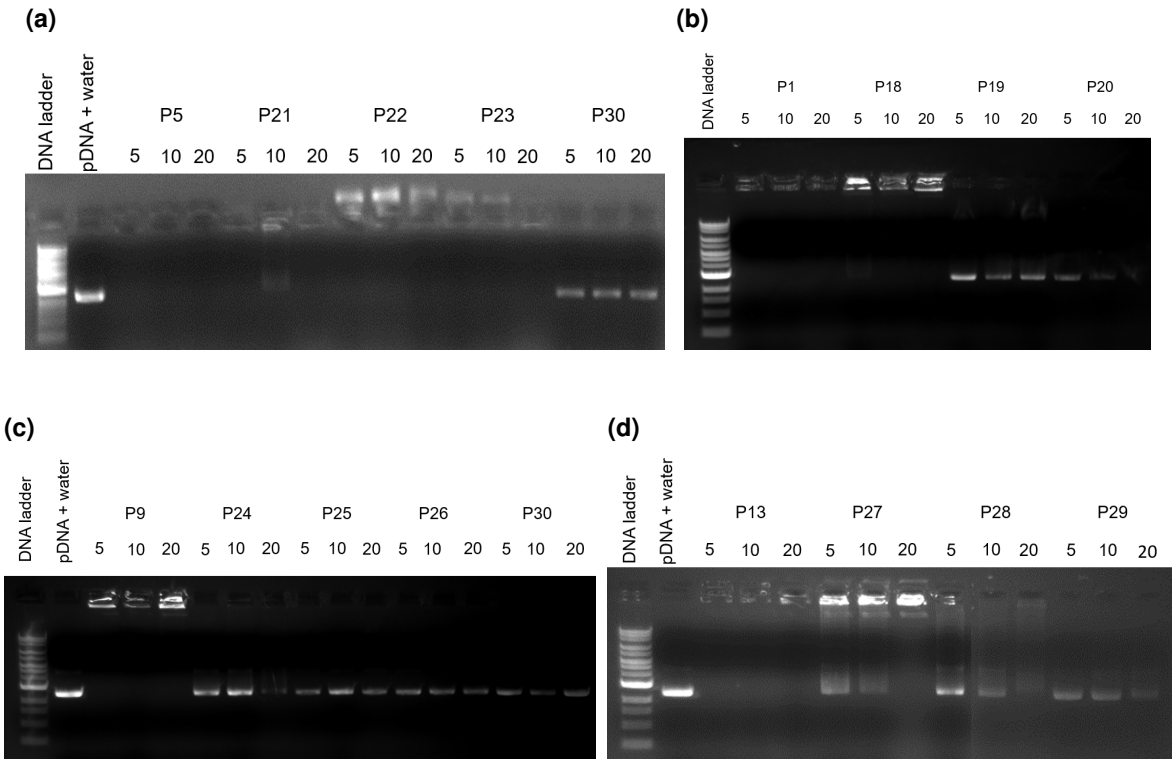

Figure S6. Gel migration assays for polyplexes formed from RNPs and PEGMEMA-based polymers (P18 through P30). Numbers denote N/P ratios.

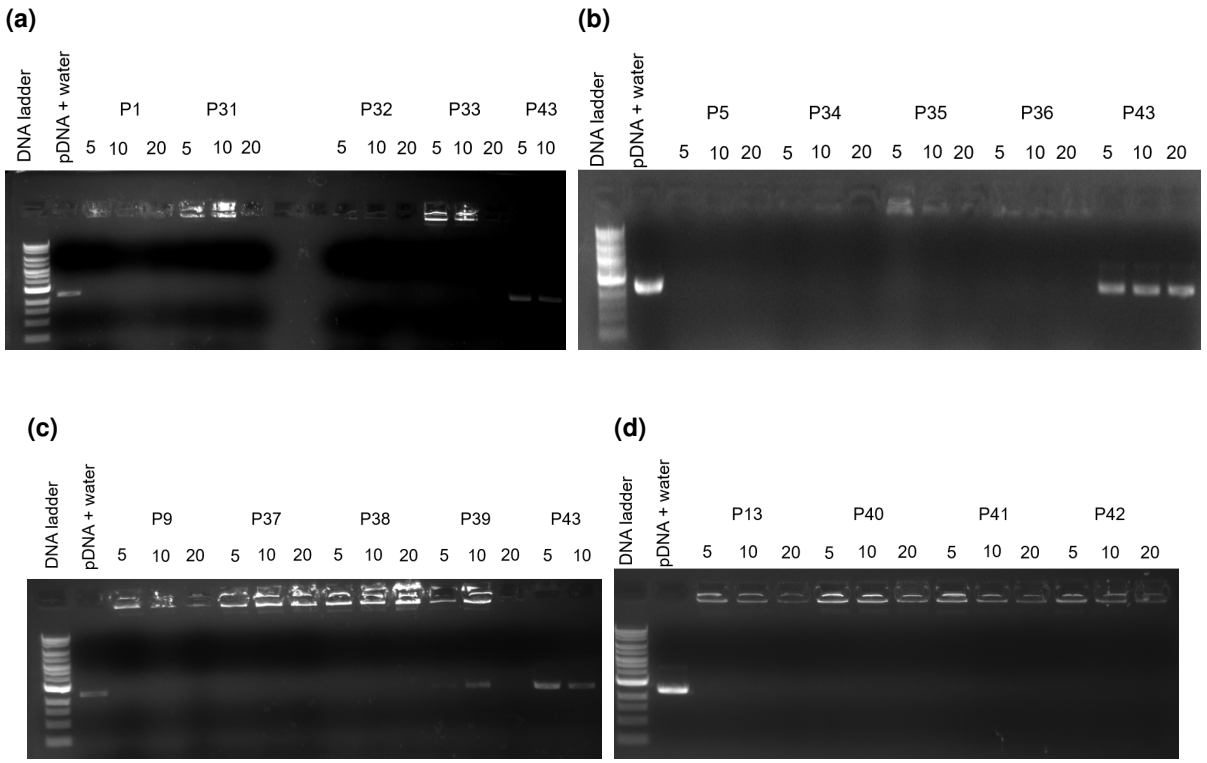

Figure S7. Gel migration assays for polyplexes formed from RNPs and HEMA-based polymers (P31 through P43). Numbers denote N/P ratios.

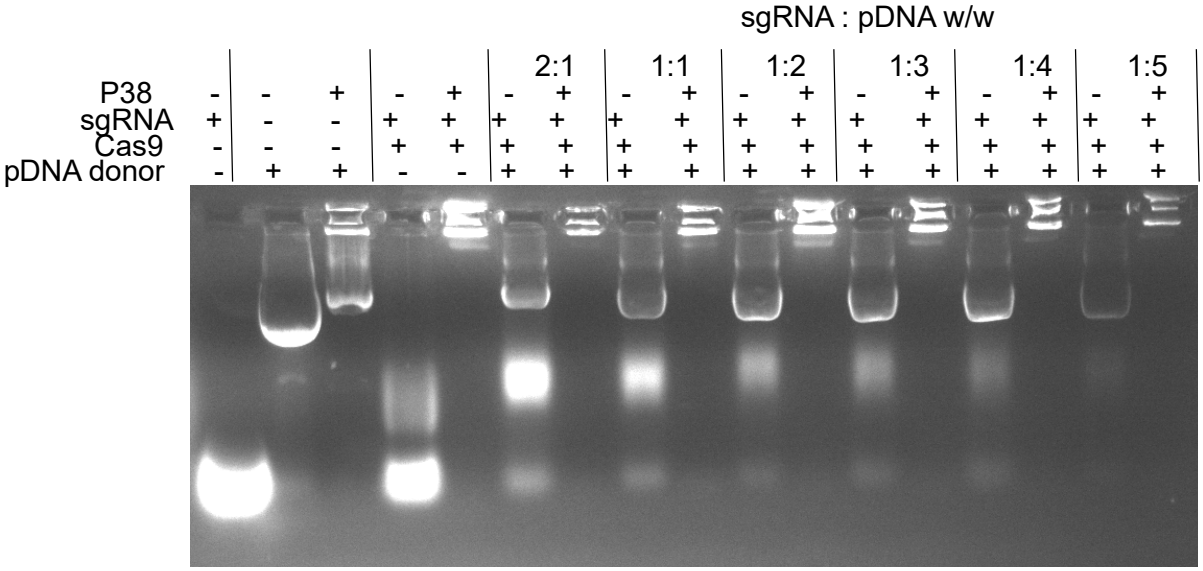

Figure S8. Gel migration assays for polyplexes formed from HDR payloads and P38 at an N/P ratio of 2. While P38 was unable to inhibit the migration of the pDNA donor, strong binding was observed between P38 and the ternary payload composed of sgRNA, spCas9, and pDNA donor.

3 Comparing surface charge distribution of pDNA and RNP payloads

We employed the Poisson-Boltzmann Equation Solver (PBEQ) developed by **?** to calculate the electrostatic potential of spCas9e. The protein data bank structures of spCas9 and B-DNA (PDB files 4OO8 and 1BNA) were used as inputs for solver. Using the tool described in **?** we mapped the electrostatic potential in an aqueous environment with physiological pH and ionic strength.

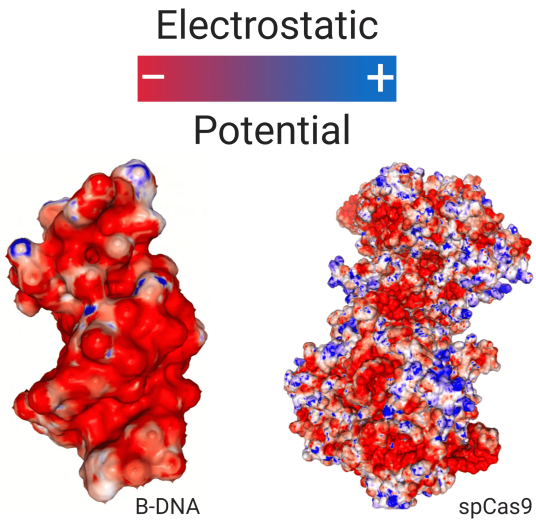

Figure S9. Map of electrostatic potential of the solvent-accessible surface of spCas9 and a dodecamer of B-DNA. This visualization was generated using the tool developed by **?**

4 P38 polyplex characterization

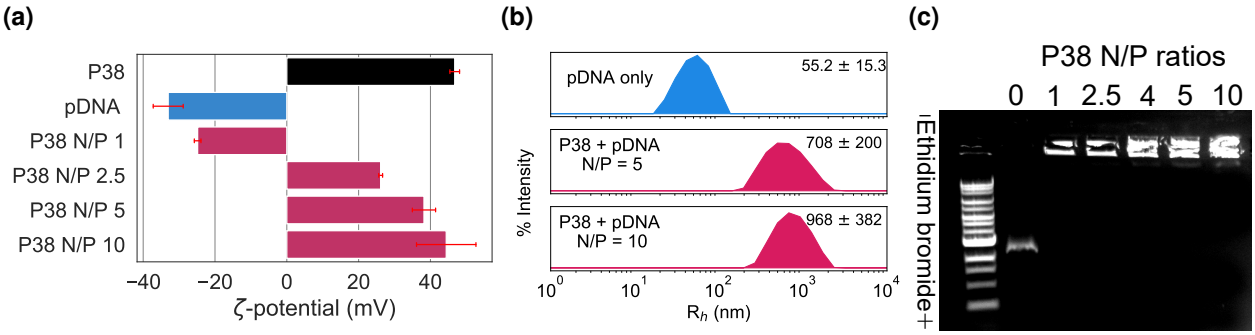

Figure S10. (a) ζ-potential measurements of unpackaged pDNA and polyplexes formulated over a range of N/P ratios. The charge density of polyplexes was found to be highly N/P-dependent, with the magnitude of surface potential increasing with the N/P ratio. (b) Dynamic light scattering measurements (n = 3) of pDNA and polyplexes formulated in PBS revealed monomodal populations with hydrodynamic radii approaching 1 μm. (c) Gel migration assays indicated that polyplexes formed between pDNA and P38 were strongly bound at N/P ratios as low as one.

Observing that the transfection efficiency of P38 was N/P-dependent, we performed gel migration assays to study the extent of pDNA–P38 complexation across a range of N/P ratios (Figure S10(a)). While unpackaged pDNA was able to migrate unhindered, the hit polymer retarded pDNA mobility at all N/P ratios studied, even at the lowest N/P of 1. The data from gel migration disproved our initial conjecture that polymer-plasmid binding was not sufficiently strong enough to package and protect pDNA during delivery at low N/P ratios where poor transfection efficiencies were recorded. To further probe N/P dependence, we conducted ζ-potential measurements (Figure S10(b)). P38 polyplex ζ-potentials ranged from -40 mV for unpackaged plasmid to -25 mV for the N/P 1 formulation to 50 mV for N/P ratios of 2.5 and above. Significantly, the negative surface potential of N/P 1 formulations may have inhibited membrane association and subsequent cellular uptake.

Next, we examined the effects of polymer composition and N/P ratio on polyplex size distributions and found that polymer hydrophilicity and polyplex diameter were negatively correlated. While unimodal populations with diameters approaching 1 μm in radius were formed when P38 was complexed with pDNA (Figure S10(b)), polyplex sizes were greatly reduced when MPC or PEGMEMA co-monomers were employed in the place of HEMA (Figures S1, S2, and S3). In some of these hydrophilic variants of P38 (P10, P11, P12, P24, P25, P26), we either observed unhindered pDNA migration, or discerned two distinct populations: the free polymer and the unbound pDNA. Indeed, gel migration assays confirmed that highly hydrophilic co-monomers inhibit polymer-plasmid binding (Figure S4). In contrast to the strong binding displayed by HEMA-based copolymers ((Figure S10(c)) and Figure S7), most MPC and PEGMEMA copolymers did not form tightly condensed polyplexes (Figures S5 and S6).

5 Supplemental flow cytometry data

5.1 Transient transfection

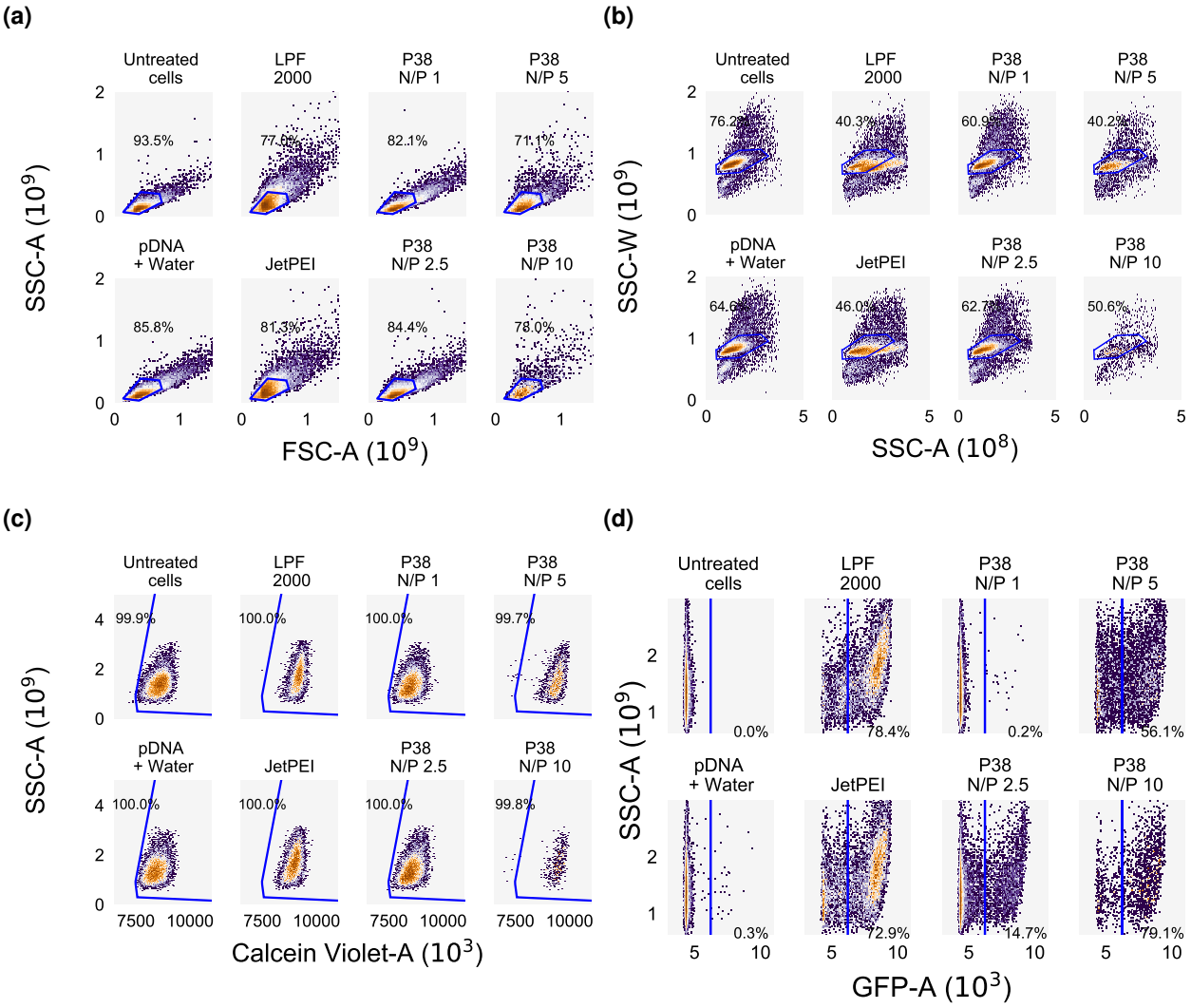

Figure S11. Representative gating schemes for (a) Cells, (b) Single cells, (c) Live cells, and (d) GFP$^+$ cells. These schemes are from the same transfection experiment as Figure 4(A) in the main manuscript.

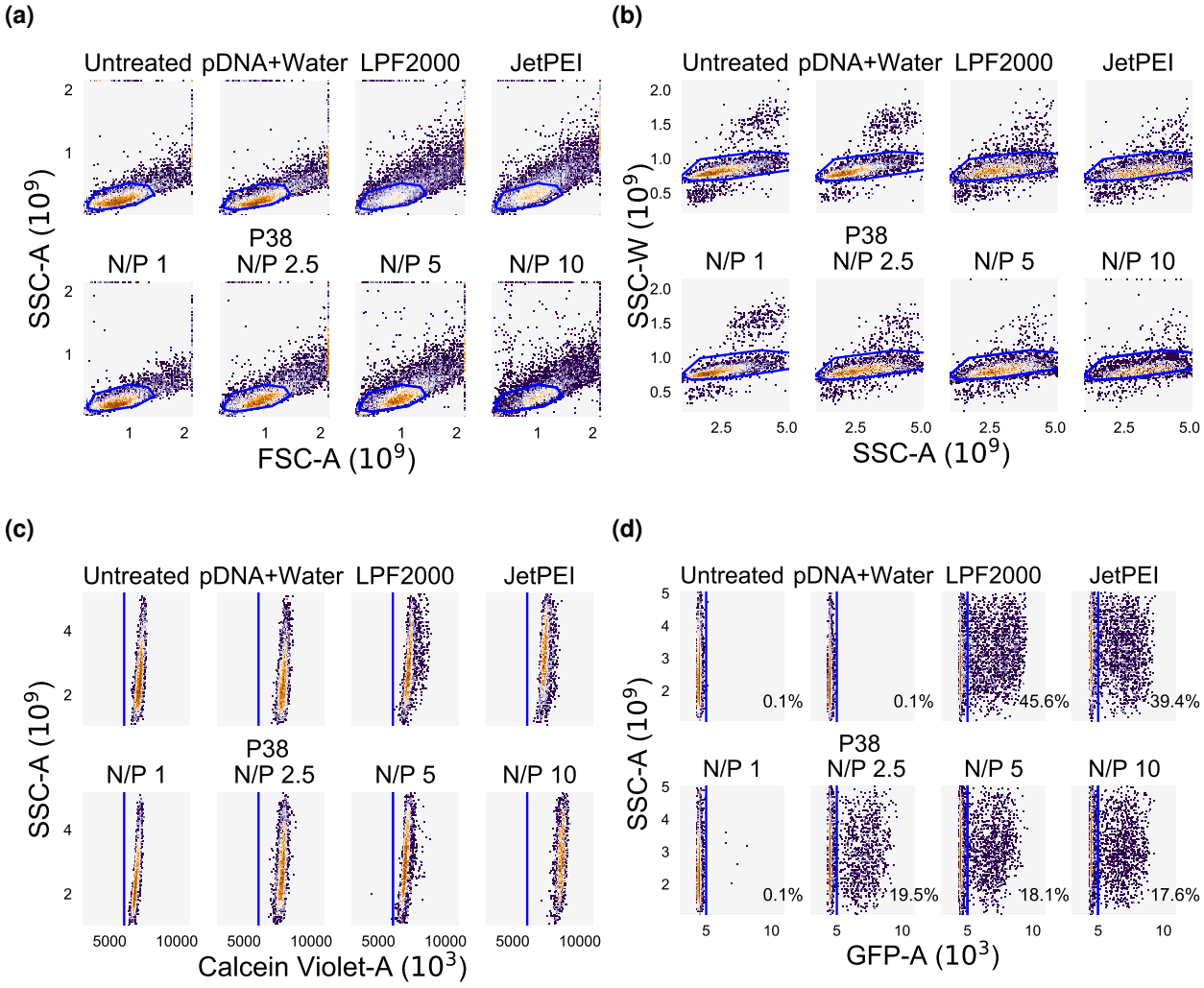

Figure S12. Representative gating schemes for (a) Cells, (b) Single cells, (c) Live cells, and (d) GFP$^+$ cells. These schemes are from the same transfection experiment as Figure 4(B) in the main manuscript.

5.2 HDR experiments

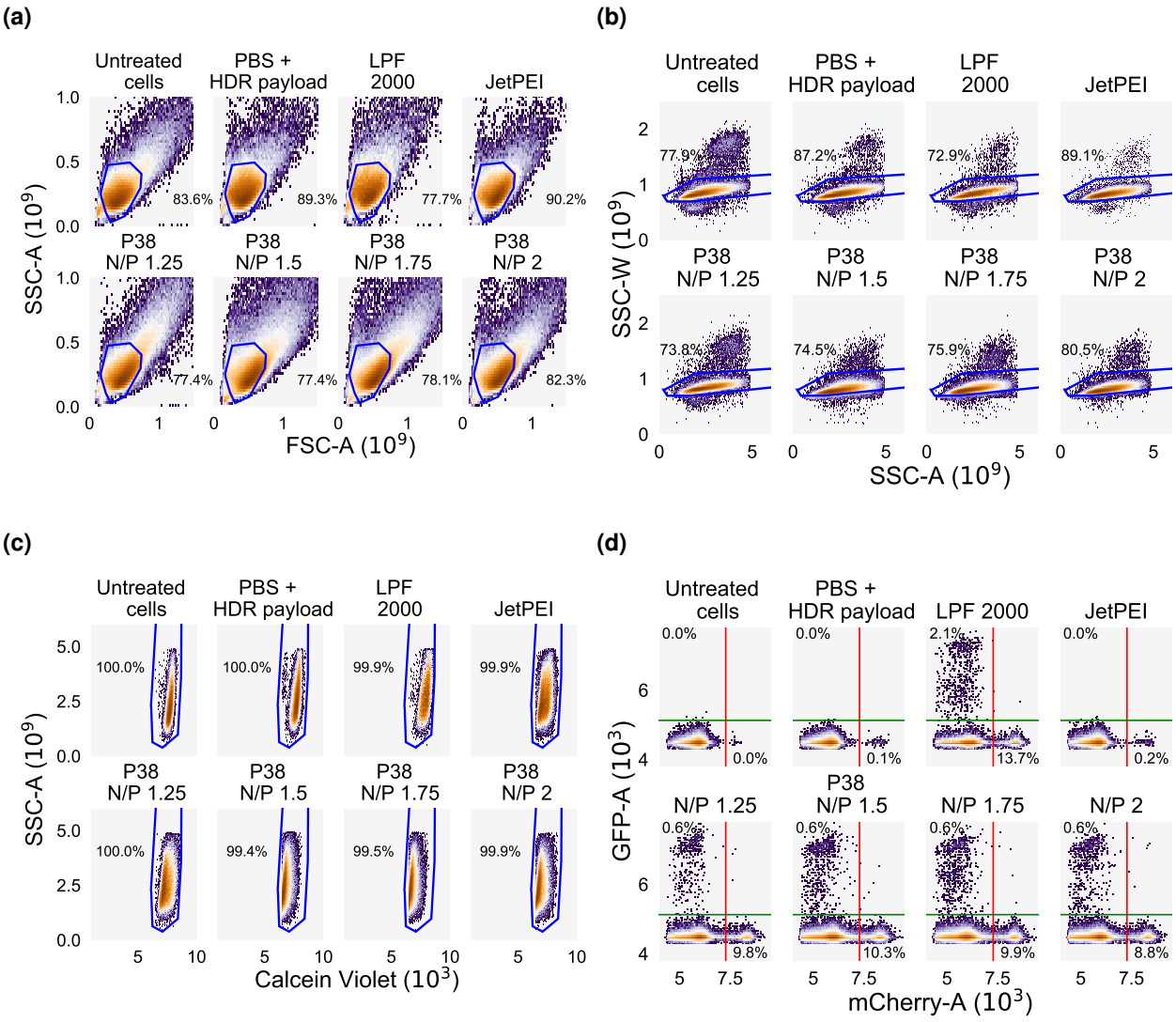

Figure S13. Representative gating schemes for (a) Cells, (b) Single cells, (c) Live cells, and (d) GFP$^+$ and mCherry$^+$ cells. These schemes are from the same transfection experiment as Figure 5 in the main manuscript.

5.3 Internalization experiments

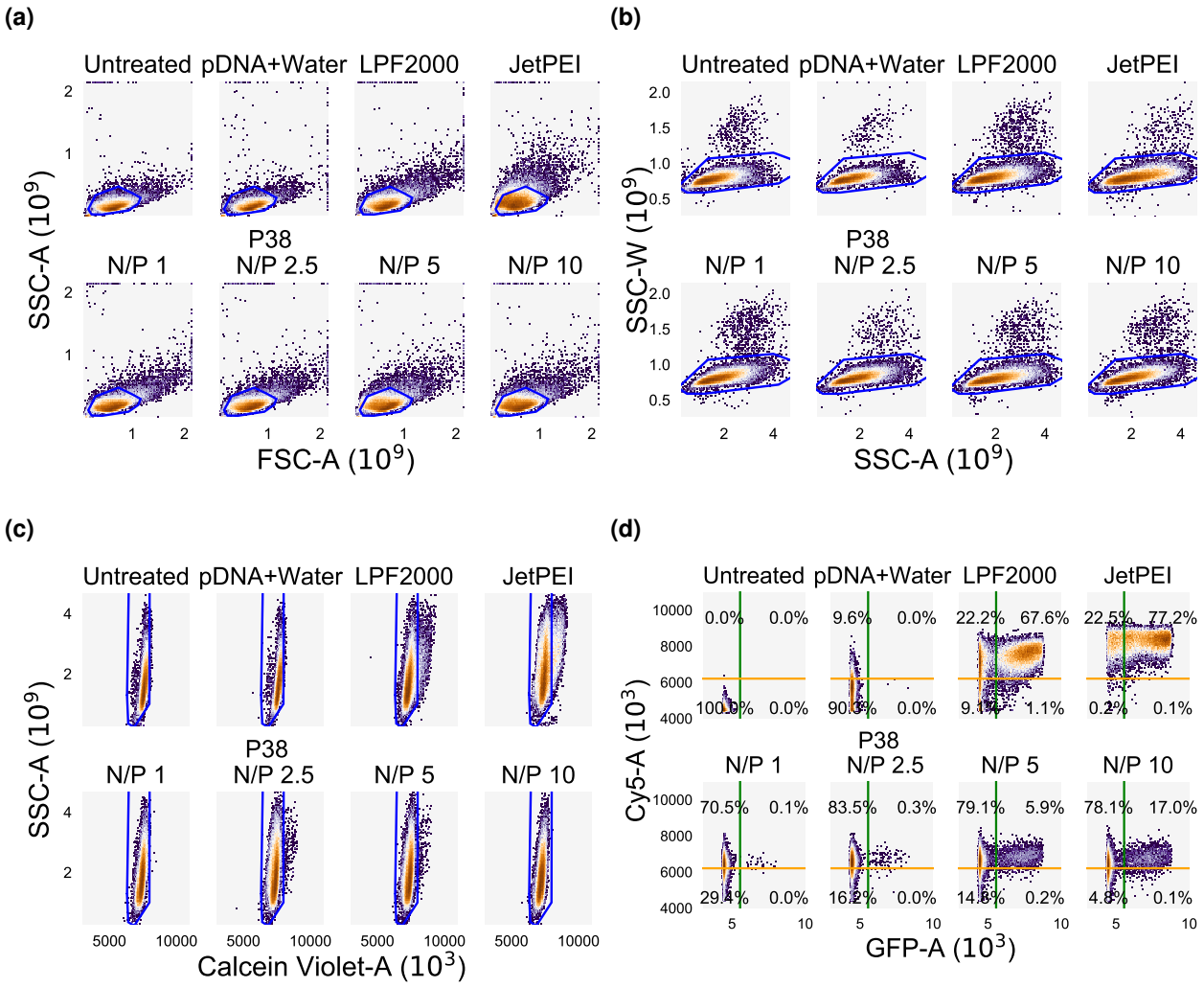

Figure S14. Representative gating schemes from for (a) Cells, (b) Single cells, (c) Live cells, and (d) GFP⁺ cells. Schemes are from experiments measuring uptake of Cy5-labeled polyplexes of P38 and commercial transfection reagents in HEK293T cells.

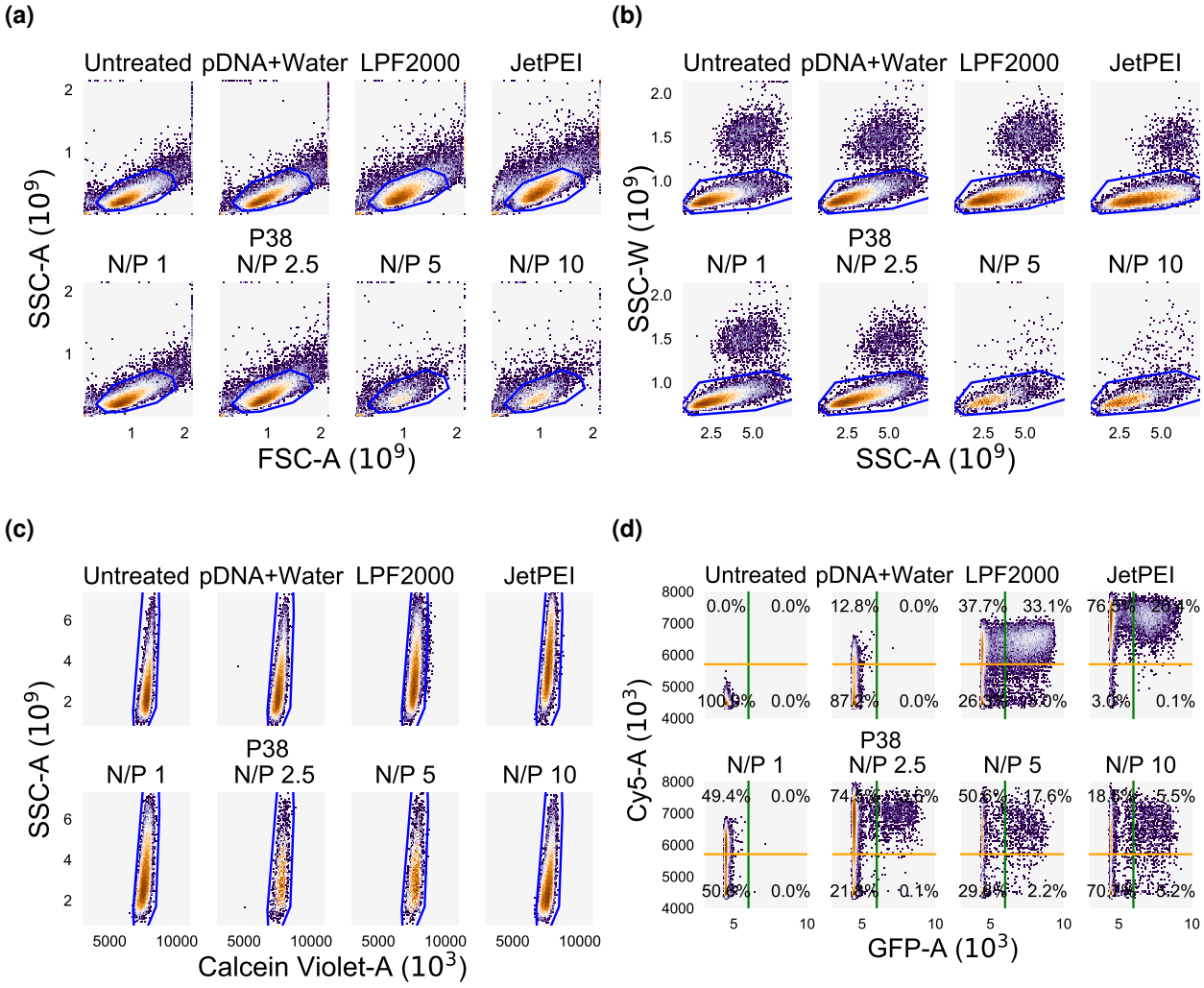

Figure S15. Representative gating schemes for (a) Cells, (b) Single cells, (c) Live cells, and (d) GFP$^+$ cells. Schemes are from experiments measuring uptake of Cy5-labeled polyplexes of P38 and commercial transfection reagents in ARPE-19 cells.

6 Cytotoxicity evaluation

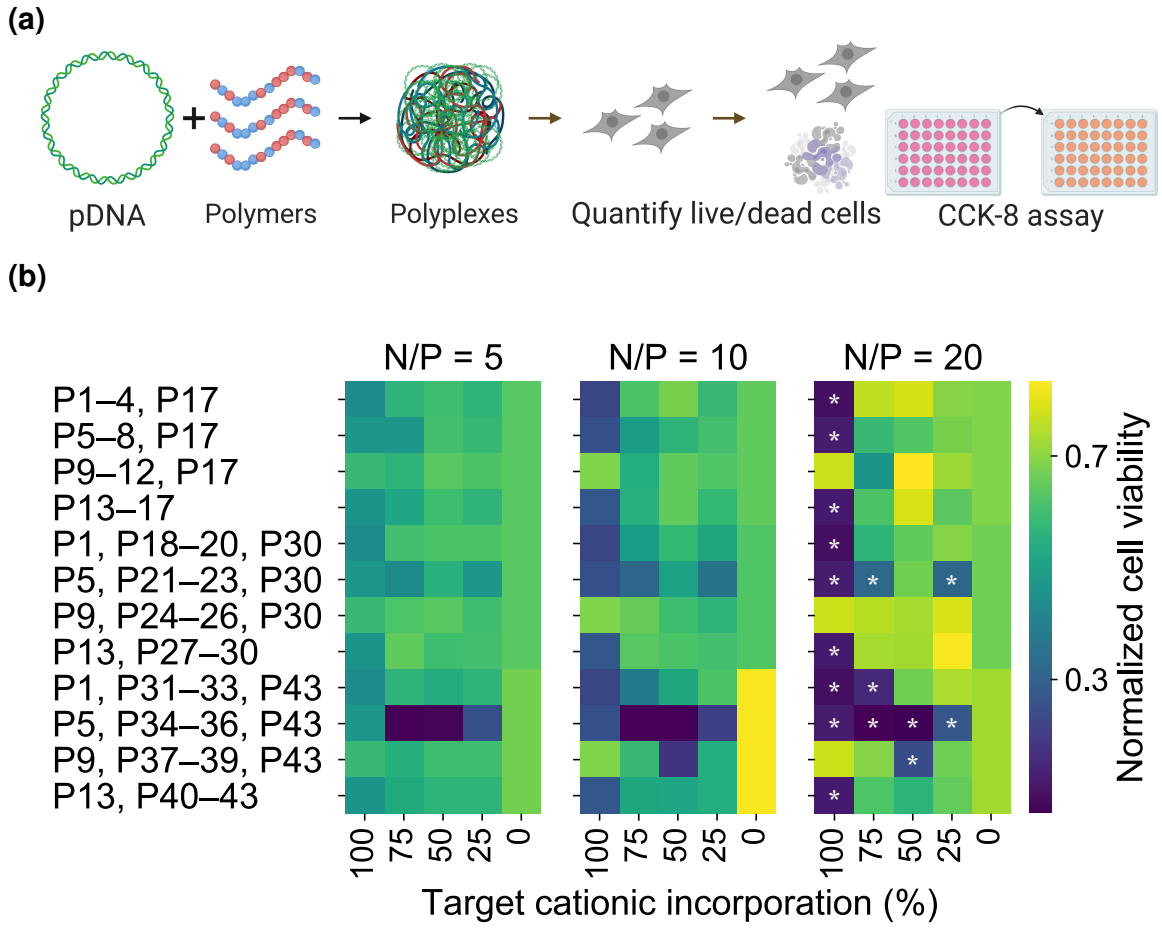

Figure S16. (a) A library-wide assessment of cytotoxicity was performed using the CCK-8 assay (b) Ten polymers (highlighted with white stars), including the hit polymer, exhibited the lowest cellular viability in the library (bottom 5 percentile). Among these toxic formulations (P1, P5, P13, P21, P23, P31, P34-36, and P38) only P1, P13, and P31 exhibited sub-par transfection efficiencies, underscoring the trade-off between pDNA delivery efficiency and cytotoxicity.

Through our library-wide evaluation of cellular viability, we elucidated the roles of polymer composition, protonation equilibria or pK_a, ζ-potential, and hydrophobic-hydrophilic balance in determining the balance between delivery performance and polymer-mediated toxicity. Viability declined with increasing incorporation of cationic repeat units, an effect that was particularly pronounced at higher N/P ratios. This suggests that membrane penetration by cationic polymers and the resulting loss of membrane integrity might be the primary mechanism through which polymers induce cell death. For P34–P36, we observed a much earlier onset of polymer-mediated toxicity at an N/P of 5. Overall, we identified ten polymers as the most toxic within our library (bottom 5 percentile, highlighted with white stars in Figure S16). As expected, a strong correlation was evident between GFP expression in Figure 2 in the main manuscript, and toxicity levels mapped in Figure S16, with only three of the ten most toxic formulations proving to be ineffective vehicles for pDNA delivery. We also compared the toxicity caused by P38 with that of JetPEI and LPF2000 (Figure S17) and found that the viability of cells treated with P38 were comparable to that of JetPEI at an N/P ratio of 10 but was significantly lower than that of Lipofectamine 2000. We attribute the lower toxicity of Lipofectamine 2000 to its reliance on membrane fusion rather than membrane penetration to facilitate cellular entry. These trends in polymer-mediated toxicity cannot be explained through the lens of a single design attribute, be it composition, cationicity, or hydrophobicity, suggesting that polymer-induced toxicity is a complex function of multiple design attributes.

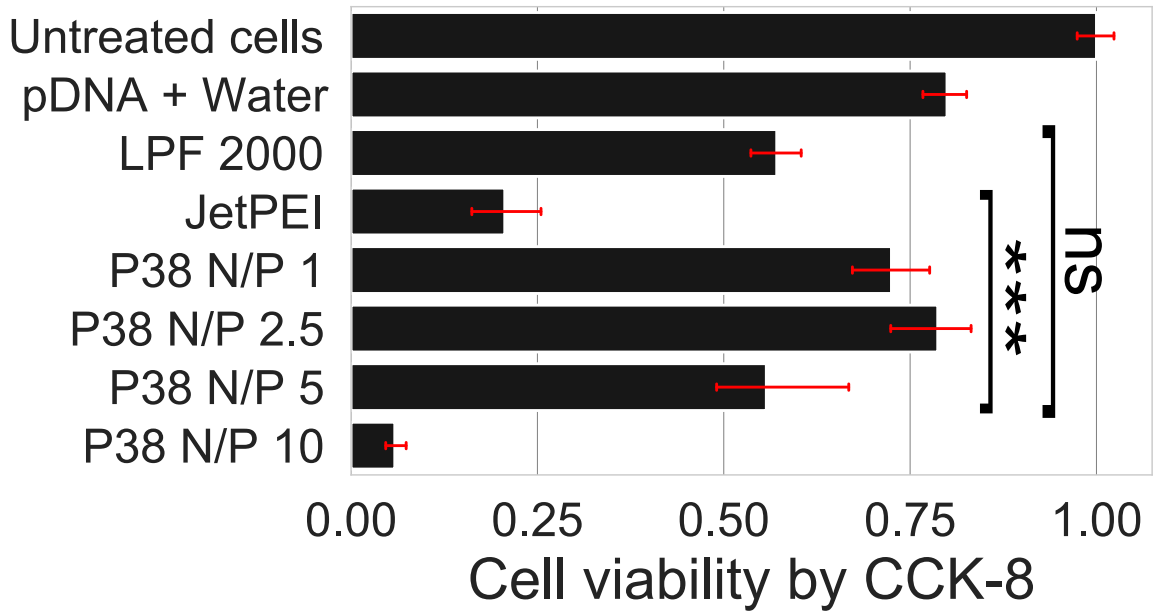

Figure S17. CCK-8 assays were used to evaluate toxicity during transient transfection in HEK293 cells. At N/P ratios of 2.5 and 1, P38 was far less toxic than LPF 2000 and JetPEI. However, polyplexes formed at higher N/P ratios proved to be extremely toxic.

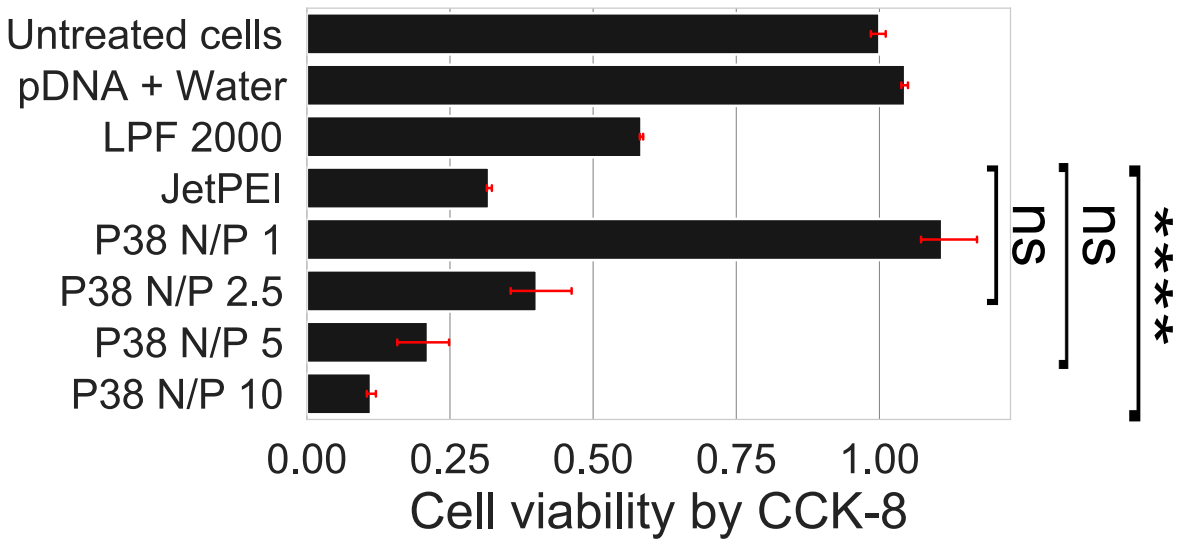

Figure S18. CCK-8 assays were used to evaluate toxicity in ARPE-19 cells resulting from polymer-mediated pDNA delivery. Although JetPEI achieved transfection efficiencies approaching 40%, this was at the cost of substantial cytotoxicity. P38, on other hand, far less toxic than JetPEI at an N/P ratio of 2.5, while exhibiting a transfection efficiency approaching 20%

7 Supplemental confocal data

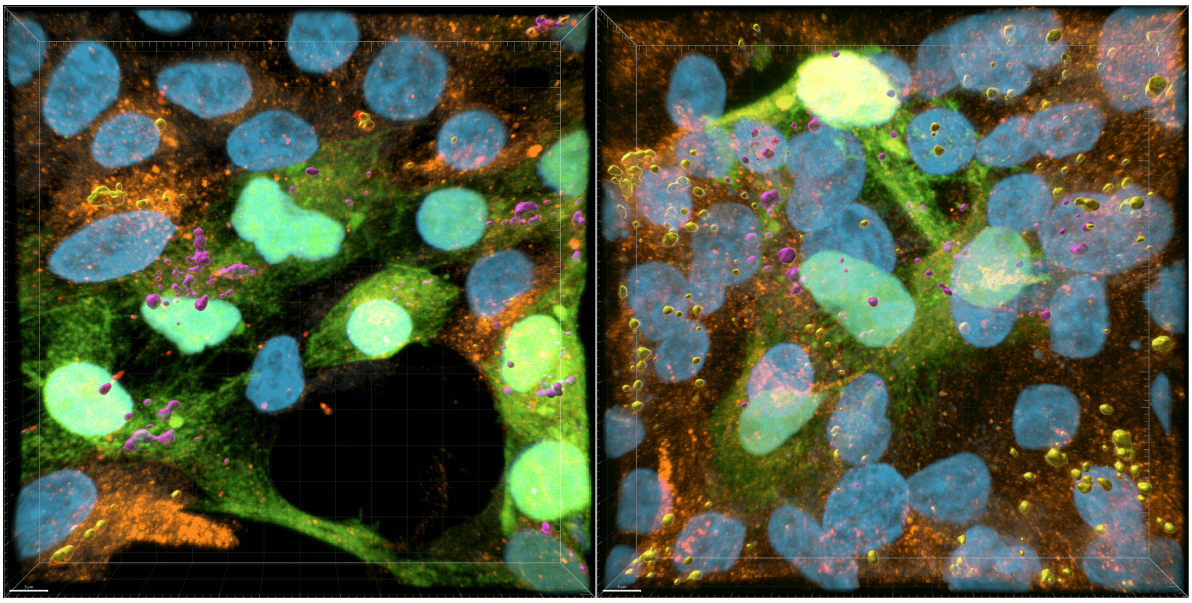

Figure S19. Confocal renderings of cells transfected with P38 polyplexes (left) and P41 polyplexes (right) at an N/P ratio of 5. Scale bar is 5 μm. Yellow surface renderings represent colocalized Alexa Fluor 568 and Cy5 signals in GFP^- cells. Magenta surface renderings represent colocalized Alexa Fluor 568 and Cy5 signals inside GFP^+ cells. Among GFP^- cells, a larger proportion of DNA polyplexes in the P41-treated group colocalizes with lysosomes than in the P38-treated group.

8 Statistical modeling: RNP causal effects and Cross-validation

Figure S20 summarizes average causal effects calculated for RNP payloads.

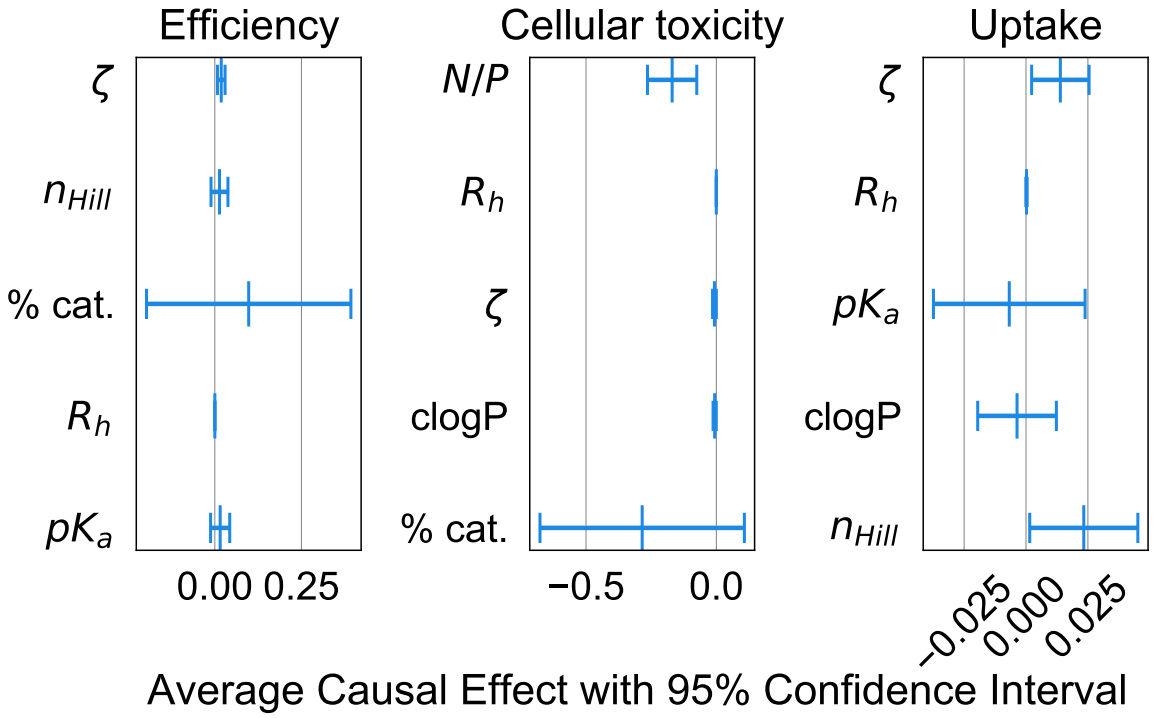

Figure S20. Direct causal effects of the top five features from SHAP analysis of the RNP dataset were computed along with 95% confidence intervals.

Figure S21 presents the mean ROC-AUC value for 5-fold cross validation for each cargo using a balanced random forest model. Two groupings for the cross validation splits were considered:

1. Stratified: the relative distribution of positive and negative examples remains constant across folds.

2. Polymer: there are no common polymer examples across training and testing folds.

The results demonstrate good predictability using the 9 polyplex descriptors (AUC > 0.8)

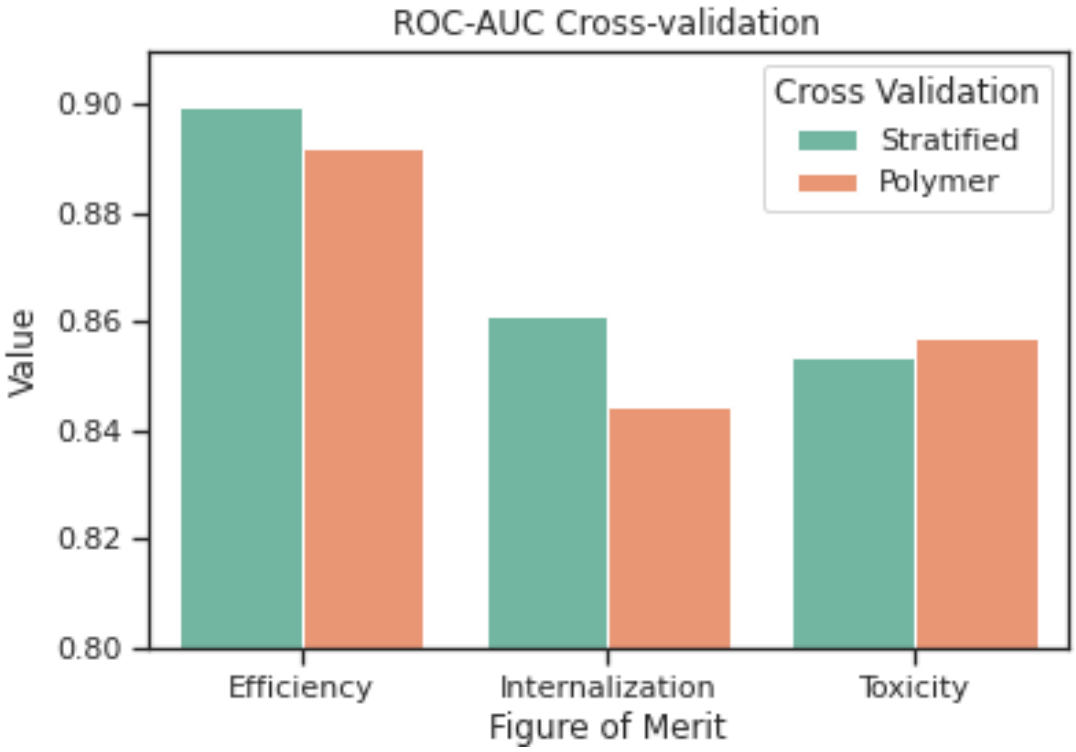

Figure S21. Summary of cross-validation results. For both the splits (Stratified and Polymer), we observed AUC values above 0.8, indicating that the predictive power of the model is high.

9 Biological replicates

9.1 Delivery efficiency

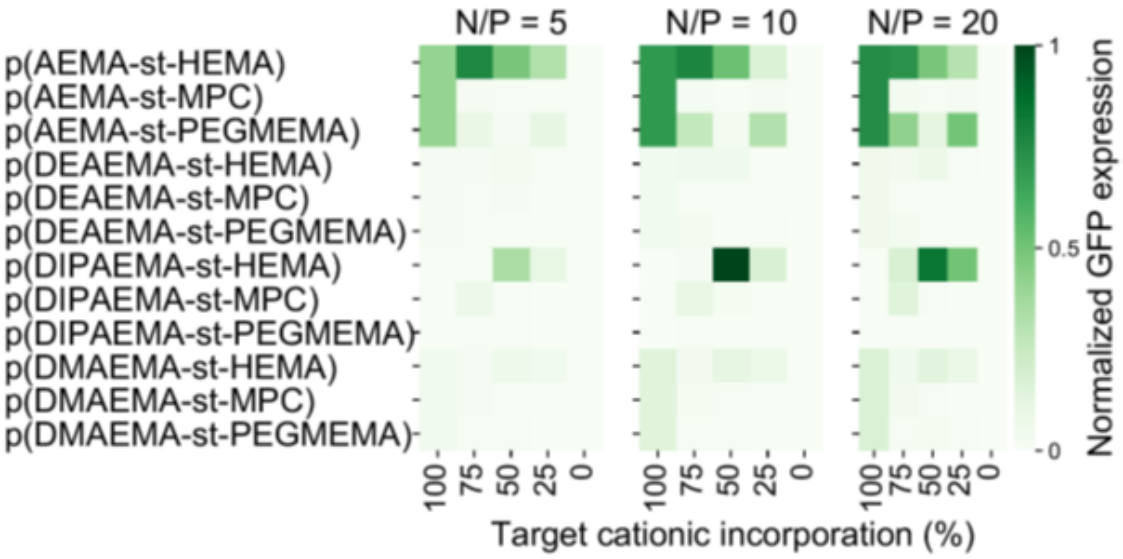

Figure S22. Summary of pDNA delivery efficiency trends across the library (first biological replicate).

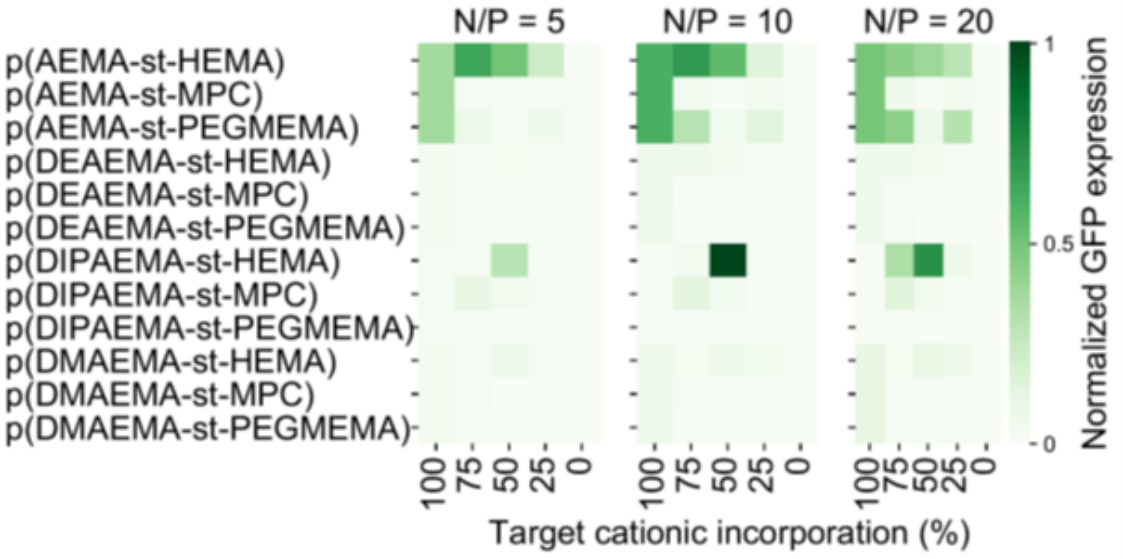

Figure S23. Summary of pDNA delivery efficiency trends across the library (second biological replicate).

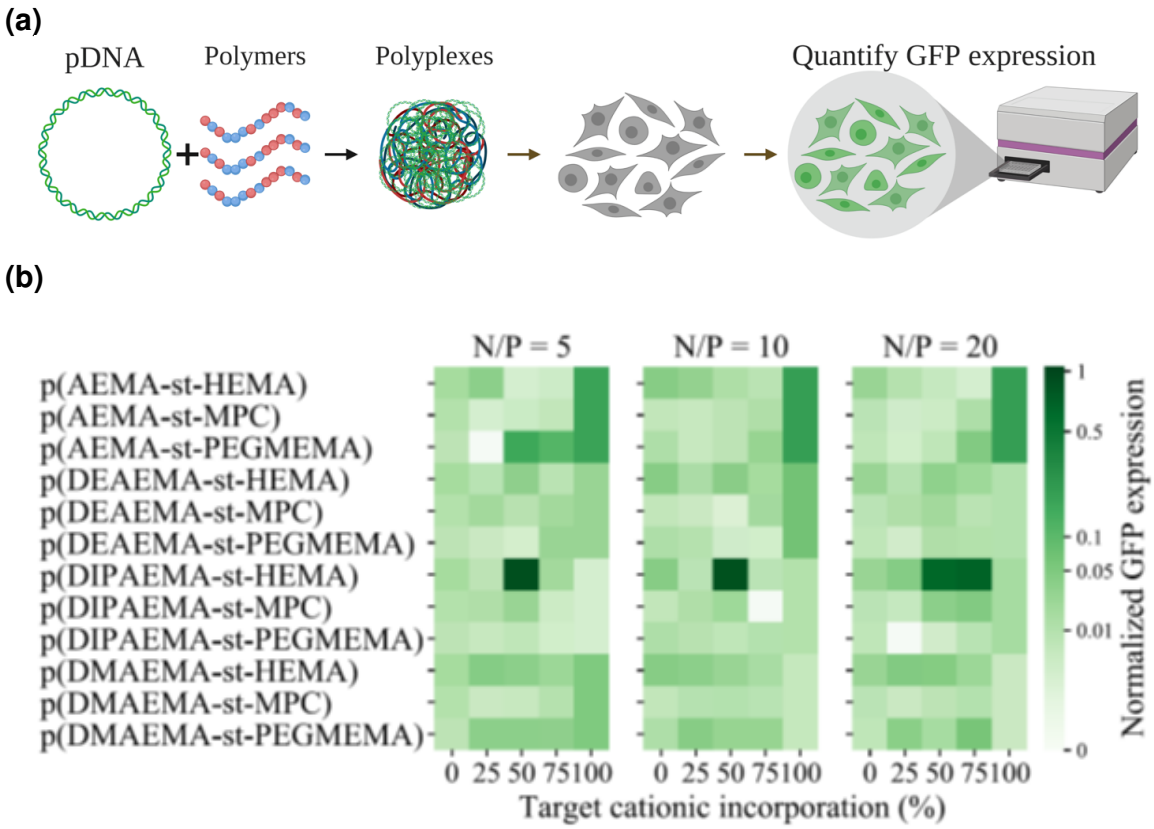

Figure S24. Summary of screening studies performed using fluorescence plate reader (instead of flow cytometry). Heat map values represent average over two wells.

9.2 Cellular toxicity

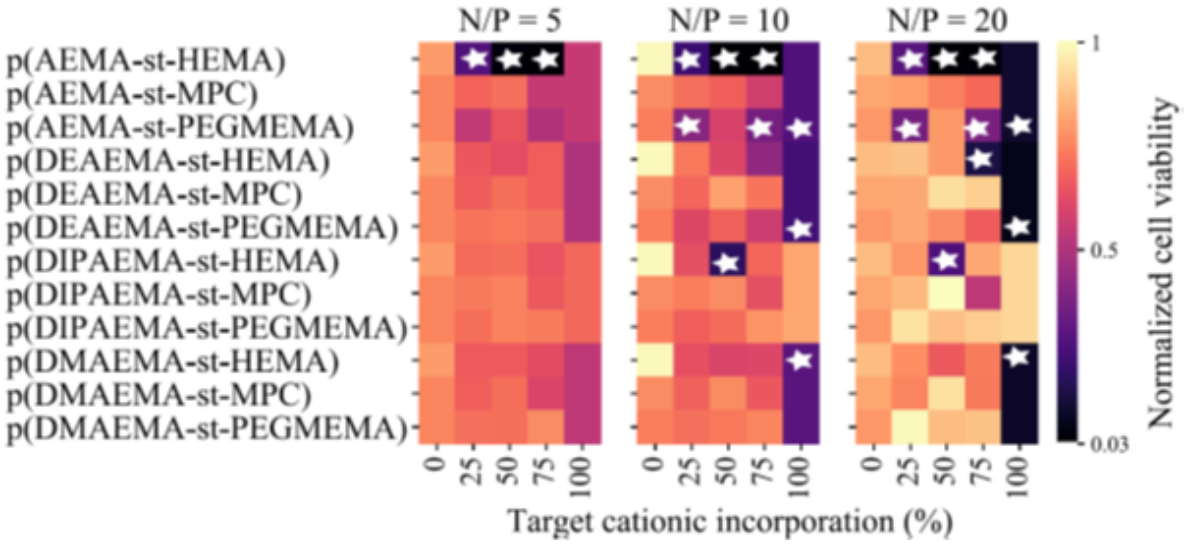

Figure S25. Summary of cellular toxicity trends across the library (first biological replicate).

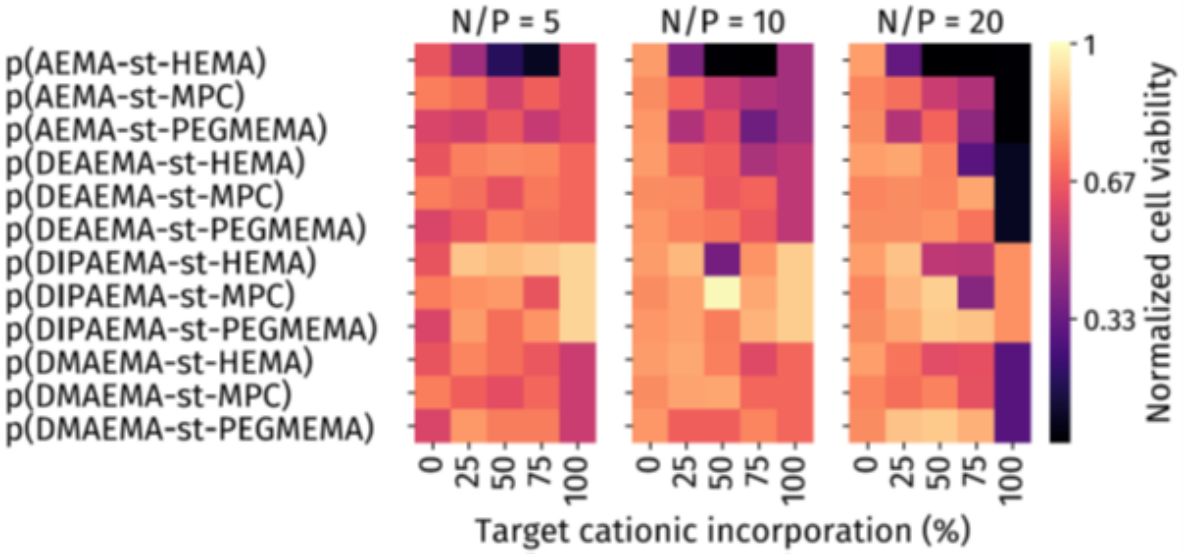

Figure S26. Summary of cellular toxicity trends across the library (second biological replicate).

9.3 Cellular uptake

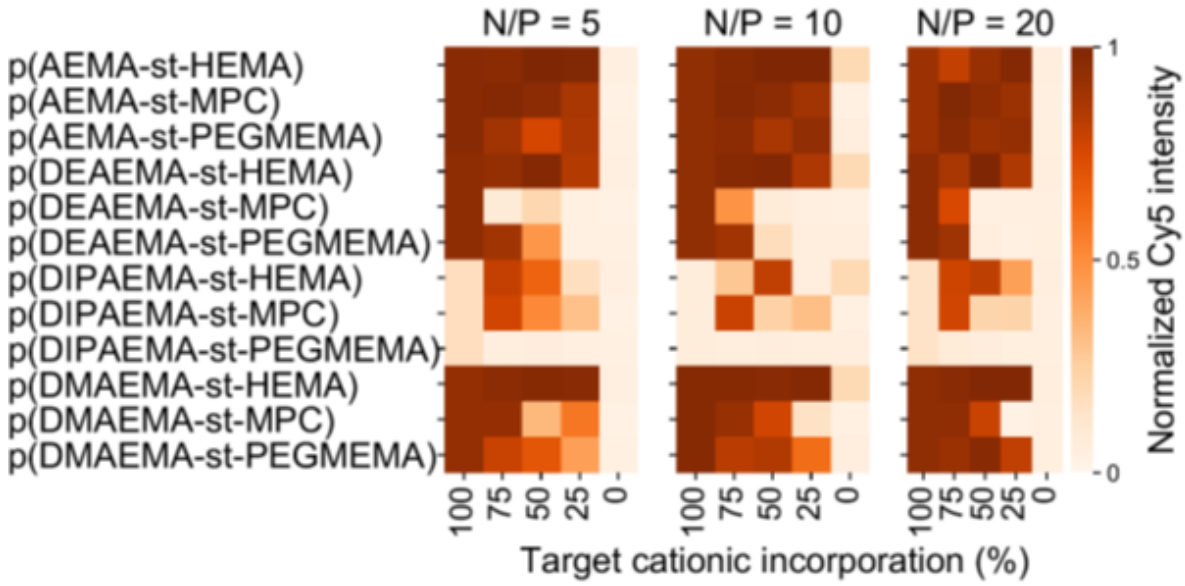

Figure S27. Summary of cellular uptake trends across the library (first biological replicate).

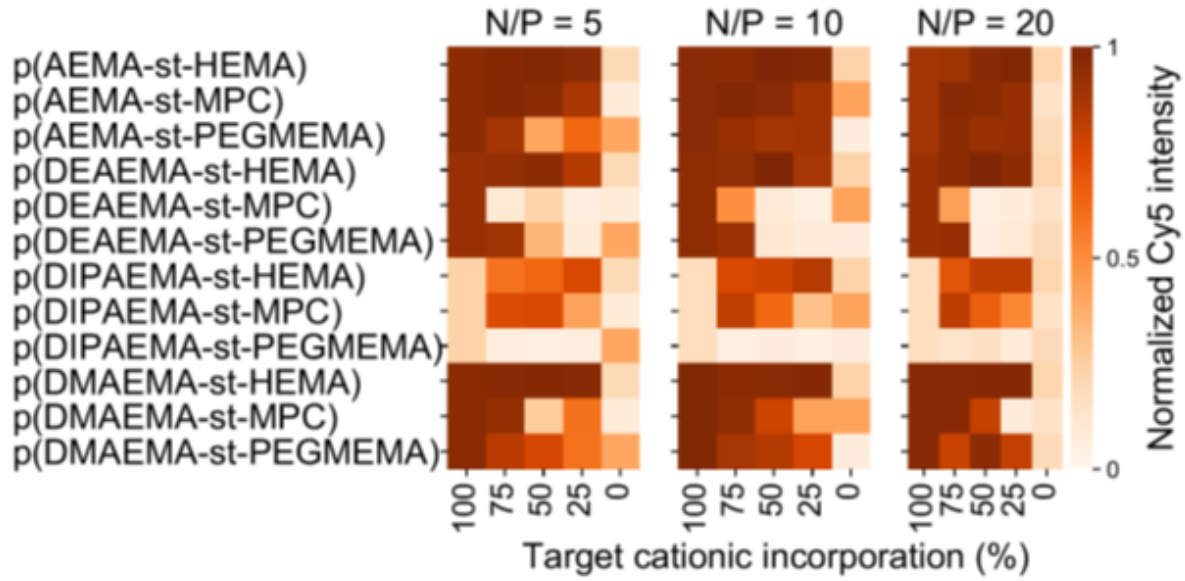

Figure S28. Summary of cellular uptake trends across the library (second biological replicate).