Electronic Supplementary Information

A machine learning approach to predict cellular uptake of pBAE polyplexes

Aparna Loecher,*[a] Michael Bruyns-Haylett,*[a] Pedro Ballester,[a] Salvador Borros,[b] and Nuria Oliva†[a,b]

a. Department of Bioengineering, Imperial College London, SW7 2AZ London (UK).

b. Department of Bioengineering, Institut Quimic de Sarria, Via Augusta 390, 08017 Barcelona (Spain).

*Equal contribution.

† Corresponding author: nuria.oliva@iqs.url.edu.
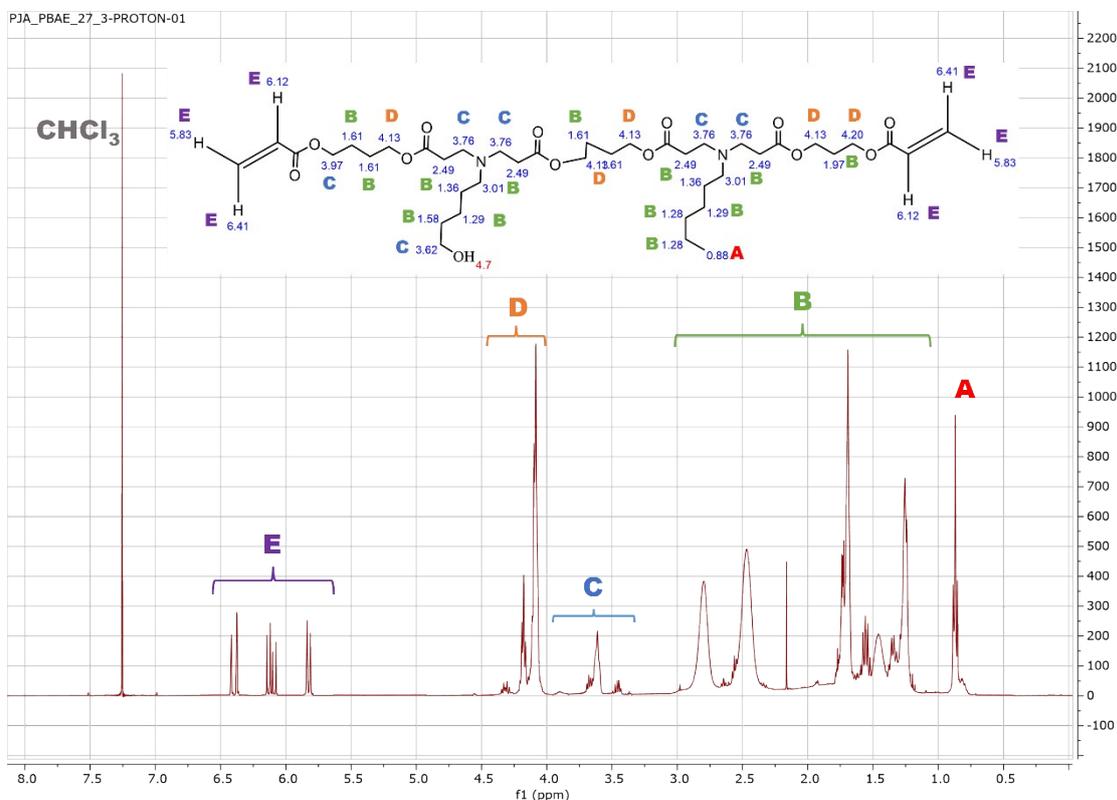
Supplementary Figure 1. C32 pBAE ¹H-NMR sample spectrum.
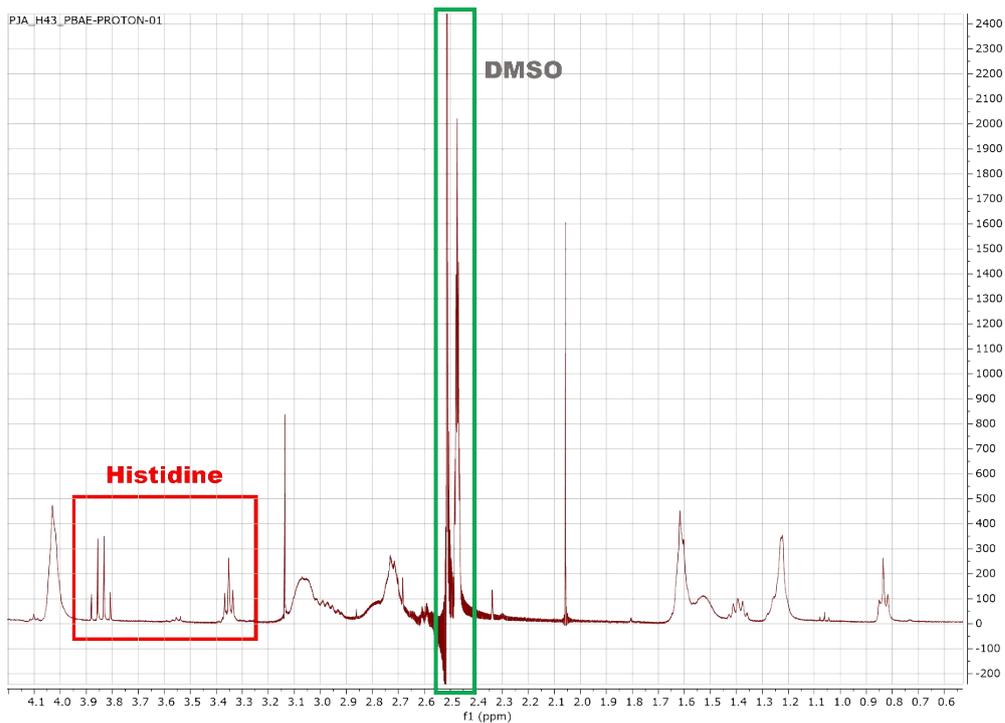
Supplementary Figure 2. C6H ^{1}H-NMR sample spectrum.
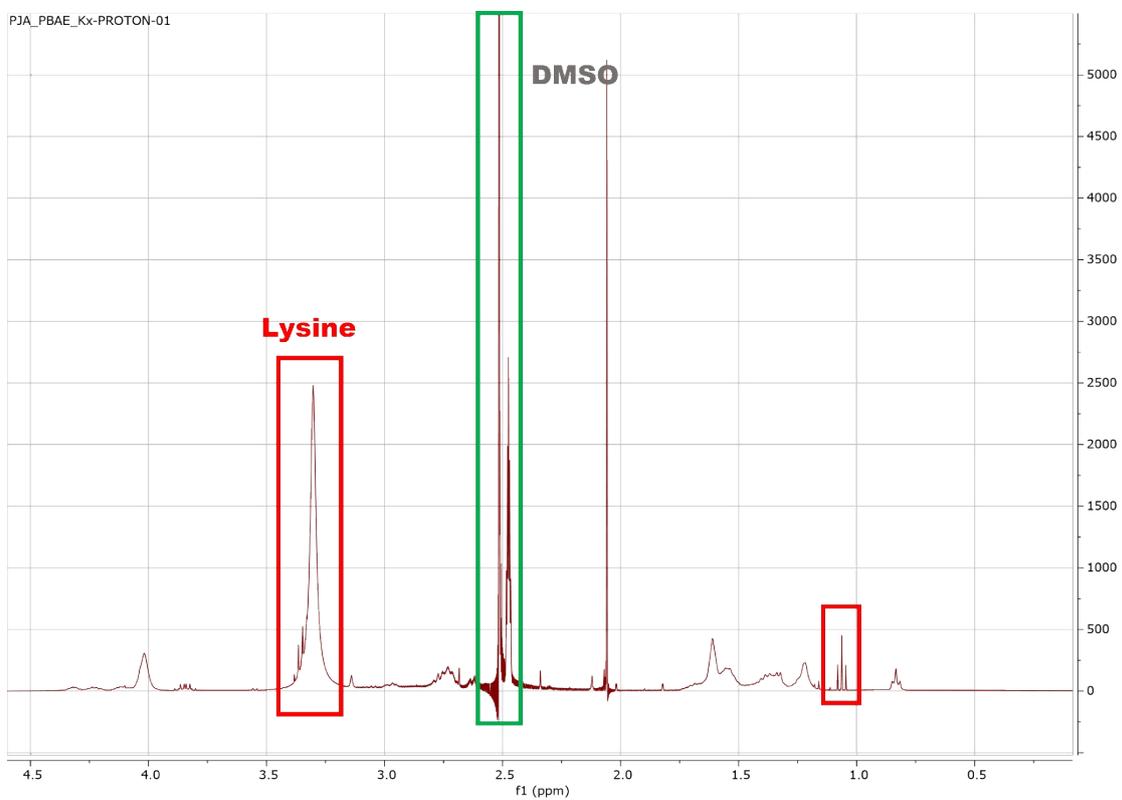
Supplementary Figure 3. C6K ^{1}H-NMR sample spectrum.
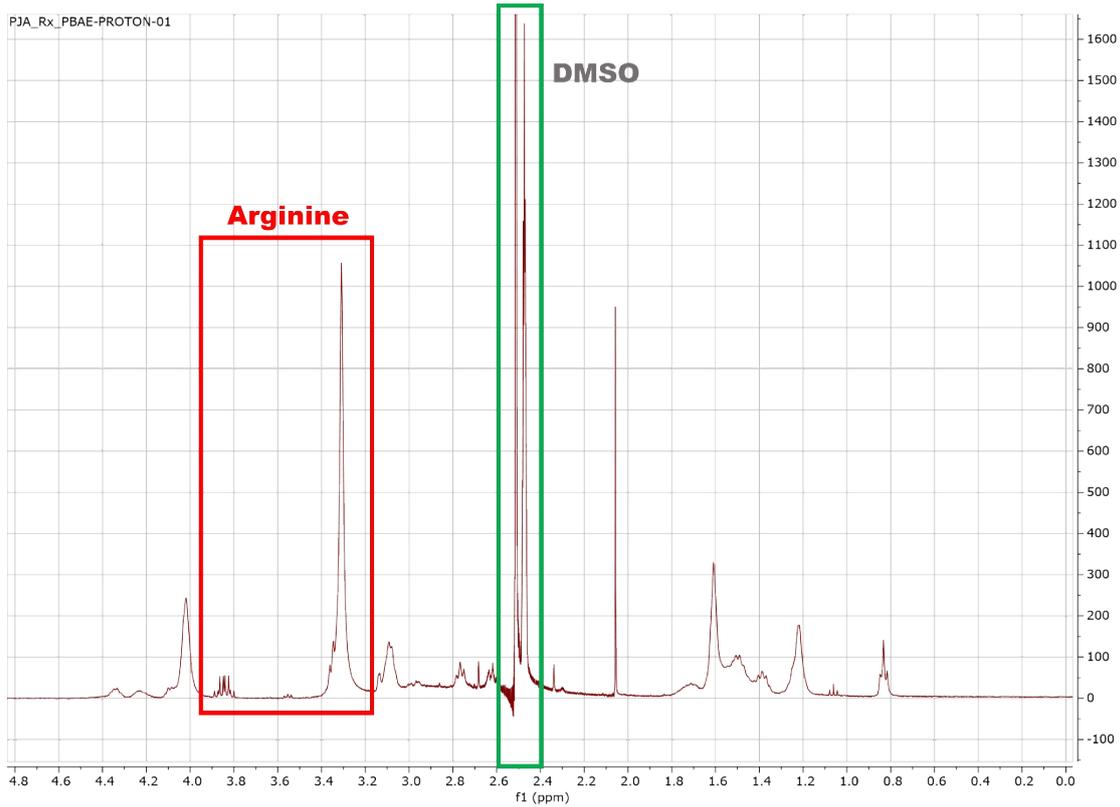
Supplementary Figure 4. C6R ^{1}H-NMR sample spectrum.
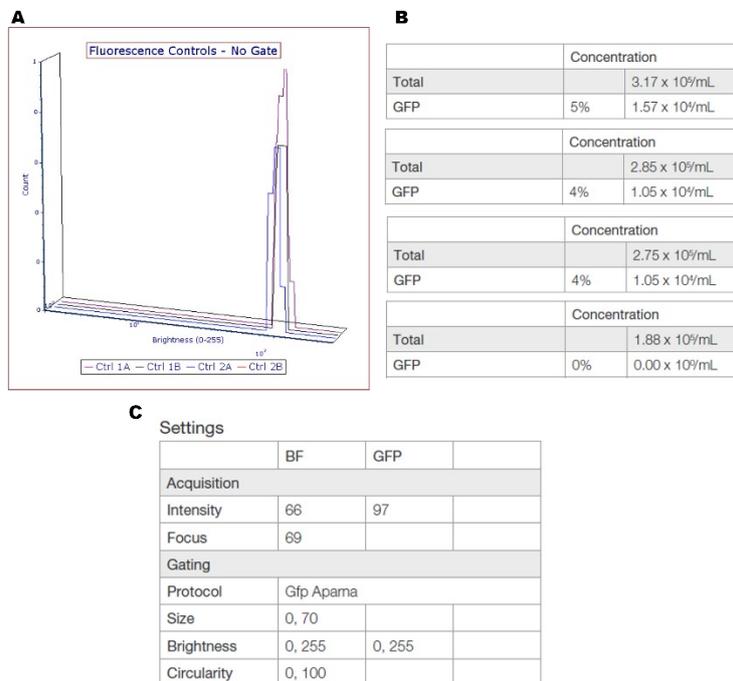
Supplementary Figure 5. Untreated controls gating information. (A) Histogram of 2 untreated controls, measured twice each (1A, 1B, 2A and 2B). (B) Concentration of GFP-positive cells in the untreated controls. (C) Gating settings for the remaining transfections experiments, where GFP intensity was set to 97 RFUs (all untreated cells that had a signal had intensities below this value, hence determined as background signal).
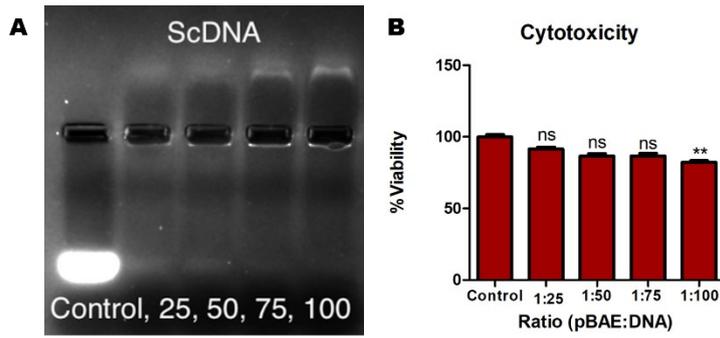
Supplemental Figure 6. (A) Electrophoresis gel of scramble DNA (control) and scDNA-loaded pBAE polyplexes at polymer:DNA ratio of 1:25, 1:50, 1:75 and 1:100. (B) Viability of OVCAR-4 cells 24-hour after treatment with pBAE polyplexes, compared to untreated control.
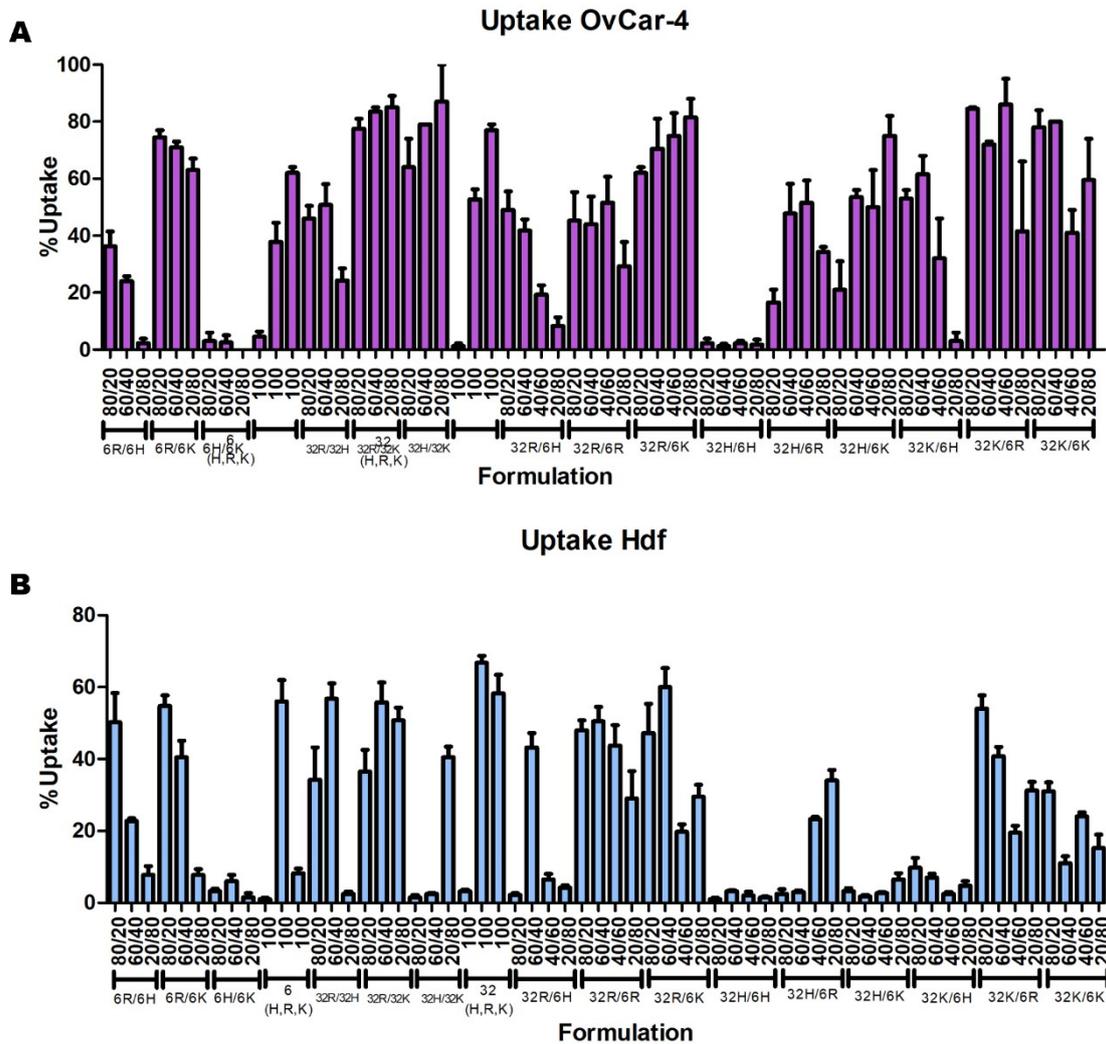
Supplemental Figure 7. Percent (A) OVCAR-4 and (B) HDF cells transfected with pBAE polyplexes loaded with fluorescent scDNA (Alexa 488-tagged).
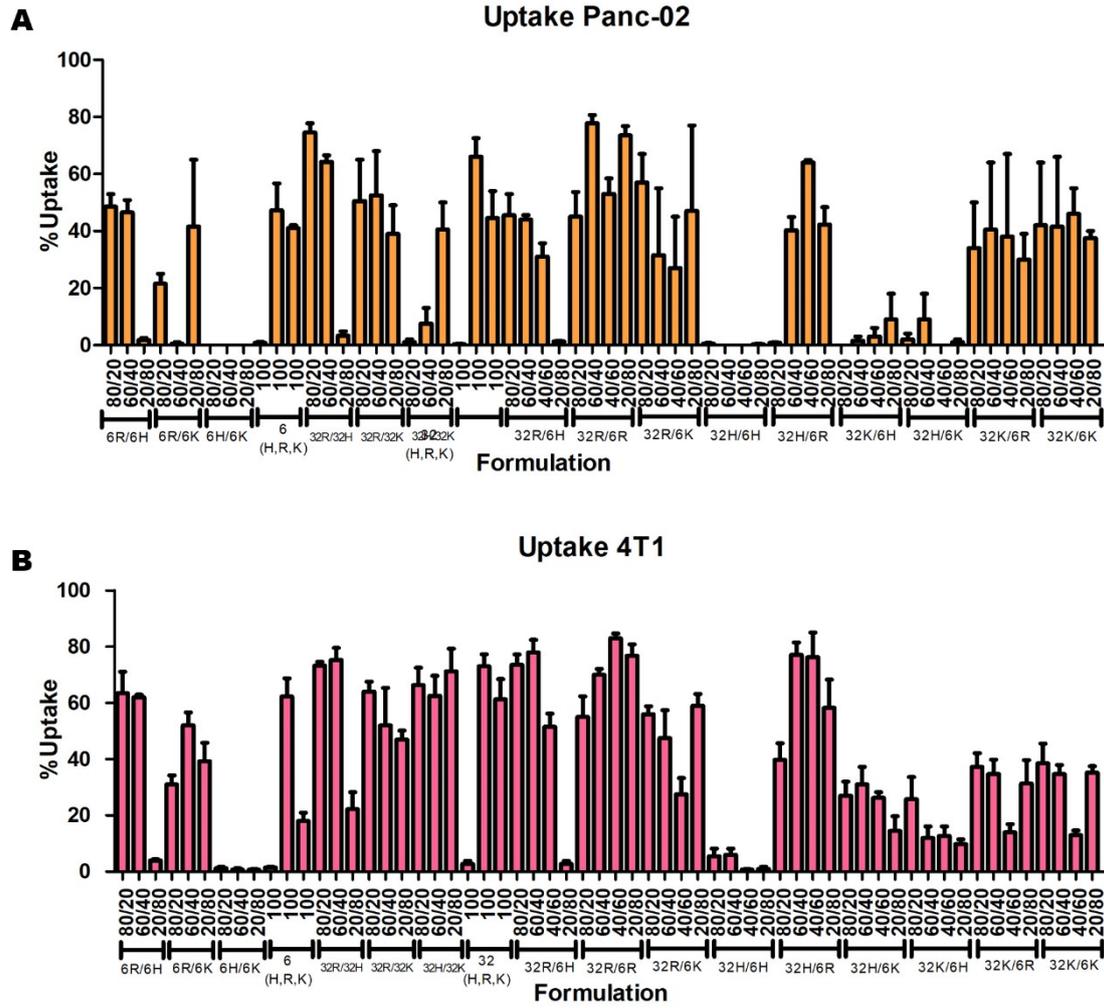
Supplemental Figure 8. Percent (A) Panc02 and (B) 4T1 cells transfected with pBAE polyplexes loaded with fluorescent scDNA (Alexa 488-tagged).
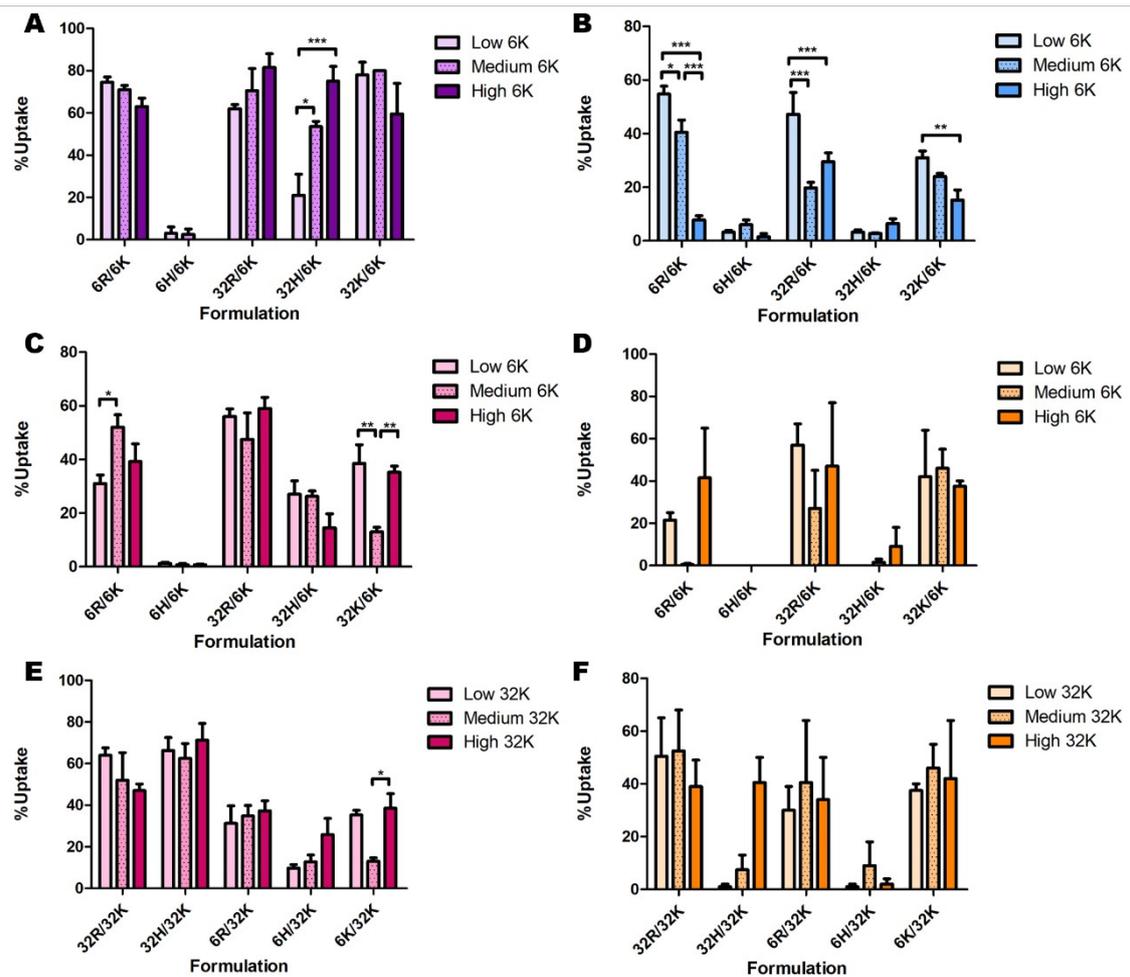
Supplemental Figure 9. Initial analysis of the effects of backbone chemistry and terminal oligopeptide on cellular uptake of 6K and 32K low, medium and high ratios in OVCAR-4 (A, purple bars), HDF (B, blue bars), 4T1 (C & E, pink bars) and Panc02 (D & F, orange bars).
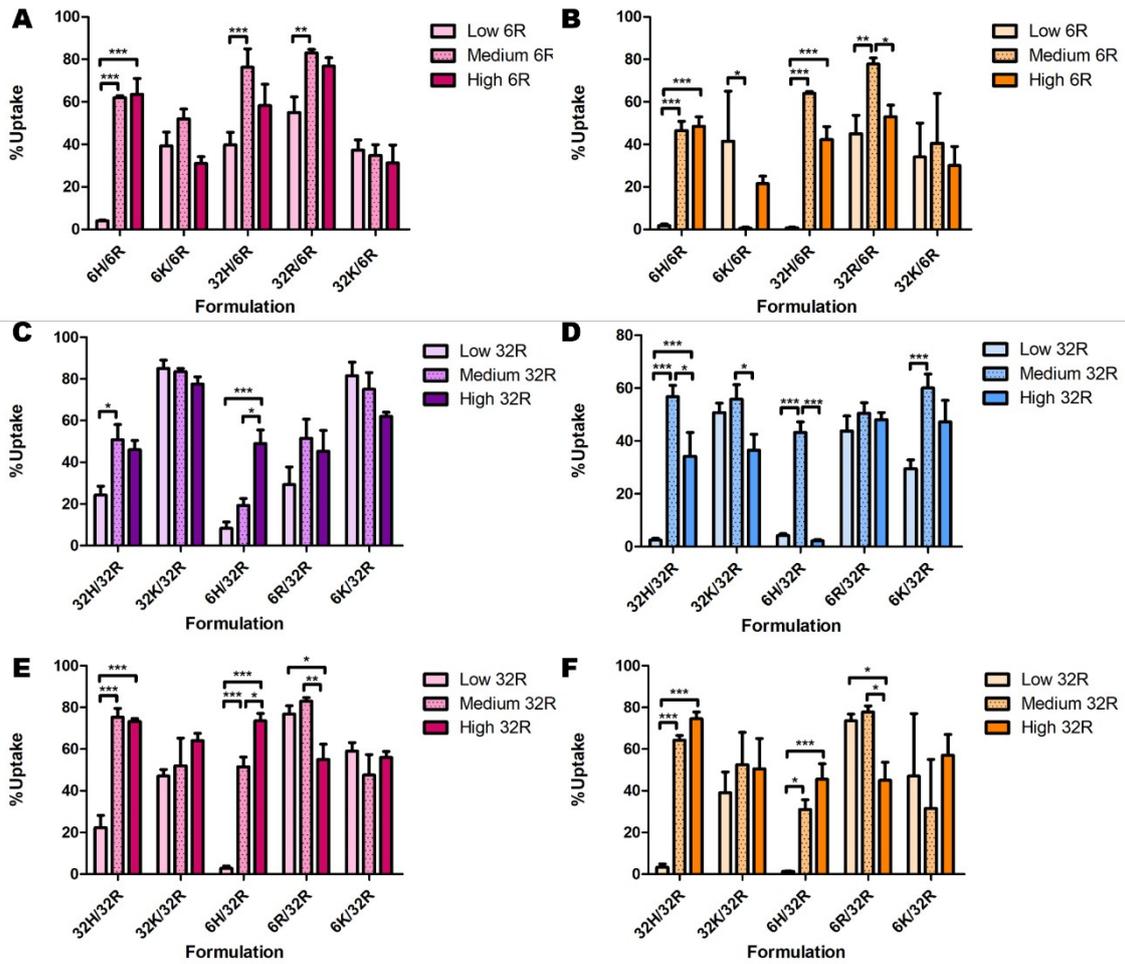
Supplemental Figure 10. Initial analysis of the effects of backbone chemistry and terminal oligopeptide on cellular uptake of 6R and 32R low, medium and high ratios in 4T1 (A & E, pink bars), Panc02 (B & F, orange bars), OVCAR-4 (C, purple bars) and HDF (D, blue bars).
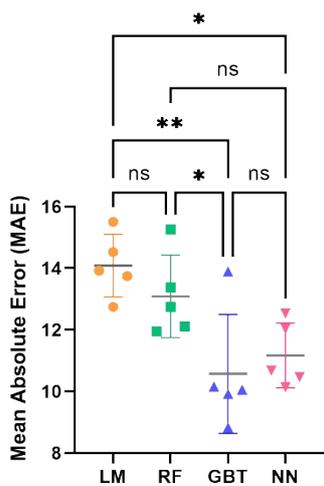
Supplemental Figure 11. Average Mean Absolute Error (MAE) of five runs for each model: multi-linear regression (LM), random forest (RF), gradient-boosted trees (GBT) and deep neural network (NN).
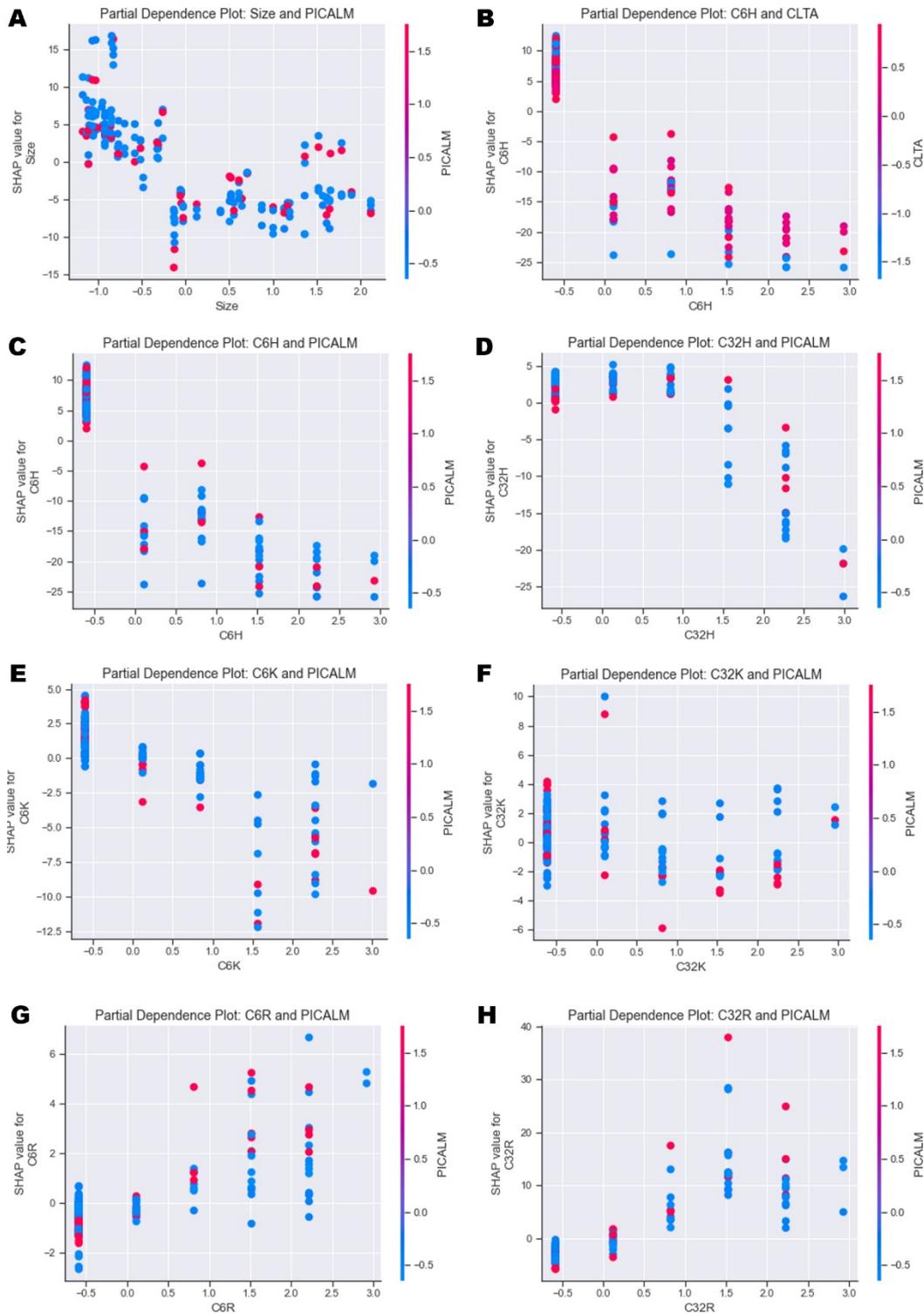
Supplementary Figure 12. Partial dependence plots investigating the individual feature interactions between PICALM expression and (A) size, (C) C6H, (D) C32H, (E) C6K, (F) C32K, (G) C6R and (H) C32R, and (B) between CLTA and C6H.
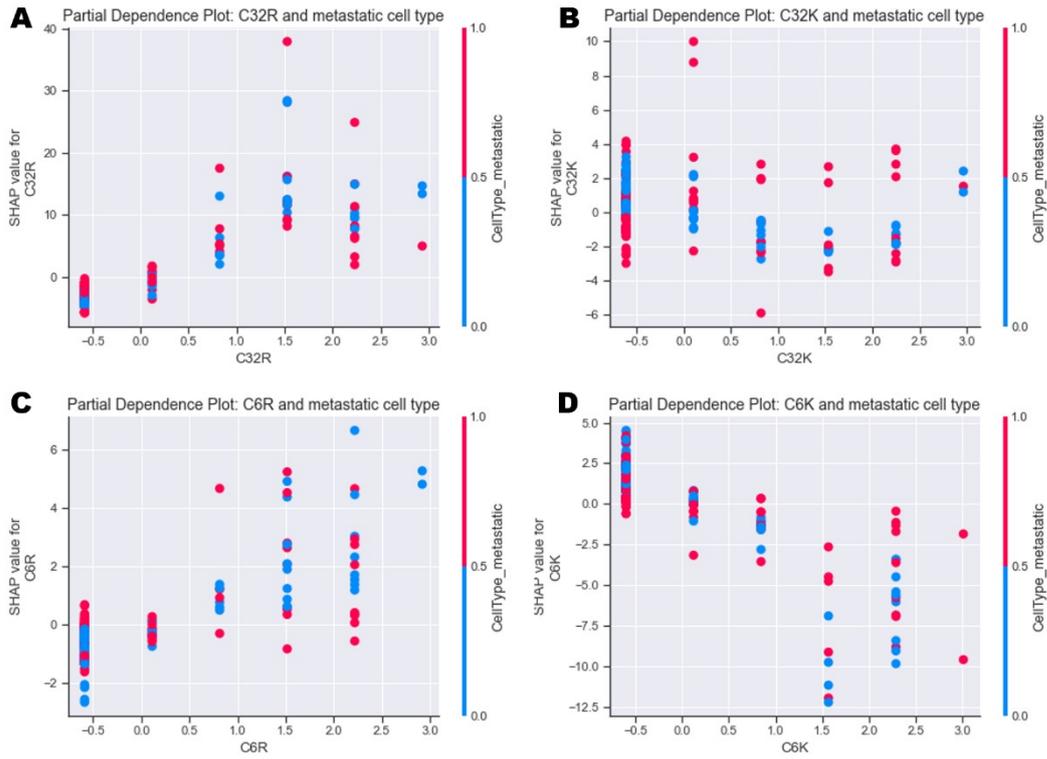
Supplementary Figure 13. Partial dependence plots investigating the individual feature interactions between cell metastatic potential and (A) C32R, (B) C32K, (C) C6R, (D) C6K. Input was indexed to have a value of 1 for metastatic cells (OVCAR-4 and 4T1) and 0 for non-metastatic cells (HDF and Panc02).