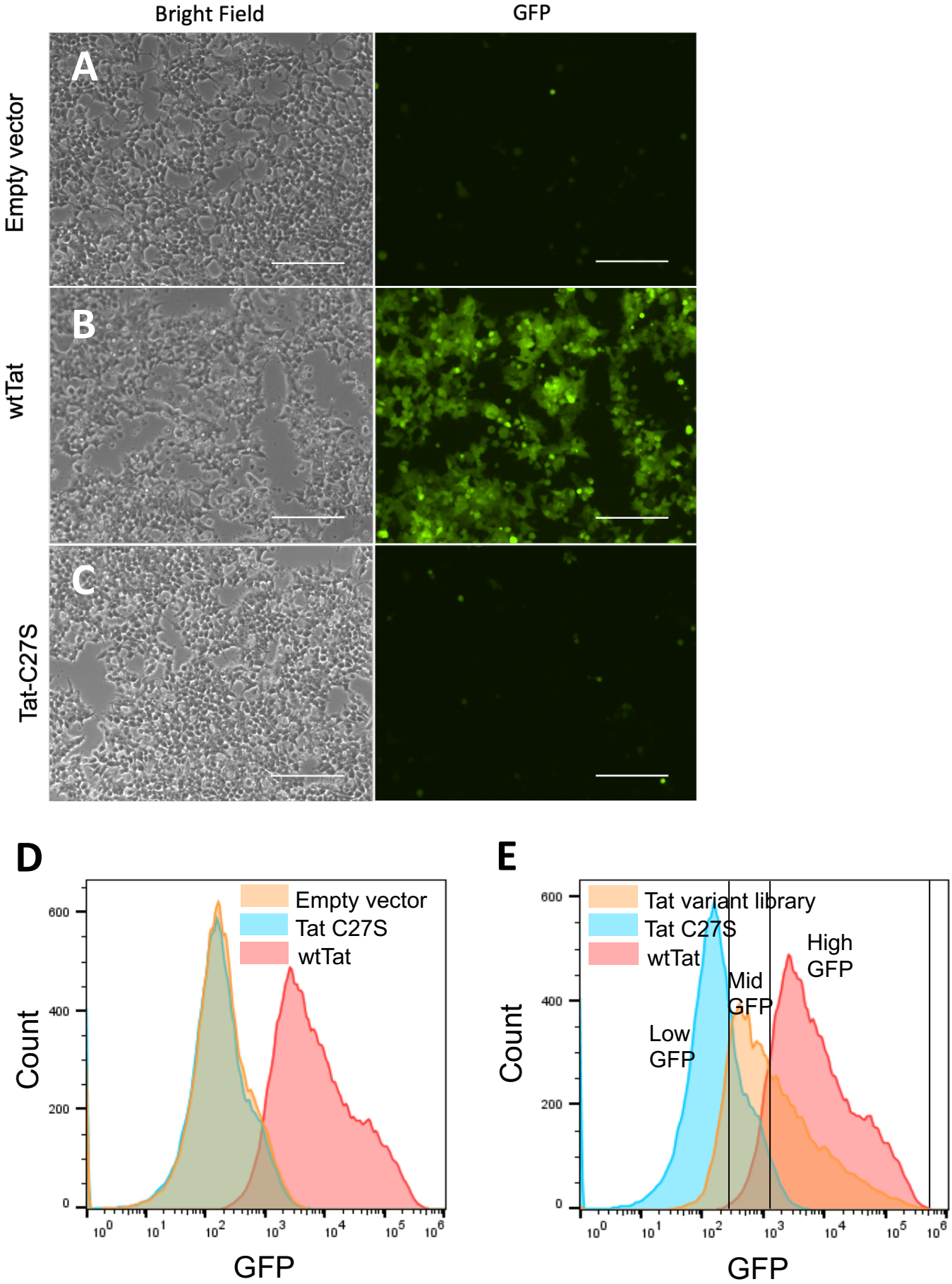
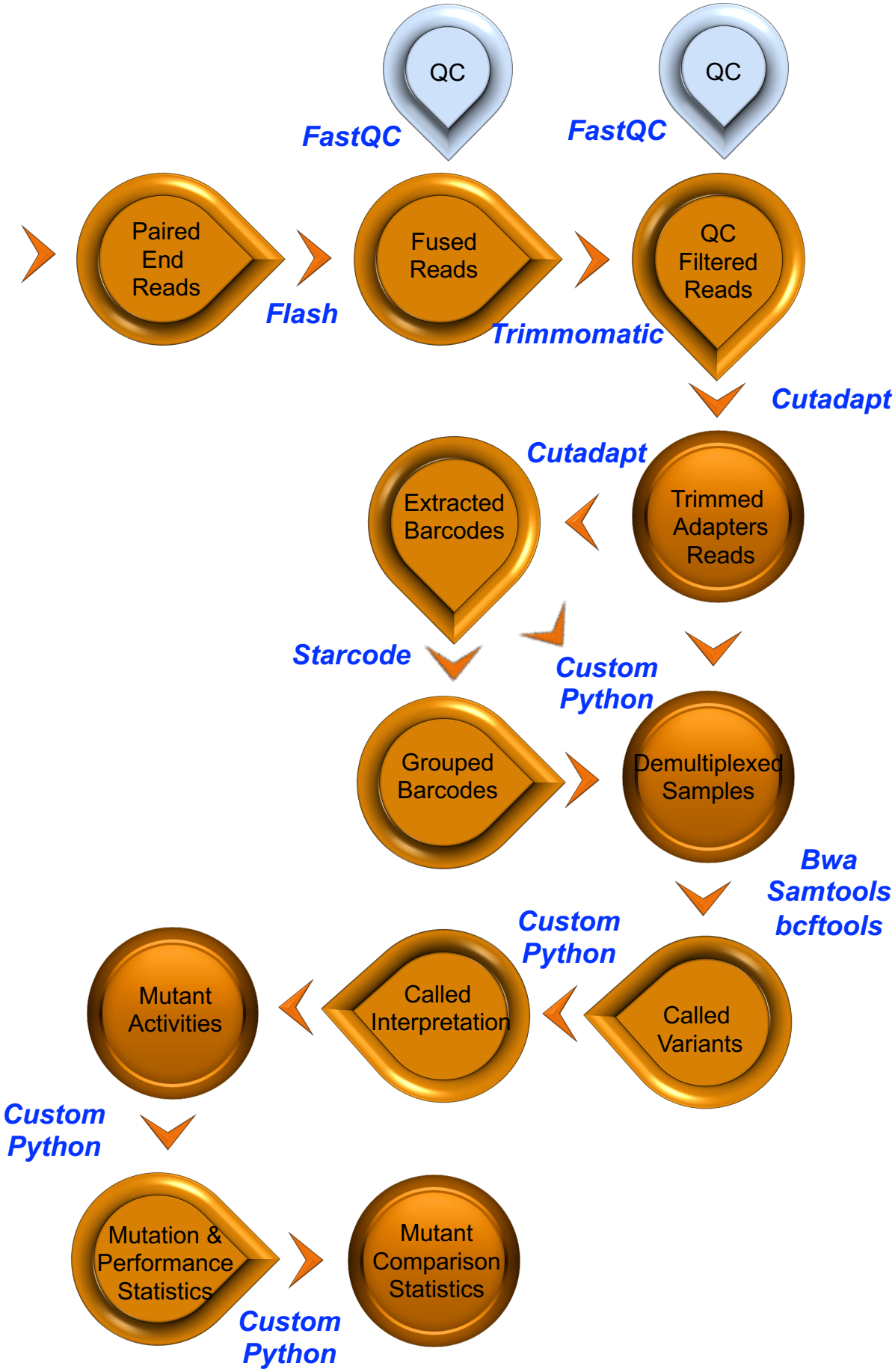
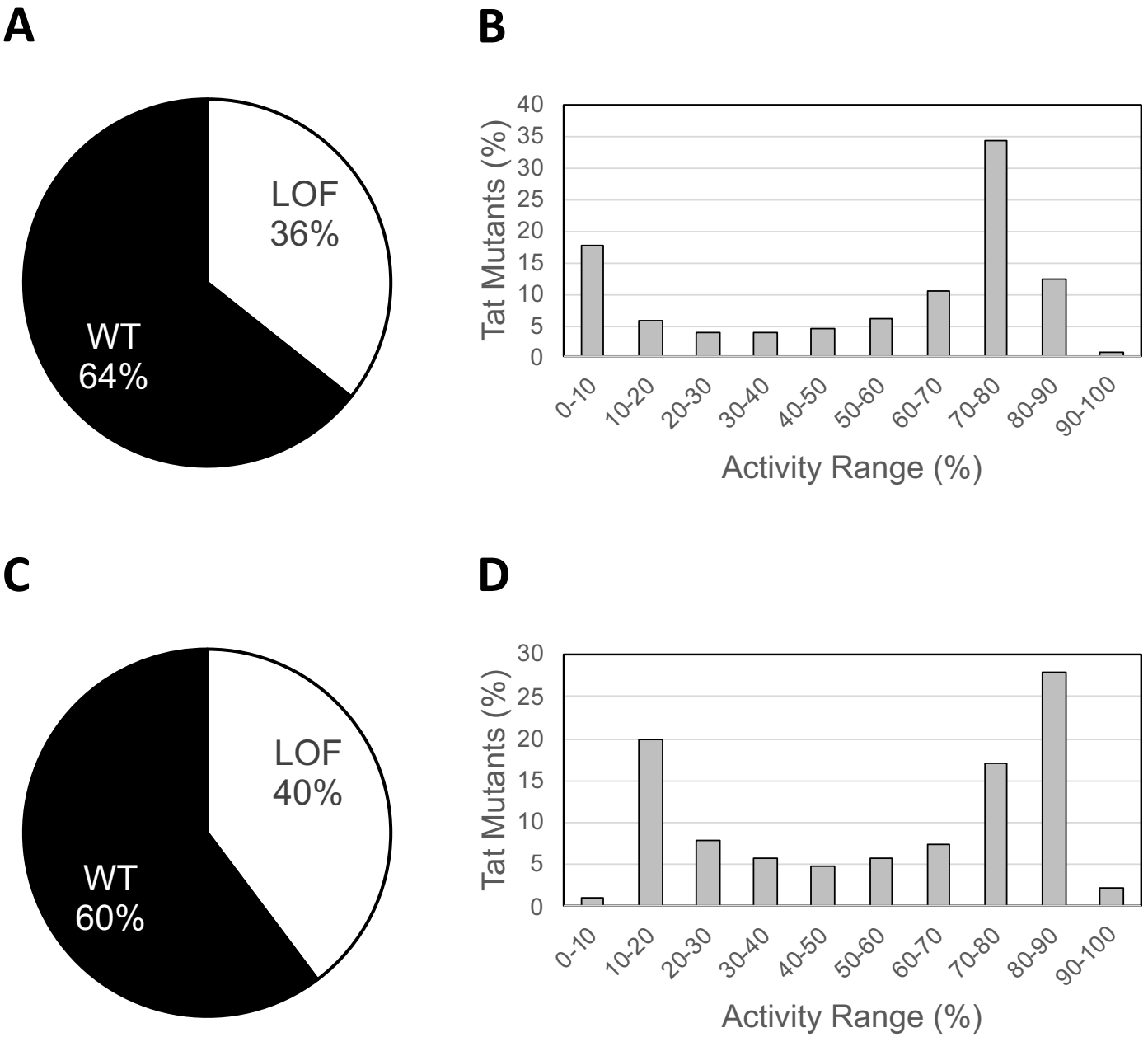


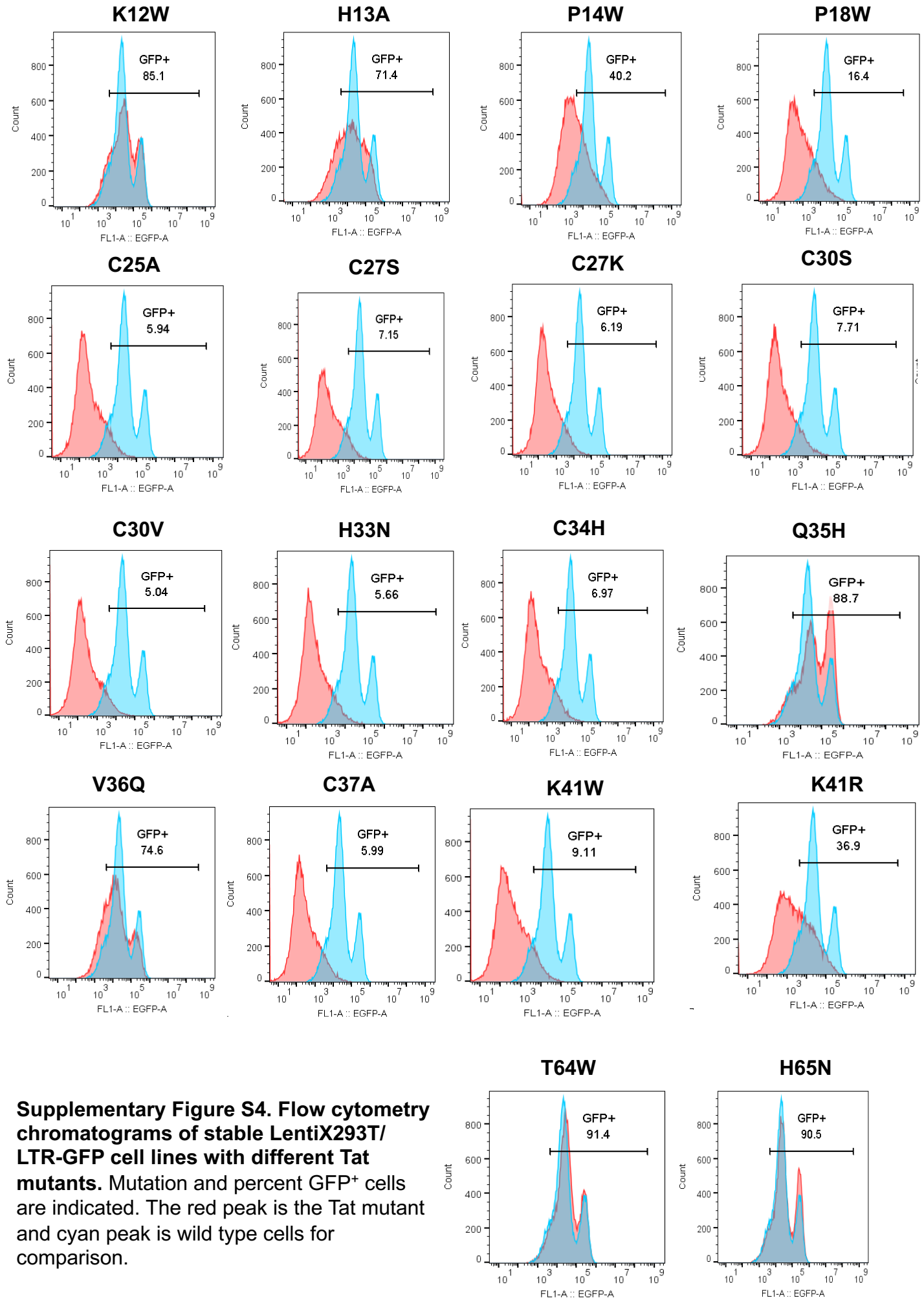
Supplementary Figure S1. Verification of GigaAssay. A-C. Epifluorescence images of controls setting up the GigaAssay in LentiX293T/LTR-GFP cells using lentiviral infection for delivery. The scale bar is 100 µm. **D. E** Flow cytometry optimization for Tat transactivation of LTR-GFP and flow cytometry sorting of the *Tat* variant library in Jurkat/LTR-GFP cells. Keys are shown.



Supplementary Figure S2. Flowchart of GigaAssay bioinformatics pipeline. Programs are indicated in blue italic font.

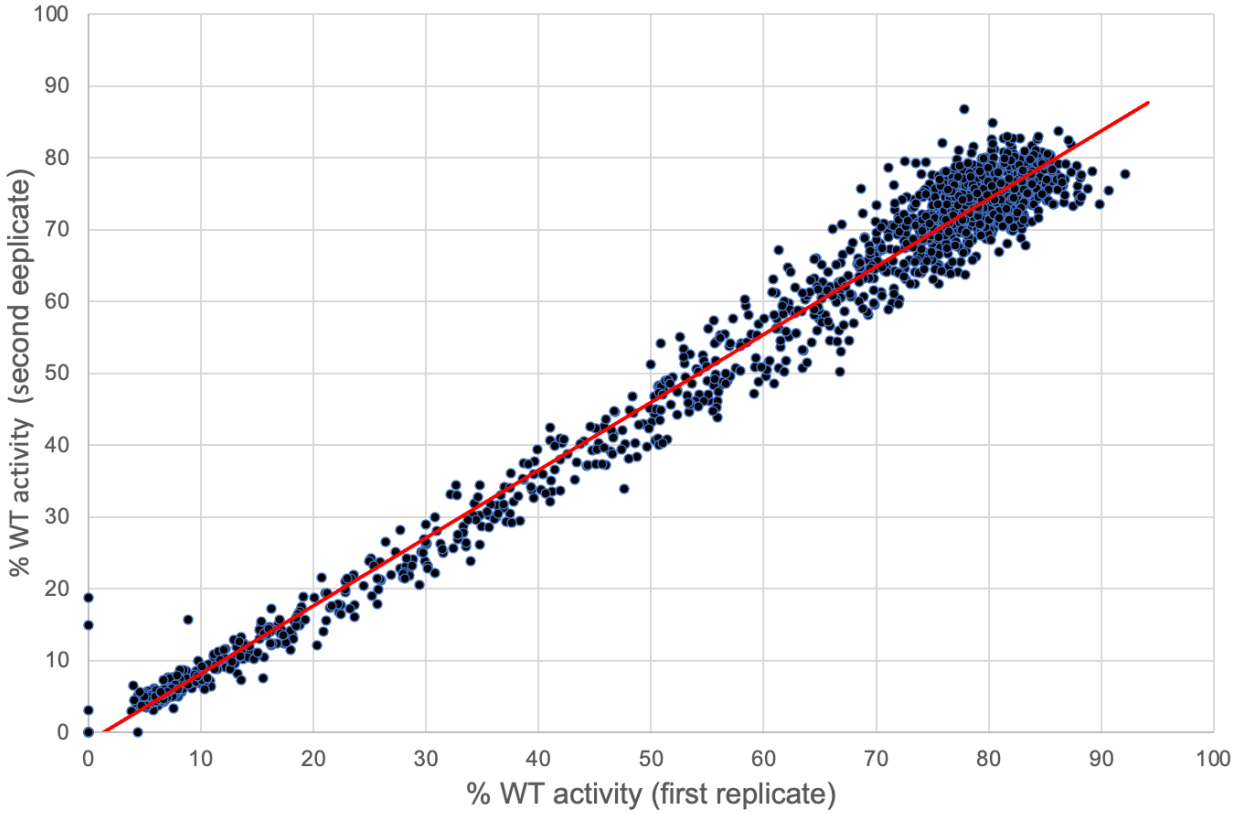


Supplementary Figure S3. Activity summary of Tat mutants. A, B. Tat mutant activity distribution in LentiX293T/LTR-GFP cells (**A, B**) on a pie graph (**A**) and bin plot (**B**). **C, D.** Tat mutant activity distribution in Jurkat/LTR-GFP cells on a pie graph (**C)** and bin plot (**D**).

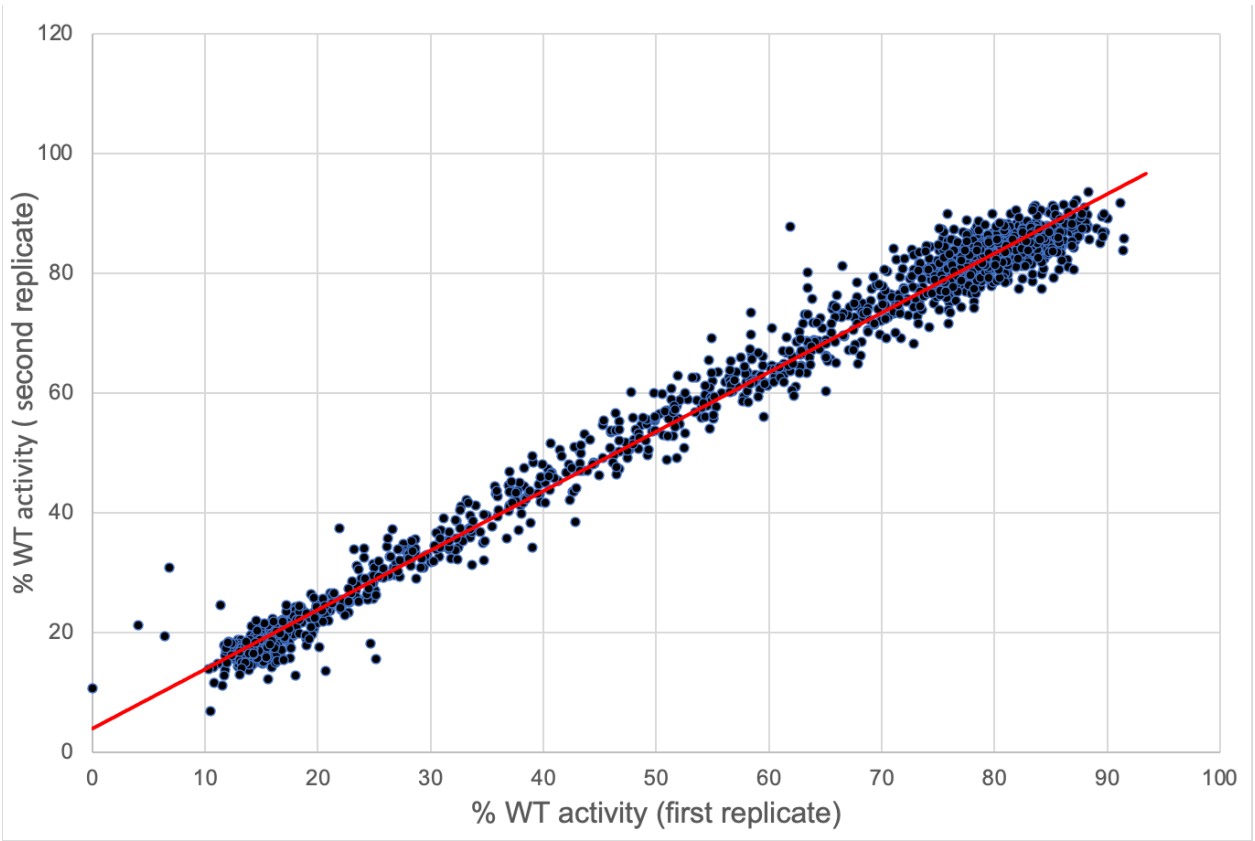


Supplementary Figure S4. Flow cytometry chromatograms of stable LentiX293T/ LTR-GFP cell lines with different Tat mutants. Mutation and percent GFP⁺ cells are indicated. The red peak is the Tat mutant and cyan peak is wild type cells for comparison.

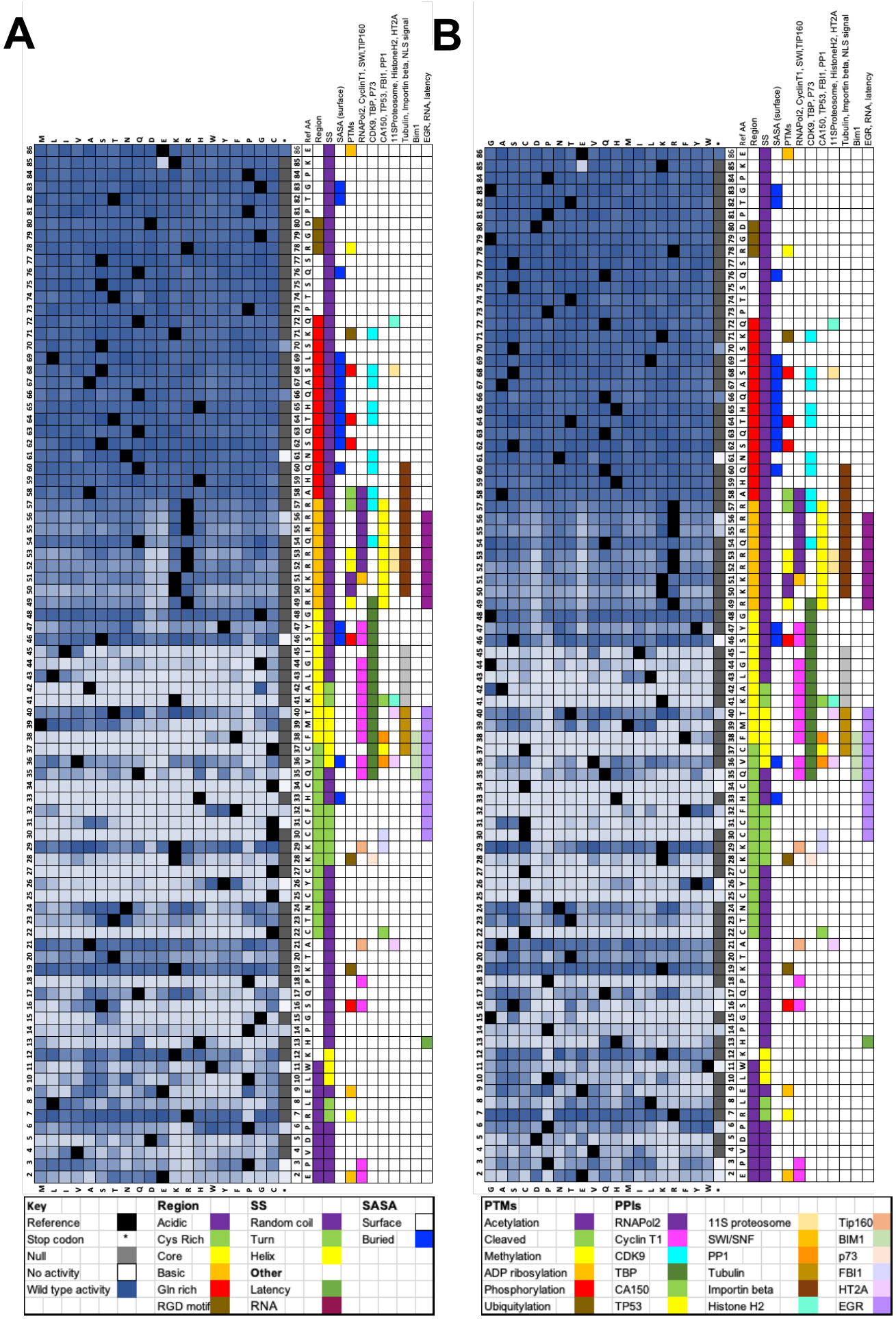
A



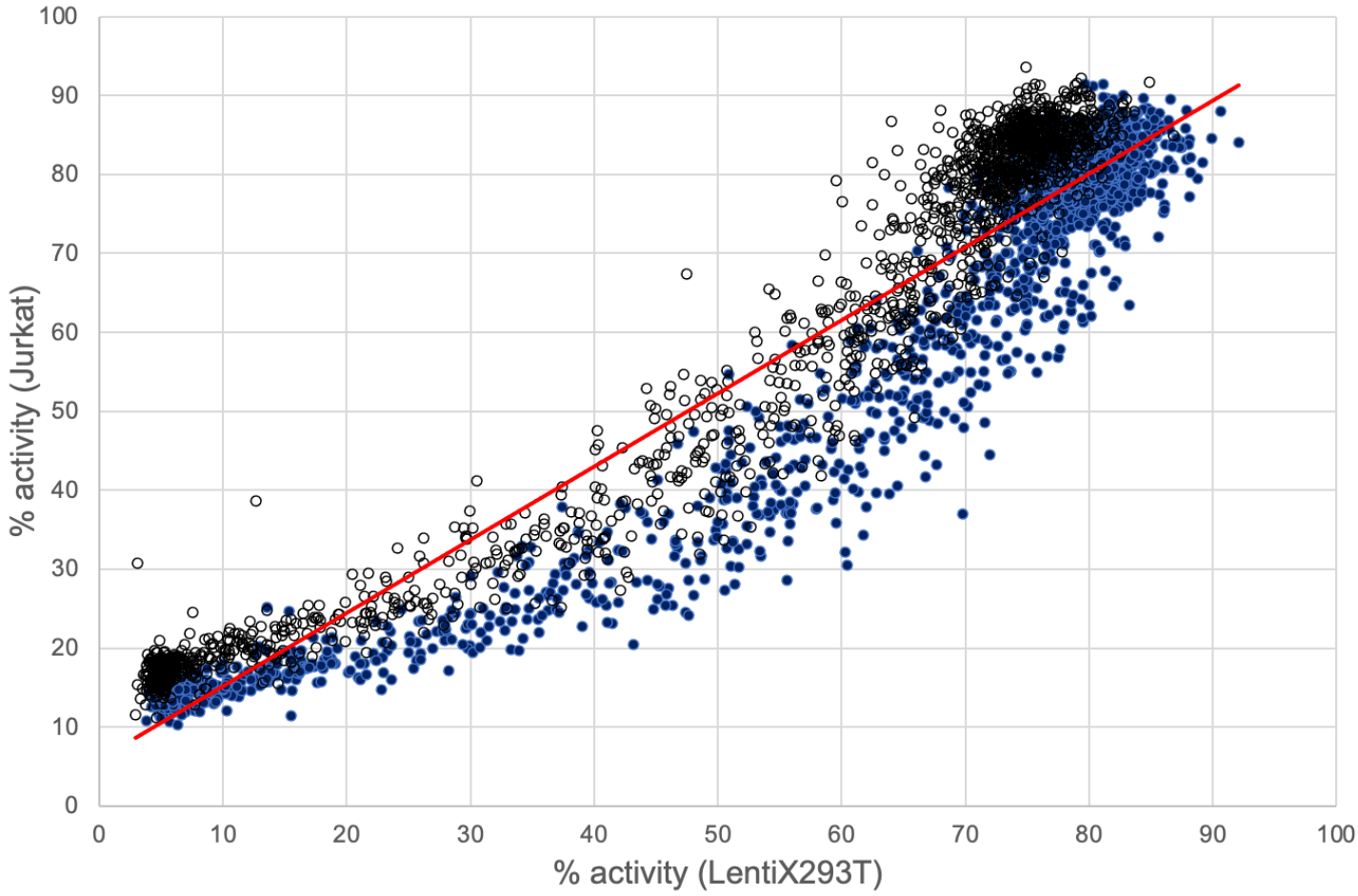
B



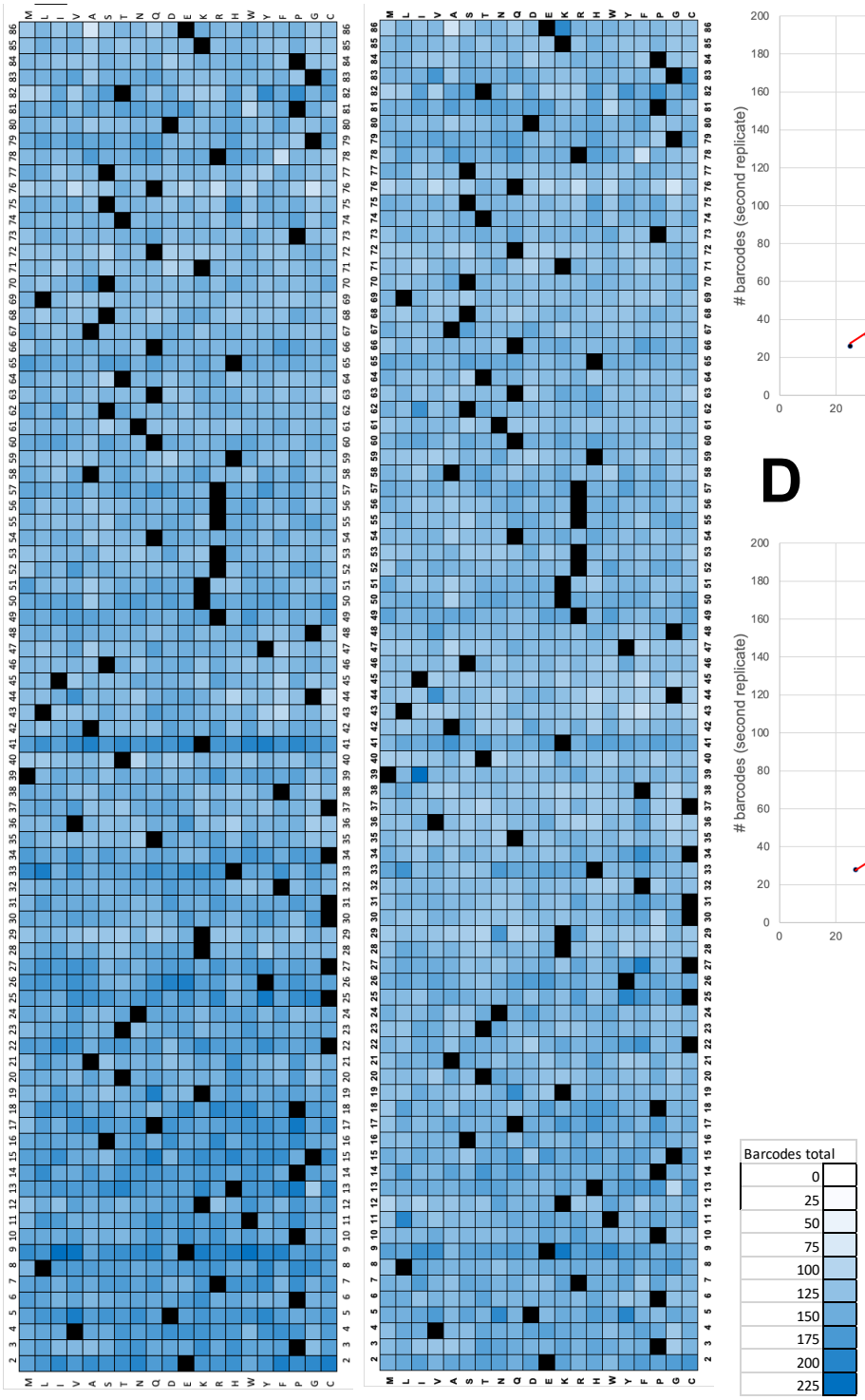
Supplementary Figure S5. Scatter plots for replicates. A,B. Transcriptional activity [$GFP^+/(GFP^- + GFP^+)$] correlation among replicate GigaAssays in LentiX293T/LTR-GFP (**A,** R^2 = 0.99) and Jurkat (**B**, R^2 = 0.99) cells.



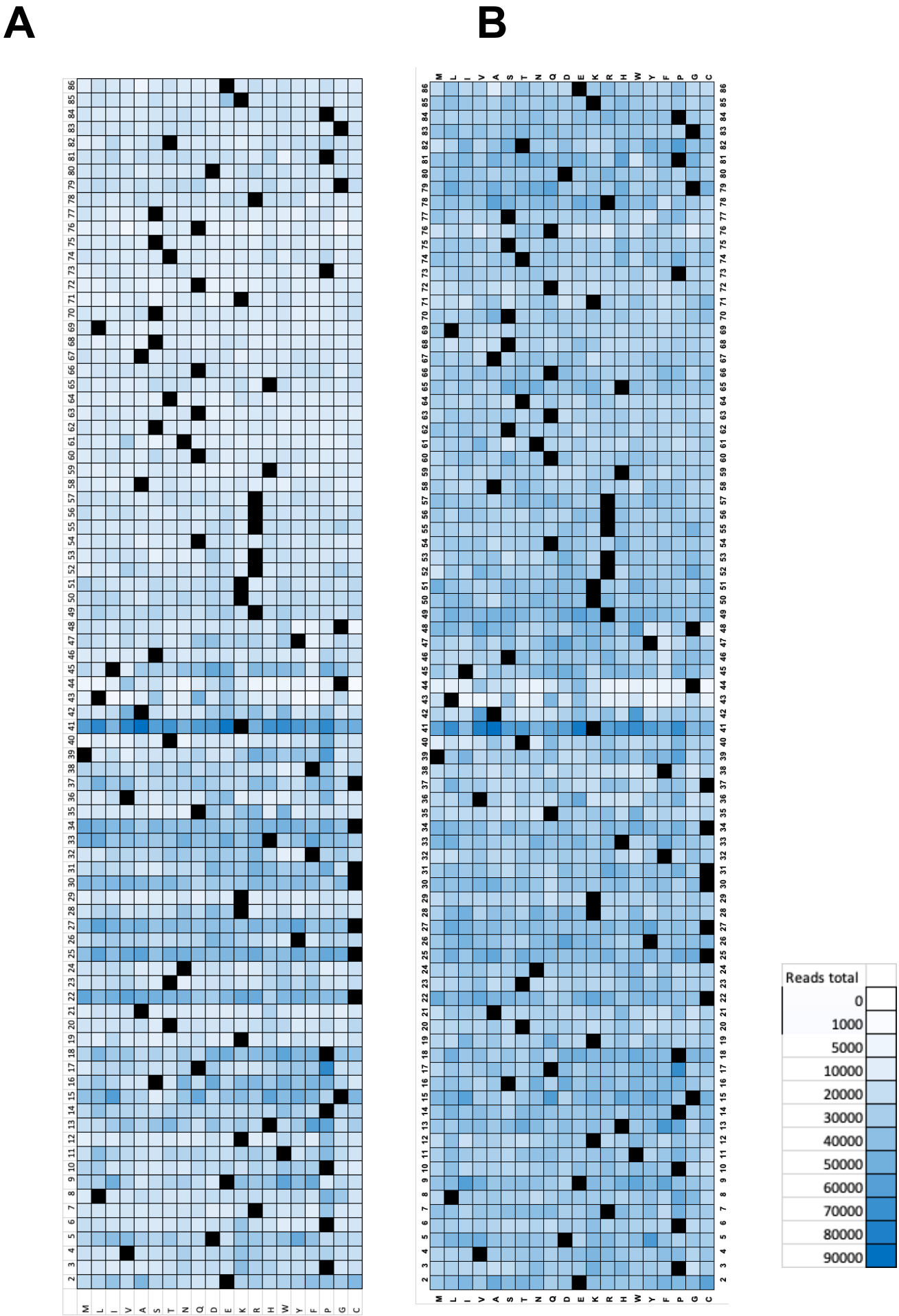
Supplementary Figure S6. Heatmaps of Tat mutant transcriptional activities in Jurkat /LTR-GFP cells. A. Amino acids organized by physiochemical properties **B.** Organized by side chain volume. The key is repeated from **Fig. 2**. Black squares = reference; Gray square = null.



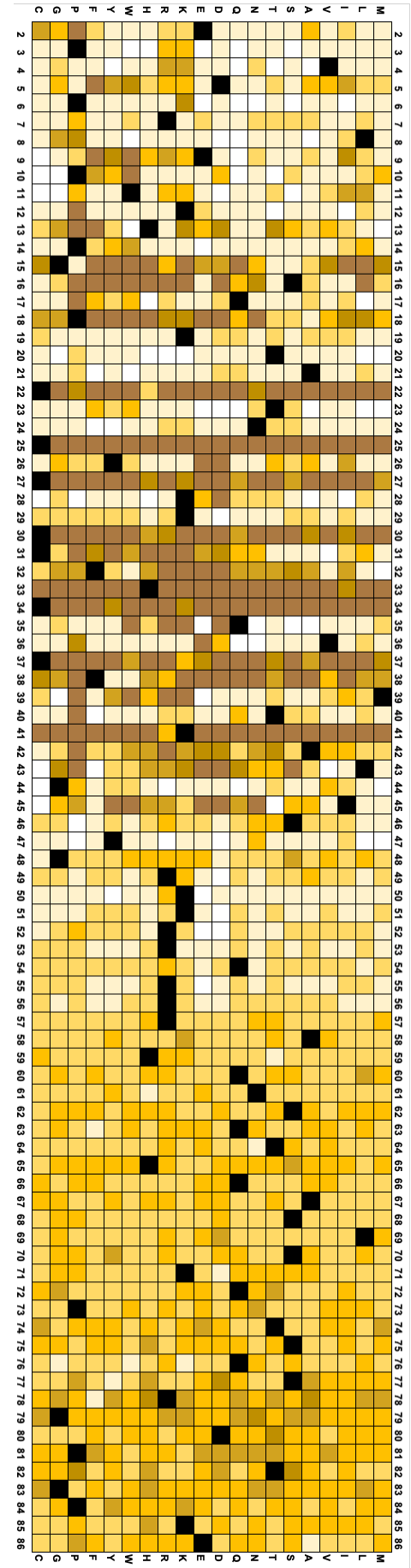
Supplementary Figure S7. Scatter plots comparing transcription activities for LentiX293T/LTR-GFP and Jurkat/LTR-GFP cells. Comparison of activities (percentage of reads for $GFP^+/(GFP^- + GFP^+)$) for matched mutants in LentiX293T/LTR-GFP (open circles) and Jurkat/LTR-GFP cells (blue filled circles); $R^2 = 0.93$.

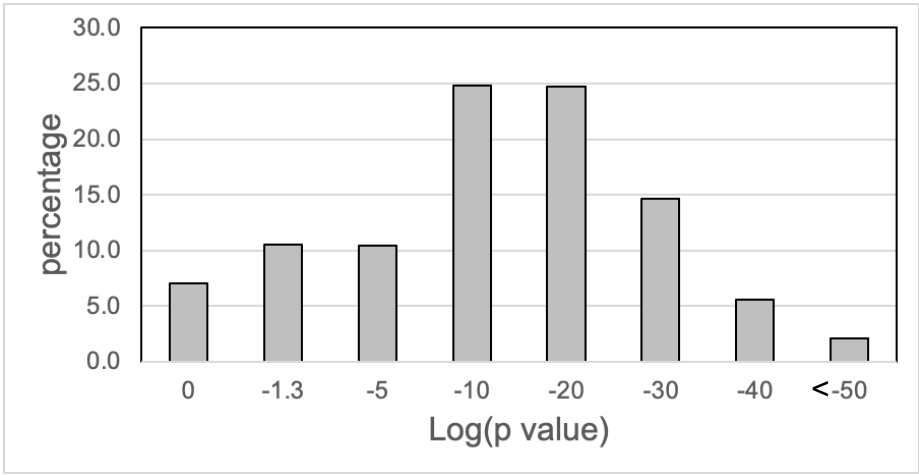


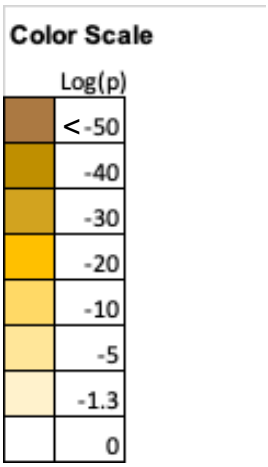
Supplementary Figure S8. Quantitation of GigaAssay barcodes. A, B. Heatmap of barcodes for Tat mutants in LentiX293T/LTR-GFP and Jurkat /LTR-GFP cells respectively. A key for the heatmap colors is shown. Black cells indicate the reference sequence. **C,D.** Barcode correlation for replicate GigaAssay samples. Each point on the scatter plot is for matched mutants. Barcode correlation for replicate GigaAssays in LentiX293T/LTR-GFP (**C**, R^2 = 0.95) and Jurkat/LTR-GFP (**D**, R^2 = 0.96) cells.



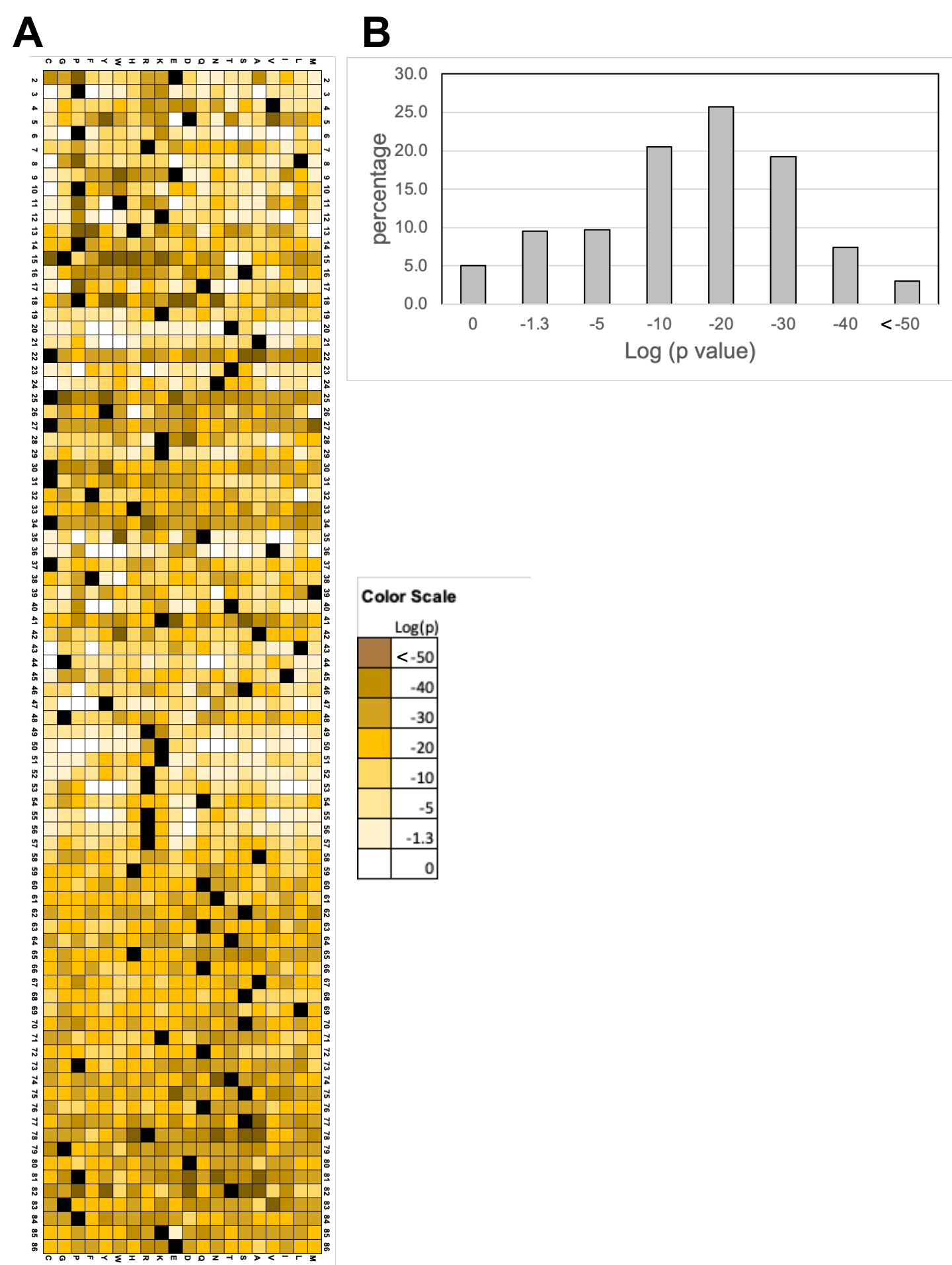
Supplementary Figure S9. Heatmaps of sequence reads. Sequence reads for Tat mutants in LentiX293T/LTR-GFP **(A)** and Jurkat/LTR-GFP **(B)** cells are shown. A key for the heatmap colors is shown. Black squares indicate the reference sequence.



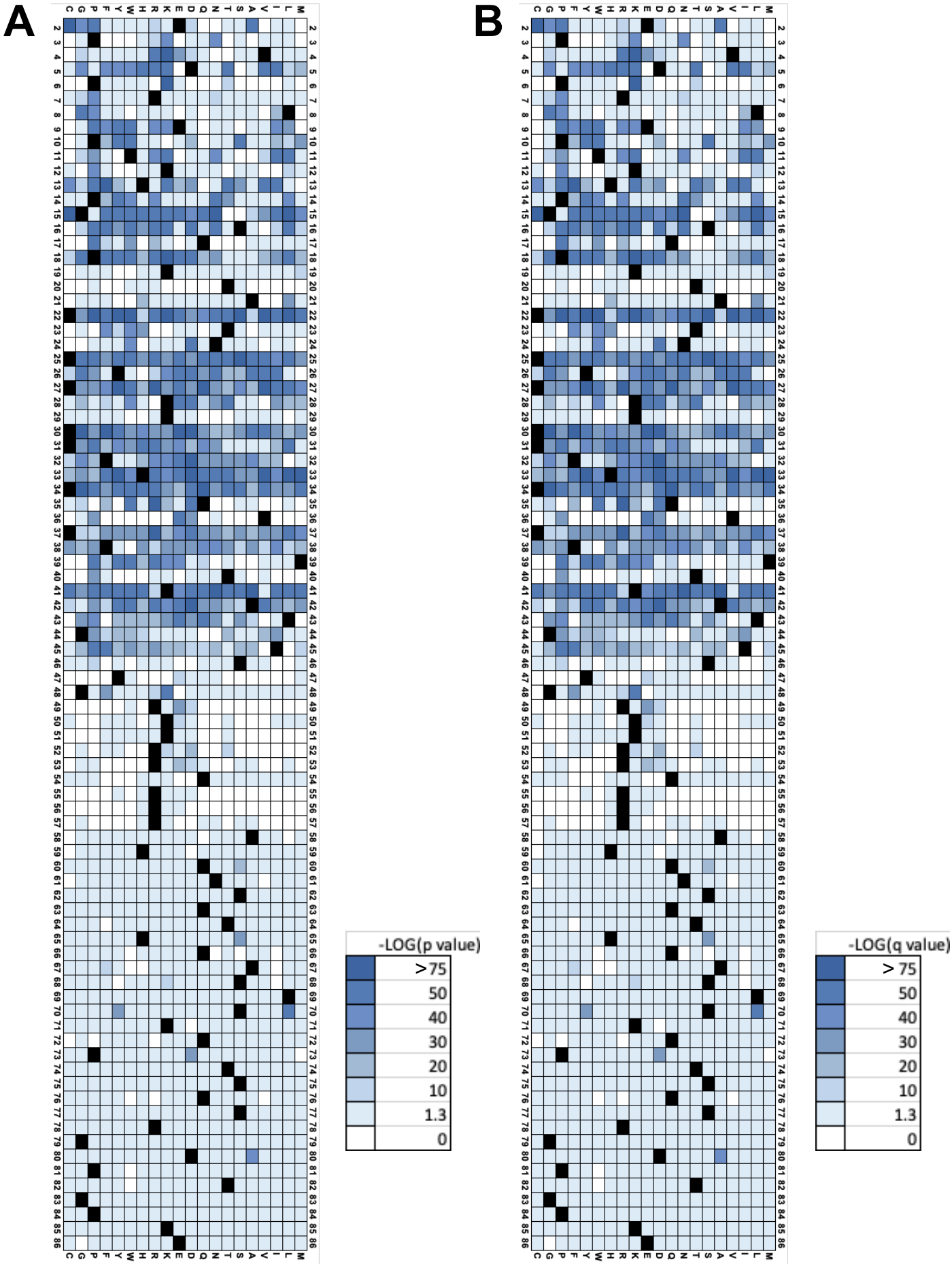




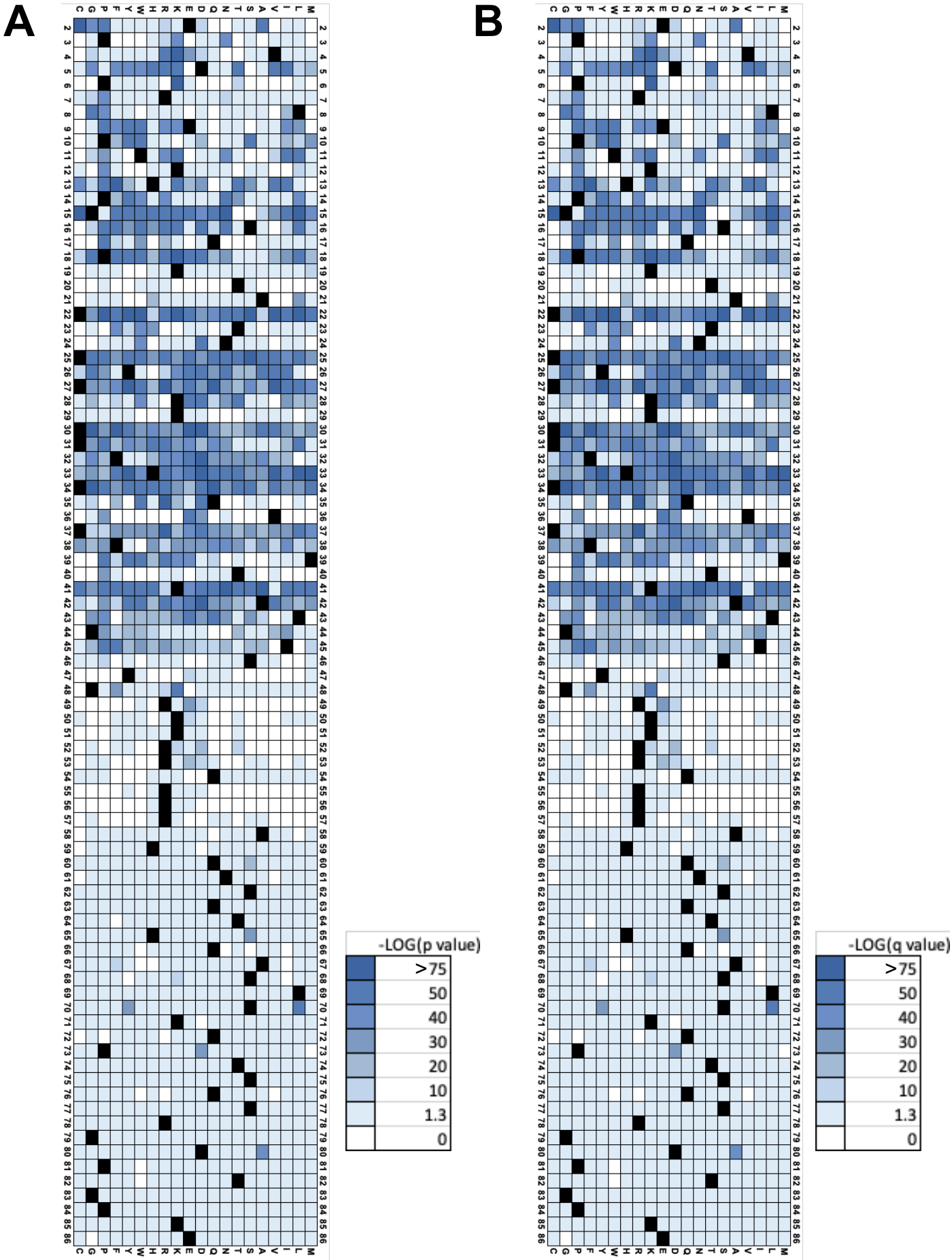
Supplementary Figure S10. Statistical significance of activities of Tat mutants in LentiX293T/LTR-GFP cells. The hypothesis tested is whether the GFP+ ratio observed for that mutant is equal to 0.5. **A.** Heatmap of –Log(p values) for Tat mutants transcriptional activities in LentiX293T/LTR-GFP cells. **B.** Bin plot showing distribution of –Log (p values); (n = 1,615).



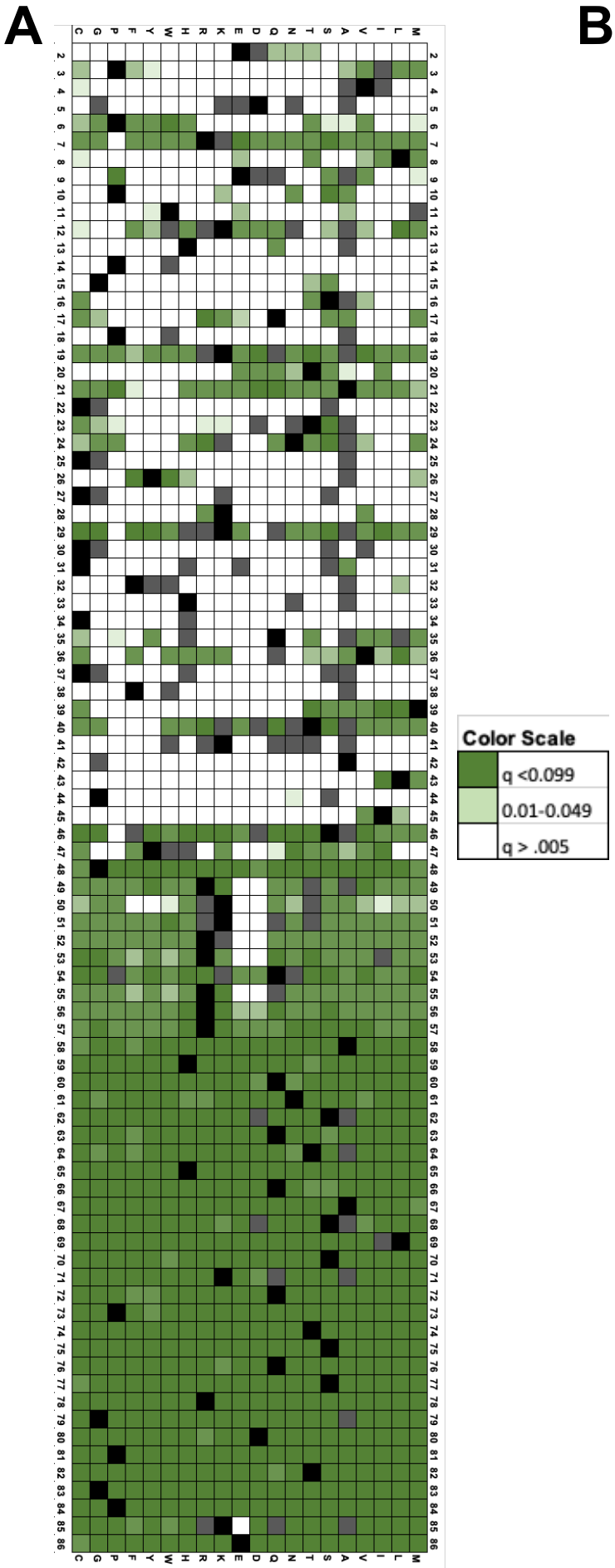
Supplementary Figure S11. Statistical significance of activities of Tat mutants in Jurkat/LTR-GFP cells. The hypothesis tested is whether the GFP+ ratio observed for that mutant is equal to 0.5. **A.** Heatmap of –Log(p values) for Tat mutants transcriptional activities in Jurkat/LTR-GFP cells. **B.** Bin plot showing distribution of –Log (p values); (n = 1,615).

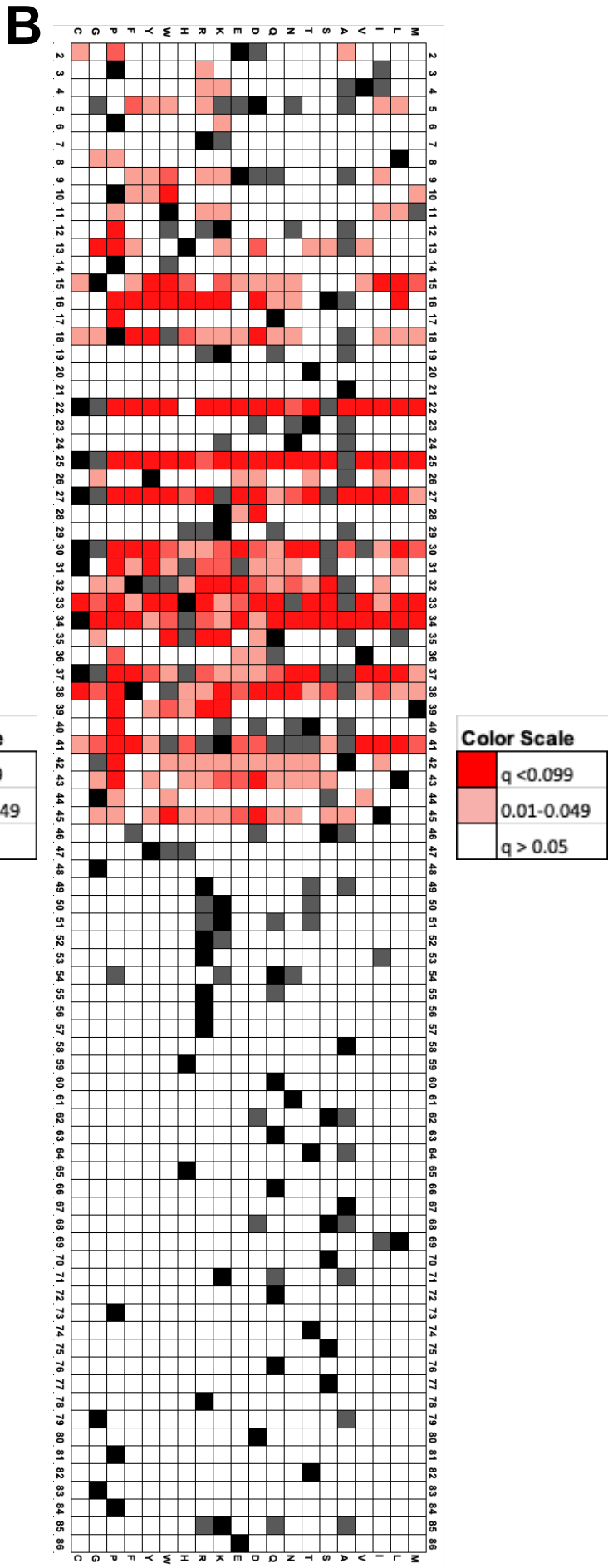


Supplementary Figure S12. Statistical significance of the effect of Tat mutants on Tat activity in LentiX293T/LTR-GFP cells. The hypothesis tested is whether the genotype (Variant/WT) has an effect on the percentage of GFP$^+$ cells. Heatmaps of **A.** –Log(p values) and **B.** –Log(q values) for the LRT test on the significance of the the genotype variable in LentiX293T/LTR-GFP cells.

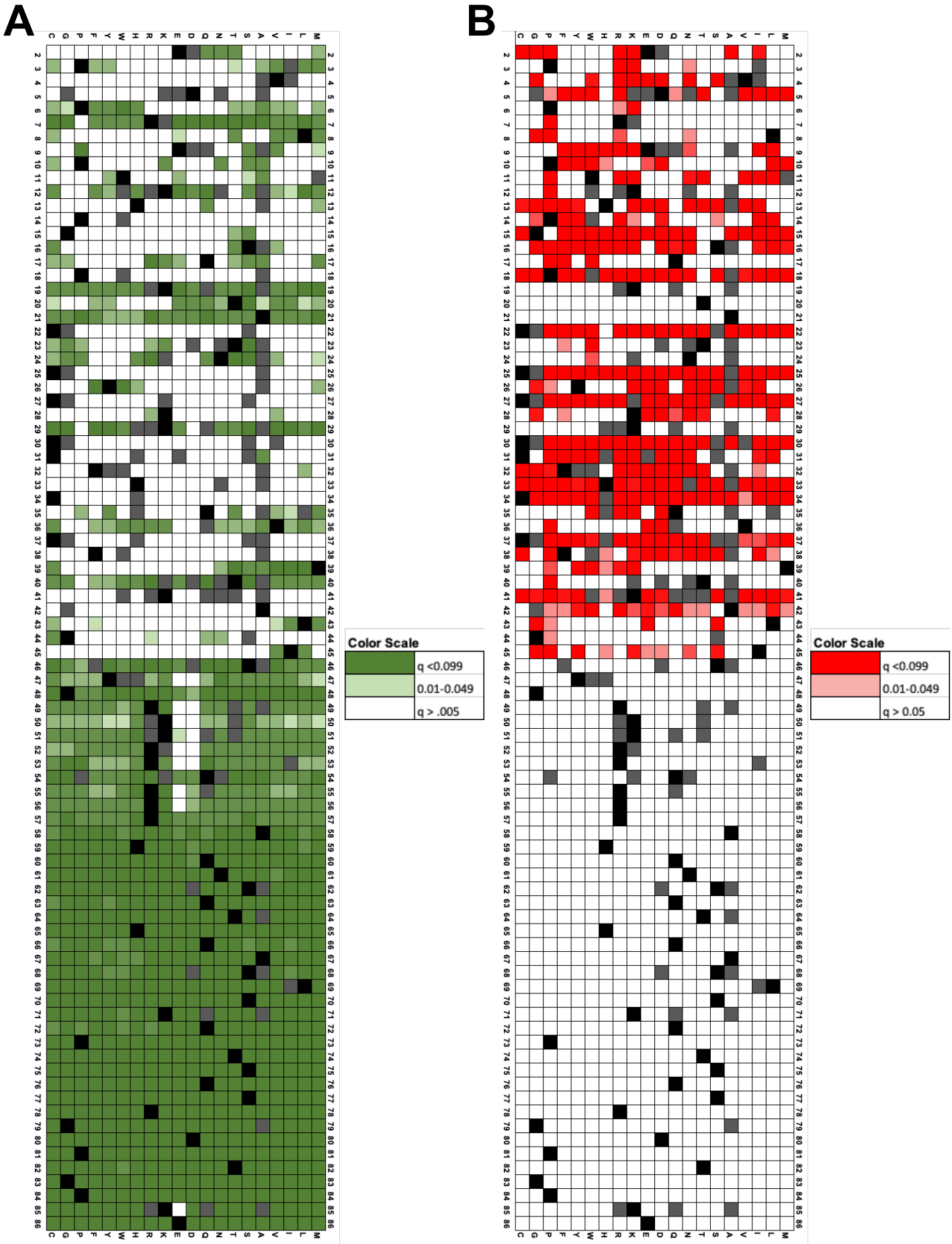


Supplementary Figure S13. Statistical significance of the effect of Tat mutants on Tat activity in Jurkat/LTR-GFP cells. The hypothesis tested is whether the genotype (Variant/WT) has an effect on the percentage of GFP$^+$ cells. Heatmaps of **A.** –Log(p values) and **B.** –Log(q values) for the LRT test on the significance of the the genotype variable in Jurkat/LTR-GFP cells.

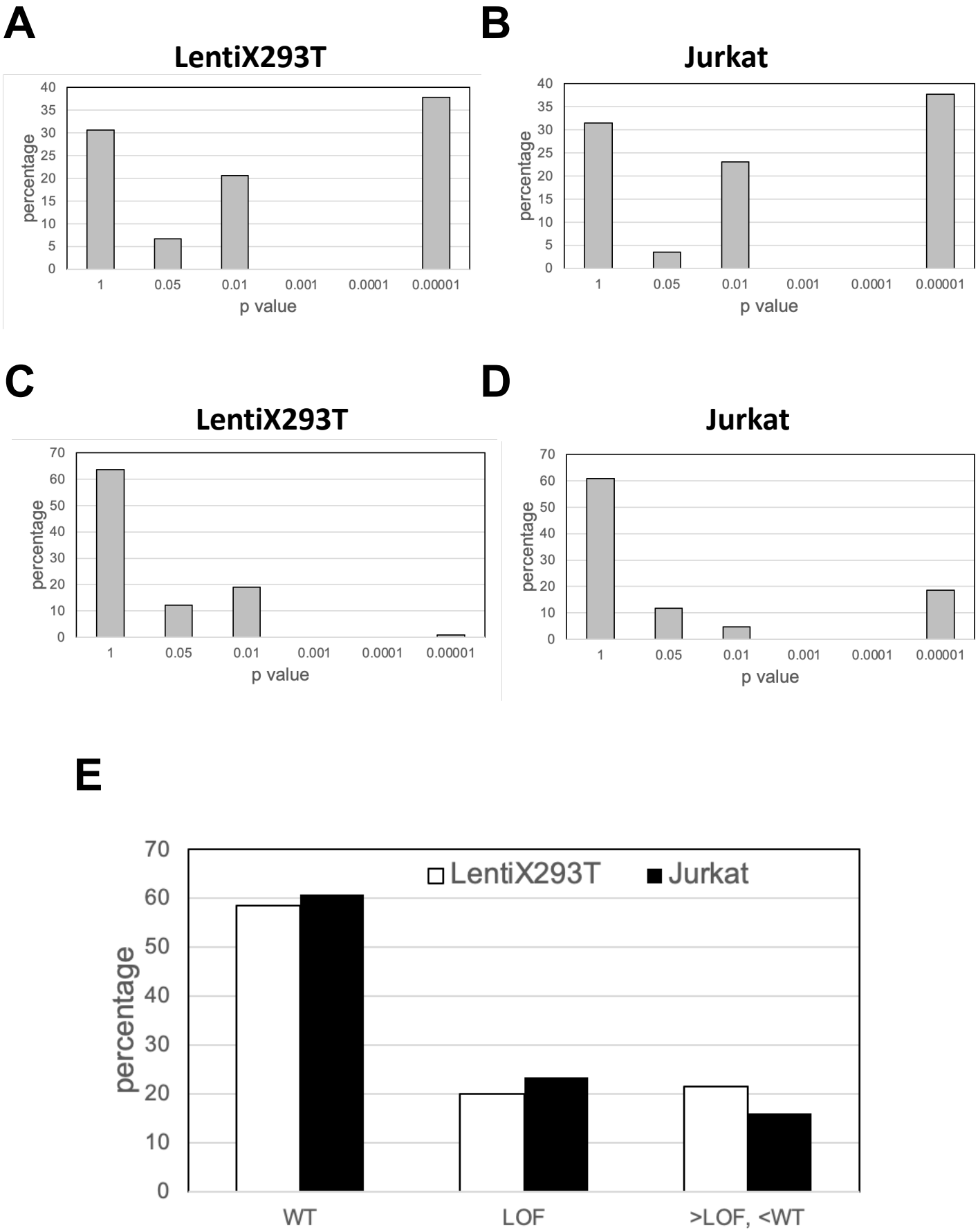




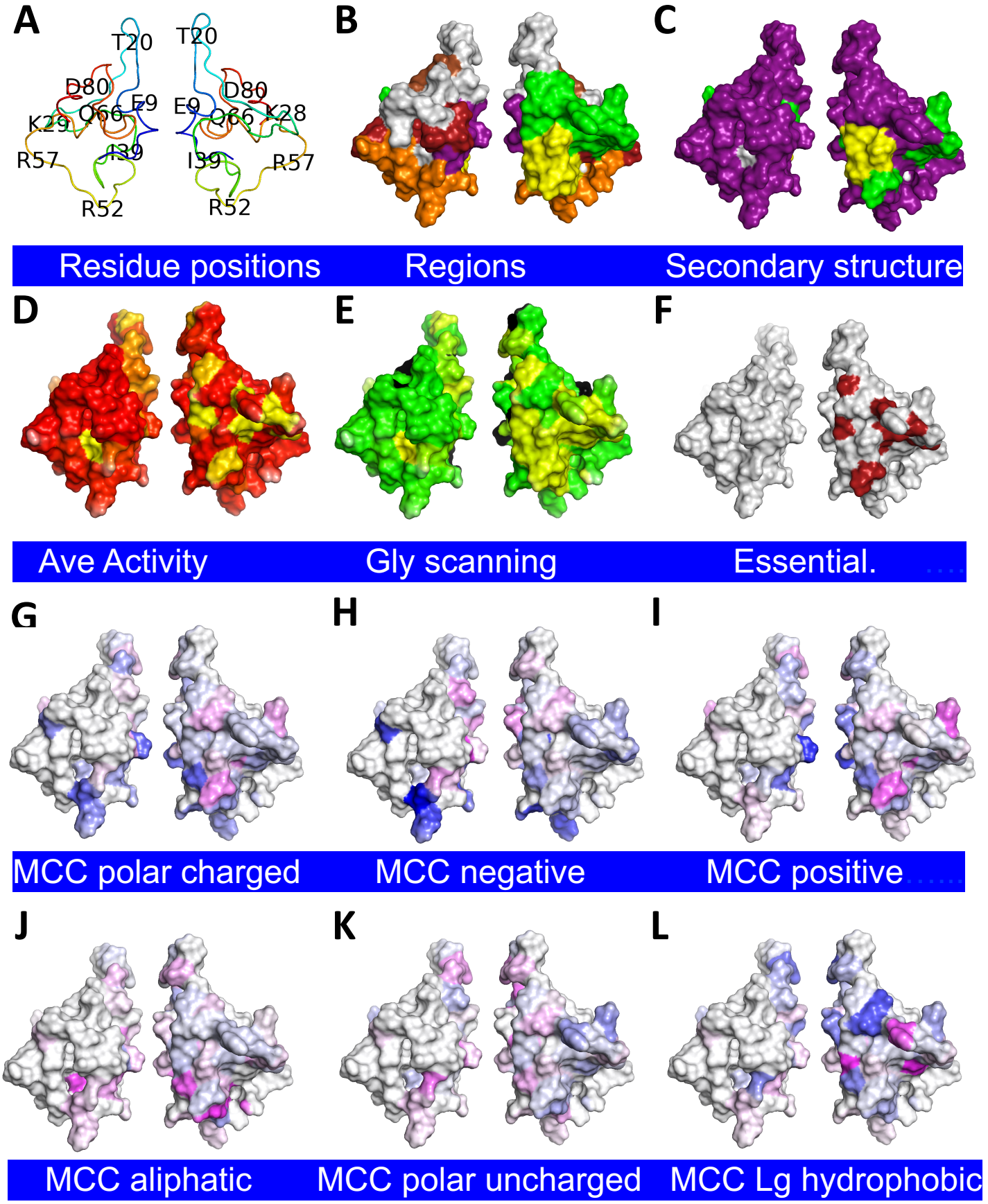
Supplementary Figure S14. Heatmaps of p values Tat mutants transcriptional activities in LentiX293T/LTR-GFP cells. p values for comparison of Tat mutant activity to sets of mutants with wild type (**A**) and LOF activity (**B**). Keys for p value colors are shown.



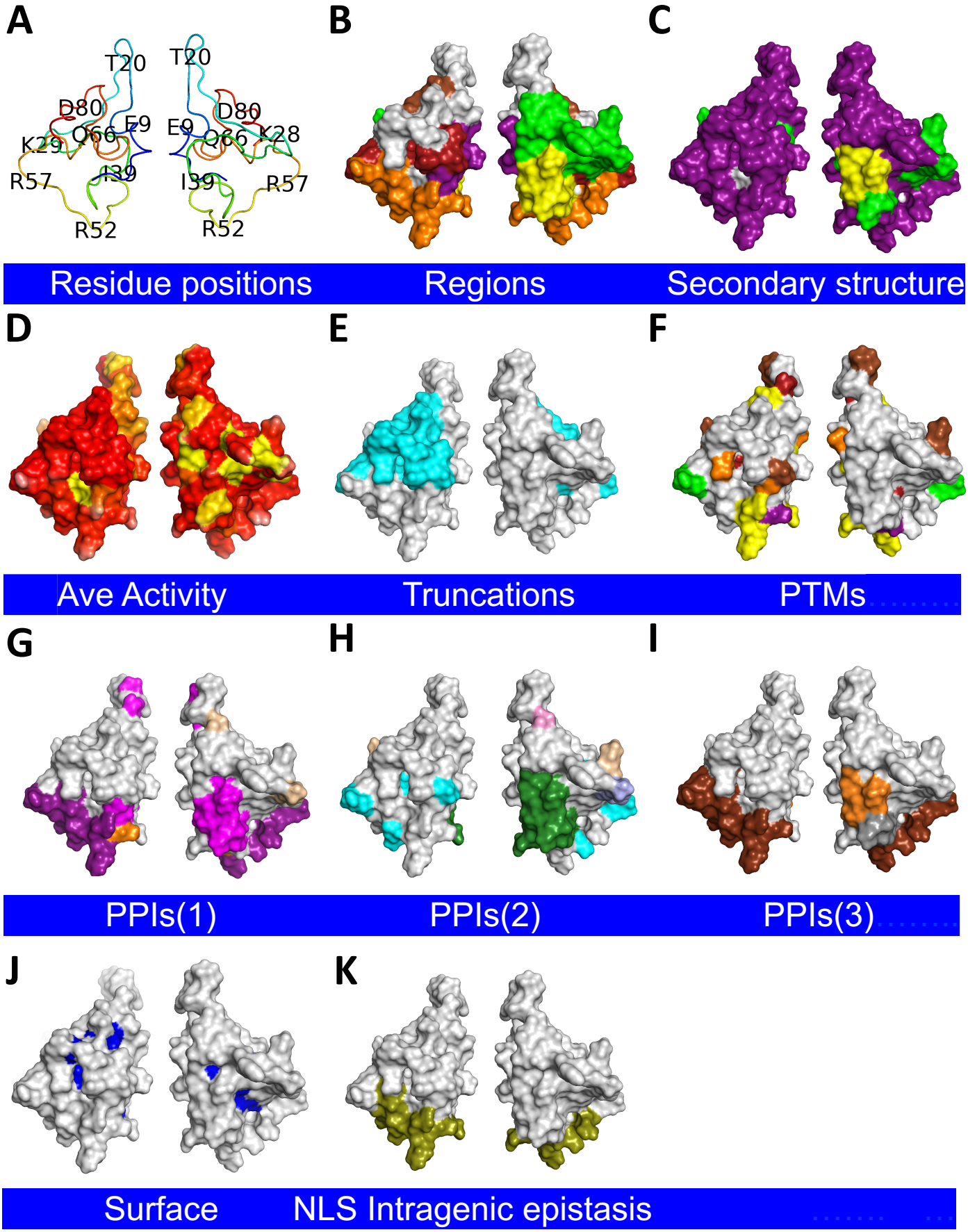
Supplementary Figure S15. Heatmaps of p values Tat mutants transcriptional activities in Jurkat/LTR-GFP cells. p values for comparison of Tat mutant activity to sets mutants with wild type (**A**) and LOF activity (**B**). Keys for p value colors are shown.



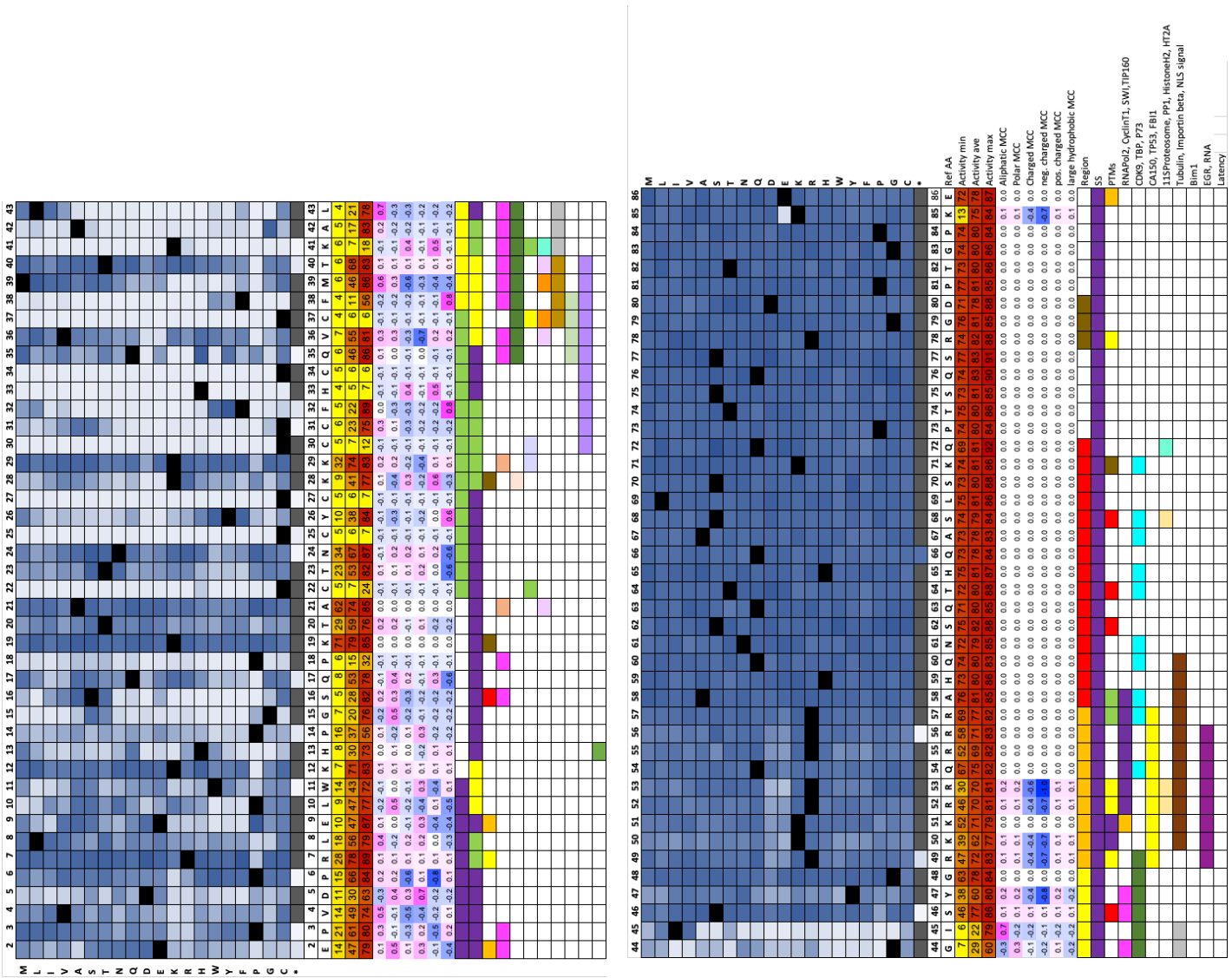
Supplementary Figure S16. Bar charts of p values for Tat mutant transcriptional activities compared to wild type Tat and LOF Tat mutants. p values for comparison of Tat mutant activity to sets of mutants with wild type activity (**A,B**) and LOF activity (**C,D**) for LentiX293T/LTR-GFP (**A,C**) and Jurkat/LTR-GFP (**B,D**) cells. WT and LOF percentages are statistically significance ($p < 0.05$).



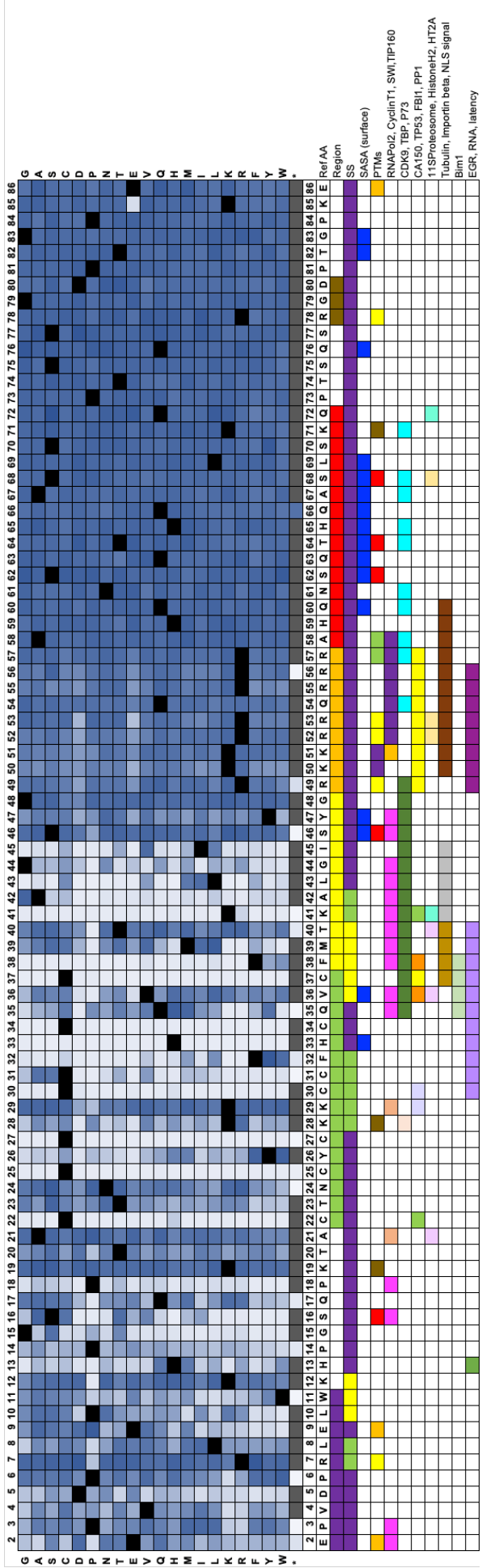
Supplementary Figure S17. 3D surface plots. All surface plots are on wild type Tat 3D structure (PDB: 1TEV): Panels **A-D** and **J-L** are repeated from **Fig. 2** here for visual comparison. **E.** Gly scanning mutagenesis of Tat; black indicates wild type Gly residues). **D-E.** yellow color indicates impaired activity. **F.** Tat positions that do not tolerate any substitution (yellow) **G-I.** Tat MCC surface plots with each position colored with a gradient of blue to white to magenta; Coloring of residues is as described in **Fig 3** and Methods. MCC = Mathews Correlation Coefficient. The color key for regions, secondary structure, PTMs, PPIs, PPVs, and Tat activity are as in **Fig. 2**.



Supplementary Figure S18. 3D structure surface plots of different properties and function of Tat. All surface maps are on wild type Tat 3D structure (PDB: 1TEV): Panels **A-D** are repeated from **Fig. 2** here for visual comparison. **E.** Regions of Tat truncation and missense mutants that lose (light grey) or retain (cyan) activity. **F.** Tat PTMs. **G-I.** Tat PPIs in 3 groups. **J.** Solvent assessable surfaces are with residues with <10% solvent exposure colored blue. **K.** Residues in the NLS that have intragenic epistasis. The color key for regions, secondary structure, surface, and Tat activity are as in **Fig. 2**.



Supplementary Figure S19. Heatmap for LentiX293T/LTR-GFP cells with scores for activities and accuracies for different physiochemical groups. The color key activity heatmap are as in **Fig. 2**. The MCC scores were used to create surface plots with scores ranging from -1 (blue) to 0 (white) to 1 (magenta). White indicates no specificity, Magenta indicates high specificity for the physiochemical group, and blue indicates high specificity for negative preference against the physiochemical group.



Supplementary Figure S20. Heatmap for LentiX293T/LTR-GFP cells with amino acids ordered by side chain volume. The color key is as in **Fig. 2**.