

Expanded View Figures

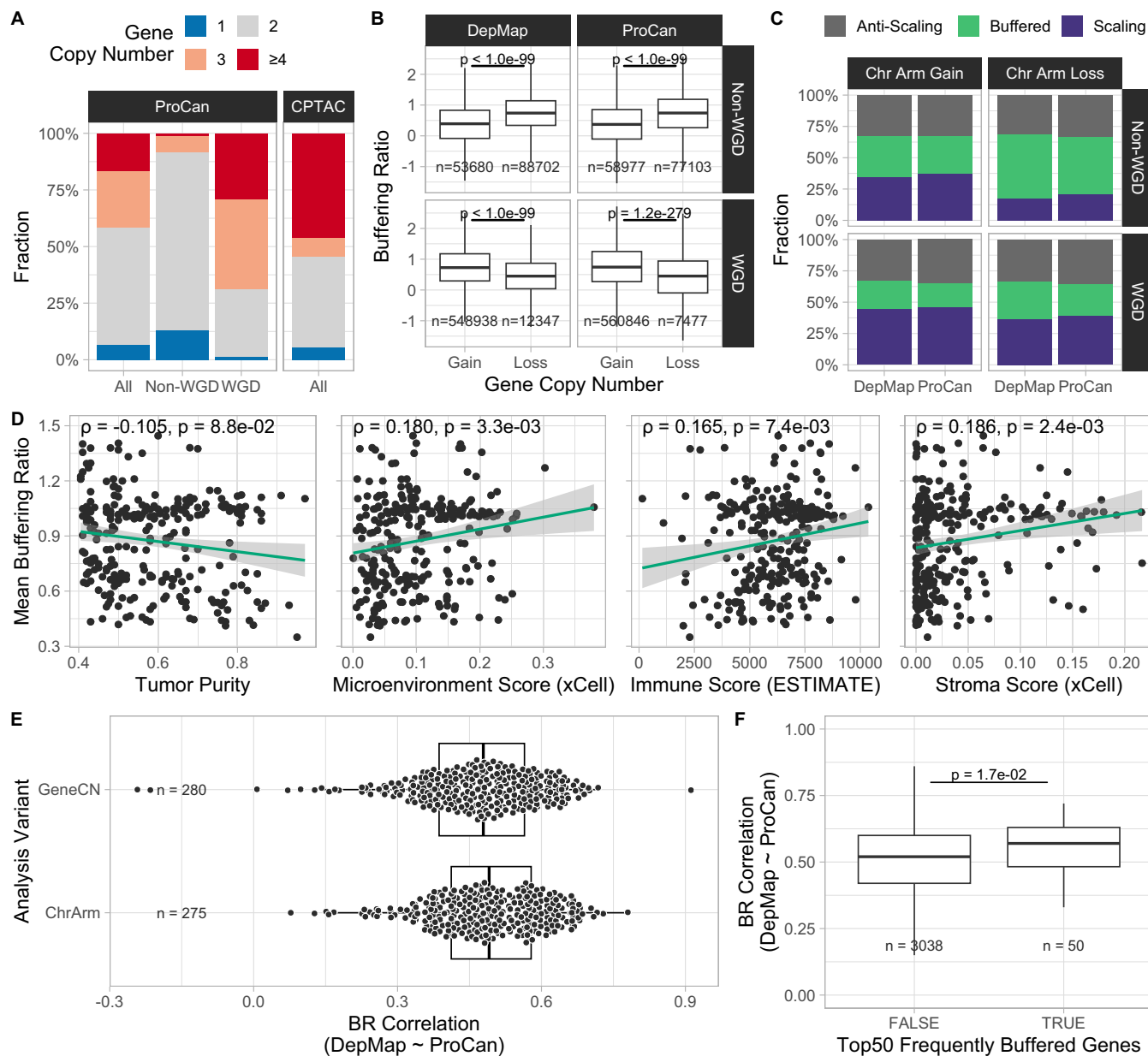


Figure EV1. Controlling for whole-genome doubling status and tumor purity confirms observed protein buffering patterns.

(A) Distribution of gene copy numbers among cell lines (ProCan) and tumor samples (CPTAC) separated by their whole-genome doubling (WGD) status. (B) Difference in buffering ratio distribution of all proteins between gene copy number gain and loss separated by the WGD status. (C) Categorical distribution of buffering classes across pan-cancer cell line datasets upon gene copy number gain and loss separated by WGD status. (D) Correlation between mean buffering ratio per tumor sample (CPTAC) and purity scores obtained from metadata provided by CPTAC (Spearman's ρ). (E) Correlation of buffering ratios per cell line between DepMap and ProCan (Spearman's ρ), separated by buffering ratios created using either gene (GeneCN) or chromosome arm copy numbers (ChrArm). (F) Gene-wise buffering ratio correlation between DepMap and ProCan (GeneCN, Spearman's ρ), separated by their presence in the list of frequently buffered genes with low BR variance. (B, E, F) Boxes represent the interquartile range (IQR) with the central line indicating the median. The whiskers extend to the data points within $1.5 \times \text{IQR}$. P values were determined using a two-tailed Wilcoxon rank-sum (Mann-Whitney U) test.

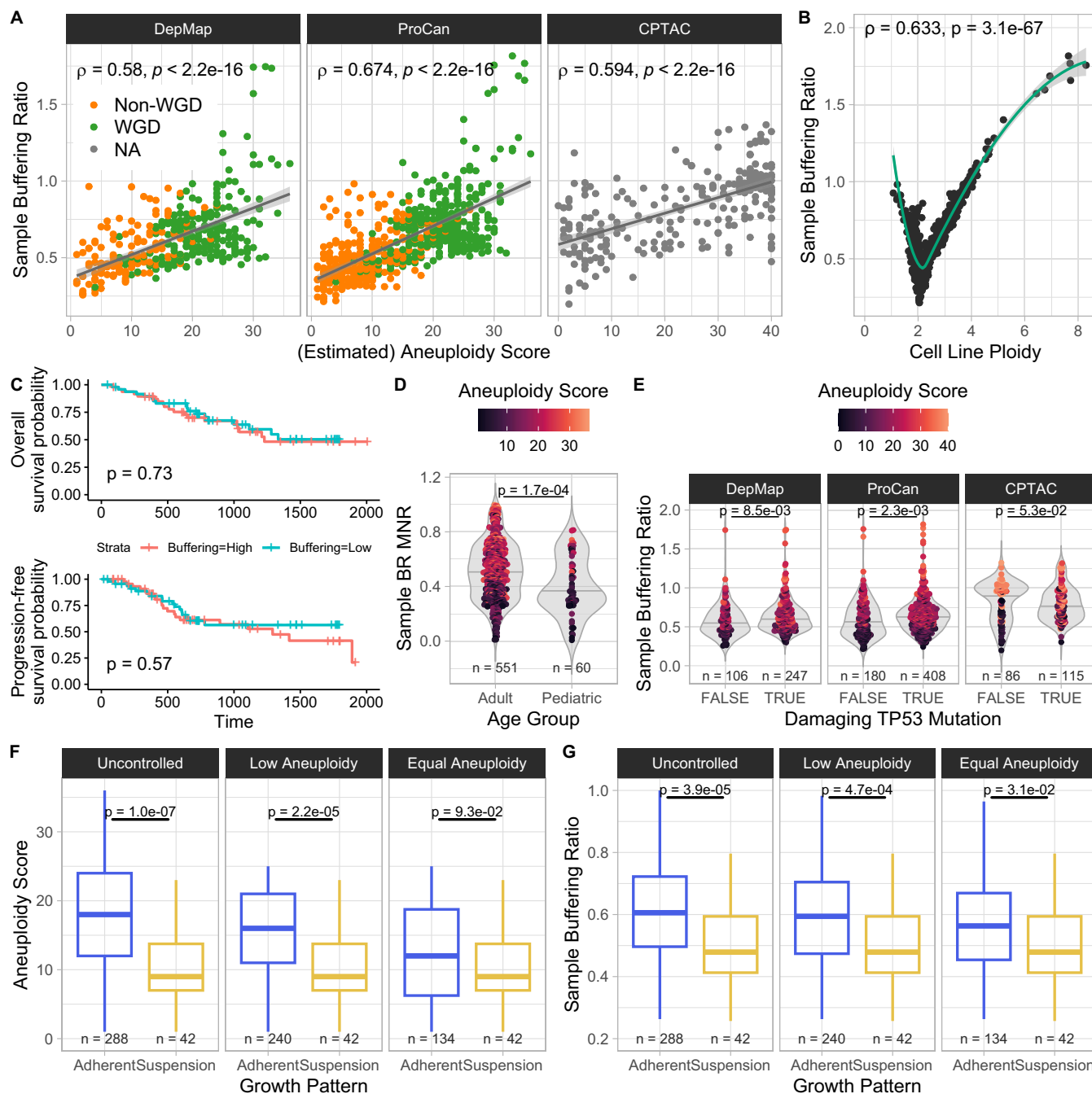


Figure EV2. Average buffering per sample shows relationship with cellular and clinical features of cancer.

(A) Scatter plot showing a positive correlation between sample buffering ratio and aneuploidy score in all pan-cancer datasets (Spearman's ρ). (B) Scatter plot showing a LOESS-smoothed non-linear relationship between sample buffering ratio and cell line ploidy (ProCan; Spearman's ρ). (C) Kaplan-Meier plots showing a decreased progression-free and overall survival probability for tumor samples with high buffering (>80% quantile of sample buffering ratio) compared to low-buffering (<20% quantile, CPTAC). (D) Difference in sample buffering ratio mean normalized ranks between age categories shows decreased sample-wide buffering in pediatric cancers (Wilcoxon rank-sum test). (E) Difference in sample buffering ratio between samples with and without damaging TP53 mutations (Wilcoxon rank-sum test). (F) Difference in aneuploidy between cell lines grown in suspension and adherent cell culture types when applying aneuploidy control methods (DepMap). Controlled for aneuploidy by removing cell lines from the adherent cohort with aneuploidy scores above the maximum of the suspension cohort (middle), and by using stratified sampling to ensure an equal distribution of aneuploidy scores (right). (G) Difference in sample buffering ratio between cell lines grown in suspension and adherent cell culture types, controlled for aneuploidy (DepMap). (F, G) Boxes represent the interquartile range (IQR) with the central line indicating the median. The whiskers extend to the data points within 1.5×IQR. (D–G) P values were determined using a two-tailed Wilcoxon rank-sum (Mann-Whitney U) test.

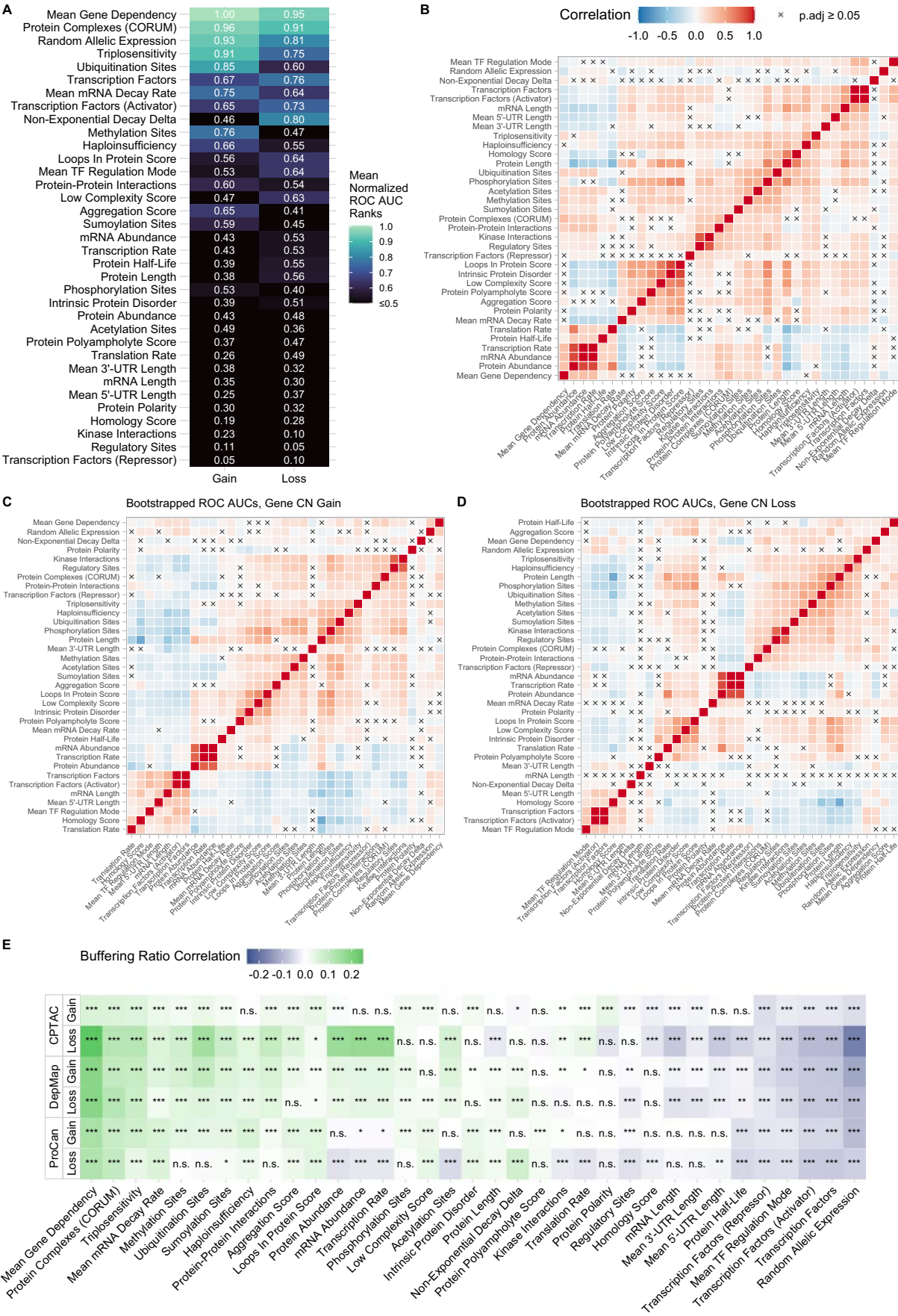


**Figure EV3. Correlation patterns of potential dosage compensation factors.**

(A) Mean normalized ranks (MNRs) of ROC AUCs of each factor used for classifying whether a gene is *Buffered* on protein level or *Scaling*, while excluding the *Anti-Scaling* class. MNRs were calculated across all analysis variants (GeneCN, ChrArm, ChrArmAvg) and pan-cancer datasets (DepMap, ProCan, CPTAC), grouped by gene or chromosome arm copy number gain and loss. WGD-controlled subsets were excluded. (B-D) Clustered heatmaps showing the correlation (Spearman's ρ) between factors. Correlations with insignificant Benjamini-Hochberg adjusted P values were crossed out ($p_{adj} \geq 0.05$). Correlations were calculated using factor values (B), bootstrapped ROC AUCs for predicting protein buffering upon gene copy number gain (C), and bootstrapped ROC AUCs upon gene copy number loss (D). (E) Spearman correlation between factors values and buffering ratios upon gene copy number loss and gain across all pan-cancer datasets (DepMap, ProCan, CPTAC).

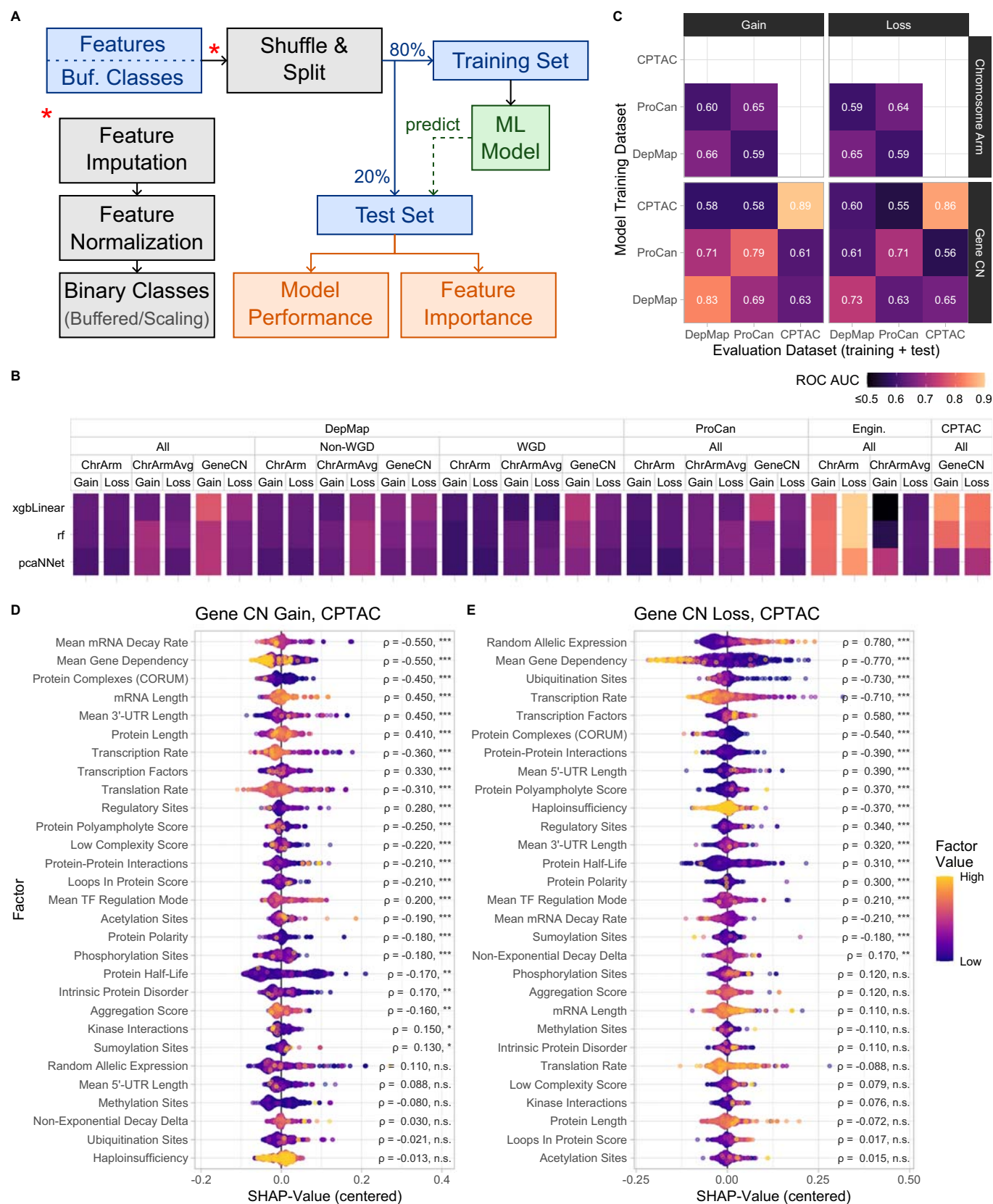


**Figure EV4. Multifactorial models improve prediction performance of protein buffering.**

(A) Illustration of training and evaluation procedures of multifactorial models. (B) Model performance (ROC AUC) of trained models, separated by model architecture (xgbLinear, rf, pcaNNet), dataset (DepMap, ProCan, CPTAC, Engineered), subset (All, WGD, Non-WGD), analysis variant (GeneCN, ChrArm, ChrArmAvg), and copy number event (Gain, Loss). (C) Out-of-sample performance of multifactorial models trained on buffering classes derived from gene and chromosome arm copy number gain and loss data from different pan-cancer datasets (DepMap, ProCan, CPTAC). Model performance was evaluated using the ROC AUC of model predictions on merged training and test datasets. (D, E) SHAP values generated by evaluating the trained CPTAC gene copy number models on a random subset of the respective test set ($n = 300$). Negative SHAP values depict a higher contribution of a model's factor towards predicting an observation as *Buffered* relative to the baseline prediction. Each dot represents a gene in a sample. Color represents the min-max-scaled value of the factor used for prediction. Spearman's ρ indicates the trend between SHAP value and feature value. SHAP values were mean-centered within each factor.



◀ **Figure EV5. Enrichment patterns suggest oncogenic effects of protein buffering.**

(A) Volcano plots showing log2 fold change between high (>80% quantile of sample buffering ratios) and low (<20% quantile) buffering cell lines and tumor samples in DepMap ($n = 71$ per group), CPTAC ($n = 52$ per group), and adherent control datasets ($n = 82$ high buffering, $n = 51$ low buffering; Student's t test, Benjamini-Hochberg adjusted P values, significance thresholds: $|\text{Log2FC}| > 0.5$, $P_{\text{adj}} < 0.05$). (B, C) Overrepresentation analysis (ORA) of genes that are significantly down- (B) and upregulated (C) in highly buffering cell lines with adherent growth patterns (ProCan, adherent control; hypergeometric test using g:Profiler). (D) Correlation between tumor purity and single-sample enrichment scores for the *Unfolded Protein Response* HALLMARK gene set using CPTAC tumor sample proteome data (Spearman's ρ). (E) Buffering ratio distribution of oncogenes (OG) significantly upregulated in high buffering cell lines and of downregulated tumor suppressor genes (TSG) upon gene copy number gain and loss (ProCan, all cell lines). Boxes represent the interquartile range (IQR) with the central line indicating the median. The whiskers extend to the data points within $1.5 \times \text{IQR}$. (F) Buffering ratio distribution of selected oncogenes upon gene copy number gain and loss (ProCan). (G) Categorical distribution of buffering classes for EGFR upon gain and loss of chromosome arm 7p (ProCan). (H) Normalized enrichment scores of MSigDB HALLMARK gene sets for pan-cancer datasets (DepMap, ProCan, CPTAC). Enrichment scores have been generated by comparing high against low-buffering samples and comparing high against low-aneuploid samples (>80% and <20% quantiles of sample buffering ratio and aneuploidy score estimates). Significance levels of Benjamini-Hochberg adjusted P values generated using fGSEA: $*P_{\text{adj}} < 0.01$, $**P_{\text{adj}} < 0.001$, $***P_{\text{adj}} < 0.0001$. Removed gene sets with no significant enrichment scores. (E, F) P values were determined using a two-tailed Wilcoxon rank-sum (Mann-Whitney U) test.

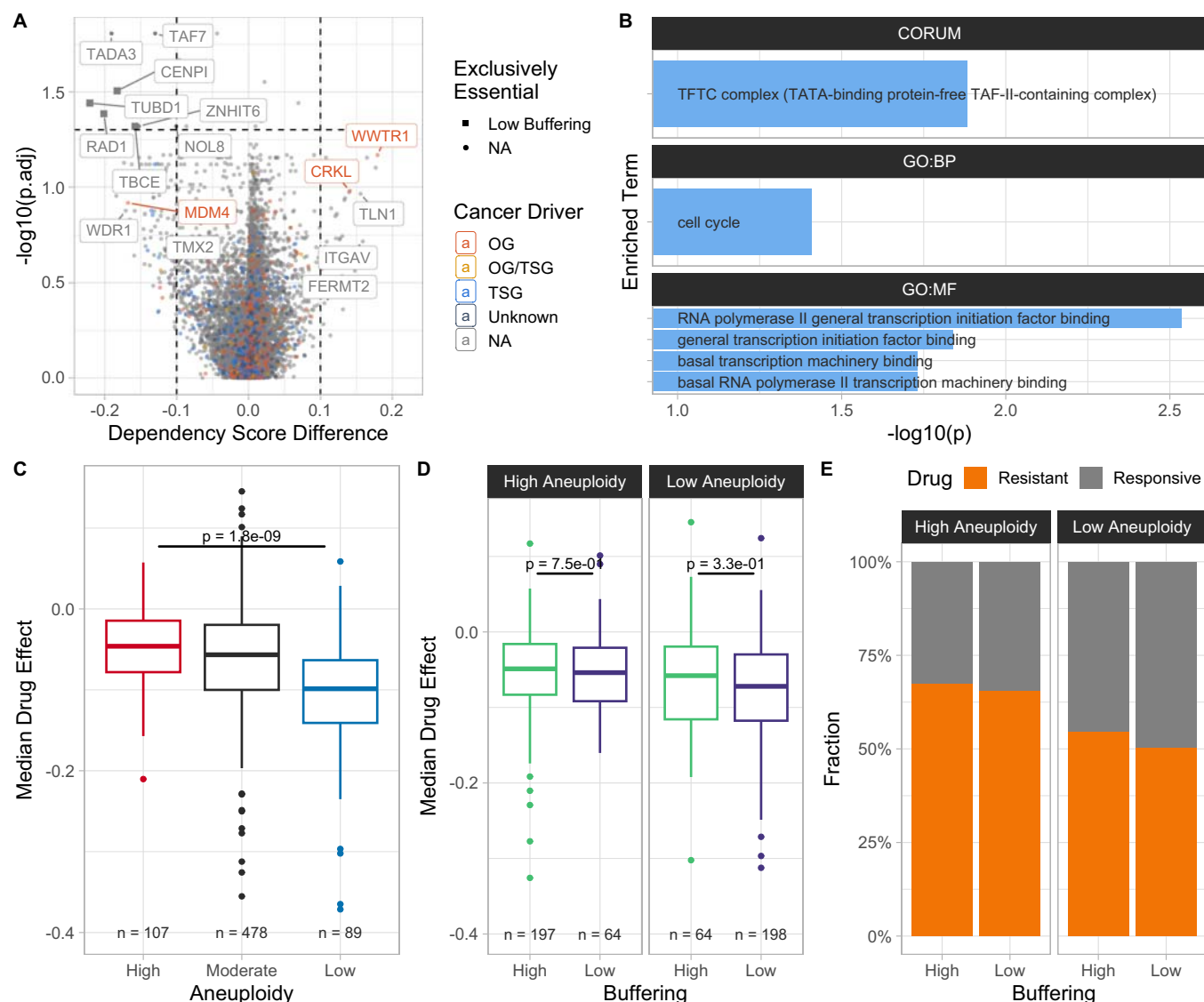


Figure EV6. Adherent and aneuploid controls for CRISPR knock-out and drug sensitivity analyses.

(A) Volcano plot showing the difference of the CRISPR-KO dependency score between high (>80% quantile of sample BR MNRs, $n = 90$) and low (<20% quantile, $n = 67$) buffering cell lines plotted against Benjamini-Hochberg adjusted Wilcoxon rank-sum test P values. Cell lines without adherent growth patterns were removed prior to the analysis. Genes were highlighted if they were significant in the adherent control or among the top 10 significant genes in Fig. 6A. (B) Over-representation analysis (ORA) of genes with significantly reduced CRISPR-KO dependency scores in highly buffering cell lines with adherent growth patterns (hypergeometric test using g:Profiler). (C) Median drug effect on cell viability across drugs from the PRISM Repurposing dataset for cell lines high (>80% quantile of aneuploidy score), moderate (20%-80%), and low aneuploidy (<20%). (D) Median drug effect for high (>50% sample BR MNR) and low-buffering cell lines, separated by median aneuploidy score. (E) Categorical distribution of drug-resistant (>50% median drug effect) and responsive cells between high (>50% sample BR MNR) and low-buffering cell lines, separated by median aneuploidy score. (C, D) Boxes represent the interquartile range (IQR) with the central line indicating the median. The whiskers extend to the data points within $1.5 \times \text{IQR}$. P values were determined using a two-tailed Wilcoxon rank-sum (Mann-Whitney U) test.

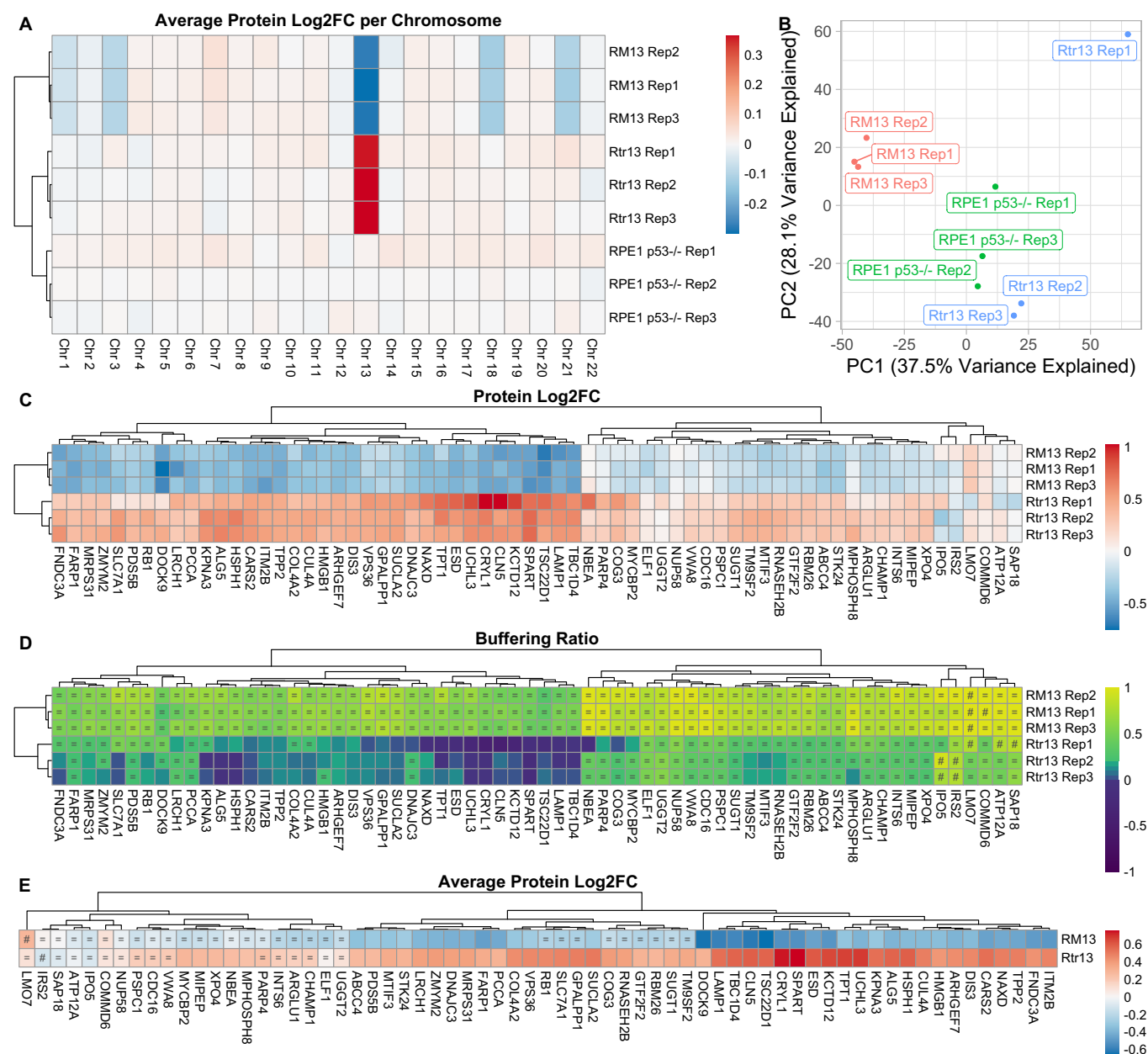


Figure EV7. Proteome and protein buffering analysis of RPE-1 cell lines with engineered aneuploidies on chromosome 13 (P0211).

(A) Mean log2 fold change of protein abundance per chromosome between cell line replicates and disomic baseline (median protein abundance of unaltered RPE-1 replicates per protein). (B) Principal components analysis (PCA) of P0211 replicates after normalization (first two principal components are displayed). (C) Protein abundance log2 fold change of proteins encoded on chromosome 13 between aneuploid replicates and disomic baseline. (D) Chromosome arm copy number-derived buffering ratios (BR) of proteins encoded on chromosome 13 in aneuploid replicates. Proteins are classified as *Buffered* (=) and *Anti-Scaling* (#) using BR-based thresholds (see "Methods"). (E) Mean protein abundance log2 fold-changes of proteins encoded on chromosome 13 between aneuploid and non-aneuploid cell lines. Proteins were classified as *Buffered* (=) and *Anti-Scaling* (#) using Log2FC-based thresholds (see "Methods").