

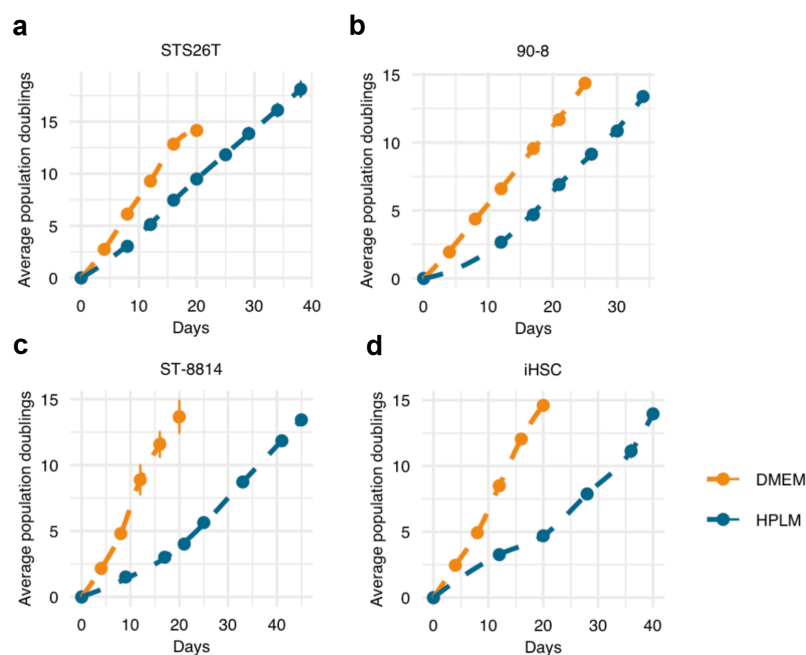
Variation in guide RNA library representation results in gene effect score bias in genome-wide CRISPR screens.

Paul Metz^{1†}, Sofia Alves-Vasconcelos¹, Richard Wallbank¹, Joey Riepsaame^{1,2}, Stuart Brown¹ and A. Bassim Hassan^{1,¶}.

¹Oxford Molecular Pathology Institute, Sir William Dunn School of Pathology, University of Oxford, South Parks Road, Oxford, OX1 3RE, United Kingdom.

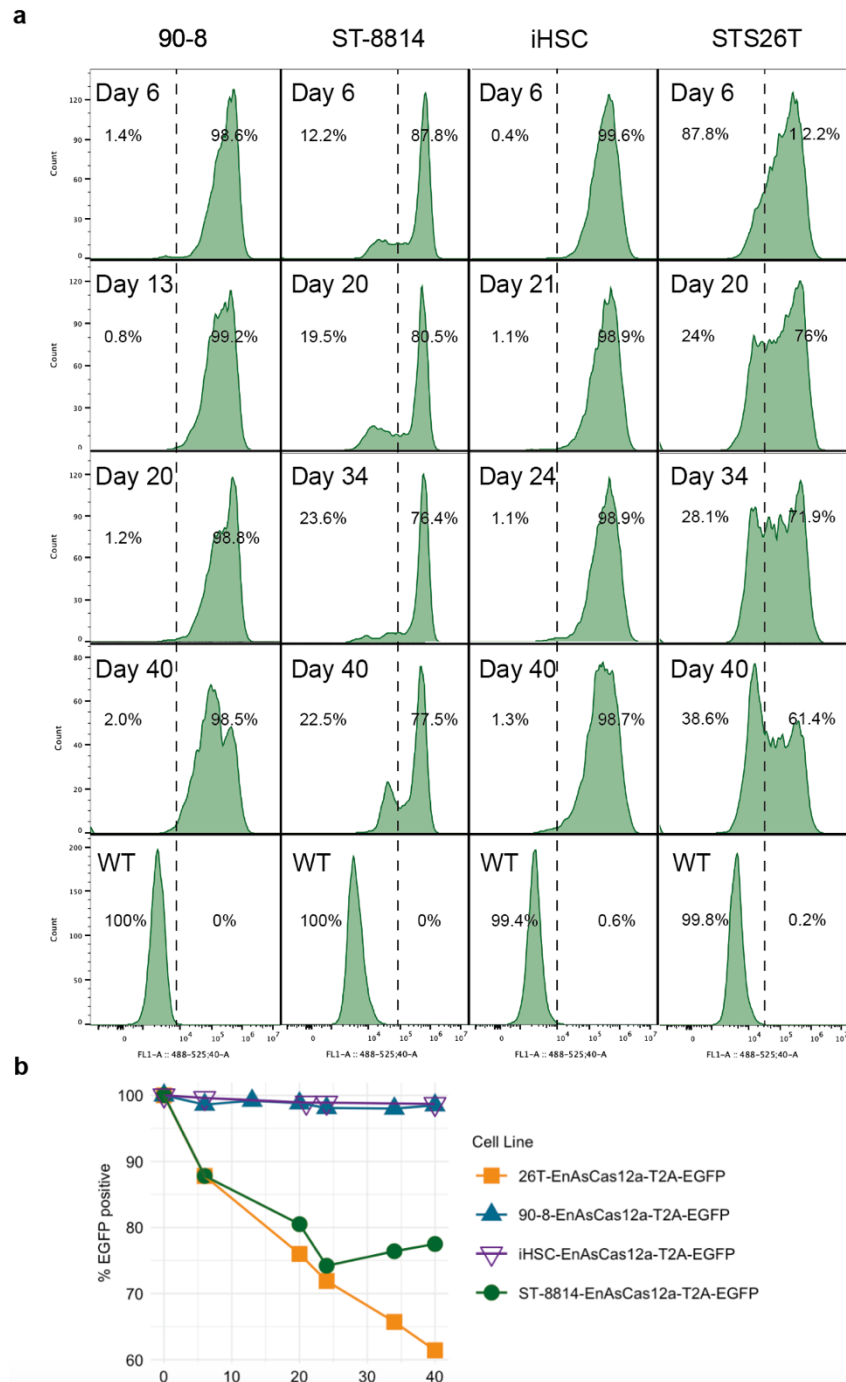
²Genome Engineering Oxford, Sir William Dunn School of Pathology, University of Oxford, South Parks Road, Oxford, OX1 3RE, United Kingdom.

Supplementary Figures

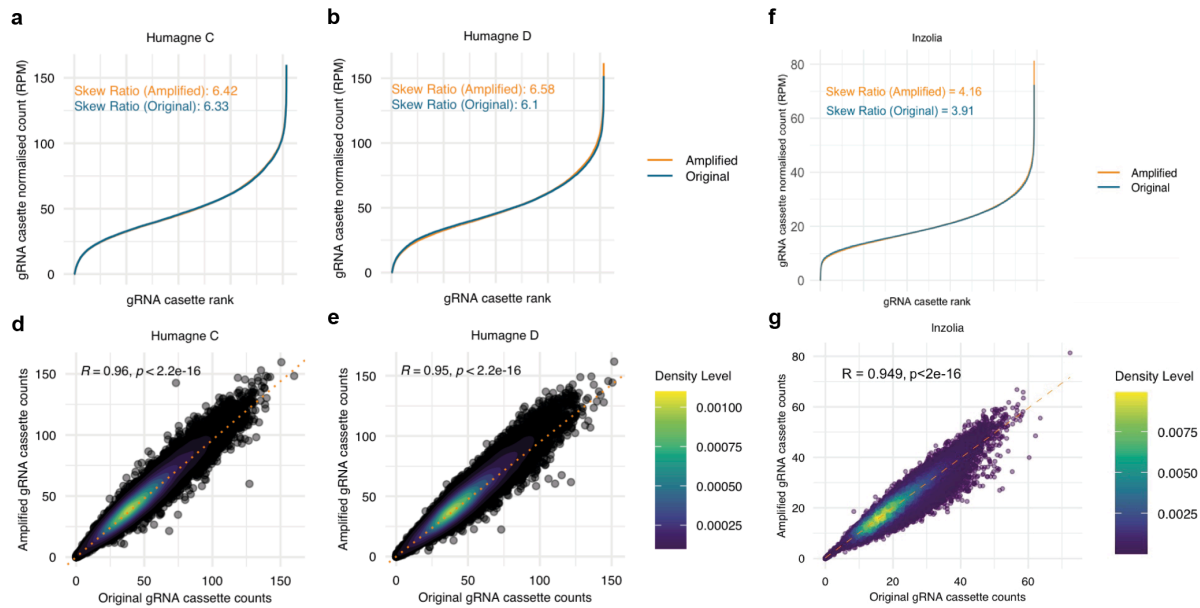


Supplementary Figure 1. Growth curves of target cell lines in DMEM and HPLM.

a-d. Growth curves of target cell lines STS26T (a.), 90-8 (b.), ST-8814 (c.) and iHSC (d.) with average population doublings (PDs) over time. Data points and dashed lines indicate PD measurements and Locally Estimated Scatterplot Smoothing (LOESS) of cells grown in DMEM (orange) and HPLM (blue). Each dot represents two biological replicates with vertical lines representing standard deviation.

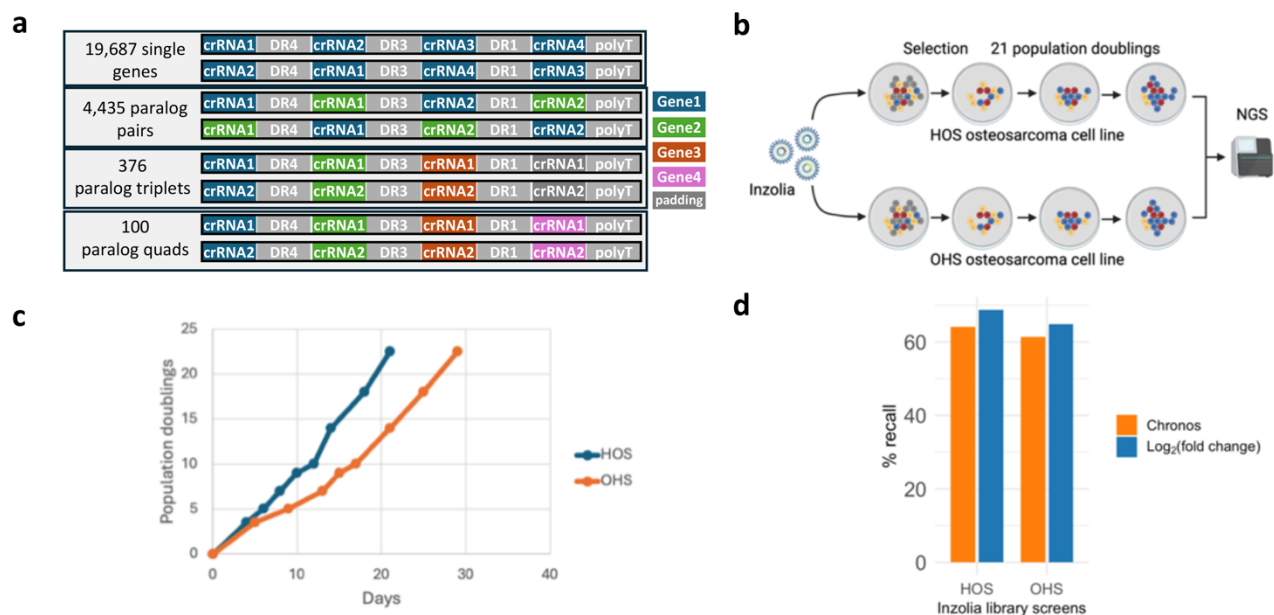


Supplementary Figure 2. Growth curves of target cell lines in DMEM and HPLM. a. Histograms of flow cytometry measuring EGFP in 90-8, ST-8814, iHSC and STS26T (26T) cells complemented with enAsCas12a-T2A-EGFP using lentiviral transgene delivery. Cells were monitored for up to 40 days post transduction. EGFP positivity was gated stringently well above negative control populations (WT) to be able to separate true positives from pseudo positives, especially for ST-8814 and STS26T (26T) cells. **b.** Line plot with time series data as depicted showing the percentage of EGFP positive cells based on the gating strategy as depicted in (a) over time for 90-8, ST-8814, iHSC and STS26T (26T) cells complemented with enAsCas12a-T2A-EGFP. Each data point represents one biological replicate.

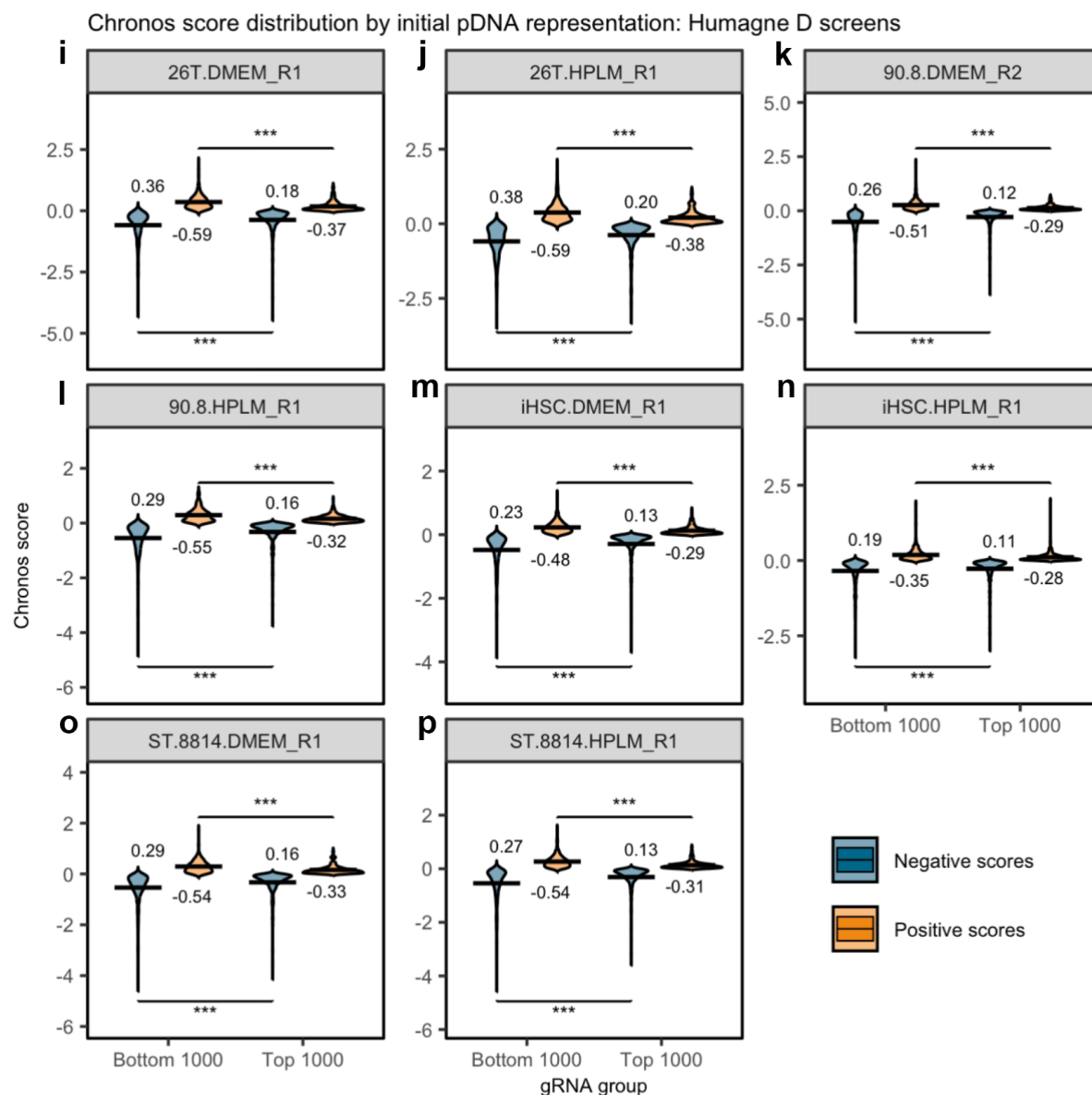


Supplementary Figure 3. CRISPR library amplification shows minimal skewing of gRNA cassette distribution.

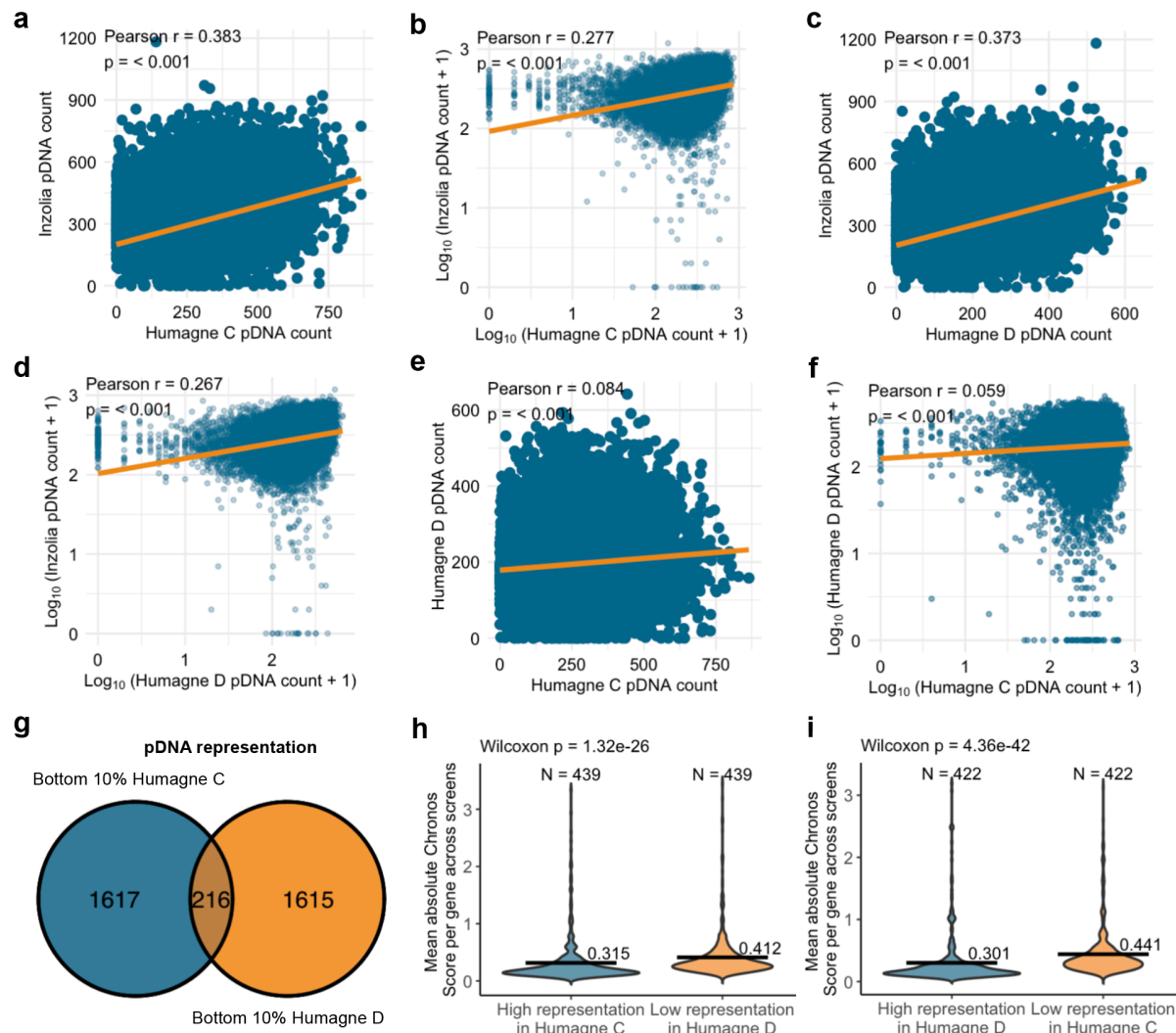
a. gRNA cassette distribution with normalised counts (RPM) per gRNA cassette for original and amplified Humagne C CRISPR libraries. The skew ratio indicates the ratio between mean normalised counts of the top 10% gRNA cassettes divided by the mean normalised counts of the bottom 10% gRNA cassettes. **b.** Same as in (a.) but for Humagne D. **c.** Same as in (a.) but for Inzolia. **d.** Dot and density plot of the relationship between the normalised counts of original and amplified gRNA cassettes. The spearman's rho correlation value and p-value are displayed on the plot. The Spearman's rho correlation coefficient is also depicted by an orange dashed line. **e.** Same as (d.) but for Humagne D. **f.** Same as (d.) but for Inzolia.



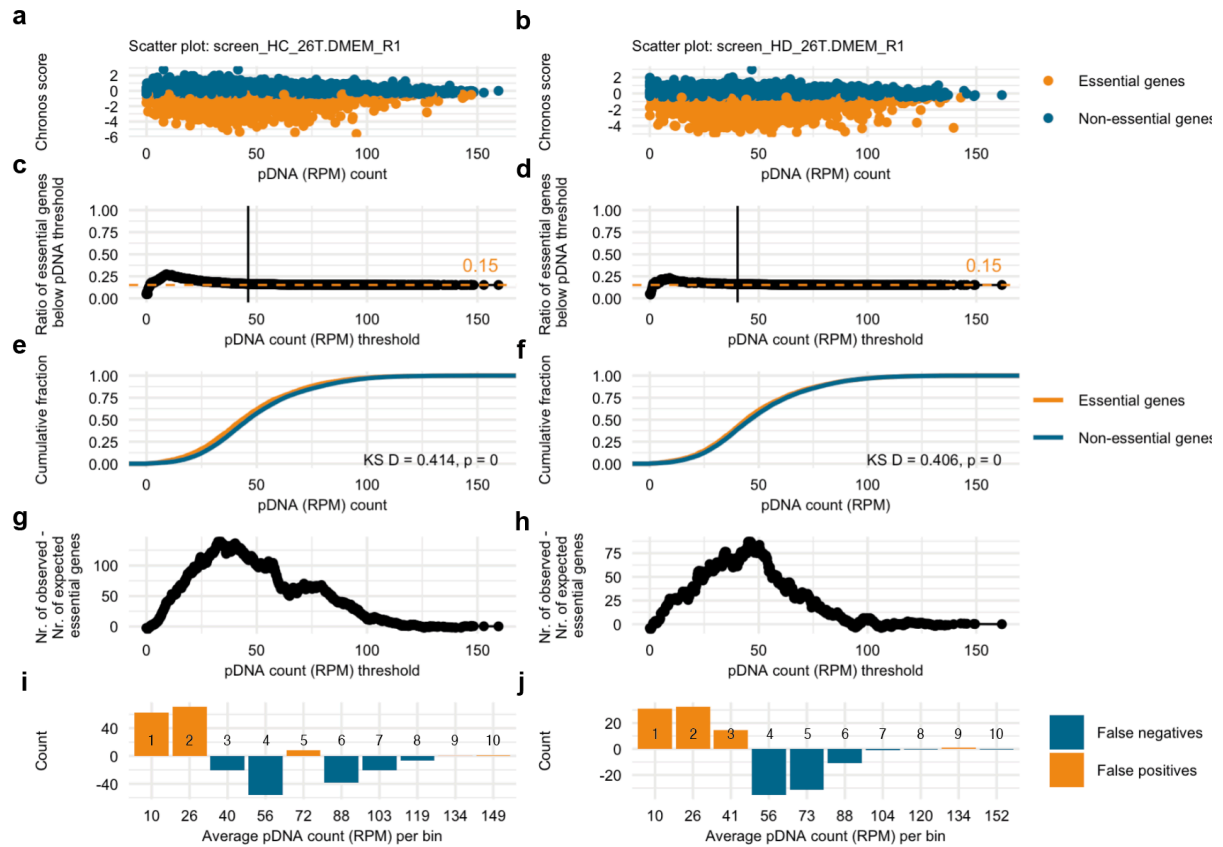
Supplementary Figure 4. Inzolia-based CRISPR screen workflow and quality control. **a.** Overview of the Inzolia library design, which includes crRNAs for 19,687 single-gene guides, 4,435 paralog-pair guides, 376 triplet guides, and 100 quadruplet guides³². **b.** CRISPR screen workflow for enAsCas12a Inzolia all-in-one library for human osteosarcoma cell lines HOS and OHS. **c.** Average population doublings (PDs) over time for HOS and OHS. **d.** Essential gene recall (Chronos and \log_2 (fold change)) for HOS and OHS.



Supplementary Figure 5. Comparison of positive and negative Chronos score distributions at high and low pDNA representation. a-p. Violin plots show the distribution of positive and negative Chronos scores for the top and bottom 1000 gRNAs ranked by initial pDNA abundance. Each figure corresponds to a specific screen performed with either Humagne C (a-h) or Humagne D (i-p). Scores are separated by sign: positive scores (orange) and negative scores (blue). Horizontal black lines indicate the mean Chronos score within each group and values are indicated alongside. For each screen, Wilcoxon rank-sum tests were used to compare Chronos score distributions between Bottom 1000 and Top 1000 gRNAs separately for positive and negative scores. $P < 0.05$ (*), $P < 0.01$ (**), $P < 0.001$ (***), or not significant (NS).

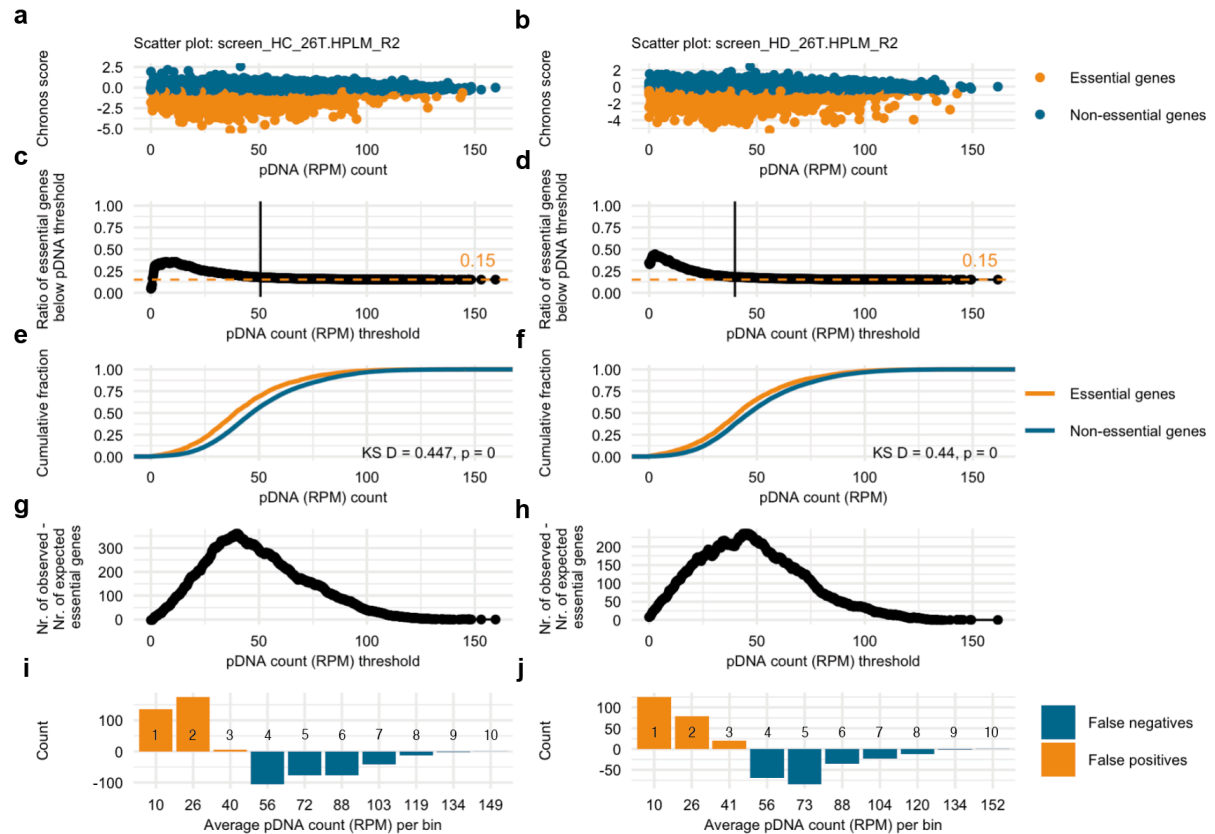


Supplementary Figure 6. Comparison of initial pDNA representation and gene effect scores between Humagne C, Humagne D, and Inzolia libraries. **a-d.** Scatter plots show raw (a, c) and log-transformed (b, d) pDNA counts for Humagne C vs. Inzolia (a, c) and Humagne D vs. Inzolia (b, d). Pearson correlation coefficients (r) and p -values are indicated. Linear regression lines are shown in orange. **e-f.** Scatter plots show raw (e) and log-transformed (f) pDNA counts for Humagne C versus Humagne D pDNA counts. Pearson correlation coefficients (r) and p -values are indicated. Linear regression lines are shown in orange. **g.** Venn diagram showing the **intersection** of gRNAs in the bottom 10% **based on** initial representation in Humagne C (blue) and Humagne D (orange). **h-i.** Violin plots of mean absolute Chronos scores per gene across Humagne C or Humagne D screens comparing gRNAs with high representation (within highest 50%) in one library (h: Humagne C, i: Humagne D) to different gRNAs targeting the same genes when **poorly** represented (within lowest 5%) in the other library (h: Humagne D, i: Humagne C). Wilcoxon test p -values are indicated on each plot. Means are indicated by horizontal black line with values stated alongside. The number of samples in each group is stated above each violin.

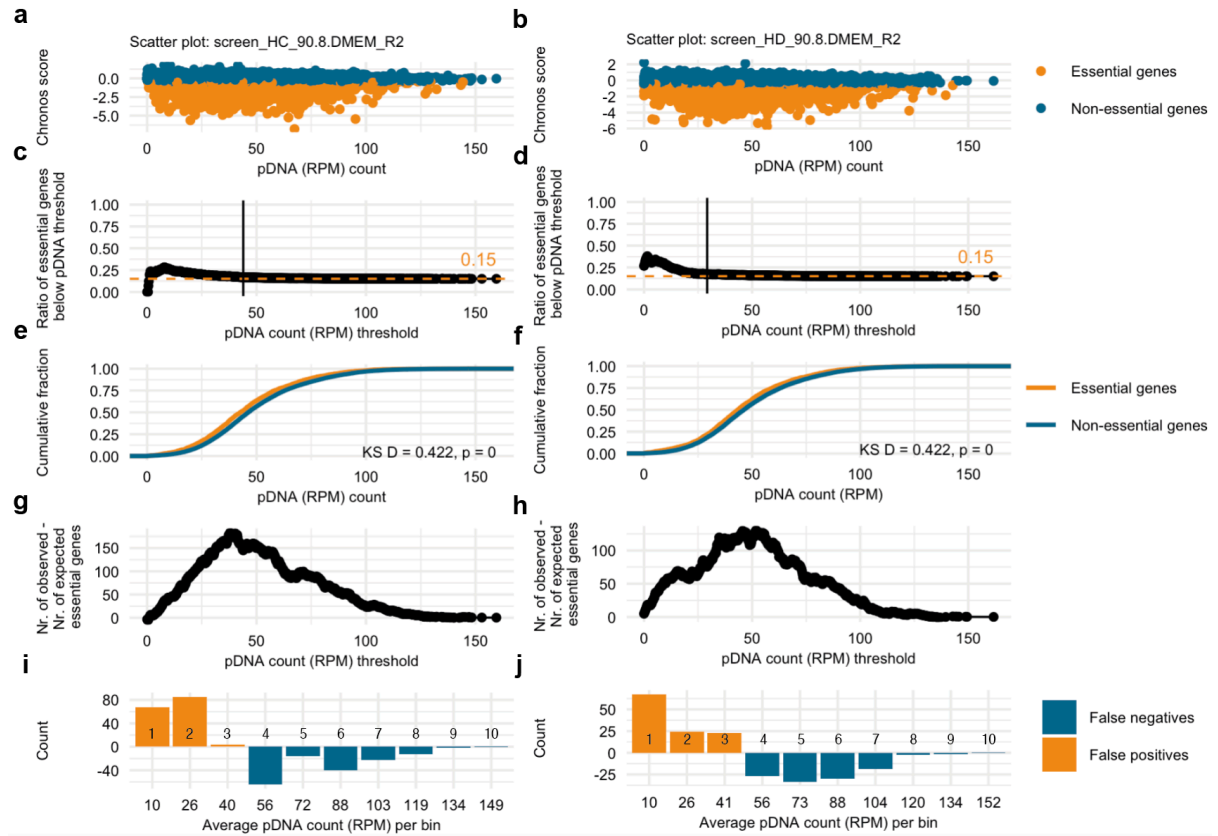


Supplementary Figure 7. Estimation of false positive and negative results coming from pDNA representation bias. **a.** Dot plot of the relationship between Chronos scores and normalised pDNA counts (RPM) in 26T, DMEM, replicate 1, Humagne C CRISPR screen. The bottom 15% genes based on Chronos score are highlighted in orange. **b.** Same as in (a.) but for the equivalent Humagne D screen. **c.** Dot plot depicting the ratio of found essential genes and the total number of genes below each possible pDNA threshold for the 26T, DMEM, replicate 1, Humagne C CRISPR screen. The dashed orange line indicates the predicted 15% essential genes threshold if essential genes were equally distributed across pDNA representation. The vertical black line represents the natural elbow as found using the maximum perpendicular distance method. **d.** Same as in c but for the equivalent Humagne D screen. **e.** The empirical cumulative distribution functions are shown for essential genes (orange) and non-essential genes (blue) for the 26T, DMEM, replicate 1, Humagne C CRISPR screen. The x-axis represents normalised pDNA counts, and the y-axis indicates the cumulative fraction of essential/non-essential genes. The Kolmogorov–Smirnov (KS) statistic and associated p value are annotated on the plot to highlight the difference between the distributions. **f.** Same as (e.) but for the equivalent Humagne D screen. **g.** Dot plot showing the predicted number of false positives below each possible normalised pDNA count (RPM) threshold, by subtraction of the number of expected essential genes from the number of observed essential genes, for the 26T, DMEM, replicate 1, Humagne C CRISPR screen. **h.** Same as in (g.) but for the equivalent Humagne D screen. **i.** Bar graph of the estimated number of false positive and negative essential genes per pDNA bin for the 26T, DMEM, replicate 1, Humagne C CRISPR screen, with the x axis showing the average number of normalised pDNA

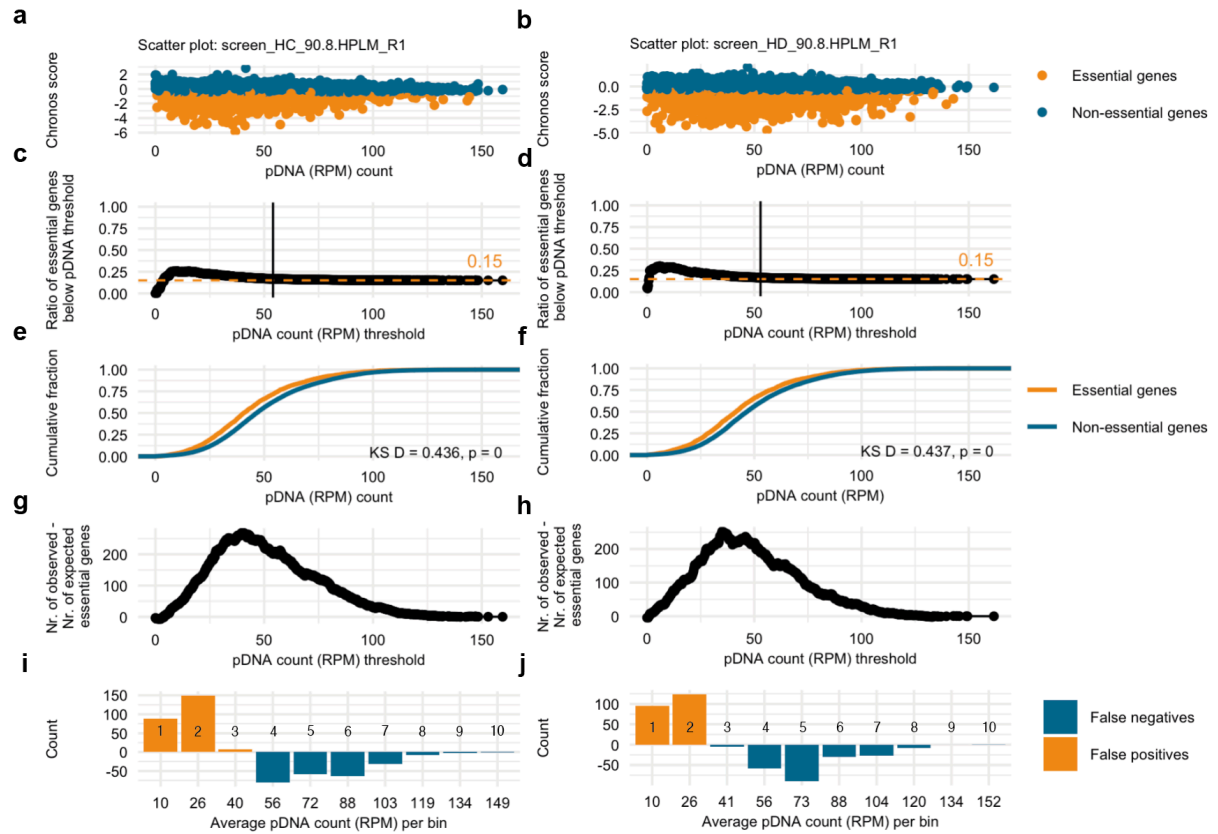
counts (RPM) associated with each bin and the number on each bar indicating the bin number. j. same as (i.) but for the equivalent Humagne D screen.



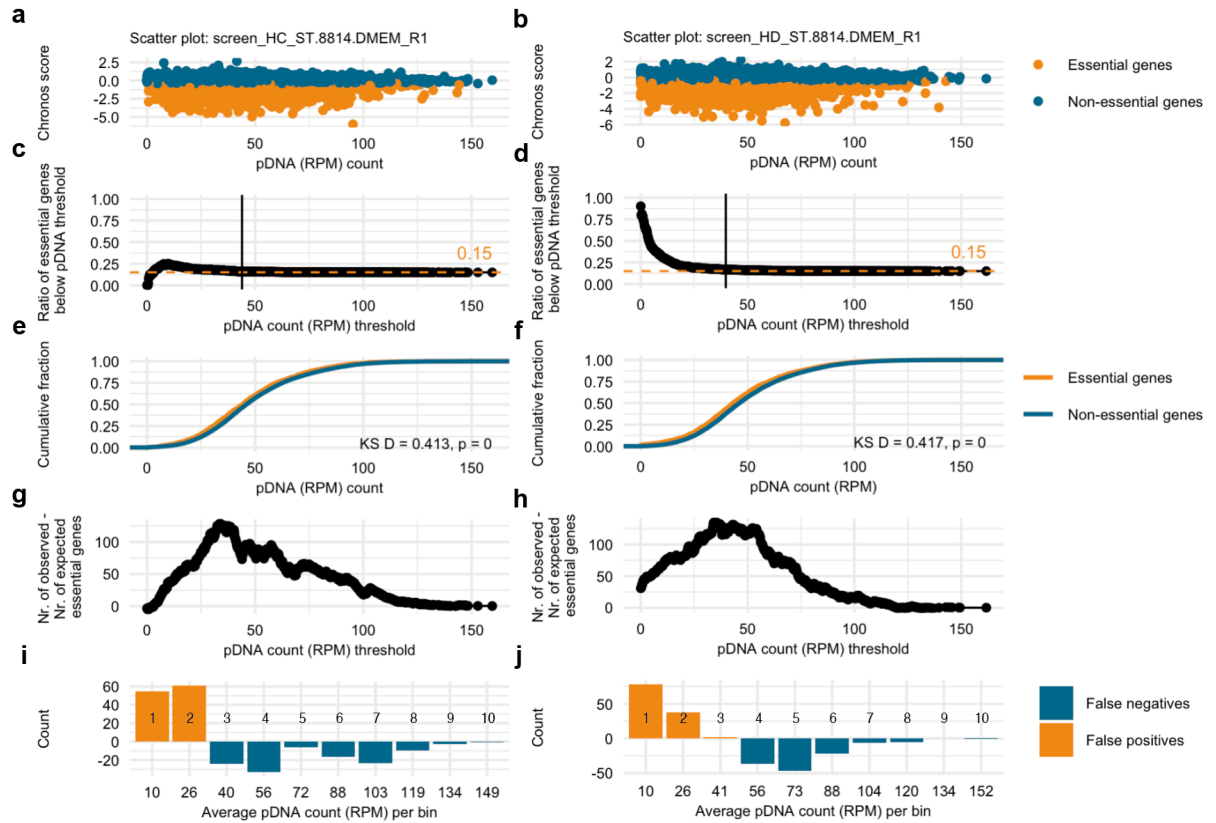
Supplementary Figure 8. Estimation of false positive and negative results coming from pDNA representation bias. a-j: same as Supplementary Figure 7, but for the STS26T, HPLM, replicate 2 screen.



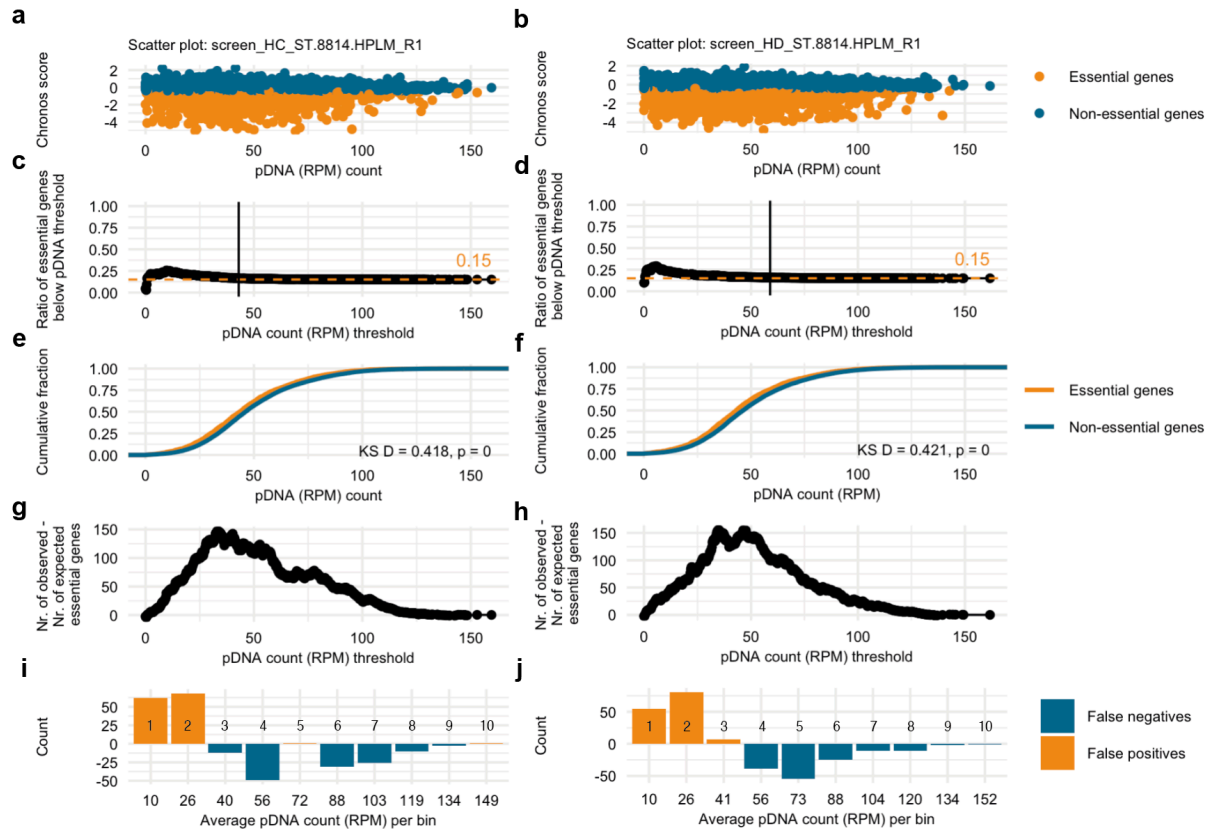
Supplementary Figure 9. Estimation of false positive and negative results coming from pDNA representation bias. a-j: same as Supplementary Figure 7, but for the 90-8, DMEM, replicate 2 screen.



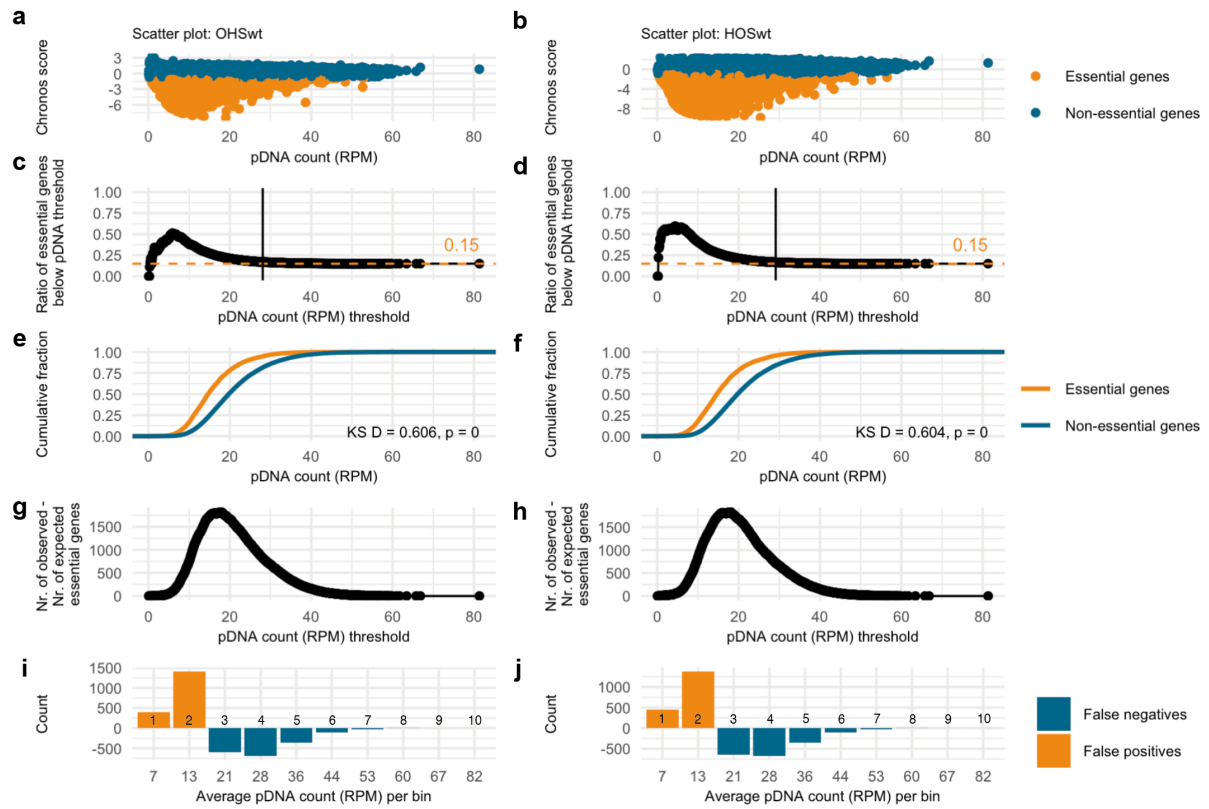
Supplementary Figure 10. Estimation of false positive and negative results coming from pDNA representation bias. a-j: same as Supplementary Figure 7, but for the 90-8, HPLM, replicate 1 screen.



Supplementary Figure 11. Estimation of false positive and negative results coming from pDNA representation bias. a-j: same as Supplementary Figure 7, but for the ST-8814, DMEM, replicate 1 screen.



Supplementary Figure 12. Estimation of false positive and negative results coming from pDNA representation bias. a-j: same as Supplementary Figure 7, but for the ST-8814, HPLM, replicate 1 screen.



Supplementary Figure 14. Estimation of false positive and negative results coming from pDNA representation bias. a-j: same as Supplementary Figure 7, but for Inzolia-based CRISPR screens in HOS and OHS cell lines.