

Supplementary material to NMDtxDB: Data-driven identification and annotation of human NMD target transcripts

Thiago Britto-Borges, Niels Gehring, Volker Boehm, and Christoph Dieterich

Overview of significantly and PTC containing features.

Supplemental_Table_S1.xlsx: Excel spreadsheet containing 20 tables, as follows:

- Dge_top20: Table with the top 20 annotated genes containing any transcript with premature stop codon (PTC) per contrast and adjusted p -value < 0.05 , ordered log2FoldChage.
- Dte_top20: Table with the top 20 PTC-containing transcripts per contrast, ordered log2FoldChage.
- Followed by 9 tables, one for each contrast with genes containing any PTC transcript and significantly called by the DGE modeling (adjusted p -value < 0.05 and $|\log2FoldChage| > 1$). Contrast names were limited to 30 characters to fit MS Excel limitations.
- The last 9 tables, ones for each contrast, containing PTC transcripts significantly called by the DTE method.

Comparison of UTR length of Ensembl transcripts by DTE status

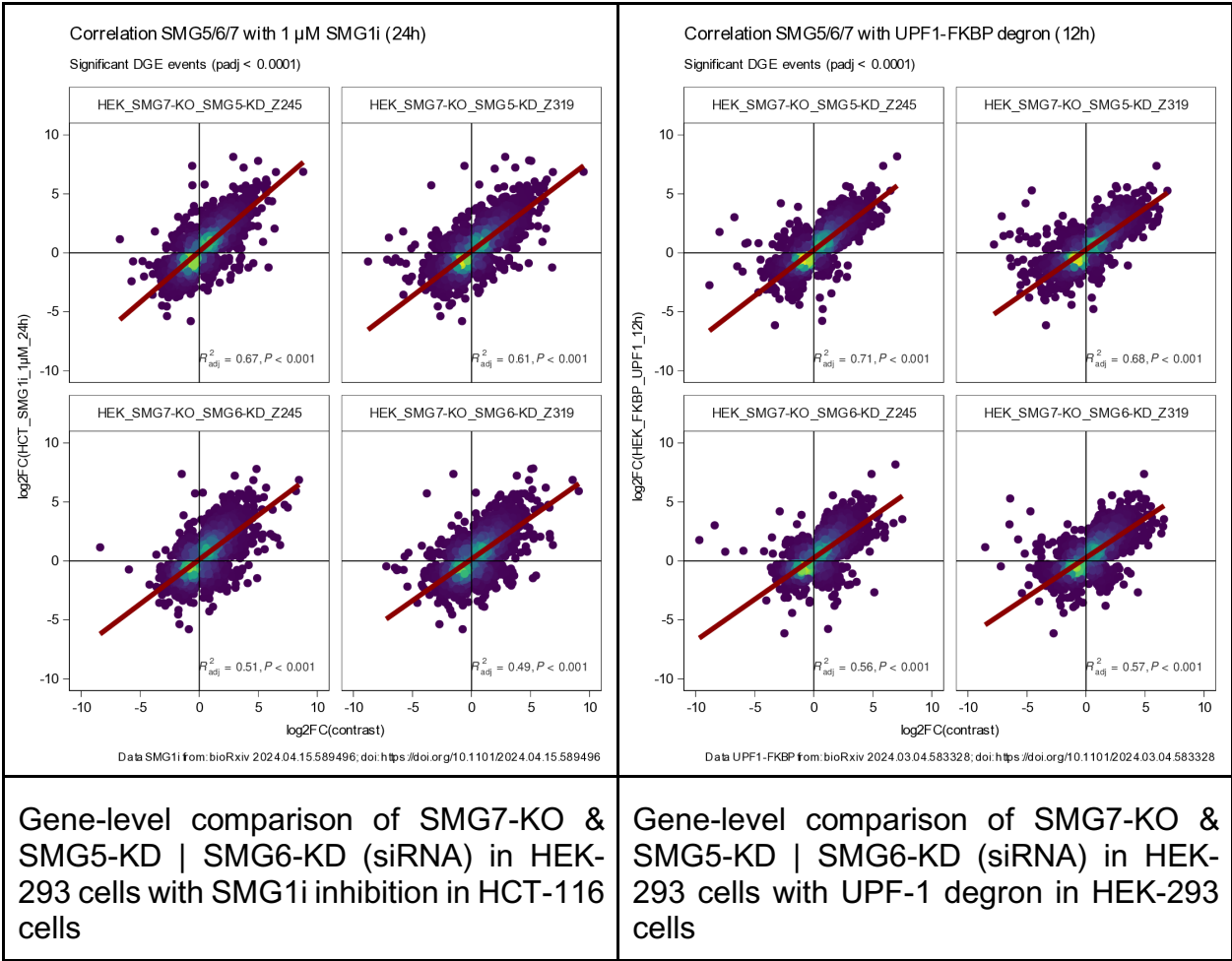
Supplemental Table S2

Category	Median UTR length (bp)	Wilcoxon rank sum test p
Down-regulated (n=2030)	976	0.39
Unchanged (n=4859)	1.062	n.a.
Up-regulated (n=1902)	883	9.22e-08

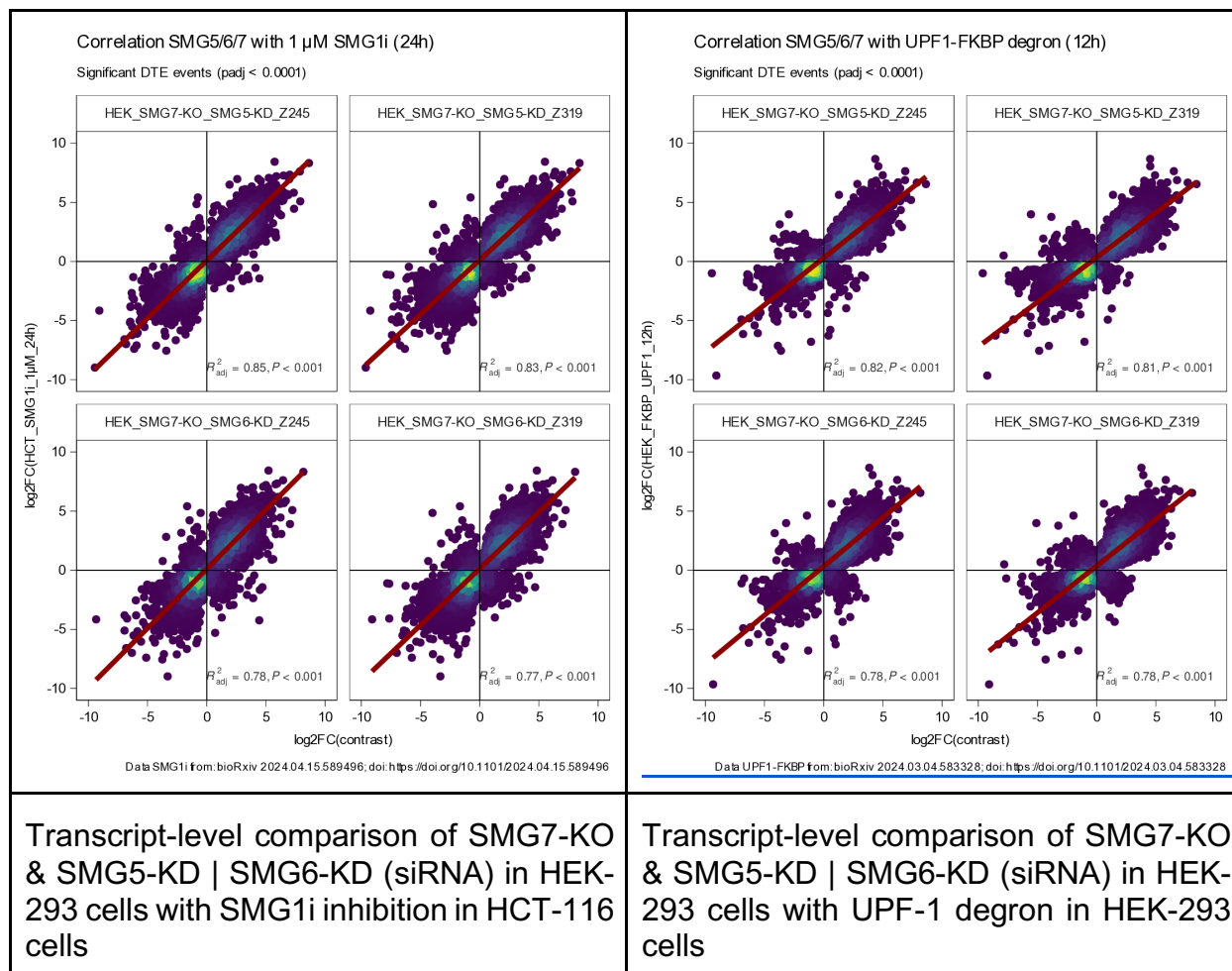
Comparison of NMD perturbation by siRNA, degron or chemical inhibition.

Supplemental Figure S1

Complementary to our siRNA-based approach of NMD inhibition, we present two alternative strategies below: a UPF1 degron based system and a SMG1 inhibitor. The first data comprise unpublished data obtained from Kueckelmann, et al., 2024 (<https://doi.org/10.1101/2024.04.15.589496>) that uses SMG1 inhibition. The second data, which is also unpublished, was obtained from Boehm et al., 2024 (<https://doi.org/10.1101/2024.03.04.583328>) and uses an UPF1 degron approach. By using gene expression and transcript expression analysis of the annotated transcriptome, we observe a high concordance ($R^2 > 0.7$) of the log2 fold change estimates of differential gene expression (DGE) and differential transcript expression (DTE) among our and both external datasets.



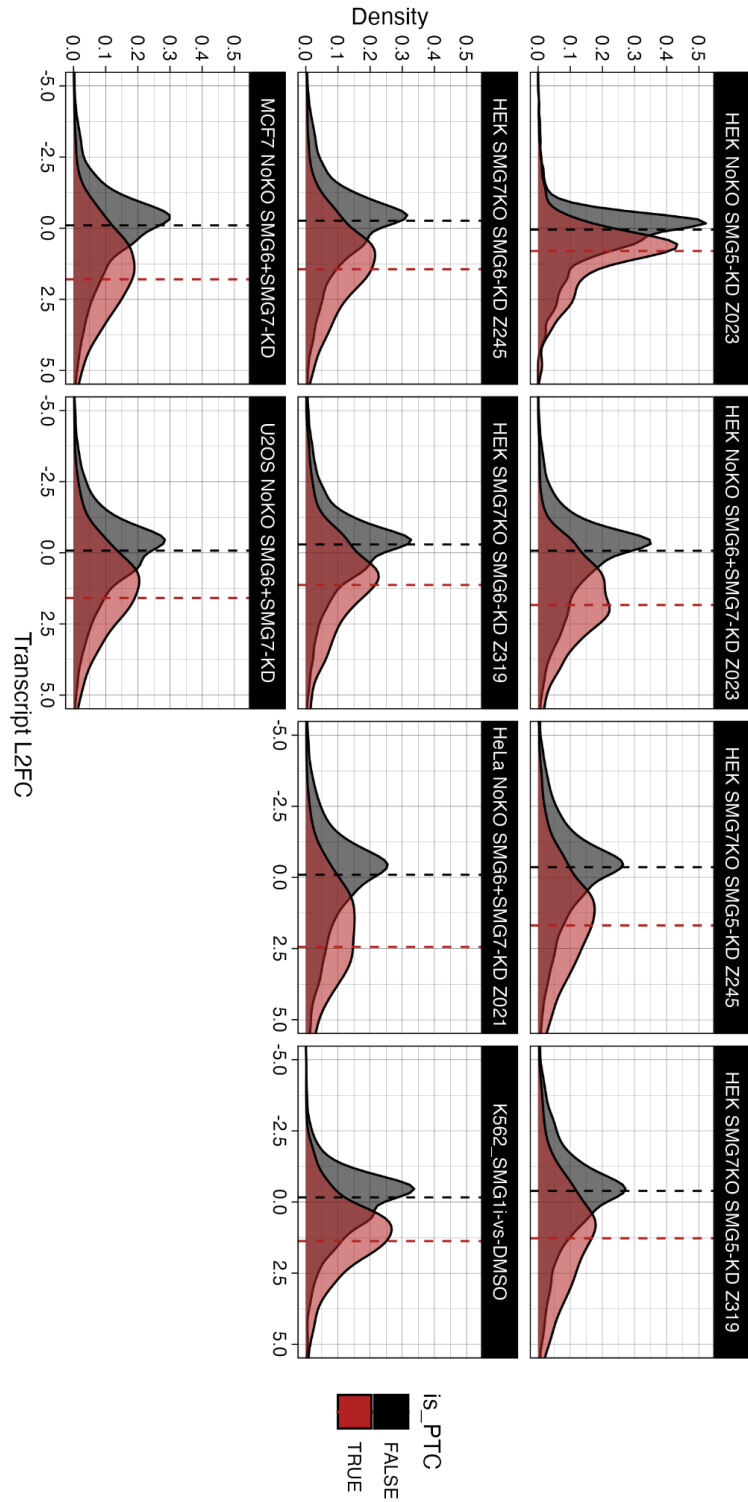
Moreover, these effects were also concordant at the transcript expression level:



NMD transcriptome expression across our data set and one external SMG1i experiment in K562 cells:

Supplemental Figure S2

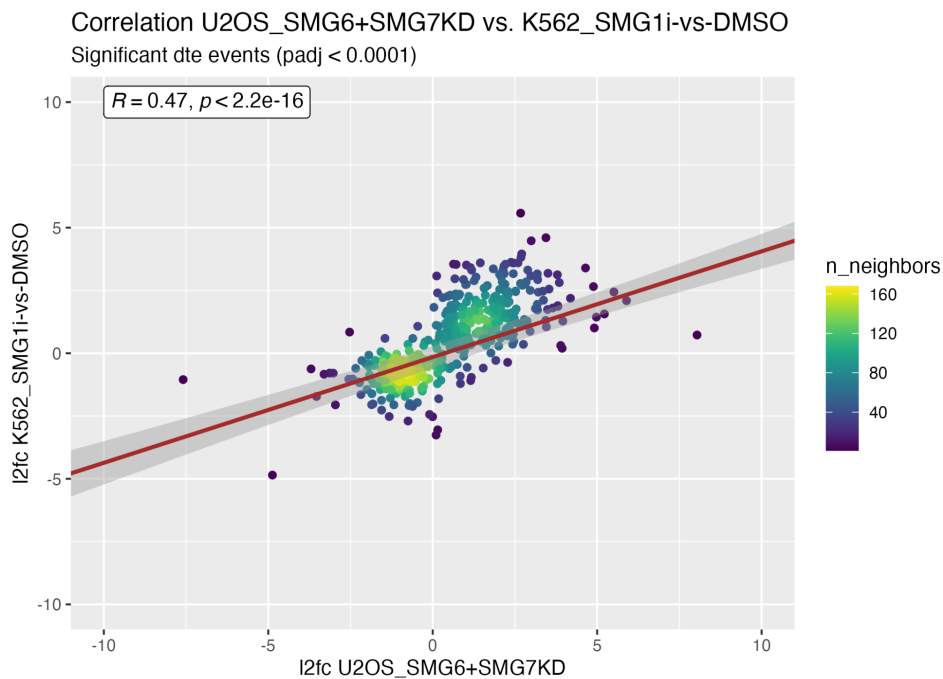
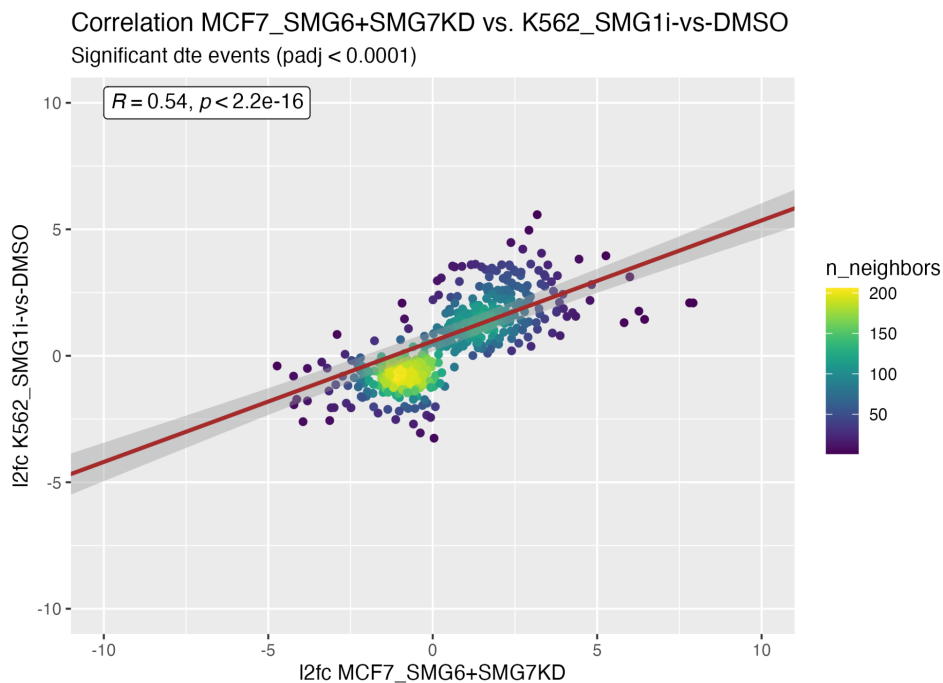
Next, we apply the transcriptome obtained *de novo* to a publicly available and independent data set provided by Zinshteyn, *et al.* 2021 (10.1371/journal.pgen.1009813), obtained from the sequence read archive (accession SRR13016575, SRR13016578, SRR13016587, SRR13016610). We provide a reproducible workflow available at https://github.com/dieterich-lab/nmd-wf/tree/master/smgi1_analysis. Here, we show that our *de novo* transcriptome annotation approach can be used to identify NMD transcripts from a different cell line. To this end, we reproduce Fig. 4 of the manuscripts and additionally include an external new dataset, showing that SMG1i treatment of K562 cell leads to an enrichment of PTC containing transcripts:



Supplemental Figure S3

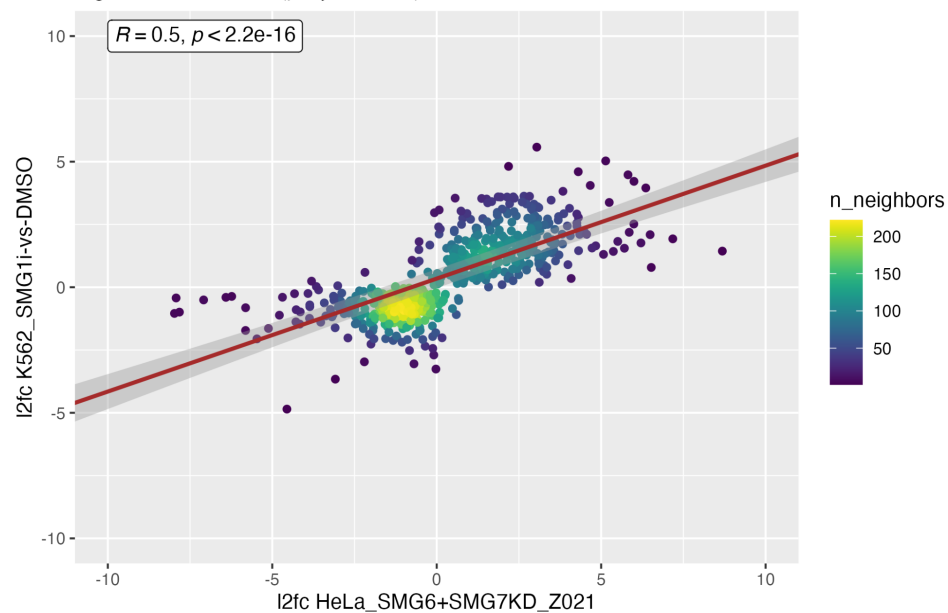
Finally, we compare the $\log_2$ fold changes of common transcripts between SMG1i-vs-DMSO comparison and all comparisons presented in NMDtxDB. We observe that, despite differences in

cellular transcriptome and methodological approach, log₂ fold changes correlate positively with an R² ranging from 0.23 to 0.56.



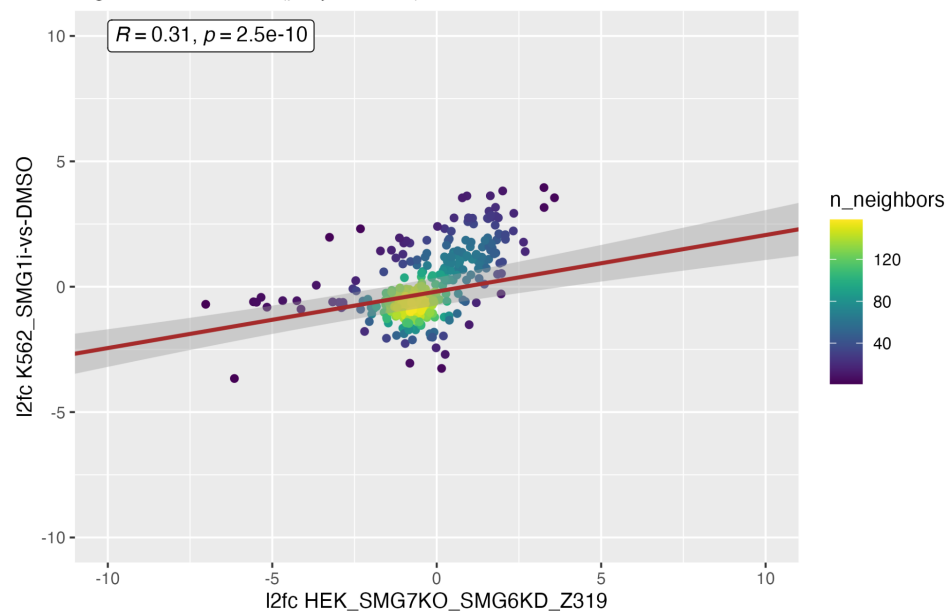
Correlation HeLa_SMG6+SMG7KD_Z021 vs. K562_SMG1i-vs-DMSO

Significant dte events (padj < 0.0001)



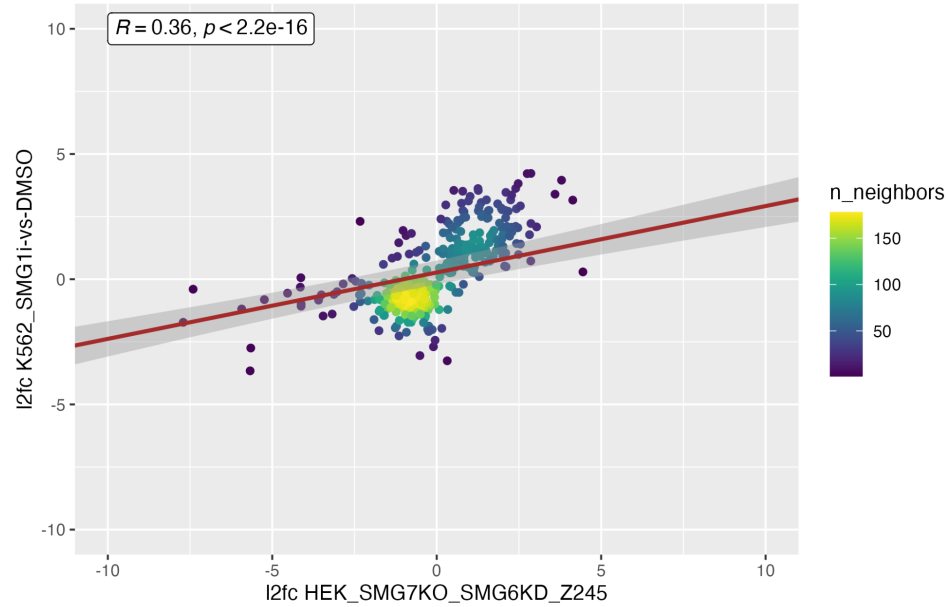
Correlation HEK_SMG7KO_SMG6KD_Z319 vs. K562_SMG1i-vs-DMSO

Significant dte events (padj < 0.0001)



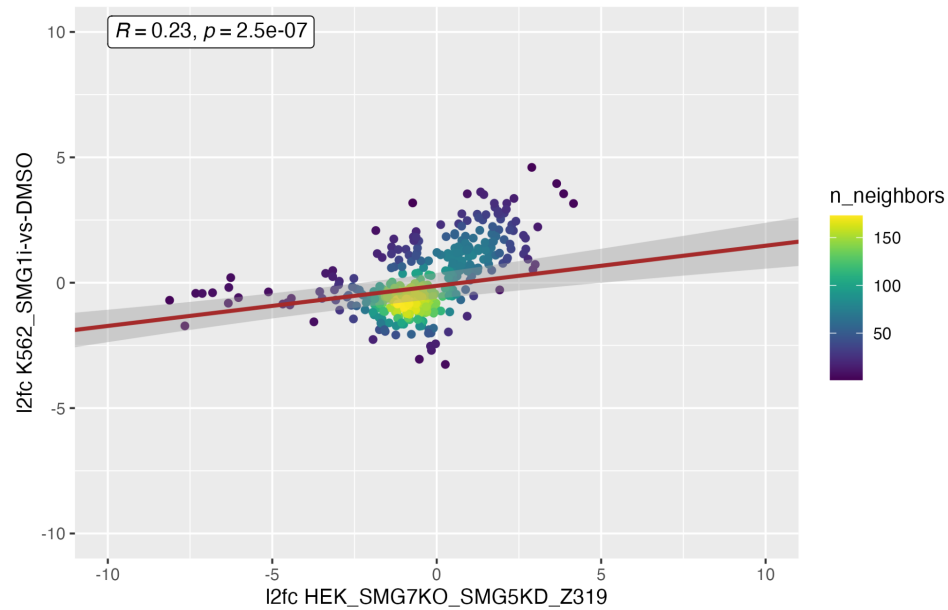
Correlation HEK_SMG7KO_SMG6KD_Z245 vs. K562_SMG1i-vs-DMSO

Significant dte events (padj < 0.0001)



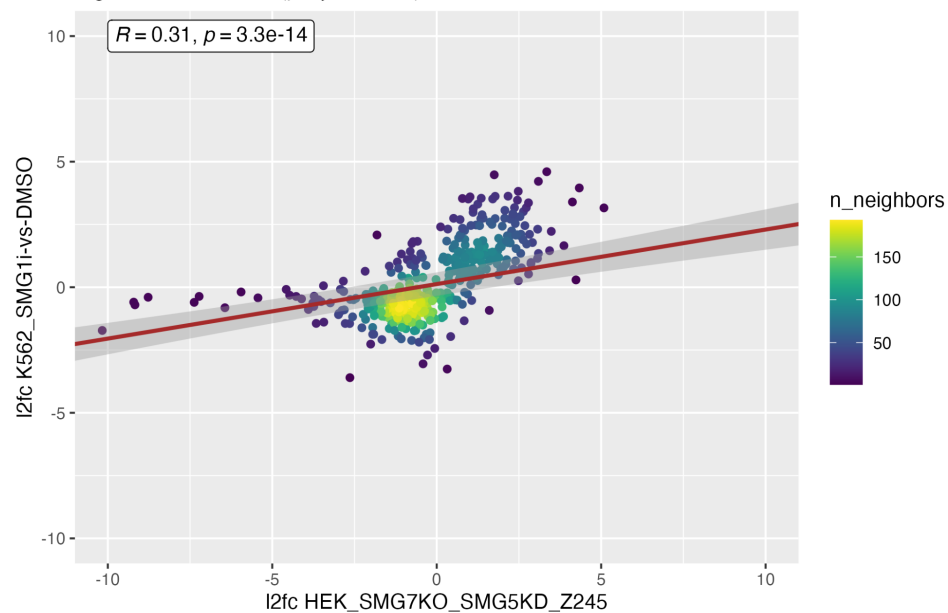
Correlation HEK_SMG7KO_SMG5KD_Z319 vs. K562_SMG1i-vs-DMSO

Significant dte events (padj < 0.0001)



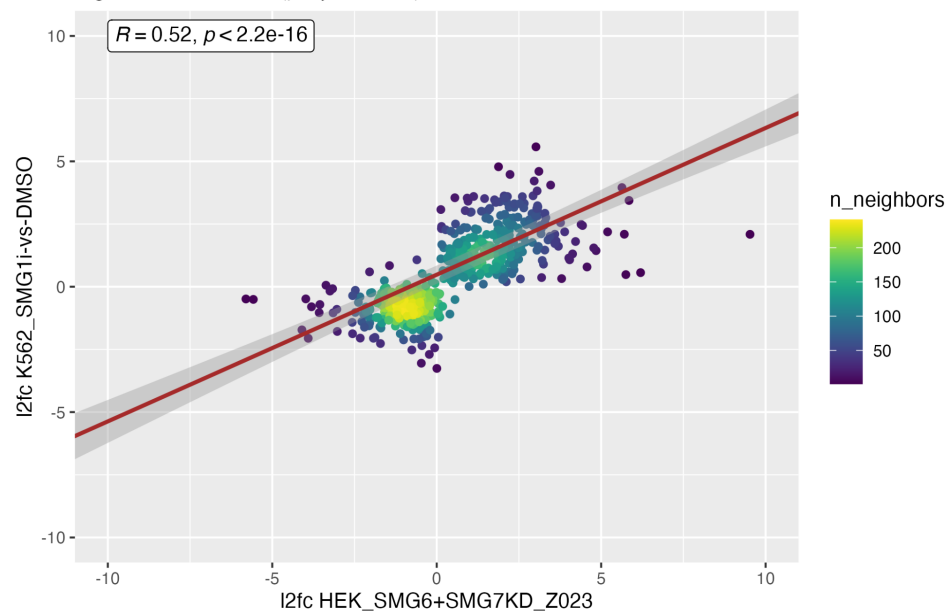
Correlation HEK_SMG7KO_SMG5KD_Z245 vs. K562_SMG1i-vs-DMSO

Significant dte events (padj < 0.0001)



Correlation HEK_SMG6+SMG7KD_Z023 vs. K562_SMG1i-vs-DMSO

Significant dte events (padj < 0.0001)



Correlation HEK_SMG5KD_Z023 vs. K562_SMG1i-vs-DMSO

Significant dte events (padj < 0.0001)

