**ORI initiation and blocking of ORI derived transcripts**

a. The core-promoter motif content is shown for the ORI of the pGL4 family. The track above is taken from Figure 1a, panel for SCP1. b. Percent of sites within the indicated plasmid elements (backbone, ORI, SCP1) that initiate at the indicated threshold counts (1,2: background; 3-5: very weak; compare Figure 1a, Supplementary Figure 1a). c. Representative STARR-seq enhancer activity profiles obtained for reporter setups from Figure 1a are shown. The Refseq (GRCh37) gene track is indicated above, the dashed boxes indicate luciferase-validated enhancers, the red asterisk high background signal. d. Representative enhancer activity profiles over three gene loci (indicated above; H3K27ac data from ENCODE, Supplementary Table 3) for an SCP1 STARR-seq screen (see Supplementary Figure 1c and text). The upper profile is obtained from all reporter transcripts and the lower two from transcripts that either initiated in the ORI (top) or the SCP1 (bottom, see methods for details on stratification). e. Scatter plot of STARR-seq signal-over-background between screens employing SCP1 or the ORI as a core-promoter. Shown are all peaks called in either screen, highlighting predicted enhancers (n=39; green) and luciferase validated enhancers (n=4; red; compare to Figure 1d). f. STARR-seq signal-to-noise between screens employing SCP1 or the ORI as a core-promoter over 10 luciferase tested candidates (4 positive over 6 negative).
stated (two-sided paired t-test, lines indicate pairing). g, Scatter plot of STARR-seq signal-over-background between screens employing SCP1 or the ORI as a core-promoter in HCT-116 cells (details as in Supplementary Figure 1e; predicted enhancers: n=27; compare to Figure 1d). h, Proposed plasmid-based reporter setups for luciferase assays, barcode-based MPRAs, and STARR-seq (for results using the proposed luciferase and STARR-seq setups see main text). For all setups, we propose to position the enhancer candidate downstream, which allows the specific assessment of enhancer- rather than promoter activity, and the use of the ORI as the core promoter. i, Proposed cloning strategies for different reporter setups employing the ORI as a core-promoter. j, Alternative strategy to block ORI-derived transcripts based on the insertion of poly-adenylation sites and splice acceptors downstream of the ORI. Note that the introduction of poly-adenylation sites alone – as present for example in pGL3 and pGL4 – are ineffective due to extensive splicing of the ORI-derived transcript (compare blocking constructs 1 and 2 to 3 and 4). The right panel depicts signal over background for the indicated constructs over predicted enhancers (n=39). Bars represent mean signal, error bars indicate 75% confidence intervals, P-values as stated (two-sided paired Fisher’s LSD test).
Most ENCODE cell lines are likely capable of mounting an IFN-I response to cytosolic nucleic acids

**a.** Hierarchical clustering of ENCODE cell lines based on their expression profiles of four genes involved in DNA- and RNA-triggered innate immunity. Blue and orange stars indicate a functioning or inactive (cGAS/STING pathway, respectively. **b, c, qPCR-based assessment of interferon stimulated genes (ISGs) induction after plasmid transfection in cells likely to have a non-functional (B, orange) or functioning DNA sensing pathway (C, blue), respectively (TUBB serves as control). Colored bars represent the mean fold change in**

---

Nature Methods: doi:10.1038/nmeth.4534
endogenous mRNA expression levels (log$_2$) after plasmid transfection across three independent transfections (grey dots). d, e, qPCR-based assessment of mRNA induction of endogenous ISGs in HeLa-S3 cells electroporated with different plasmid types (d) and by different means of transfection (e). In E, RNA extraction was performed at two different time points to account for the different kinetics of DNA delivery between electroporation and chemical transfection. In all cases, mean fold change (colored bars) in mRNA expression levels (log$_2$) after transfection is shown across three independent transfections (grey dots).
Supplementary Figure 3

TBK1/IKK/PKR inhibition improves false-positive and false-negative errors

**a.** The five most significant disease ontology terms and their enrichments reported by GREAT²³ for the top 500 accessible peaks in STARR-seq screens without and with TBK1/IKK/PKR-inhibitors. Shown are log_{10} transformed FDR-adjusted P-values (as reported by Nature Methods: doi:10.1038/nmeth.4534)
GREAT) and fold-enrichments (shades of purple). b, The 10 most significantly enriched GO terms for genes proximal to the top 1000 peaks in the HeLa-S3 STARR-seq screen with TBK1/IKK/PKR-inhibitors. Shown are log_{10} transformed FDR-adjusted P-values (one-sided Fisher's exact test) and fold-enrichments (shades of purple; compare to Figure 2c). c, The 10 most significantly enriched GO terms and their enrichments among genes proximal to enhancers that are at least 5-fold down-regulated upon TBK1/IKK/PKR-inhibitor treatment (FDR-adjusted P-value < 0.001). d, HeLa-S3 genome wide STARR-seq screens with TBK1/IKK/PKR inhibitors vs. without inhibitors. Differential peaks are highlighted as indicated at fold change cutoffs of 2 and 5 (adjusted P-value < 0.001). 39 peaks close to the TSS of IFN-I stimulated genes ('ISG'; based on GO annotation) are shown as black circles. e, Representative STARR-seq enhancer activity profiles for canonical ISGs (H3K27ac data from ENCODE, Supplementary Table 3). Note that their differential activity in screens without (red) and with (green) TBK1/IKK/PKR inhibition is a good indicator for their inducibility by IFN-I. The peak ranks are listed below each strong enhancer, indicating that these regions are among the most highly active in STARR-seq screens performed without inhibitors. f, g, Bar graphs and scatter plots of average signal-over-background of STARR-seq screens in cells with or without TBK1/IKK/PKR-inhibitor treatment for HeLa-S3 cells (f) and HCT-116 cells (g). Shown are all peaks called in either screen, highlighted are predicted enhancers (green, n=39 for HeLa-S3; n=27 for HCT-116) and luciferase validated enhancers (red; n=4). Bars represent mean signal, error bars indicate 75% confidence intervals, P-values as stated (two-sided paired t-test). h, Percent recovery of peak calls in focused STARR-seq screens in HeLa-S3 cells employing the ORI or SCP1 as core-promoters as a fraction of peak calls in the ORI setup with inhibitor treatment. i, Number of peak calls at the indicated enrichment cutoffs in focused STARR-seq screens in HeLa-S3 cells employing the ORI or SCP1 as core-promoters.
**Supplementary Figure 4**

**STARR-seq enhancers are mostly intergenic or intronic and are enriched in enhancer-like chromatin states**

**a.** Average percentages of genomic annotations for the human genome (top) and STARR-seq peaks (bottom). CDS = coding sequence, 3' / 5' UTR = 3' / 5' untranslated region, n.c. exon = non-coding exon, upstream = -2 kb from the TSS. **b.** Enrichment of different ChromHMM states within non-overlapping bins of 500 STARR-seq enhancers (ranked by corrected fold-enrichment, left) and
shifted control regions (+50kb, right). c, Normalized enrichment scores for different HeLa-S3 ChIP-seq datasets (NES, i-cisTarget\textsuperscript{(i)}) for chromHMM strong enhancers (‘Enh’) without STARR-seq support and open STARR-seq enhancers that do not overlap chromHMM strong enhancers (‘Enh’) and the respective fold-differences (right, log\textsubscript{2}). d, Odds of motif occurrence (odds > 8 in any condition; FDR-adjusted P value < 10\textsuperscript{-5}, two-sided Fisher’s exact test) in proximal, distal intergenic or distal intronic STARR-seq enhancers over random regions. e, Average ENCODE ChIP-seq signal of ELK1, ELK4, JUN and FOS for proximal, distal intergenic or distal intronic STARR-seq enhancers compared to random regions (grey). f, Average ENCODE ChIP-seq signal (read coverage) of H3K4me1, H3K4me3 and H3K27ac for proximal, distal intergenic or distal intronic STARR-seq enhancers compared to random regions (grey).
Supplementary Figure 5

ERV elements are enriched within endogenously silenced enhancers

**a**, Cumulative percentage of STARR-seq peaks explained as DNase-I-hypersensitive sites in ENCODE cell lines (blue; grey indicates chance expectation based on random regions). From left to right, each additional dataset is intersected with regions not yet explained by the previous dataset (HeLa-S3 first, followed by additional cell types in alphabetical order). **b**, Odds ratios (FDR-adjusted P-value < 0.05, two-sided Fisher’s exact test) of indicated transposable elements in STARR-seq enhancers inaccessible in HeLa-S3 cells (DHS, P-value < 0.05, binomial test) vs. 1x10^6 random control regions.