Supplementary Information

Manuscript Title

A modular synthetic promoter toolkit for programmable gene control in activated and chronically stimulated T cells

Authors

Giuliano Bonfá[1,*], Giovanna Martino[1,2,*], Assunta Sellitto[1], Juan Luis Blazquez-Roman[1], Fabiana Tedeschi[1], Antonio Rinaldi[1], Caio Raony Farina Silveira[3], Francesca Piro[1], Fabio Caliendo[4], Loris Melchiorri[1], Daniela Perna[1], Rita Tarnai[1], Christoph Heuser-Loy[3,5], Fabio Mastrogiovanni[3], Timea Vadász[3], Freddie Starkey[6], Evangelos Nikolados[6], Filippo Menolascina[6], Luca Gattinoni[3,7] and Velia Siciliano[1,#]

[1] Center for Advanced Biomaterials for Healthcare, Istituto Italiano di Tecnologia - IIT, Largo Barsanti e Matteucci, 53, Naples, 80125, Italy.

[2] Department of Electrical Engineering and Information Technology (DIETI), University of Naples Federico II, Naples, 80131, Italy.

[3] Division of Functional Immune Cell Modulation, Leibniz Institute for Immunotherapy, Regensburg, Germany.

[4] Synthetic Biology Center, Department of Biological Engineering, Massachusetts Institute of Technology - MIT, 500 Technology Square, Cambridge-MA, 02139, USA.

[5] University Hospital Bonn, Medical Clinic and Polyclinic III, Internal Medicine with a focus on Oncology, Hematology, and Rheumatology, Bonn, Germany.

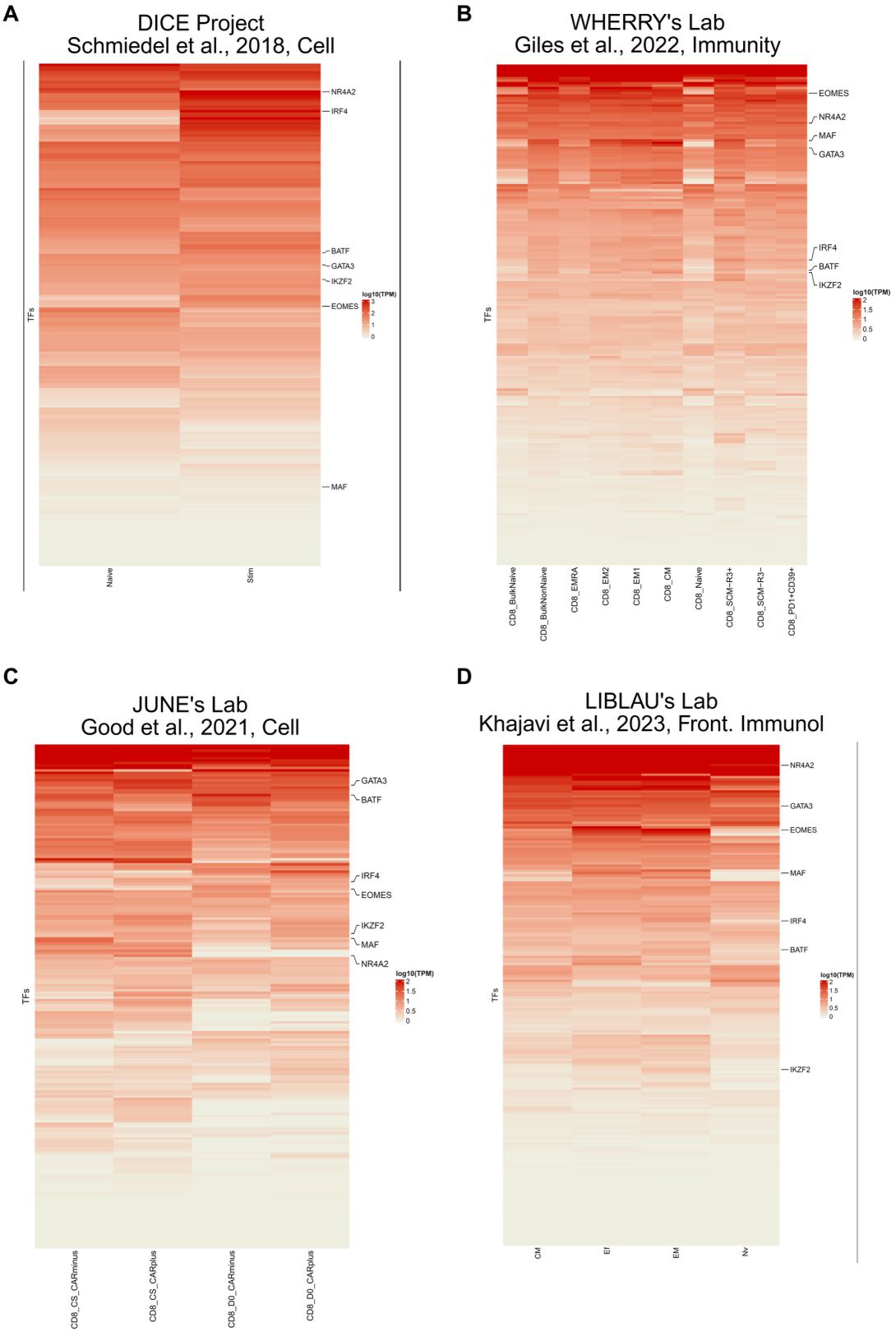
[6] Institute for Bioengineering, School of Engineering, The University of Edinburgh, Edinburgh, EH9 3BF, United Kingdom.

[7] University of Regensburg, Regensburg, Germany.

* Equal contribution.

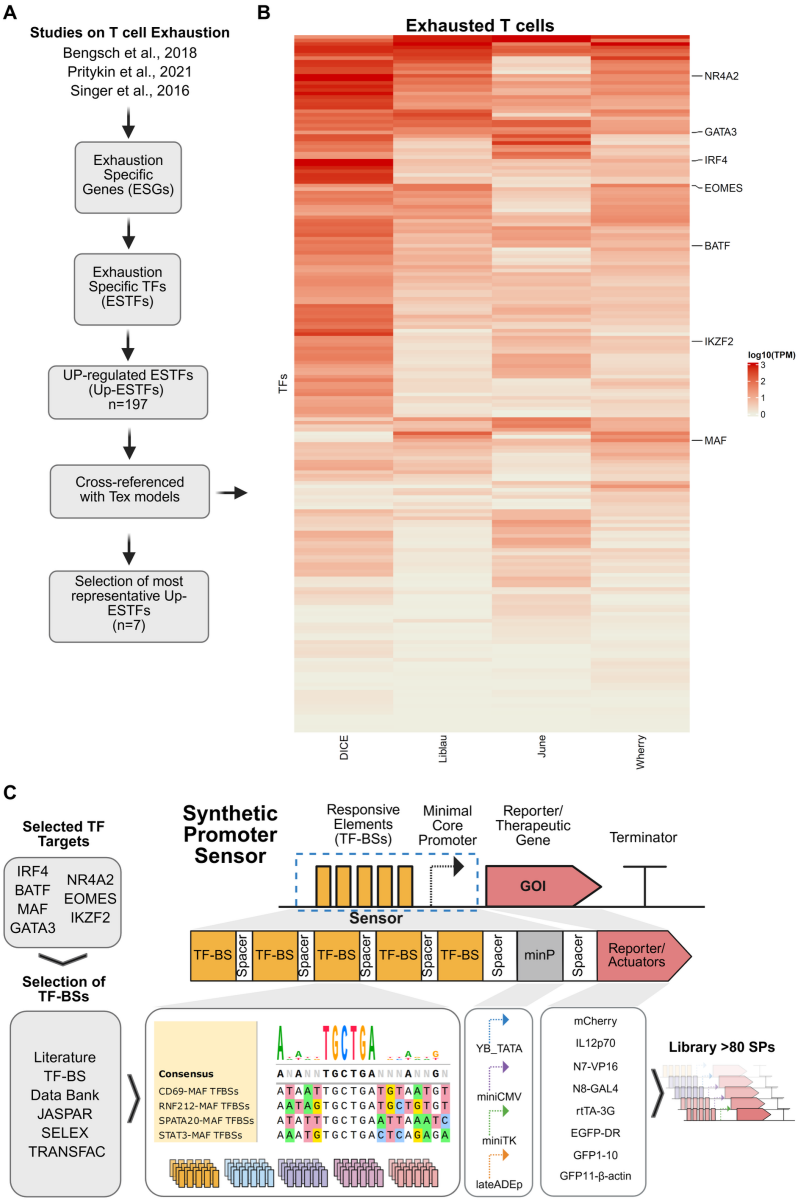
[#] To whom correspondence should be addressed: Velia Siciliano, Email: velia.siciliano@iit.it.

Supplementary Figures



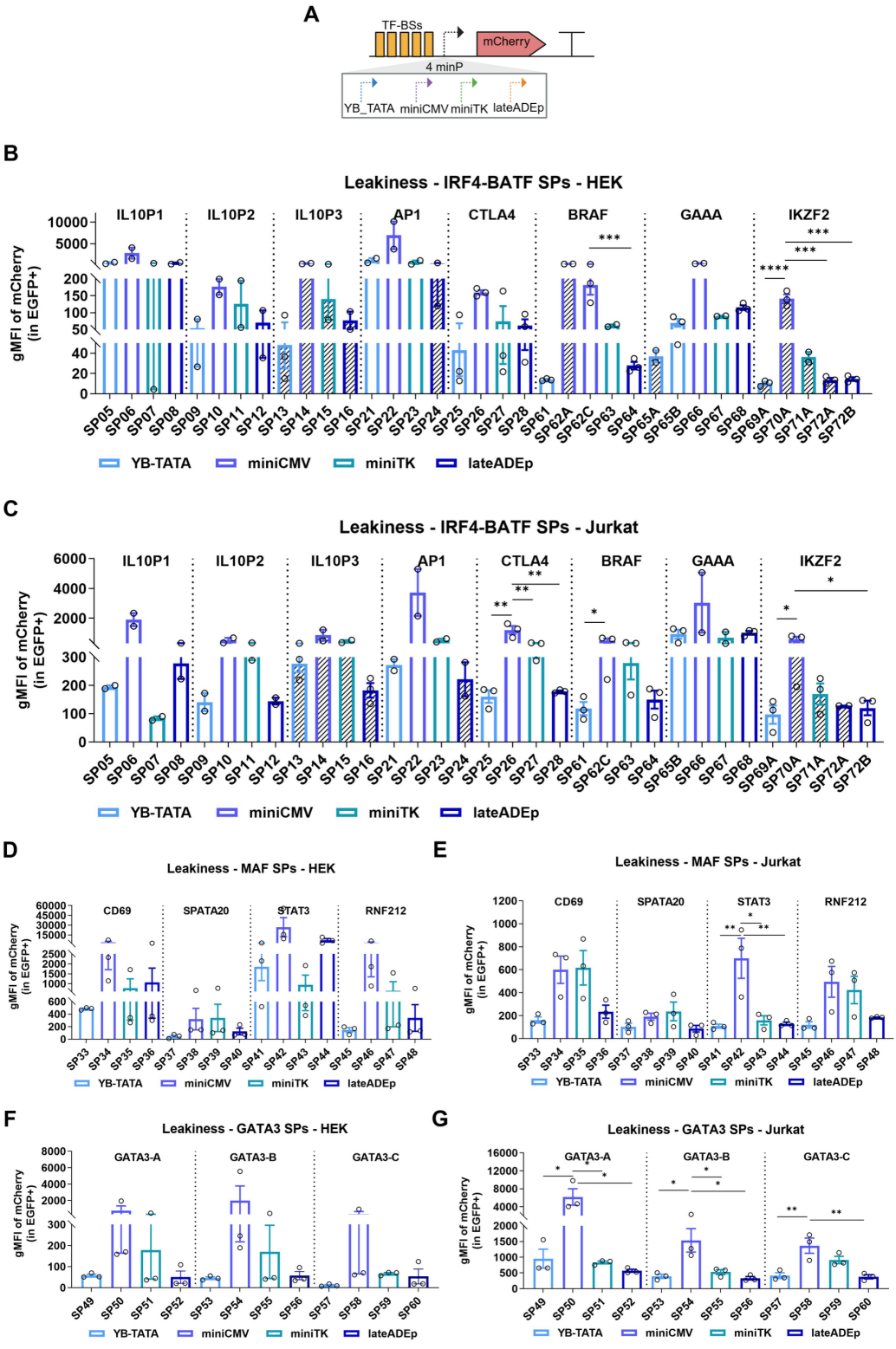
Supplementary Figure 1. Bioinformatics selection of transcription factors (TFs) differentially expressed in activated and exhausted T cells. (**A-D**) Heatmaps showing the ranking of TFs expression levels across different studies (expression values are reported

in Supplementary Data 1).



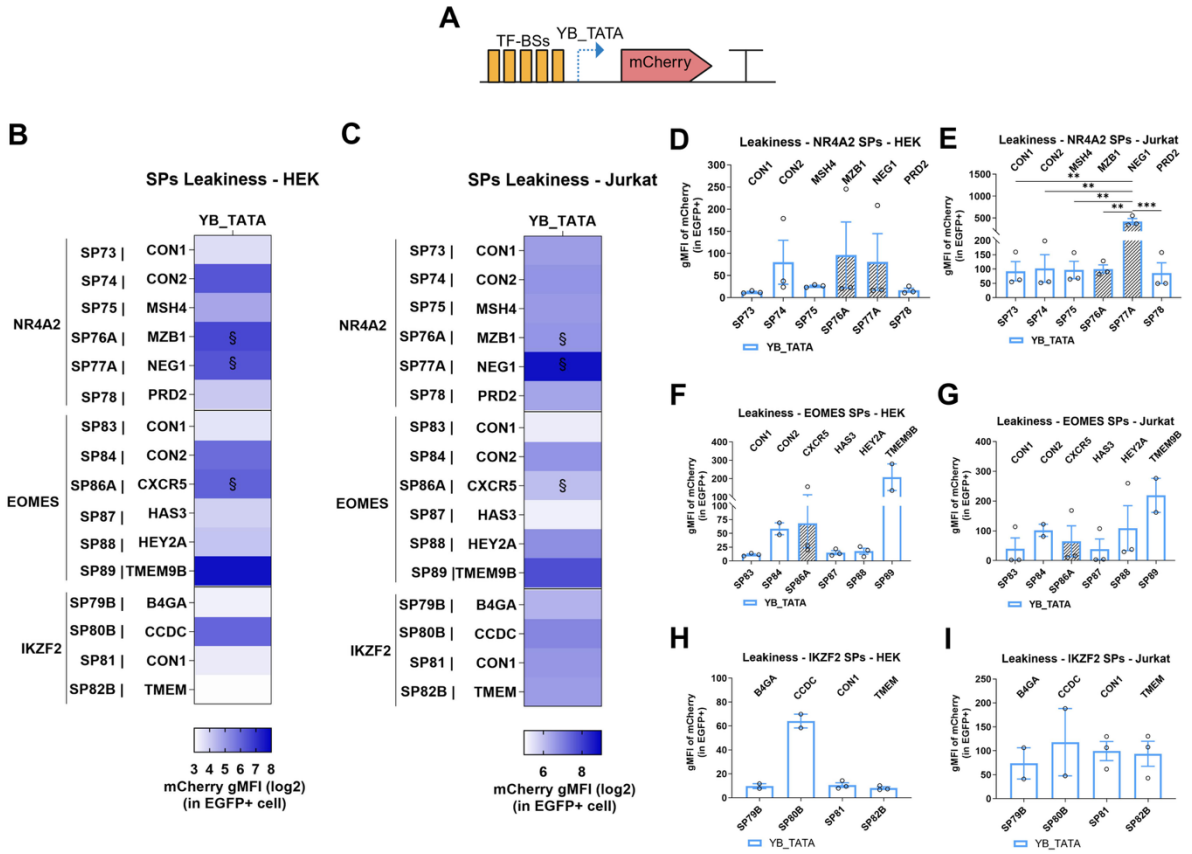
Supplementary Figure 2. Bioinformatics-guided selection of transcription factors and modular design of TF-responsive synthetic promoters. (**A**) Workflow used to identify transcription factors (TFs) associated with T cell states by selecting an Upregulated Exhaustion-Specific TF (Up-ESTF) subset from the broader set of Exhaustion-Specific Genes (ESGs) using previously published transcriptomic datasets. (**B**) Heatmap showing median expression values ($\log_{10}$ TPM, transcripts per million) of 197 TFs in activated and exhausted T cells. Genes were clustered using Euclidean distance (expression values are reported in Supplementary Data 1). (**C**) General architecture of TF-responsive synthetic promoters (SPs) designed in this study. Each SP consists of a TF-responsive region containing five tandem repeats of TF binding sites (TF-BSs), separated by defined spacer sequences (10–20 bp), positioned upstream of a minimal core promoter (minP) with an additional spacer (20 bp). TF-BSs were selected from literature and curated TF-binding site repositories (e.g., JASPAR, SELEX, TRANSFAC, TFBS-DB http://tfbsdb.systemsbiology.net/). Four minimal promoters (YB_TATA, miniCMV, miniTK, and lateADEp) were used to drive expression of fluorescent reporters (mCherry, EGFP-DR,

split EGFP), components for logic-gate circuits (rtTA3G, N7-VP16, N8-GAL4), or therapeutic payload IL12p70.



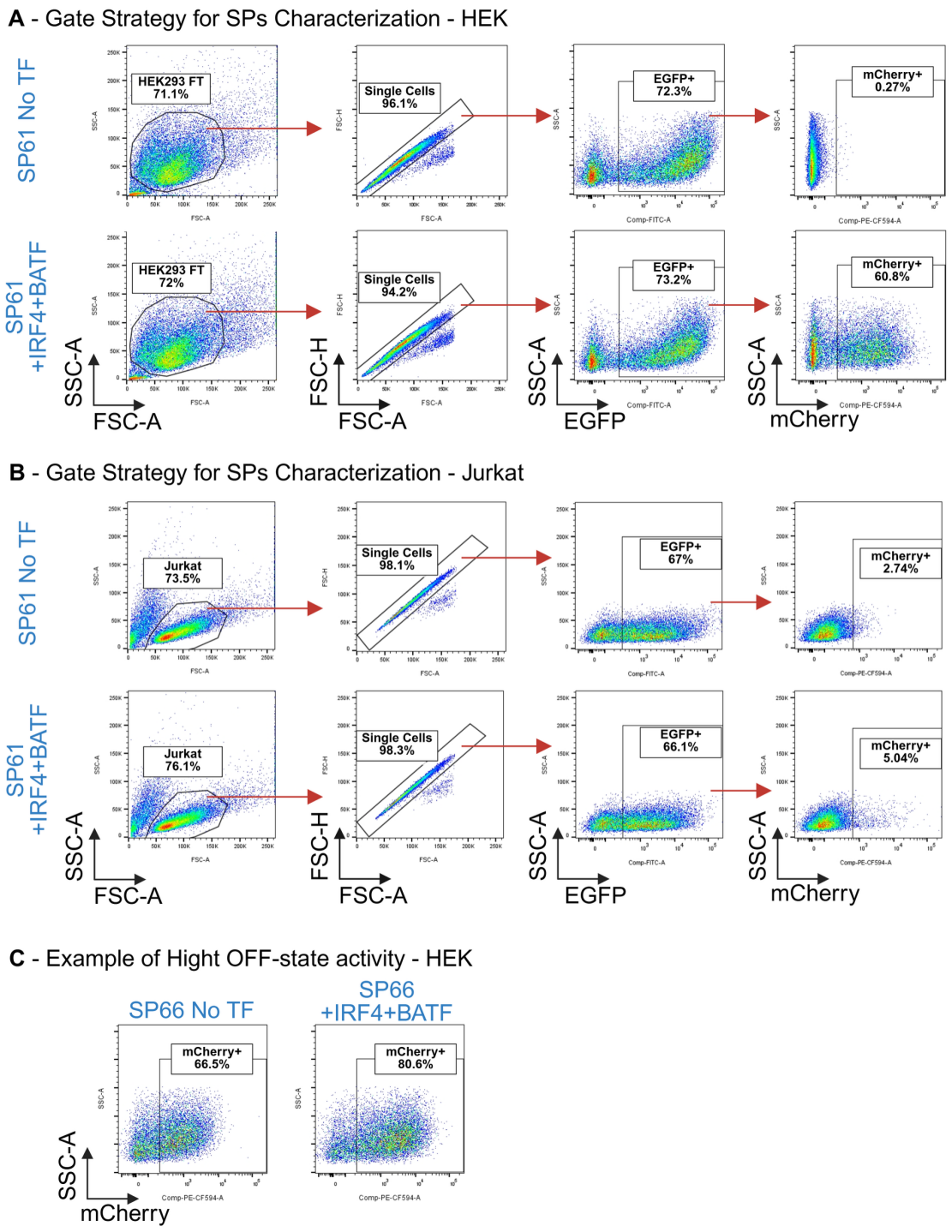
Supplementary Figure 3. Minimal promoter choice influences OFF-state basal activity of TF-responsive synthetic promoters. (**A**) Schematic representation of different minimal promoters (minP) combined with transcription factor binding sites (TF-BSs) to generate synthetic promoters. (**B-C**) OFF-state basal activity ("leakiness") of synthetic promoters

containing IRF4/BATF TF-BSs combined with different minimal promoters, quantified as mCherry geometric mean fluorescence intensity (gMFI) 48 h post-transfection or electroporation in HEK293 and Jurkat cells. (**D-G**) OFF-state basal activity of synthetic promoters containing MAF or GATA3 TF-BSs combined with different minimal promoters, quantified as mCherry gMFI 48 h post-transfection or electroporation in HEK293 and Jurkat cells. Across all TF-BS combinations and cell types, higher basal activity is consistently associated with the miniCMV promoter. Data are represented as mean ± SEM of *n* = 3 independent biological replicates. Statistical test: one-way ANOVA with Tukey's multiple comparison test. *P* value: *<0.05; **<0.005. Source data are provided as a Source Data file.



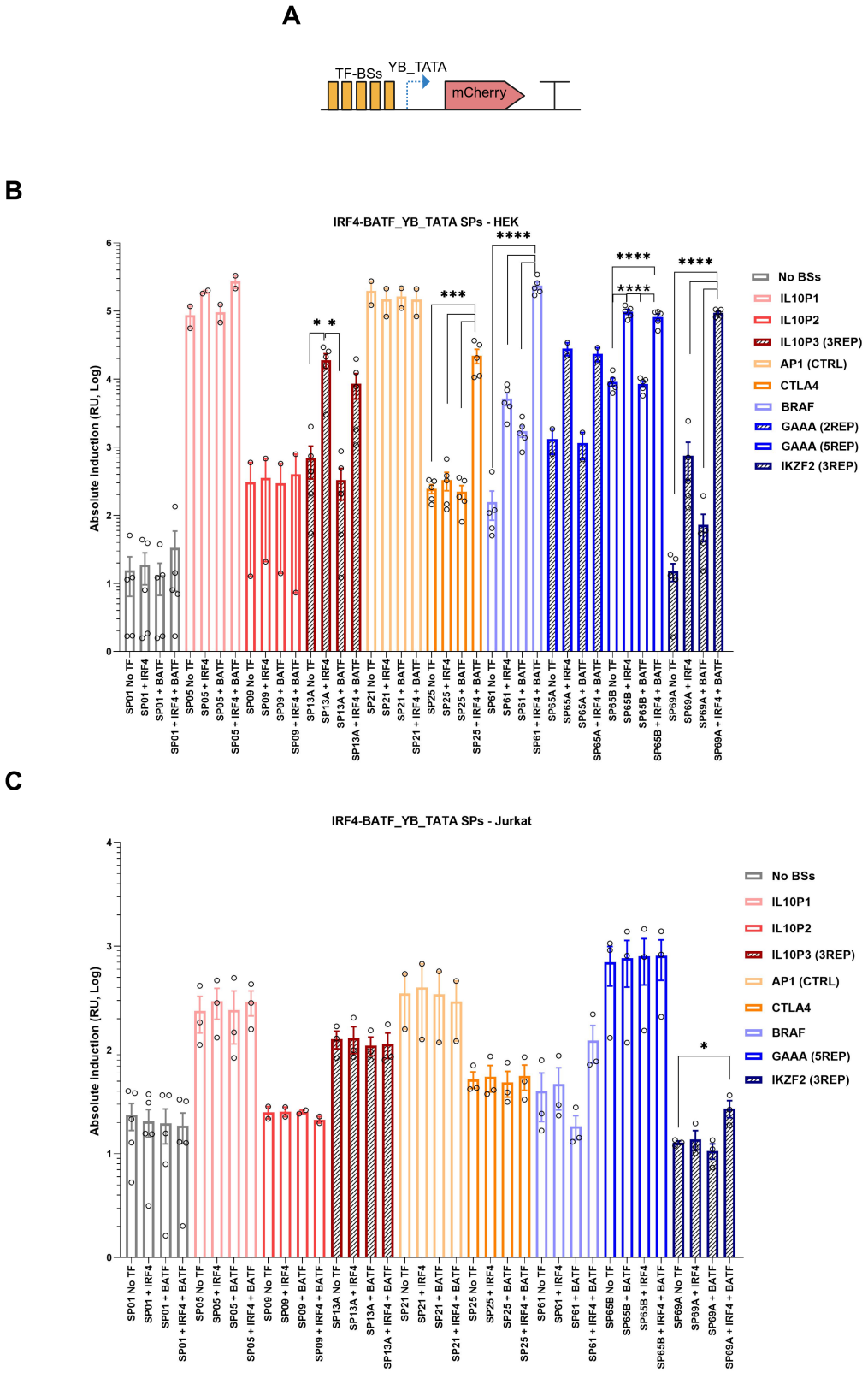
Supplementary Figure 4. YB_TATA-based synthetic promoters exhibit low OFF-state basal activity across multiple transcription factors.
(**A**) Schematic representation of synthetic promoters composed of the YB_TATA minimal promoter combined with different transcription factor binding sites (TF-BSs). (**B–C**) Heatmaps showing OFF-state basal activity ("leakiness") of YB_TATA-based synthetic promoters containing TF-BSs for NR4A2, EOMES, and IKZF2, quantified as mCherry geometric mean fluorescence intensity (gMFI) 48 h post-transfection or electroporation in HEK293 and Jurkat cells. (**D–I**) Bar plots showing mCherry gMFI for YB_TATA-based synthetic promoters containing TF-BSs for NR4A2, EOMES, and IKZF2, measured 48 h post-transfection or electroporation in HEK293 and Jurkat cells.Hatched bars indicate synthetic promoters containing fewer than five TF-BS repeats. Data are represented as mean ± SEM of *n* = 2-3 independent biological replicates. Statistical test: one-way ANOVA with Tukey's multiple comparison test. *P* value: **<0.005; ***<0.0005. Source data are

provided as a Source Data file.



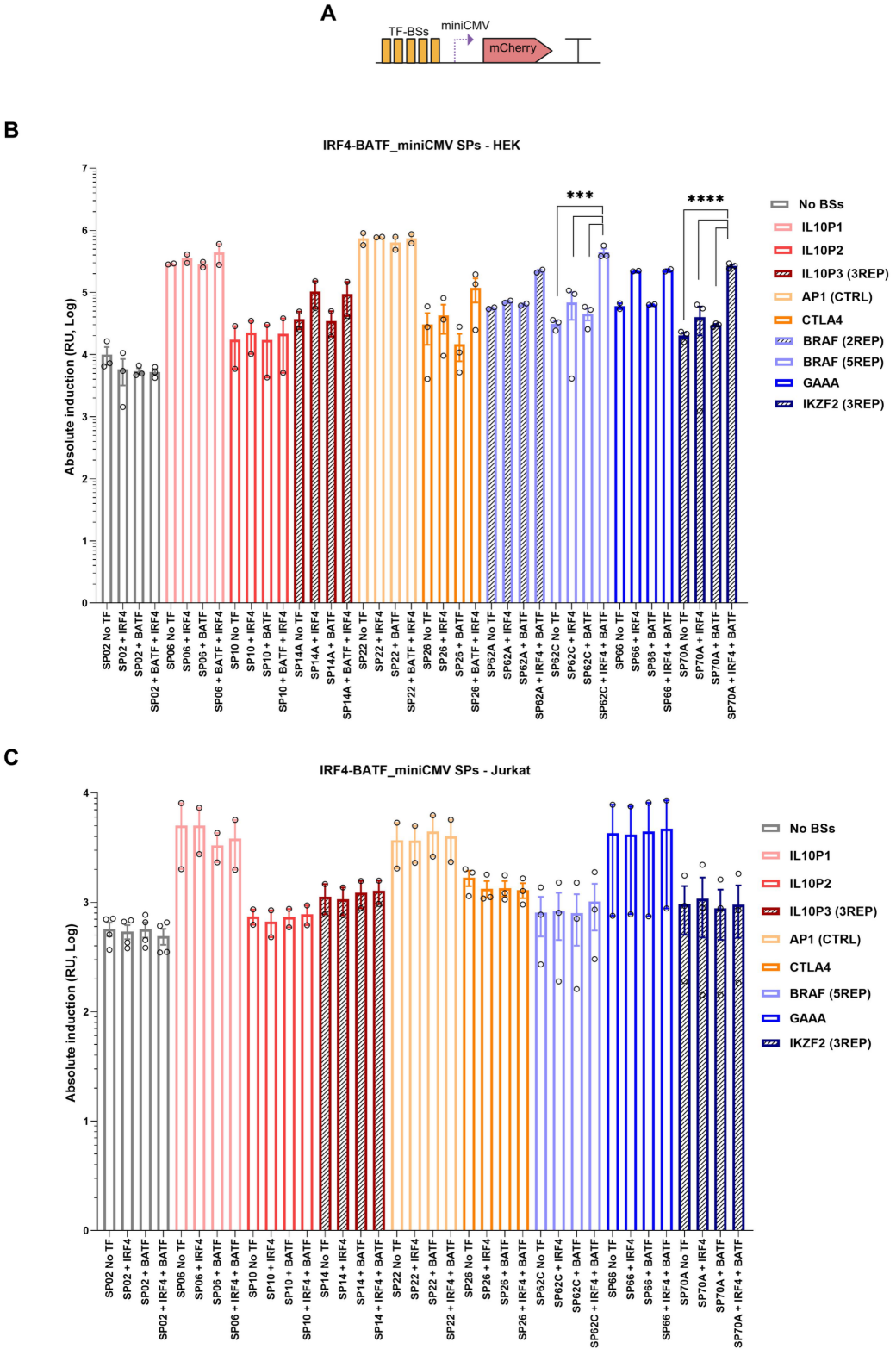
Supplementary Figure 5. Flow cytometry gating strategy and representative example of high OFF-state activity during synthetic promoter screening. (**A-B**) Gating hierarchy applied to transfected HEK293 and electroporated Jurkat cells. In both cases, analysis was performed by sequentially selecting live cells followed by singlets. (**C**) Representative example of elevated OFF-state basal activity observed for a subset of synthetic promoters during screening in HEK293 cells. Reporter expression driven by synthetic promoters

(mCherry) was quantified within the EGFP-positive population, which served as a transfection or electroporation marker. OFF-state basal activity was assessed by measuring mCherry expression in the absence of the cognate transcription factor (SPx, no TF).



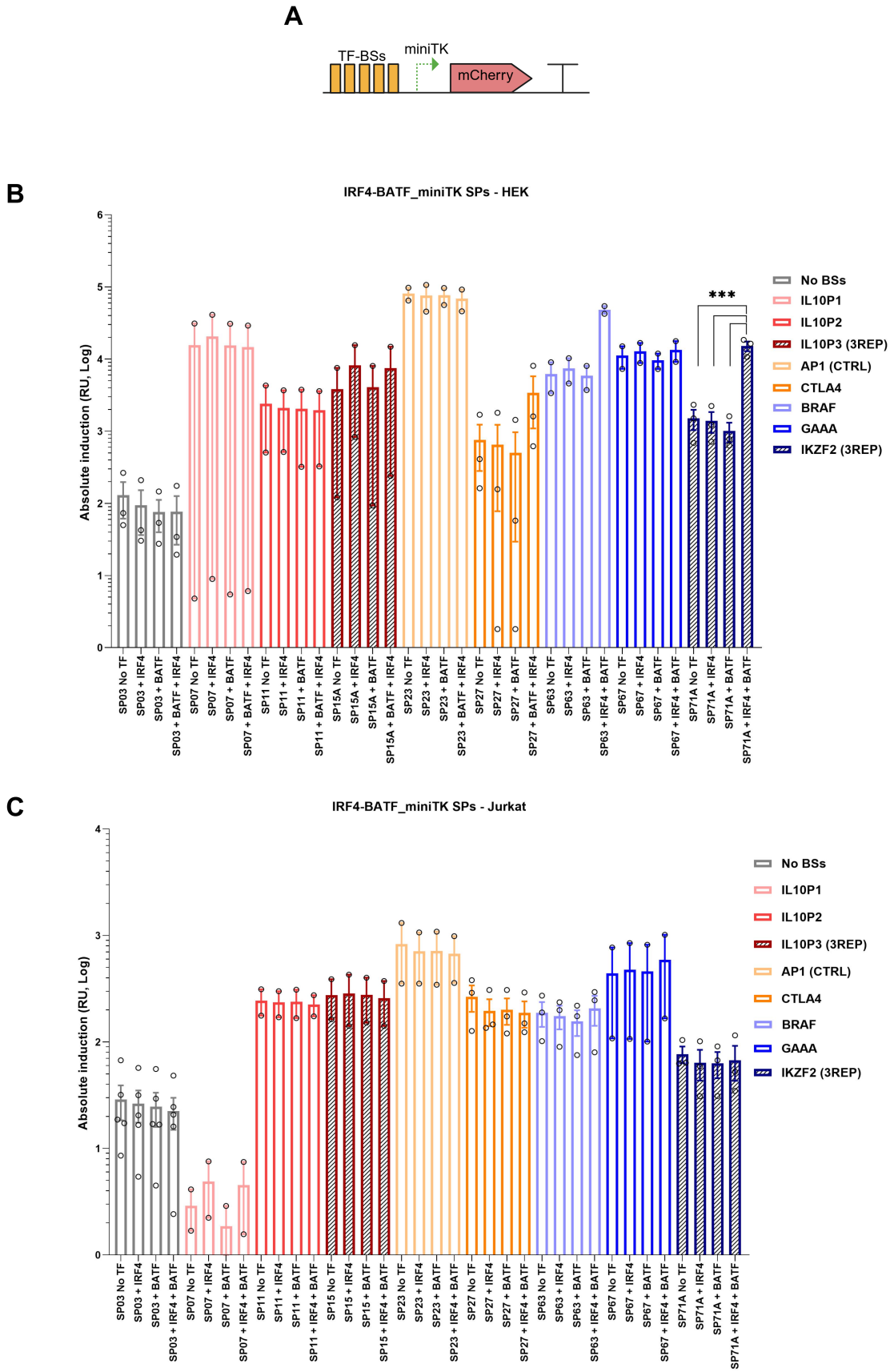
Supplementary Figure 6. IRF4/BATF-responsive synthetic promoters show inducible

activity across promoter architectures. (**A**) Design of different IRF4/BATF TF-BS combined to YB_TATA minimal promoter. (**B-C**) Absolute induction of IRF4/BATF-responsive synthetic promoters in HEK293 and Jurkat cells, quantified following transfection or electroporation, respectively. Hatched bars indicate synthetic promoters containing fewer than five TF-BS repeats. Data are represented as mean ± SEM of *n* = 2-5 independent biological replicates. Statistical test: one-way ANOVA with Tukey's multiple comparison test. *P* value: *<0.05; **<0.005; ***<0.0005; ****<0.0001. Source data are provided as a Source Data file.



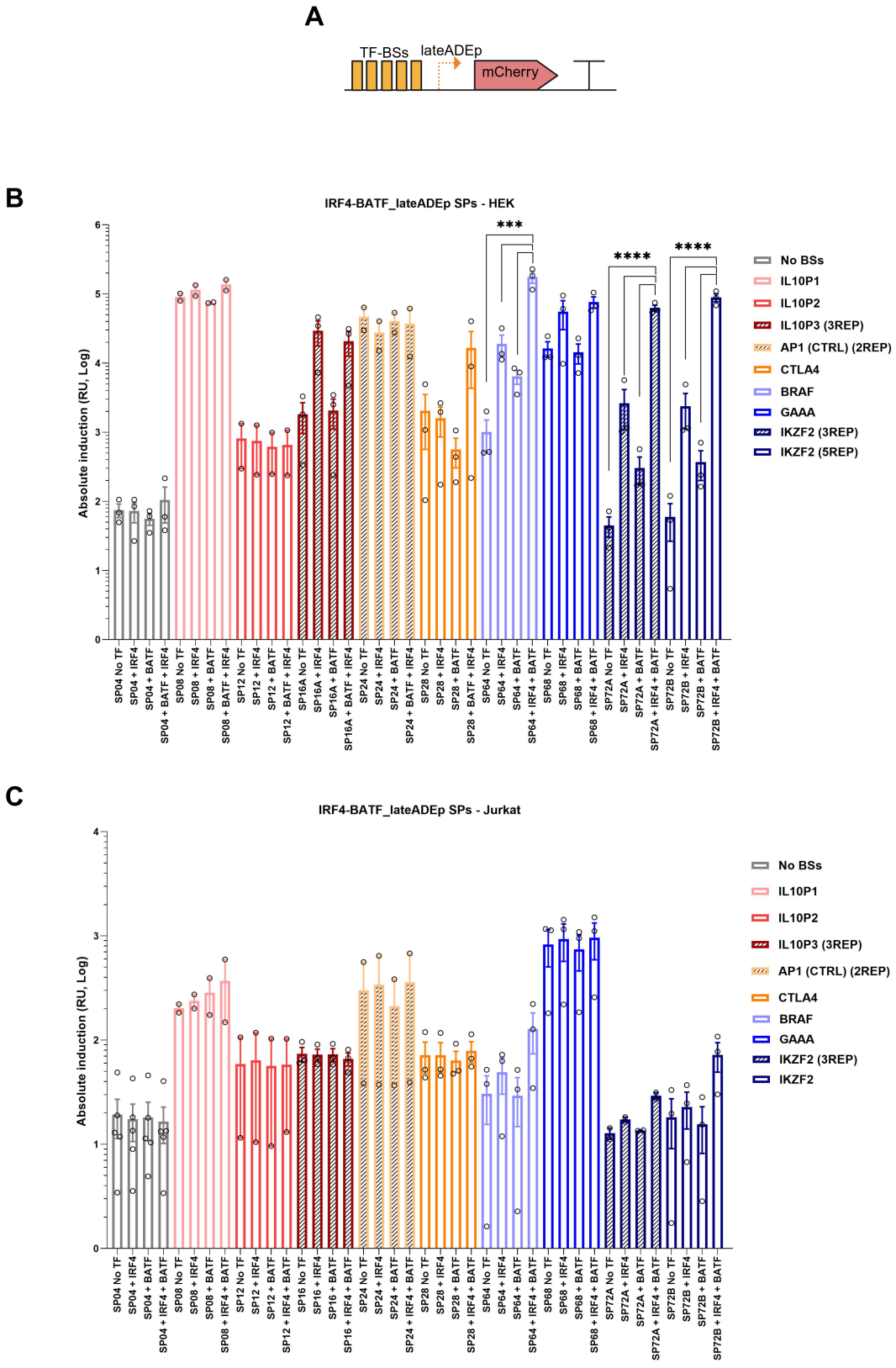
Supplementary Figure 7. IRF4/BATF-responsive synthetic promoters driven by miniCMV display inducible activity. (**A**) Schematic representation of synthetic promoters composed of the miniCMV minimal promoter combined with IRF4/BATF transcription factor binding sites (TF-BSs). (**B-C**) Absolute induction of IRF4/BATF-responsive synthetic promoters in HEK293 or Jurkat cells quantified following transfection or electroporation, respectively. Hatched bars indicate synthetic promoters containing fewer than five TF-BS

repeats. Data are represented as mean ± SEM of *n* = 2-3 independent biological replicates. Statistical analysis: one-way ANOVA with Tukey's multiple comparison test. *P* value: ***<0.0005; ****<0.0001. Source data are provided as a Source Data file.



Supplementary Figure 8. IRF4/BATF-responsive synthetic promoters driven by miniTK exhibit inducible activity. (**A**) Schematic representation of synthetic promoters composed

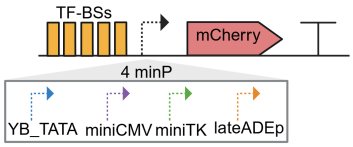
of the miniTK minimal promoter combined with IRF4/BATF transcription factor binding sites (TF-BSs). (**B-C**) Absolute induction of IRF4/BATF-responsive synthetic promoters in HEK293 and Jurkat cells, quantified following transfection or electroporation, respectively. Hatched bars indicate synthetic promoters containing fewer than five TF-BS repeats. Data are represented as mean ± SEM of *n* = 2-3 independent biological replicates. Statistical analysis: one-way ANOVA with Tukey's multiple comparison test. *P* value: ***<0.0005. Source data are provided as a Source Data file.



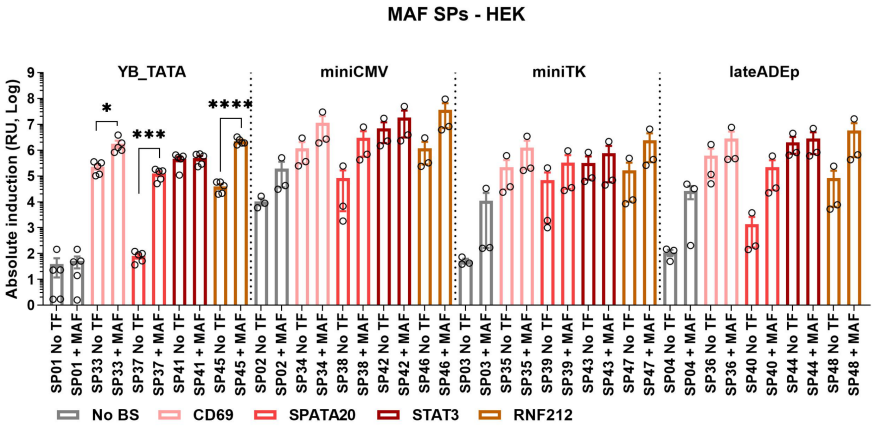
Supplementary Figure 9. IRF4/BATF-responsive synthetic promoters driven by lateADEp exhibit inducible activity. (**A**) Schematic representation of synthetic promoters composed of the lateADEp minimal promoter combined with IRF4/BATF transcription factor binding sites (TF-BSs). (**B-C**) Absolute induction of IRF4/BATF-responsive synthetic

promoters in HEK293 and Jurkat cells, quantified following transfection or electroporation, respectively. Hatched bars indicate synthetic promoters containing fewer than five TF-BS repeats. Data are represented as mean ± SEM of *n* = 2-3 independent biological replicates. Statistical analysis: one-way ANOVA with Tukey's multiple comparison test. *P* value: ***<0.0005; ****<0.0001. Source data are provided as a Source Data file.

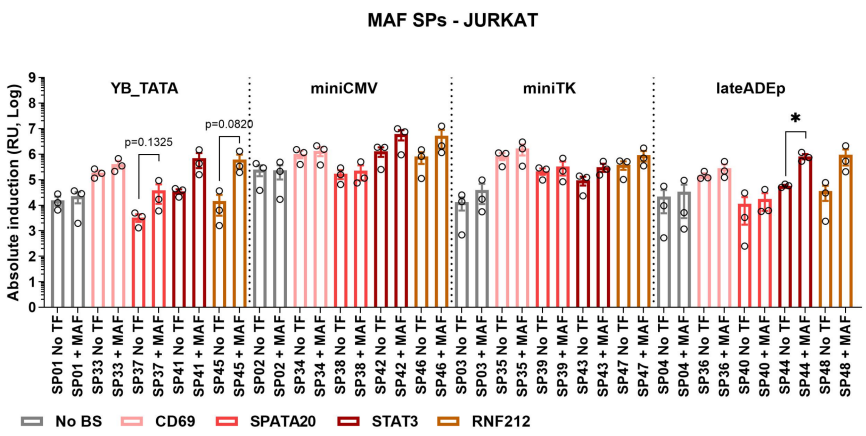
A


TF-BSs
mCherry
4 minP
YB_TATA
miniCMV
miniTK
lateADEp

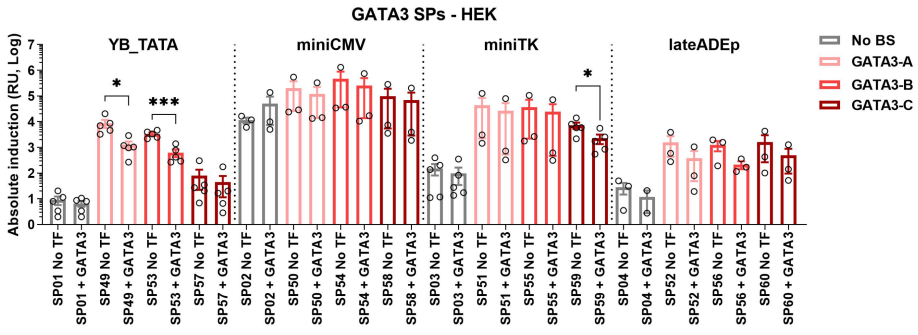
B


MAF SPs - HEK
YB_TATA
miniCMV
miniTK
lateADEp
Absolute induction (RU, Log)
No BS
CD69
SPATA20
STAT3
RNF212

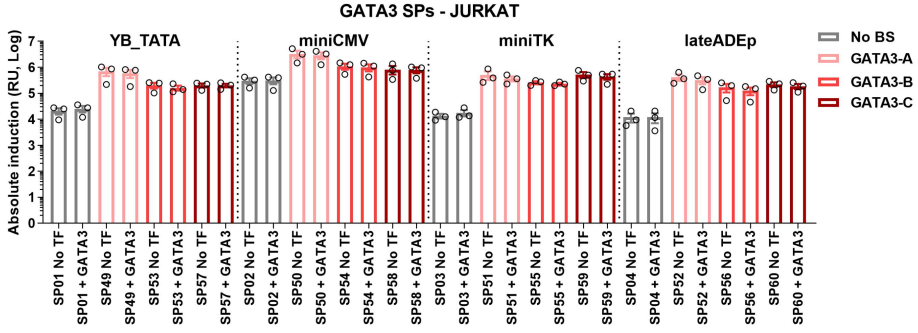
C


MAF SPs - JURKAT
YB_TATA
miniCMV
miniTK
lateADEp
p=0.1325
p=0.0820
Absolute induction (RU, Log)
No BS
CD69
SPATA20
STAT3
RNF212

D

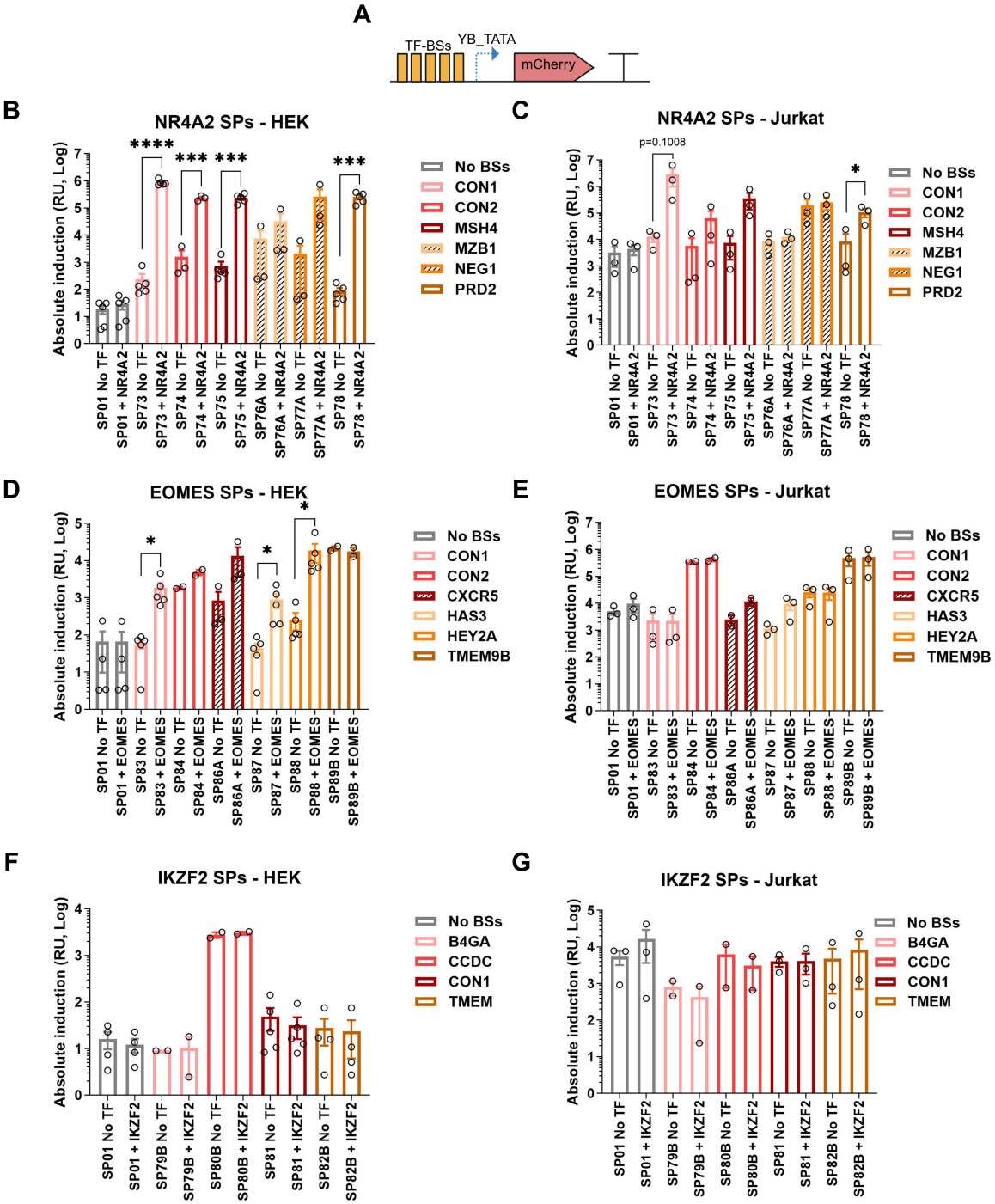

GATA3 SPs - HEK
YB_TATA
miniCMV
miniTK
lateADEp
Absolute induction (RU, Log)
No BS
GATA3-A
GATA3-B
GATA3-C

E


GATA3 SPs - JURKAT
YB_TATA
miniCMV
miniTK
lateADEp
Absolute induction (RU, Log)
No BS
GATA3-A
GATA3-B
GATA3-C

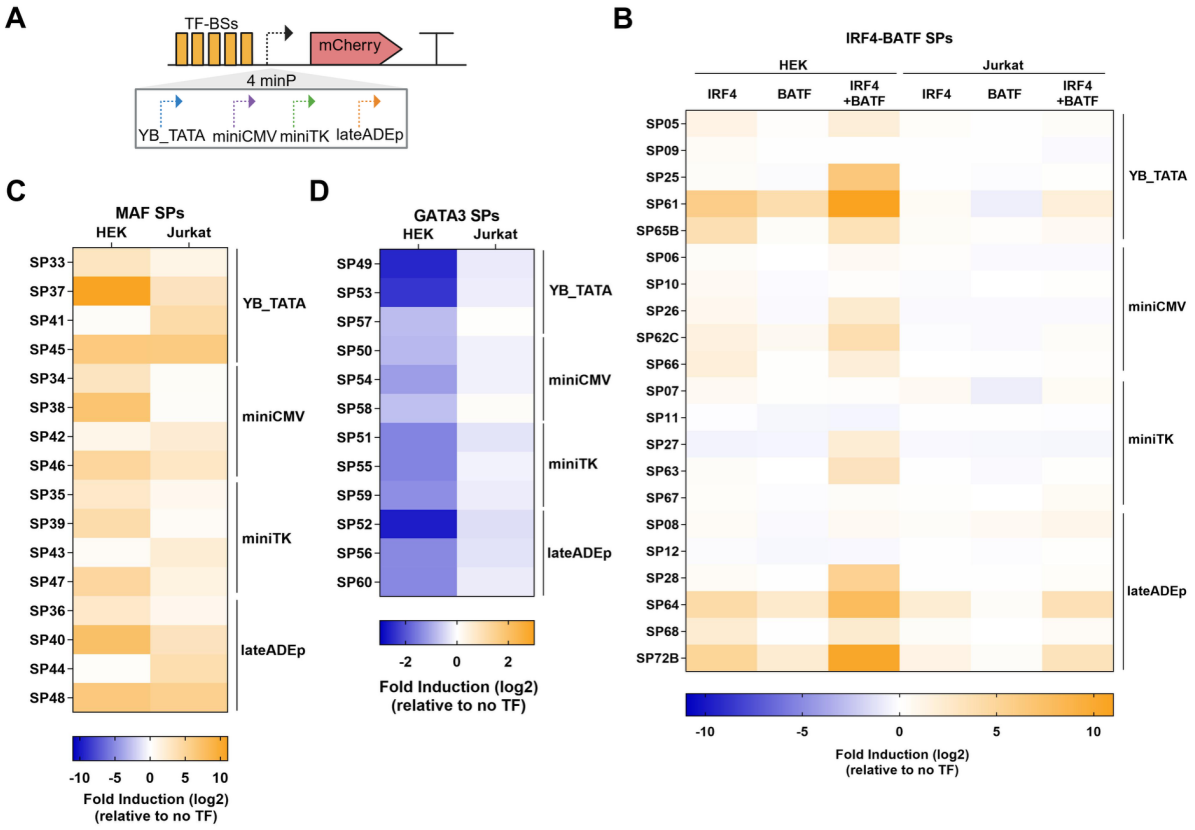
137

Supplementary Figure 10. MAF- and GATA3-responsive synthetic promoters exhibit inducible activity across minimal promoter configurations. (A) Schematic representation of synthetic promoters composed of different minimal promoters combined with MAF or GATA3 transcription factor binding sites (TF-BSs). (**B-C**) Absolute induction of MAF-responsive synthetic promoters across all minimal promoter configurations in HEK293 and Jurkat cells, quantified following transfection or electroporation. (**D–E**) Absolute induction of GATA3-responsive synthetic promoters across all minimal promoter configurations in HEK293 and Jurkat cells, quantified following transfection or electroporation. Hatched bars indicate synthetic promoters containing fewer than five TF-BS repeats. Data are represented as mean ± SEM of *n* = 2-3 independent biological replicates. Statistical analysis: unpaired two-tailed *t*-test. *P* value: *<0.05; ***<0.0005; ****<0.0001. Source data are provided as a Source Data file.

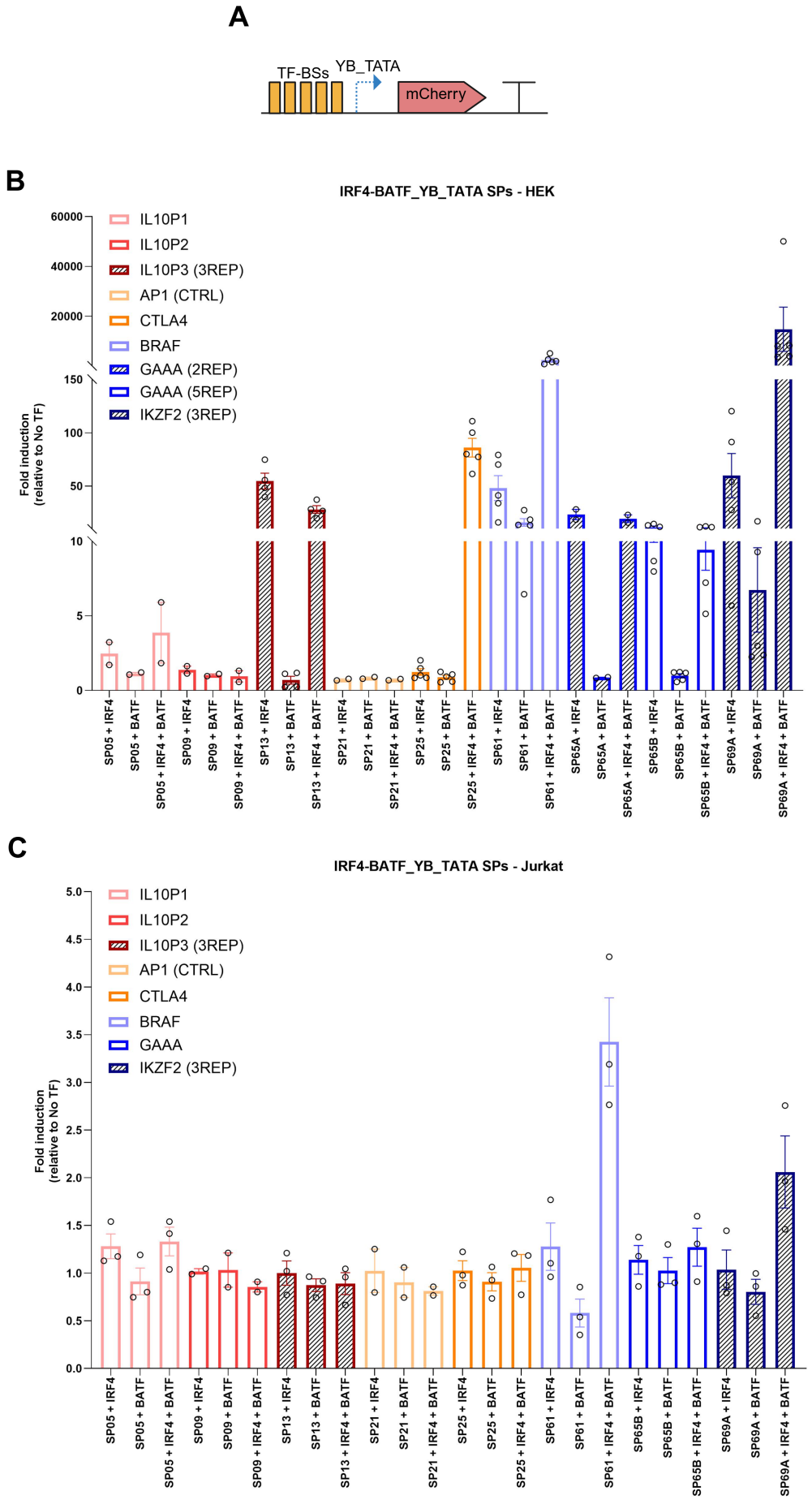


Supplementary Figure 11. NR4A2- and EOMES-responsive synthetic promoters exhibit inducible activity in YB_TATA configuration. (**A**) Schematic representation of synthetic promoters composed of the YB_TATA minimal promoter combined with transcription factor binding sites (TF-BSs) for NR4A2, EOMES, and IKZF2. (**B–C**) Absolute induction of NR4A2-responsive synthetic promoters in the YB_TATA configuration in HEK293 and Jurkat cells, quantified following transfection or electroporation, respectively. (**D–E**) Absolute induction of EOMES-responsive synthetic promoters in the YB_TATA configuration in HEK293 and Jurkat cells. (**F–G**) Absolute induction of IKZF2-responsive synthetic promoters in the YB_TATA configuration in HEK293 and Jurkat cells. Hatched bars indicate synthetic promoters containing fewer than five TF-BS repeats. Data are represented

as mean ± SEM of *n* = 2-3 independent biological replicates. Statistical analysis: one-tailed unpaired *t*-test. *P* value: *<0.05; ***<0.0005; ****<0.0001. Source data are provided as a Source Data file.

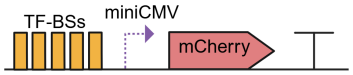


Supplementary Figure 12. Fold-induction analysis reveals differential activation of synthetic promoters across transcription factors and minimal promoters. (**A**) Schematic representation of synthetic promoters composed of different minimal promoters combined with transcription factor binding sites (TF-BSs) for IRF4/BATF, MAF, and GATA3. (**B–D**) Heatmaps showing fold induction of synthetic promoters across different TF-BSs and minimal promoter configurations for IRF4/BATF (B), MAF (C), and GATA3 (D). Fold induction was calculated from normalized conditions using mCherry and EGFP geometric mean fluorescence intensities and the percentage of positive cells to derive absolute induction values (see Methods), which were then referenced to the corresponding no-TF condition. Data are represented as mean ± SEM of *n* = 2-3 independent biological replicates. Source data are provided as a Source Data file.

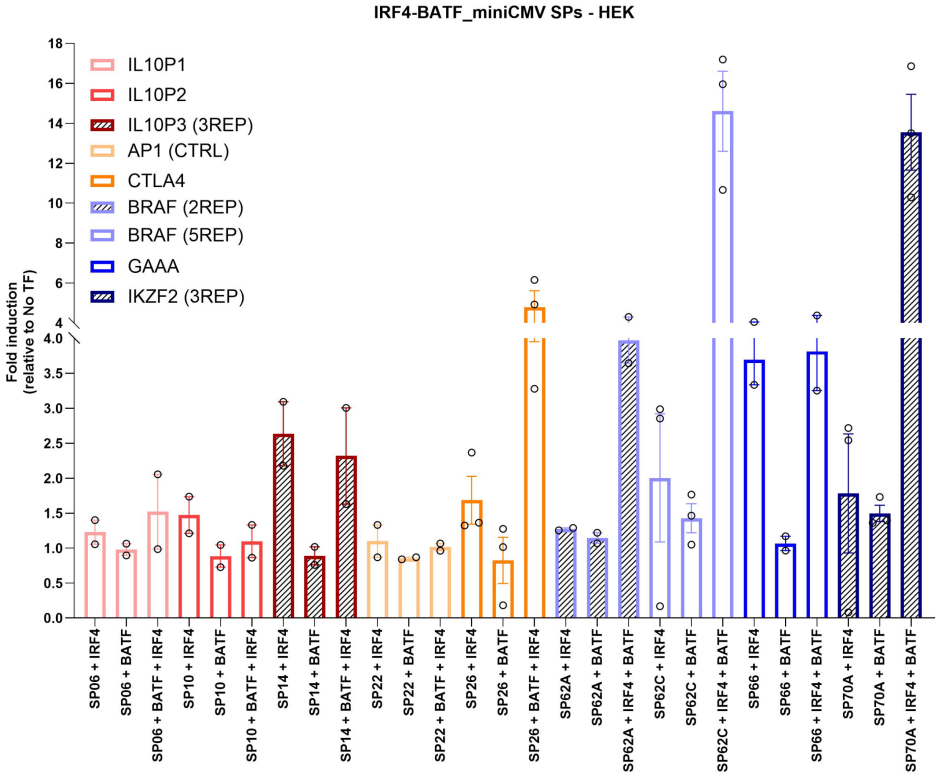

A
TF-BSs
YB_TATA
mCherry
B
IRF4-BATF_YB_TATA SPs - HEK
IL10P1
IL10P2
IL10P3 (3REP)
AP1 (CTRL)
CTLA4
BRAF
GAAA (2REP)
GAAA (5REP)
IKZF2 (3REP)
Fold induction (relative to No TF)
SP05 + IRF4
SP05 + BATF
SP05 + IRF4 + BATF
SP09 + IRF4
SP09 + BATF
SP09 + IRF4 + BATF
SP13 + IRF4
SP13 + BATF
SP13 + IRF4 + BATF
SP21 + IRF4
SP21 + BATF
SP21 + IRF4 + BATF
SP25 + IRF4
SP25 + BATF
SP25 + IRF4 + BATF
SP61 + IRF4
SP61 + BATF
SP61 + IRF4 + BATF
SP65A + IRF4
SP65A + BATF
SP65A + IRF4 + BATF
SP65B + IRF4
SP65B + BATF
SP65B + IRF4 + BATF
SP69A + IRF4
SP69A + BATF
SP69A + IRF4 + BATF
C
IRF4-BATF_YB_TATA SPs - Jurkat
IL10P1
IL10P2
IL10P3 (3REP)
AP1 (CTRL)
CTLA4
BRAF
GAAA
IKZF2 (3REP)
Fold induction (relative to No TF)

Supplementary Figure 13. Fold-induction analysis of IRF4/BATF-responsive synthetic promoters in the YB_TATA configuration. (**A**) Schematic representation of synthetic promoters composed of the YB_TATA minimal promoter combined with IRF4/BATF transcription factor binding sites (TF-BSs). (**B–C**) Fold induction of IRF4/BATF-responsive synthetic promoters in the YB_TATA configuration in HEK293 and Jurkat cells, calculated relative to the corresponding no-TF condition. Hatched bars indicate synthetic promoters containing fewer than five TF-BS repeats. Data are represented as mean ± SEM of *n* = 2-3 independent biological replicates. Source data are provided as a Source Data file.

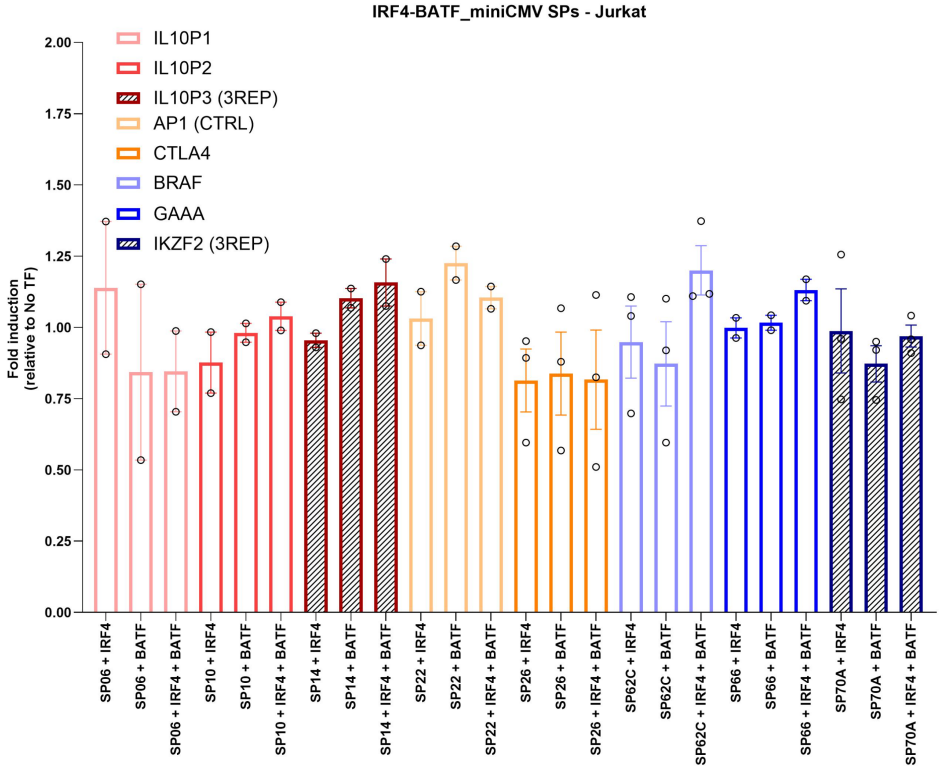
A


TF-BSs
miniCMV
mCherry

B

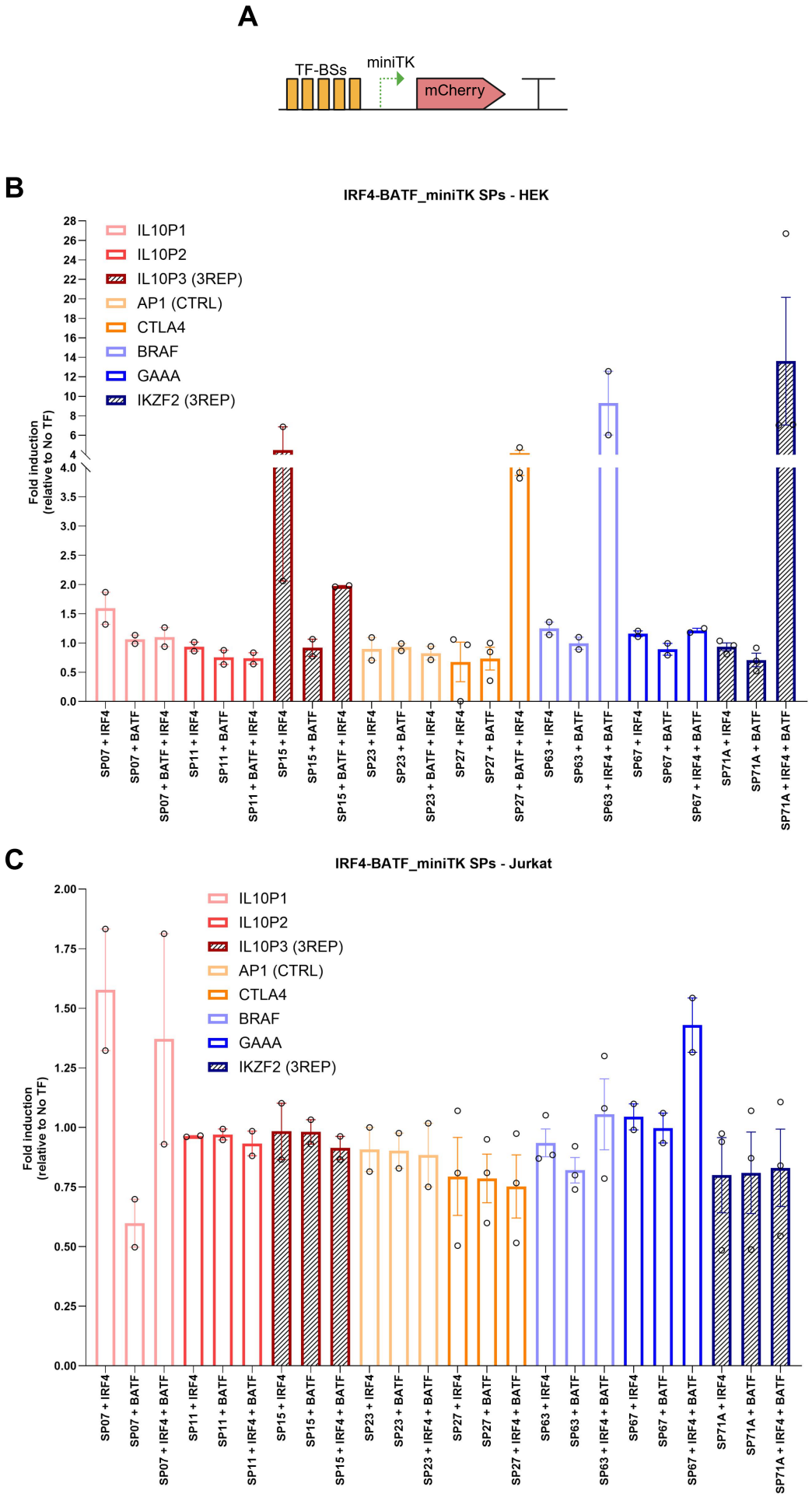

IRF4-BATF_miniCMV SPs - HEK
Fold induction (relative to No TF)
IL10P1
IL10P2
IL10P3 (3REP)
AP1 (CTRL)
CTLA4
BRAF (2REP)
BRAF (5REP)
GAAA
IKZF2 (3REP)
SP06 + IRF4
SP06 + BATF
SP06 + BATF + IRF4
SP10 + IRF4
SP10 + BATF
SP10 + BATF + IRF4
SP14 + IRF4
SP14 + BATF
SP14 + BATF + IRF4
SP22 + IRF4
SP22 + BATF
SP22 + BATF + IRF4
SP26 + IRF4
SP26 + BATF
SP26 + BATF + IRF4
SP62A + IRF4
SP62A + BATF
SP62A + IRF4 + BATF
SP62C + IRF4
SP62C + BATF
SP62C + IRF4 + BATF
SP66 + IRF4
SP66 + BATF
SP66 + IRF4 + BATF
SP70A + IRF4
SP70A + BATF
SP70A + IRF4 + BATF

C


IRF4-BATF_miniCMV SPs - Jurkat
Fold induction (relative to No TF)
IL10P1
IL10P2
IL10P3 (3REP)
AP1 (CTRL)
CTLA4
BRAF
GAAA
IKZF2 (3REP)
SP06 + IRF4
SP06 + BATF
SP06 + IRF4 + BATF
SP10 + IRF4
SP10 + BATF
SP10 + IRF4 + BATF
SP14 + IRF4
SP14 + BATF
SP14 + IRF4 + BATF
SP22 + IRF4
SP22 + BATF
SP22 + IRF4 + BATF
SP26 + IRF4
SP26 + BATF
SP26 + IRF4 + BATF
SP62C + IRF4
SP62C + BATF
SP62C + IRF4 + BATF
SP66 + IRF4
SP66 + BATF
SP66 + IRF4 + BATF
SP70A + IRF4
SP70A + BATF
SP70A + IRF4 + BATF

188

Supplementary Figure 14. Fold-induction analysis of IRF4/BATF-responsive synthetic promoters in the miniCMV configuration. (**A**) Schematic representation of synthetic promoters composed of the miniCMV minimal promoter combined with IRF4/BATF transcription factor binding sites (TF-BSs). (**B–C**) Fold induction of IRF4/BATF-responsive synthetic promoters in the miniCMV configuration in HEK293 and Jurkat cells, calculated relative to the corresponding no-TF condition. Hatched bars indicate synthetic promoters containing fewer than five TF-BS repeats. Data are represented as mean ± SEM of *n* = 2-3 independent biological replicates. Source data are provided as a Source Data file.

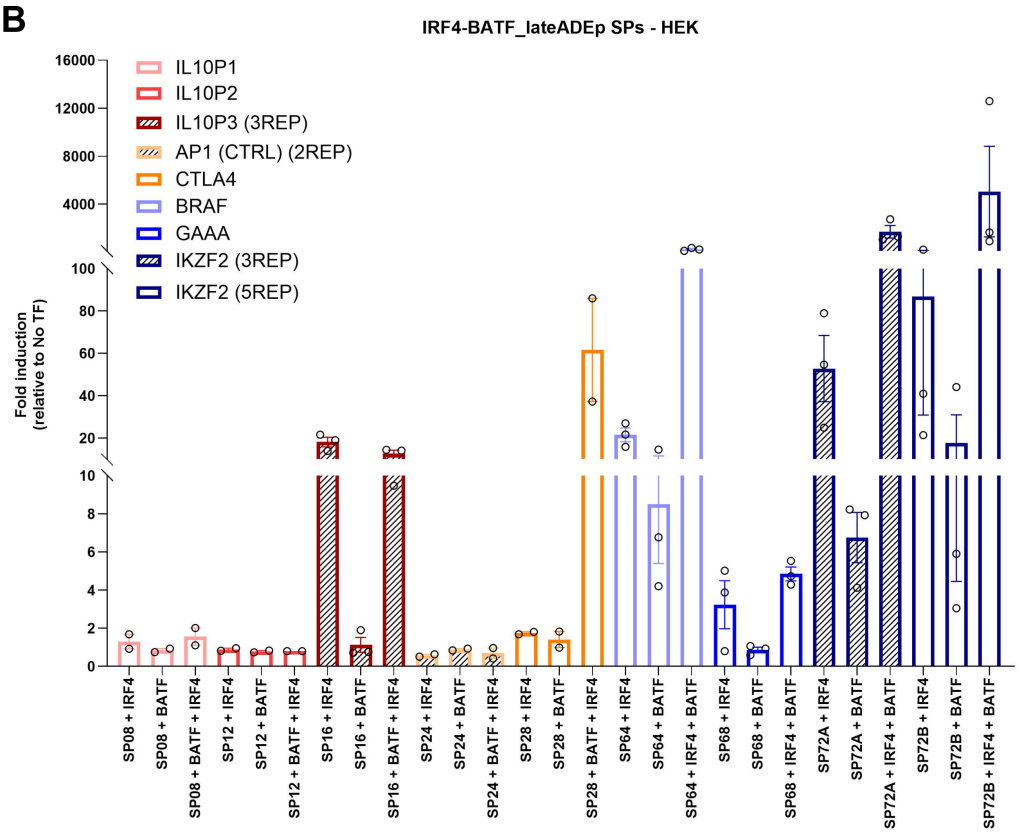

A
TF-BSs
miniTK
mCherry
B
IRF4-BATF_miniTK SPs - HEK
IL10P1
IL10P2
IL10P3 (3REP)
AP1 (CTRL)
CTLA4
BRAF
GAAA
IKZF2 (3REP)
Fold induction (relative to No TF)
SP07 + IRF4
SP07 + BATF
SP07 + BATF + IRF4
SP11 + IRF4
SP11 + BATF
SP11 + BATF + IRF4
SP15 + IRF4
SP15 + BATF
SP15 + BATF + IRF4
SP23 + IRF4
SP23 + BATF
SP23 + BATF + IRF4
SP27 + IRF4
SP27 + BATF
SP27 + BATF + IRF4
SP63 + IRF4
SP63 + BATF
SP63 + IRF4 + BATF
SP67 + IRF4
SP67 + BATF
SP67 + IRF4 + BATF
SP71A + IRF4
SP71A + BATF
SP71A + IRF4 + BATF
C
IRF4-BATF_miniTK SPs - Jurkat
IL10P1
IL10P2
IL10P3 (3REP)
AP1 (CTRL)
CTLA4
BRAF
GAAA
IKZF2 (3REP)
Fold induction (relative to No TF)
SP07 + IRF4
SP07 + BATF
SP07 + IRF4 + BATF
SP11 + IRF4
SP11 + BATF
SP11 + IRF4 + BATF
SP15 + IRF4
SP15 + BATF
SP15 + IRF4 + BATF
SP23 + IRF4
SP23 + BATF
SP23 + IRF4 + BATF
SP27 + IRF4
SP27 + BATF
SP27 + IRF4 + BATF
SP63 + IRF4
SP63 + BATF
SP63 + IRF4 + BATF
SP67 + IRF4
SP67 + BATF
SP67 + IRF4 + BATF
SP71A + IRF4
SP71A + BATF
SP71A + IRF4 + BATF

Supplementary Figure 15. Fold-induction analysis of IRF4/BATF-responsive synthetic promoters in the miniTK configuration. (**A**) Schematic representation of synthetic promoters composed of the miniTK minimal promoter combined with IRF4/BATF transcription factor binding sites (TF-BSs). (**B–C**) Fold induction of IRF4/BATF-responsive synthetic promoters in the miniTK configuration in HEK293 and Jurkat cells, calculated relative to the corresponding no-TF condition. Hatched bars indicate synthetic promoters containing fewer than five TF-BS repeats. Data are represented as mean ± SEM of *n* = 2-3 independent biological replicates. Source data are provided as a Source Data file.

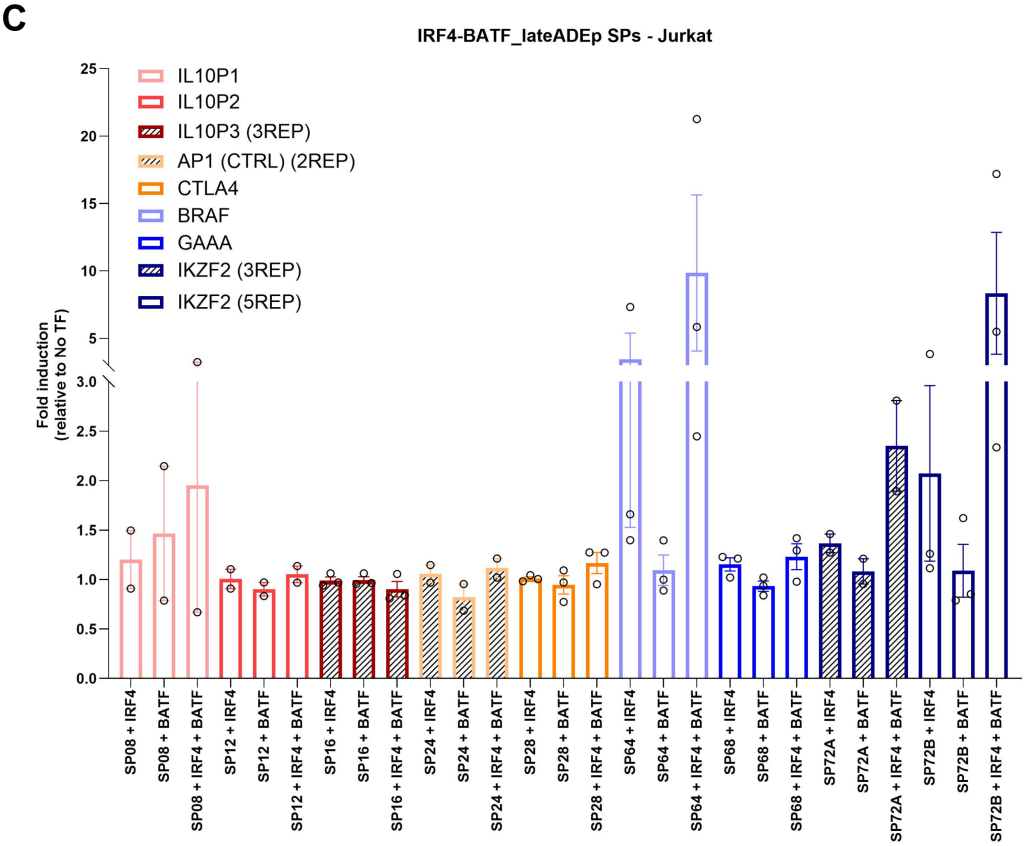
A


TF-BSs
lateADEp
mCherry

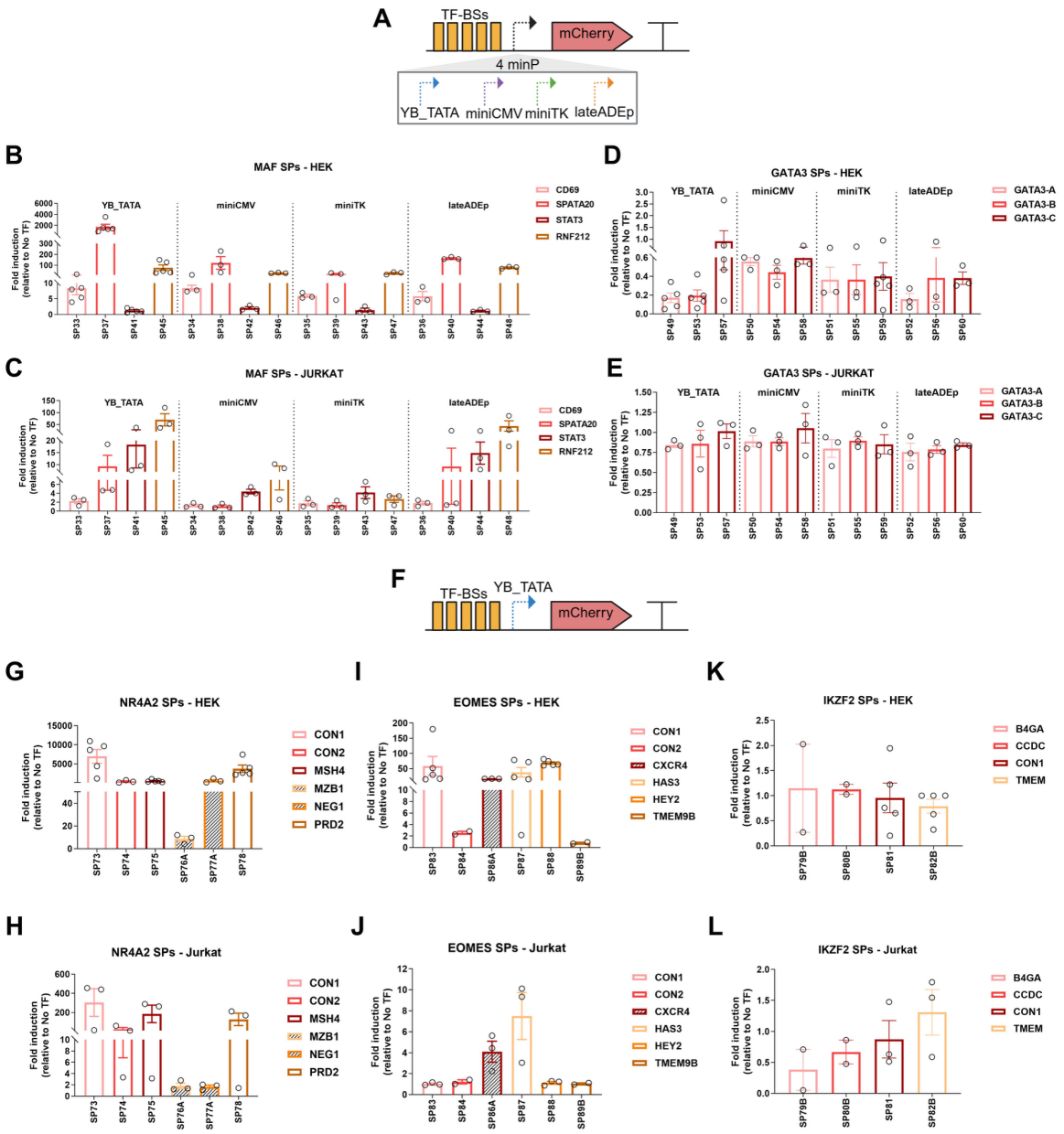
B


IRF4-BATF_lateADEp SPs - HEK
IL10P1
IL10P2
IL10P3 (3REP)
AP1 (CTRL) (2REP)
CTLA4
BRAF
GAAA
IKZF2 (3REP)
IKZF2 (5REP)
Fold induction (relative to No TF)
SP08 + IRF4
SP08 + BATF
SP08 + BATF + IRF4
SP12 + IRF4
SP12 + BATF
SP12 + BATF + IRF4
SP16 + IRF4
SP16 + BATF
SP16 + BATF + IRF4
SP24 + IRF4
SP24 + BATF
SP24 + BATF + IRF4
SP28 + IRF4
SP28 + BATF
SP28 + BATF + IRF4
SP64 + IRF4
SP64 + BATF
SP64 + IRF4 + BATF
SP68 + IRF4
SP68 + BATF
SP68 + IRF4 + BATF
SP72A + IRF4
SP72A + BATF
SP72A + IRF4 + BATF
SP72B + IRF4
SP72B + BATF
SP72B + IRF4 + BATF

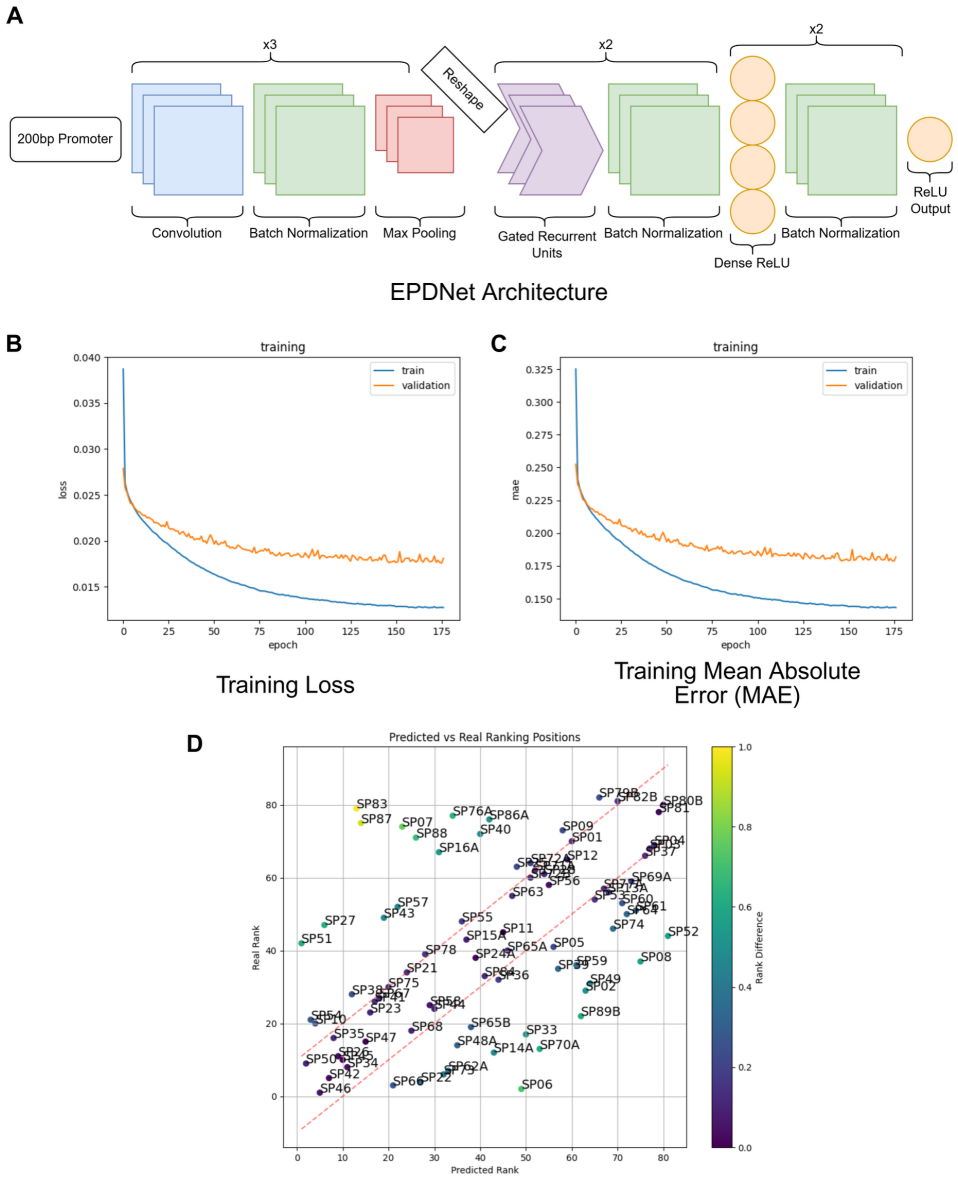
C


IRF4-BATF_lateADEp SPs - Jurkat
IL10P1
IL10P2
IL10P3 (3REP)
AP1 (CTRL) (2REP)
CTLA4
BRAF
GAAA
IKZF2 (3REP)
IKZF2 (5REP)
Fold induction (relative to No TF)
SP08 + IRF4
SP08 + BATF
SP08 + IRF4 + BATF
SP12 + IRF4
SP12 + BATF
SP12 + IRF4 + BATF
SP16 + IRF4
SP16 + BATF
SP16 + IRF4 + BATF
SP24 + IRF4
SP24 + BATF
SP24 + IRF4 + BATF
SP28 + IRF4
SP28 + BATF
SP28 + IRF4 + BATF
SP64 + IRF4
SP64 + BATF
SP64 + IRF4 + BATF
SP68 + IRF4
SP68 + BATF
SP68 + IRF4 + BATF
SP72A + IRF4
SP72A + BATF
SP72A + IRF4 + BATF
SP72B + IRF4
SP72B + BATF
SP72B + IRF4 + BATF

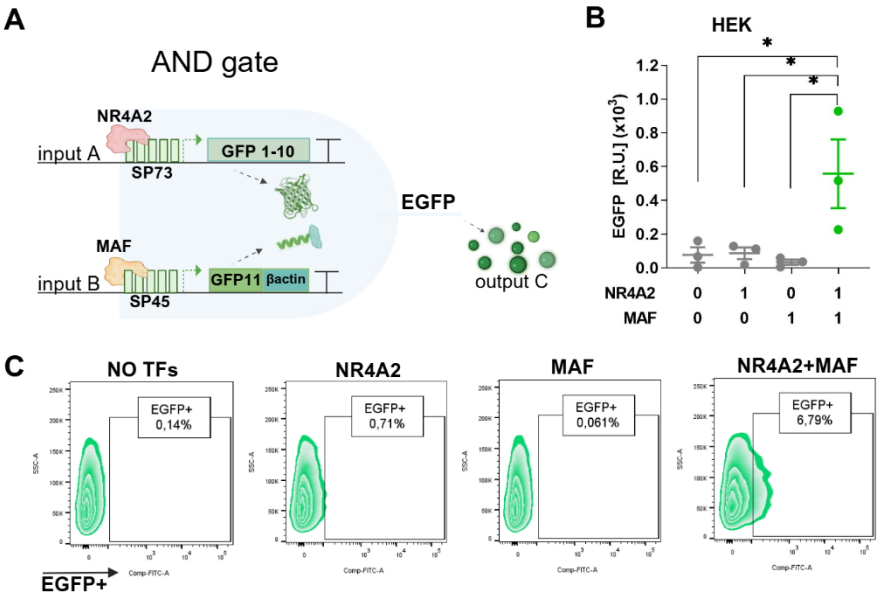
Supplementary Figure 16. Fold-induction analysis of IRF4/BATF-responsive synthetic promoters in the lateADEp configuration. (**A**) Schematic representation of synthetic promoters composed of the lateADEp minimal promoter combined with IRF4/BATF transcription factor binding sites (TF-BSs). (**B–C**) Fold induction of IRF4/BATF-responsive synthetic promoters in the lateADEp configuration in HEK293 and Jurkat cells, calculated relative to the corresponding no-TF condition. Hatched bars indicate synthetic promoters containing fewer than five TF-BS repeats. Data are represented as mean ± SEM of *n* = 2-3 independent biological replicates. Source data are provided as a Source Data file.



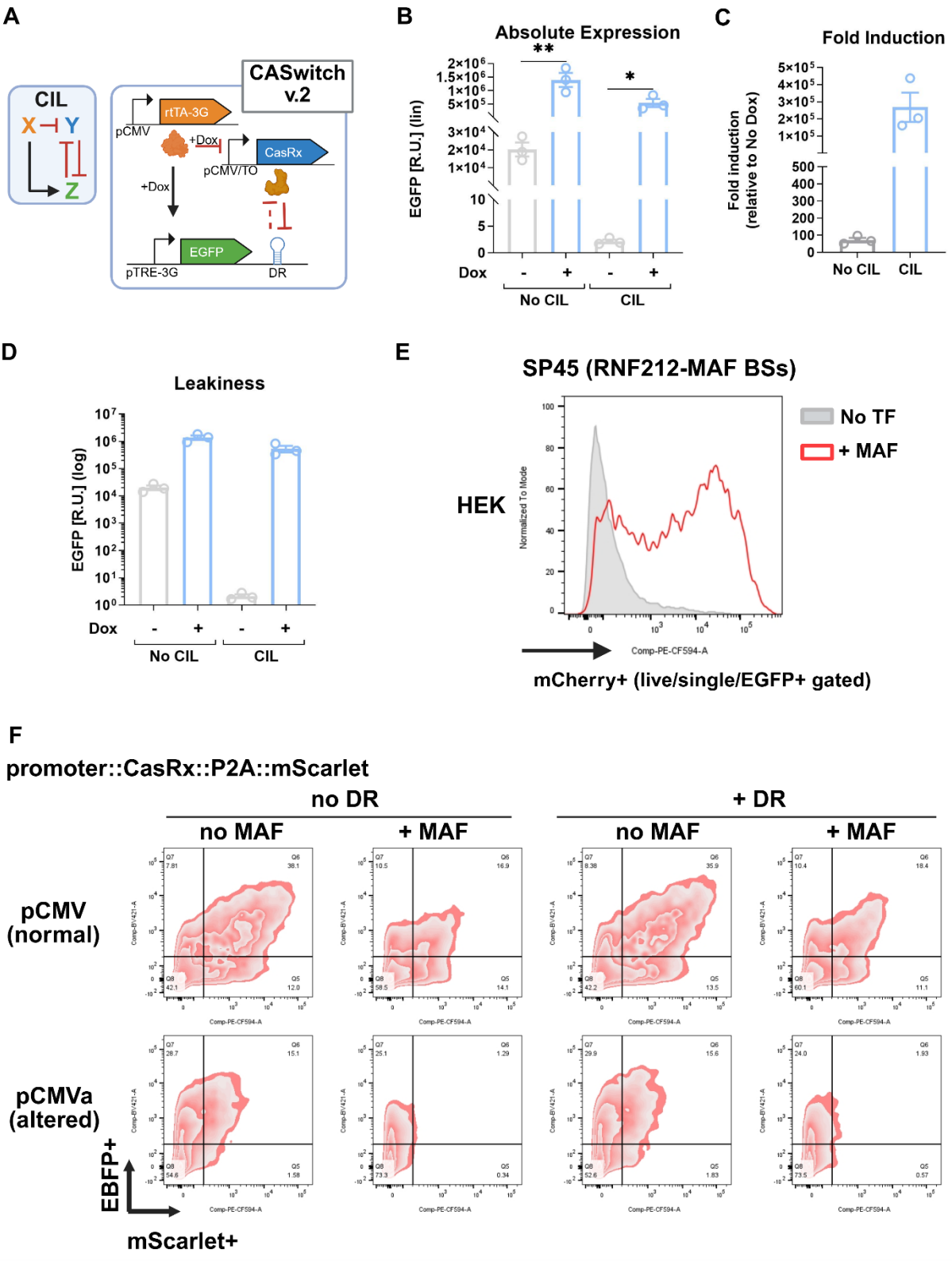
Supplementary Figure 17. Fold-induction analysis of synthetic promoters responsive to multiple transcription factors across promoter configurations. (**A**) Schematic representation of synthetic promoters composed of different minimal promoters combined with MAF or GATA3 transcription factor binding sites (TF-BSs). (**B–E**) Fold induction of MAF- or GATA3-responsive synthetic promoters across all minimal promoter configurations in HEK293 and Jurkat cells, calculated relative to the corresponding no-TF condition. (**F**) Schematic representation of synthetic promoters composed of the YB_TATA minimal promoter combined with TF-BSs for NR4A2, EOMES, and IKZF2. (**G–L**) Fold induction of NR4A2-, EOMES-, or IKZF2-responsive synthetic promoters in the YB_TATA configuration in HEK293 and Jurkat cells, calculated relative to the corresponding no-TF condition. Hatched bars indicate synthetic promoters containing fewer than five TF-BS repeats. Data are represented as mean ± SEM of *n* = 2-3 independent biological replicates. Source data are provided as a Source Data file.



Supplementary Figure 18. Transfer learning model architecture and training. (**A**) EPDNet architecture. (**B-C**) Training models. (**D**) Transfer learning model performance showing measured vs predicted ranking. The model produced a predicted ranking of the sequences based on their expected 'strength'. The mean absolute induction when the promoter is at its strongest expression was used to build the ranking to highlight the exact differences in the correct and predicted ranking for each individual promoter. Detailed description can be found in Supplementary Note 1.

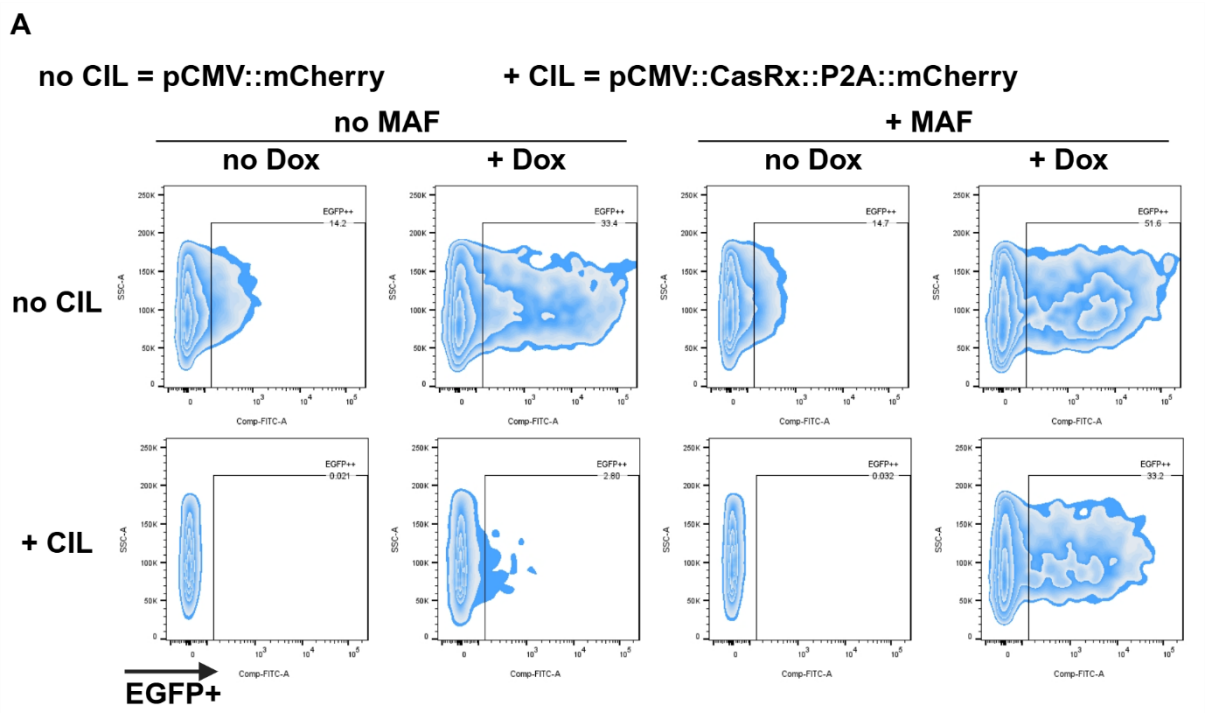


Supplementary Figure 19. Split-EGFP–based AND-gate–like circuit enables combinatorial control by MAF- and NR4A2-responsive synthetic promoters. (**A**) Schematic representation of an AND-gate–like circuit implemented using split EGFP in HEK293 cells. The GFP1–10 fragment is expressed under the control of the NR4A2-responsive synthetic promoter SP73 (input A), while the GFP11 fragment fused to β-actin is expressed under the control of the MAF-responsive synthetic promoter SP45 (input B). (**B-C**) When both transcription factors are present, the two EGFP fragments reconstitute a functional fluorescent protein (output C). GFP expression was quantified by flow cytometry 48 h post-transfection in HEK293 cells and is shown as bar plots and representative flow cytometry contour plots. Data are represented as mean ± SEM of *n* = 3 biological replicates. Statistical test: one-tailed unpaired *t*-test. *P* value: *<0.05. Source data are provided as a Source Data file. Created in BioRender. Olandesi, R. (2026) https://BioRender.com/hcktu2q.

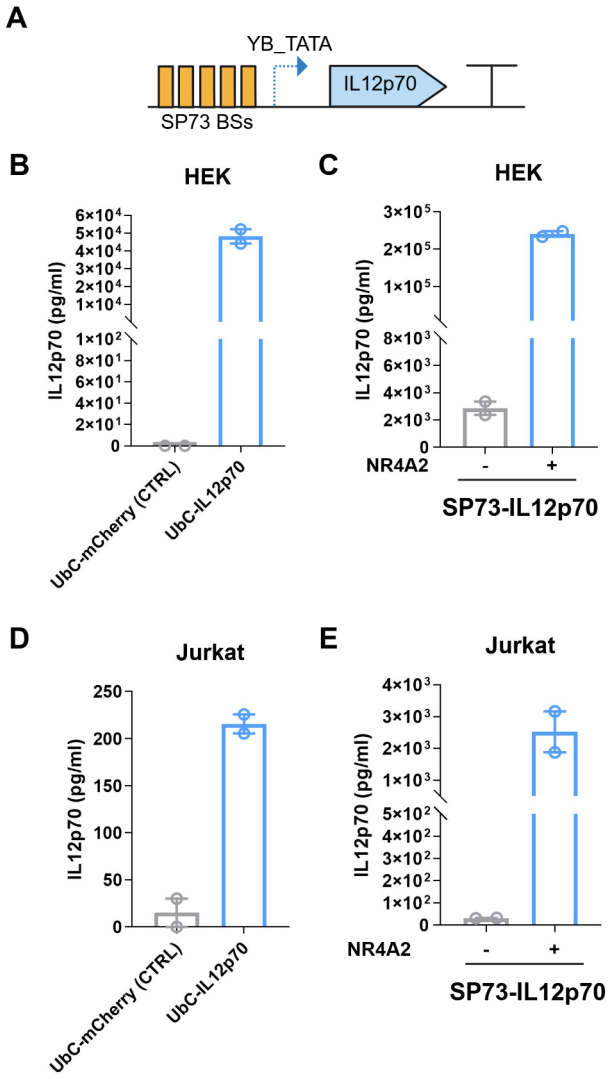


Supplementary Figure 20. Network-based circuit architectures reduce OFF-state basal activity of a MAF-responsive synthetic promoter. (**A-D**) Original design and functional characterization of the CASwitch circuit, showing absolute expression levels, OFF-state basal activity (leakiness), and fold induction of the output reporter. (**E**) Histogram showing basal activity of the MAF-responsive synthetic promoter SP45 in the absence of the cognate transcription factor (no TF) and induced activity in the presence of MAF (+MAF). (**F**)

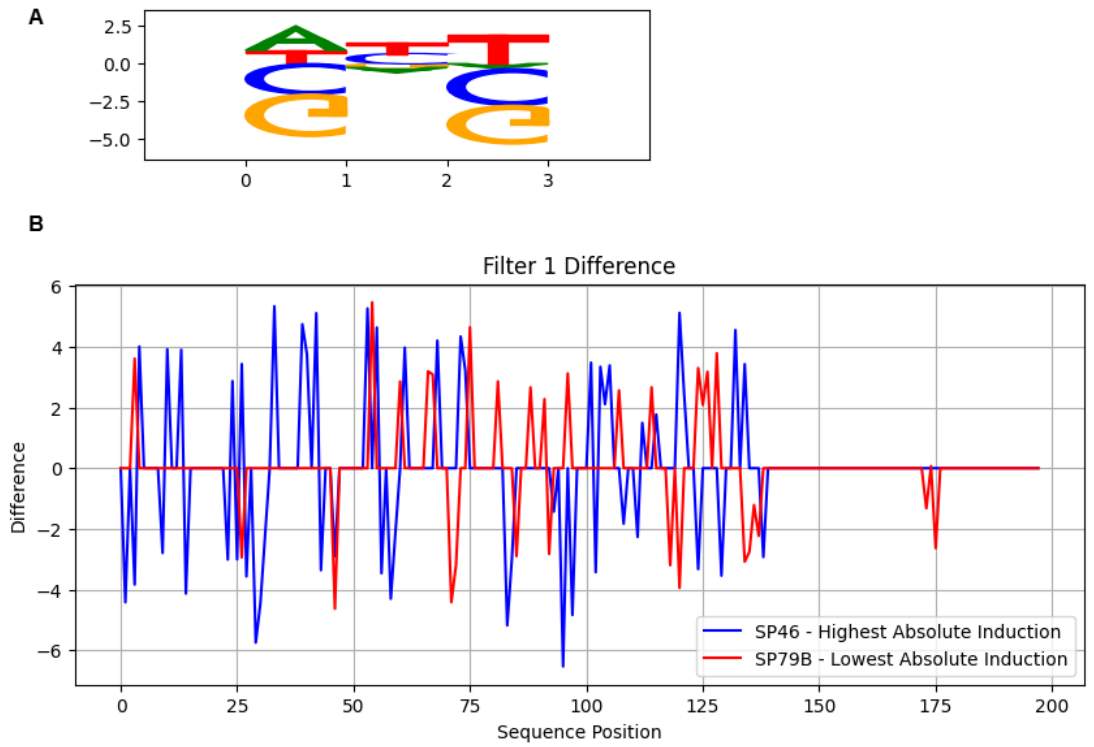
Representative flow cytometry contour plots of mScarlet expression for the SP-MI configuration, illustrating reduced expression levels when CasRx is driven by a weakened pCMV promoter variant (pCMVa) compared to the original pCMV promoter. mScarlet fluorescence serves as a proxy for CasRx expression, as the two proteins are co-expressed via a P2A sequence (CasRx–P2A–mScarlet). Data are represented as mean ± SEM of *n* = 3 independent biological replicates. Statistical analysis: one-tailed unpaired *t*-test. *P* value: *<0.05; **<0.005. Source data are provided as a Source Data file. Created in BioRender. Olandesi, R. (2026) https://BioRender.com/yzq7fdj.



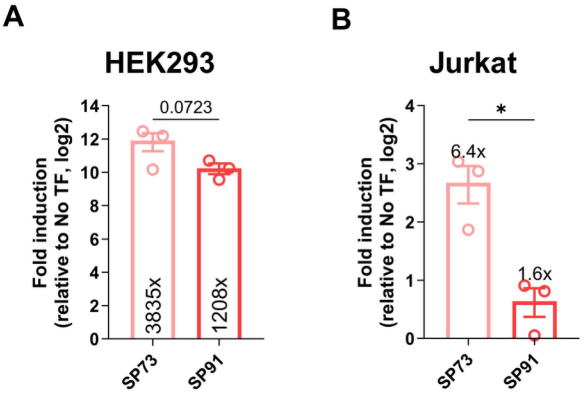
Supplementary Figure 21. SP-CIL network architecture reduces OFF-state basal activity of a MAF-responsive synthetic promoter. Implementation of the SP-CIL (Coherent Inhibitory Loop) network architecture using the MAF-responsive synthetic promoter SP45. The circuit design combines inducible activation with CasRx-mediated repression to reduce basal activity and increase the ON/OFF ratio of promoter output. Representative flow cytometry contour plots of *n* = 3 independent biological replicates show EGFP reporter expression for the SP-CIL configuration, illustrating reduced OFF-state leakiness in the presence of the inhibitory loop (+CIL) compared to the configuration lacking CasRx. EGFP expression was quantified by flow cytometry 48 h post-transfection in HEK293 cells.



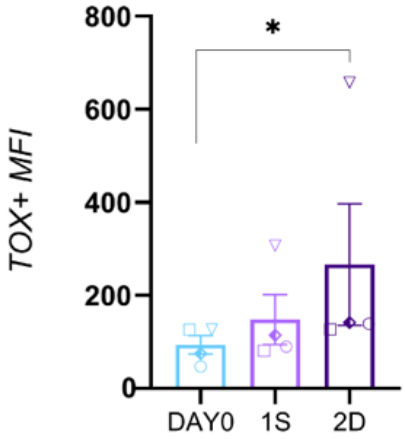
Supplementary Figure 22. NR4A2-responsive synthetic promoter SP73 enables inducible secretion of cytokinepayloads. (**A**) Schematic representation of the NR4A2-responsive synthetic promoter SP73 driving expression of the cytokine IL12p70. HEK293 and Jurkat cells were co-transfected or electroporated with SP73 and the NR4A2 expression plasmid together with a transfection marker. Transfection efficiency was assessed by flow cytometry 48hs post-delivery. (**B–E**) ELISA quantification of IL12p70 secretion in HEK293 and Jurkat cells. Panels (B, D) show constitutive expression driven by the UbC promoter, while panels (C, E) show inducible expression driven by NR4A2-responsive SP73. Data are represented as mean ± SEM of *n* = 2 independent biological replicates. Source data are provided as a Source Data file.



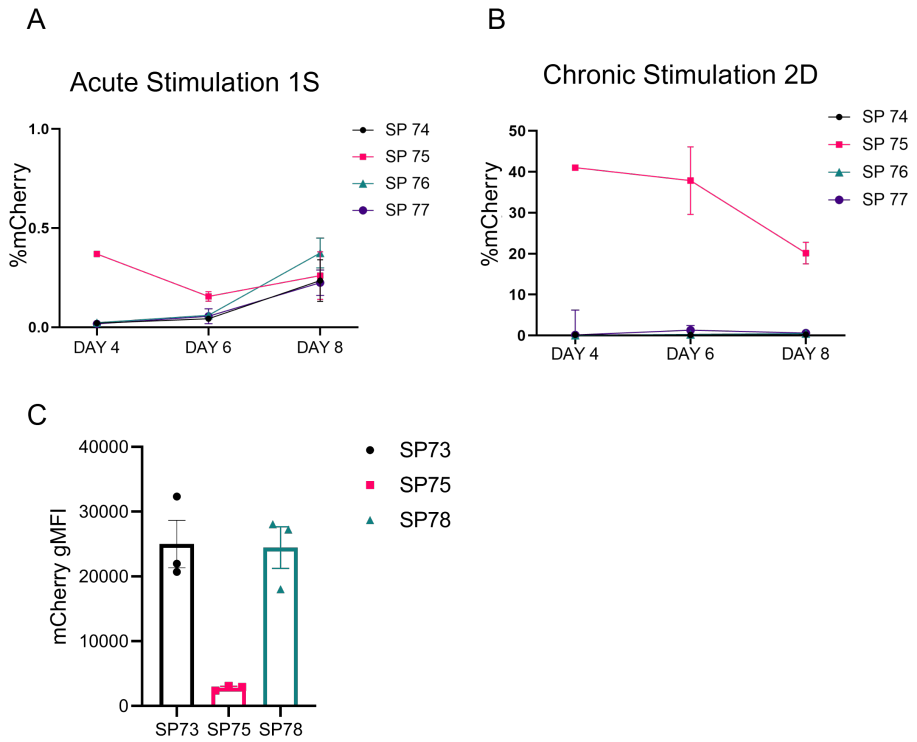
Supplementary Figure 23. Explainability of Convolutional Filter. (A) The 3 base pair filter weights from the first convolution. **(B)** Comparison of the significant activation difference in the first filter across the 'strongest' and 'weakest' promoters. At each position we plot how the activation differs from the average activation across all synthetic promoters, but only for deviations greater than one standard deviation from the average.



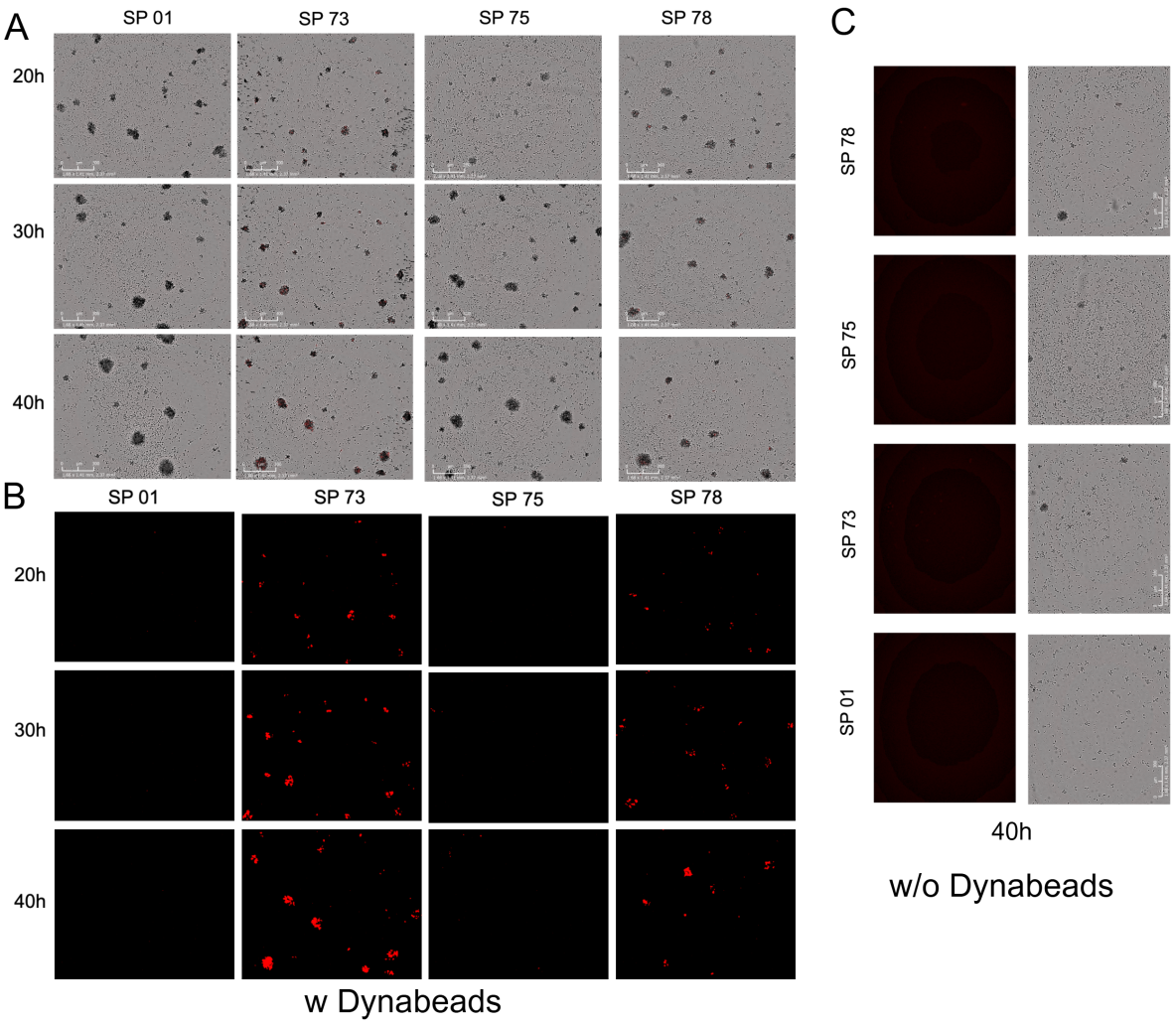
Supplementary Figure 24. Benchmarking comparison of NR4A2 SP73 promoter with previously published NR4A promoter. HEK293 (**A**) and Jurkat (**B**) cells were transfected or electroporated respectively with NR4A-responsive promoters SP73 and SP91 (promoter architecture reported in Guo et al., 2022). Bar plots show log2 fold induction of the SPs in the presence of NR4A2 expression, showing higher reporter activation with SP73 (3800 vs 1208X in HEK293 cells; 6.4 vs 1.6x in Jurkat cells). Data are represented as mean ± SEM of *n* = 3 independent biological replicates. Statistical analysis: one-tailed unpaired *t*-test. *P* value: *<0.05. Source data are provided as a Source Data file.



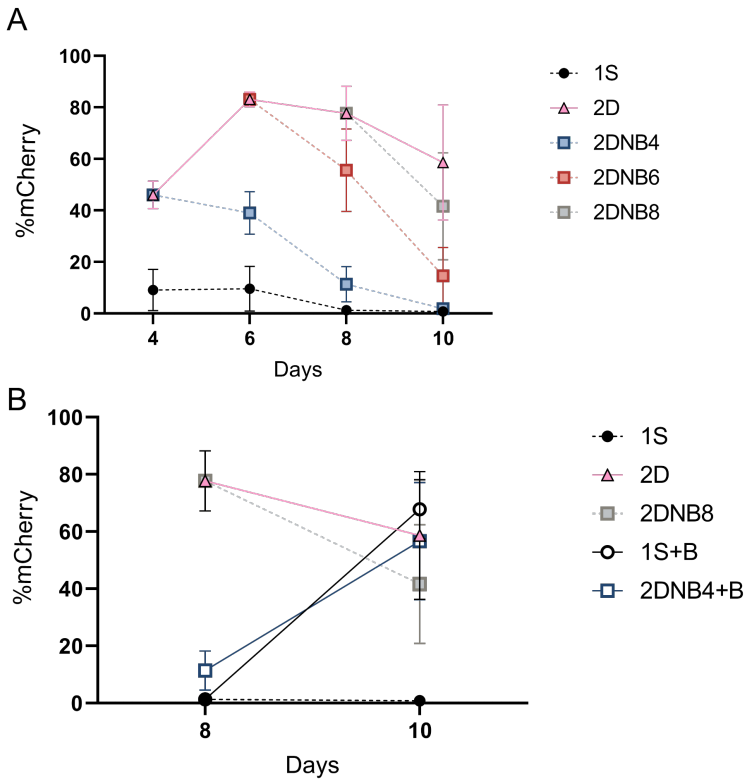
Supplementary Figure 25. TOX expression is modestly increased under persistent in vitro T cell stimulation. Flow cytometry analysis of TOX expression in primary human $CD8^+$ T cells performed 8 days after either a single acute stimulation (1S) or repeated CD3/CD28 stimulation every 2 days (2D), used here as an in vitro model of persistent stimulation. Compared to the acute condition, TOX expression is increased under persistent stimulation, although overall expression levels remain low. Day 0 indicates TOX expression levels prior to stimulation with CD3/CD28 dynabeads. Data are represented as mean ± SEM of *n* = 4 donors. Statistical analysis: one-way ANOVA with Dunn's post hoc test. *P* value: *<0.05. Source data are provided as a Source Data file.



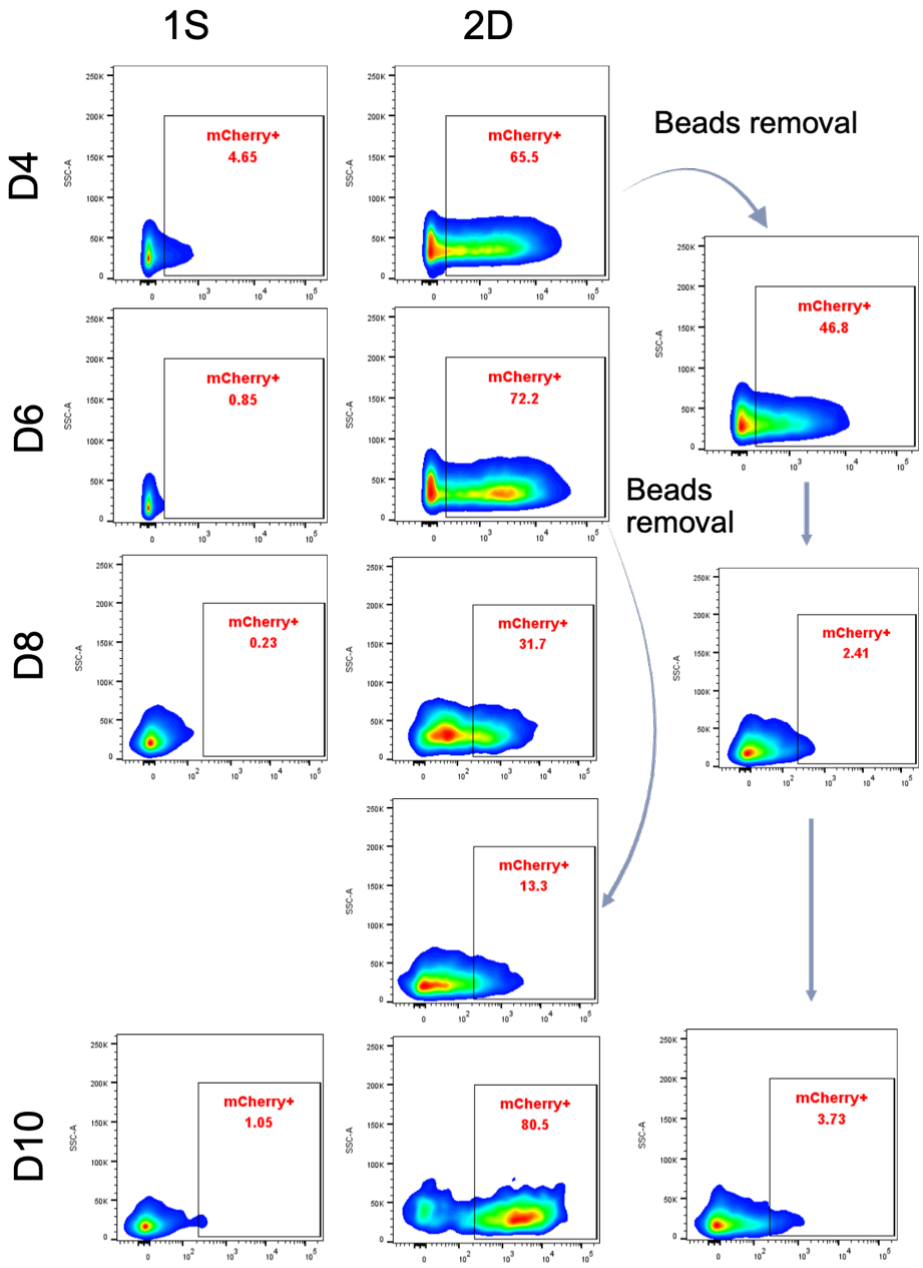
Supplementary Figure 26. Validation of NR4A2-responsive synthetic promoters with distinct binding site configurations in primary human T cells. (**A-B**) Primary human CD8$^+$ T cells were transduced with lentiviruses comprising engineered with NR4A2-responsive synthetic promoters SP74, SP75, SP76, and SP77 in the first transcriptional unit and a constitutively expressed LNGFR as transduction marker in the second transcriptional unit, and subsequently subjected to stimulation with CD3/CD28 dynabeads. Promoter activity was monitored over time under repeated stimulation conditions. Among the tested variants, SP75 exhibited detectable activation upon repeated stimulation, although a decrease in reporter expression was observed between day 6 and day 8. The remaining promoters (SP74, SP76, and SP77) did not show activation under these conditions. (**C**) Comparison of mCherry expression in SP73, SP75, SP78 following 8 days of stimulation shows lower mCherry MFI induced by SP75. Overall, these results are consistent with the screening data obtained in HEK293 and Jurkat cells, in which SP74, SP76, and SP77 also displayed lower inducible activity (see Supplementary Figure 17G–H). Data are shown as mean ± SEM from *n* = 3 independent donors. Source data are provided as a Source Data file.



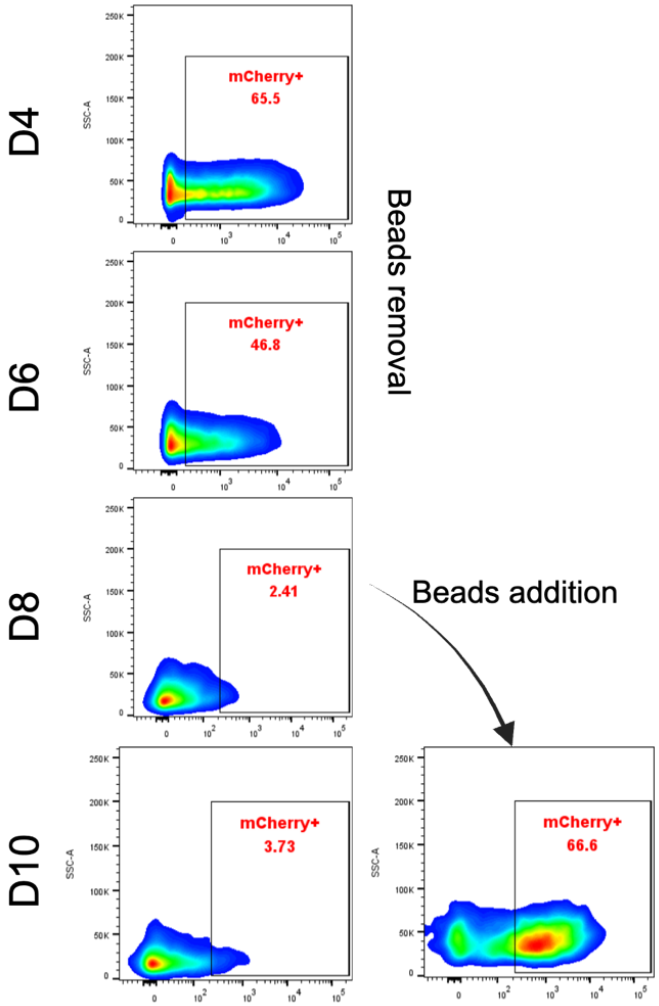
Supplementary Figure 27. Time-lapse imaging reveals distinct activation kinetics of NR4A2-responsive synthetic promoters in primary CD8⁺ T cells. (**A-B**) Representative time-lapse images showing merged bright-field and mCherry fluorescence (A) or mCherry fluorescence alone (B) in primary human CD8⁺ T cells engineered with NR4A2-responsive synthetic promoters SP73, SP75, or SP78 during >40 h of stimulation with anti-CD3/CD28 dynabeads. A synthetic promoter lacking NR4A2 binding sites (SP01) was used as a negative control. The images illustrate differences in induction kinetics among promoter variants, with SP73 showing earlier and stronger mCherry induction compared with SP78. SP75 did not display detectable mCherry expression under these imaging conditions, consistent with the low reporter signal observed by flow cytometry (Supplementary Figure 26C). (**C**) In the absence of anti-CD3/CD28 dynabead stimulation, mCherry fluorescence was not detected.



Supplementary Figure 28. NR4A2-responsive SP78 exhibits dynamic and reversible activation under repeated T cell stimulation. (**A**) Primary human CD8⁺ T cells engineered with the NR4A2-responsive synthetic promoter SP78 driving mCherry expression were subjected to acute or repeated stimulation using anti-CD3/CD28 dynabeads, following the schematic shown in Fig. 5A. Cells were either stimulated once at day 0 with bead removal at day 2 (1S, acute stimulation), or repeatedly stimulated every 48 h (2D, persistent stimulation). To assess reversibility of promoter activation, stimulation was interrupted by dynabead removal after two (2DNB4), three (2DNB6), or four (2DNB8) rounds of stimulation, and mCherry expression was monitored over time by flow cytometry. In all conditions, removal of stimulation resulted in downmodulation of mCherry expression, indicating that promoter activity decreases in the absence of continued stimulation. (**B**) To evaluate whether promoter activity could be re-induced, T cells from the 1S and 2DNB4 conditions were re-exposed to anti-CD3/CD28 dynabeads. Flow cytometry analysis shows renewed mCherry expression upon re-stimulation. Together, these results demonstrate that SP78 activity is dynamic and reversible and closely tracks the presence or absence of sustained T cell stimulation. Data are shown as mean ± SEM from *n* = 3 independent donors. Source data are provided as a Source Data file.

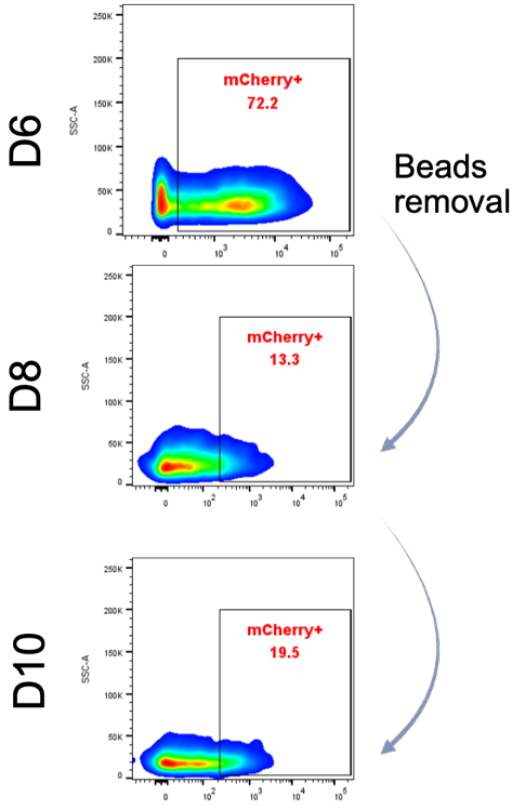


Supplementary Figure 29. NR4A2-responsive SP73 activity depends on sustained T cell stimulation and is reversible. Representative flow cytometry plots showing mCherry expression in engineered primary $CD8^+$ T cells carrying the NR4A2-responsive synthetic promoter SP73 under different stimulation regimes. Cells subjected to a single round of stimulation (1S) display minimal mCherry induction, which progressively decreases over time following stimulus withdrawal. In contrast, cells exposed to repeated stimulation every 48 h (2D) exhibit higher levels of mCherry expression. Removal of anti-CD3/CD28 stimulation at day 4 or day 6 results in a gradual reduction of mCherry-positive cells, indicating that promoter activity decreases in the absence of sustained stimulation.

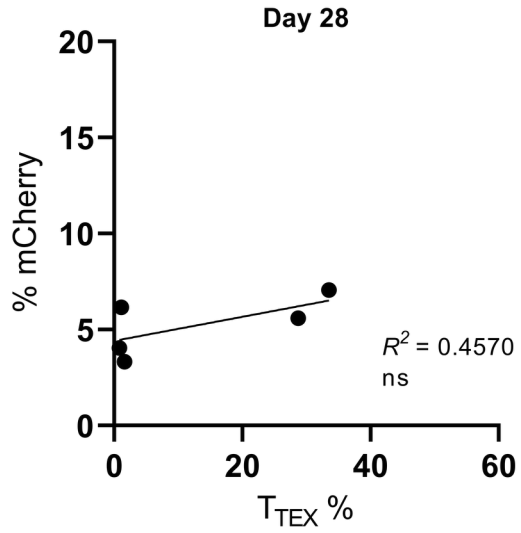


406

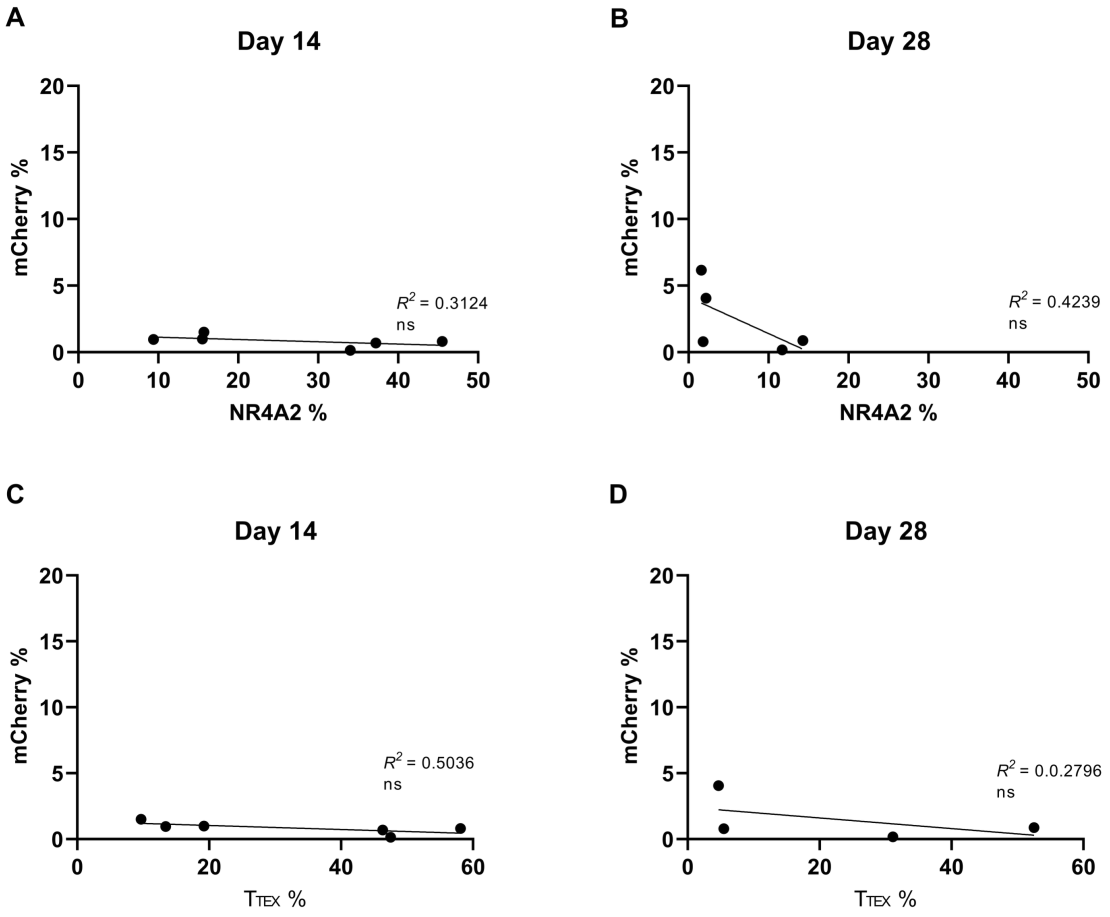
407 **Supplementary Figure 30. NR4A2-responsive promoter SP73 is reactivated upon re-**
408 **stimulation of CD8⁺ T cells.** Representative flow cytometry plots showing mCherry
409 expression in primary CD8⁺ T cells engineered with the NR4A2-responsive synthetic
410 promoter SP73. Cells were initially stimulated with anti-CD3/CD28 dynabeads for four days,
411 followed by removal of the stimulation. Upon bead withdrawal, mCherry expression
412 progressively decreased from day 4 to day 10. When cells were re-stimulated with anti-
413 CD3/CD28 dynabeads at day 8, mCherry expression increased again, indicating reactivation
414 of the synthetic promoter upon renewed T cell stimulation.

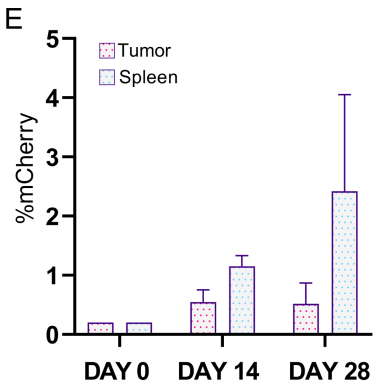


Supplementary Figure 31. NR4A2-responsive promoter SP73 switches OFF following stimulation withdrawal at later time points. Representative flow cytometry dot plots showing mCherry expression in primary $CD8^+$ T cells engineered with the NR4A2-responsive synthetic promoter SP73. Cells were stimulated with anti-CD3/CD28 dynabeads and subsequently cultured following removal of the stimulation. At later time points after bead withdrawal, mCherry expression is reduced, indicating that SP73 activity decreases in the absence of continued T cell stimulation.



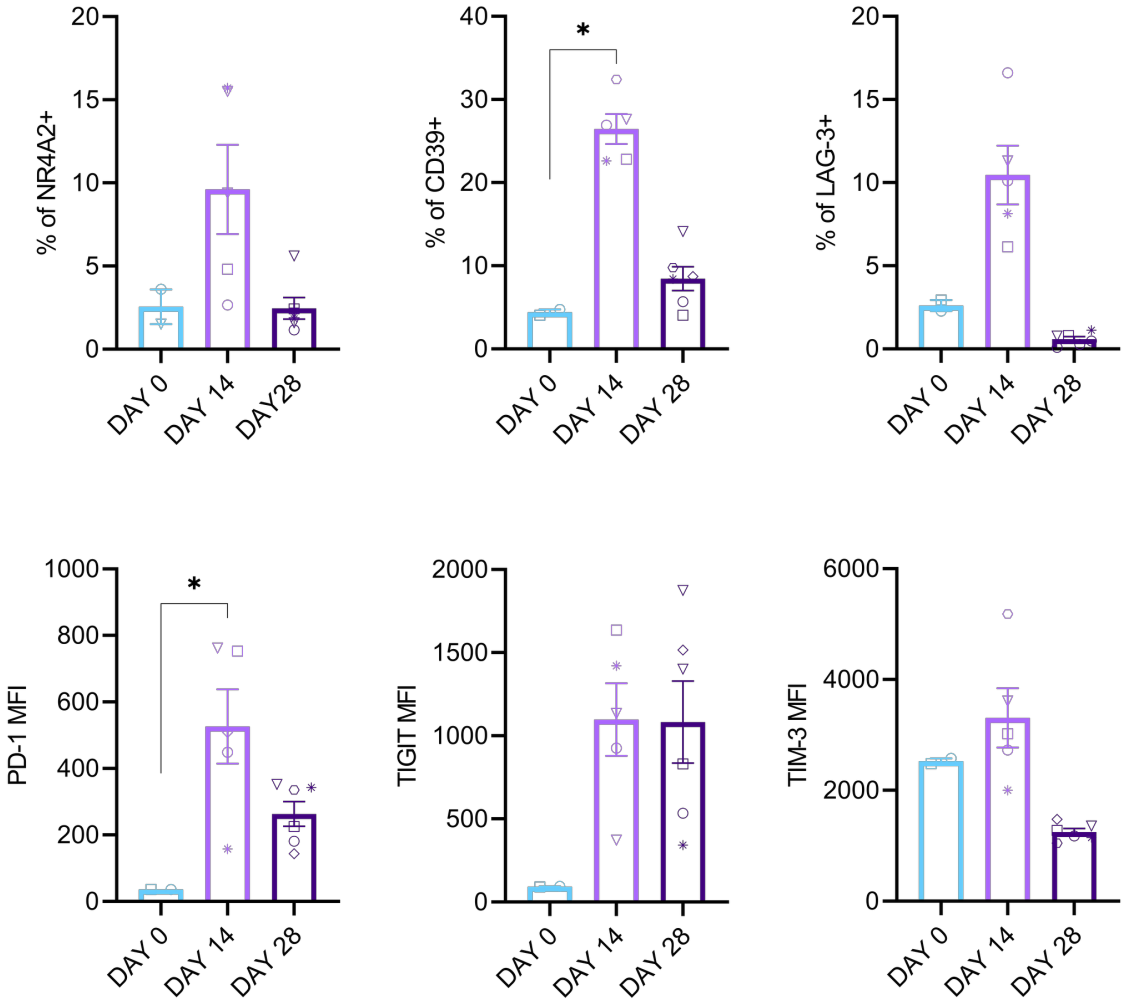
Supplementary Figure 32. mCherry expression shows limited correlation with activation and dysfunction markers during tumor regression. Correlation analysis between mCherry reporter expression and the frequency of terminally exhausted (Ttex) cells within the $CD8^+$ TCR^+ T cells population at day 28, following tumor regression in a melanoma mouse model. Ttex cells were identified using the same sequential gating strategy described in Figure 7H and Materials and Methods section. R^2 0.4570. Compared to day 14, expression levels of these markers are reduced at day 28, consistent with tumor regression (see Figure 7H). At day 28, mCherry expression shows a positive association with NR4A2 levels, whereas no consistent correlation is observed with other activation or dysfunction markers. These data indicate that mCherry reporter activity primarily reflects NR4A2-associated transcriptional activity rather than broader changes in T cell activation or dysfunction status during tumor regression. Pearson correlation (two-tailed test). R^2 = 0.4570, *P* = ns. Source data are provided as a Source Data file.



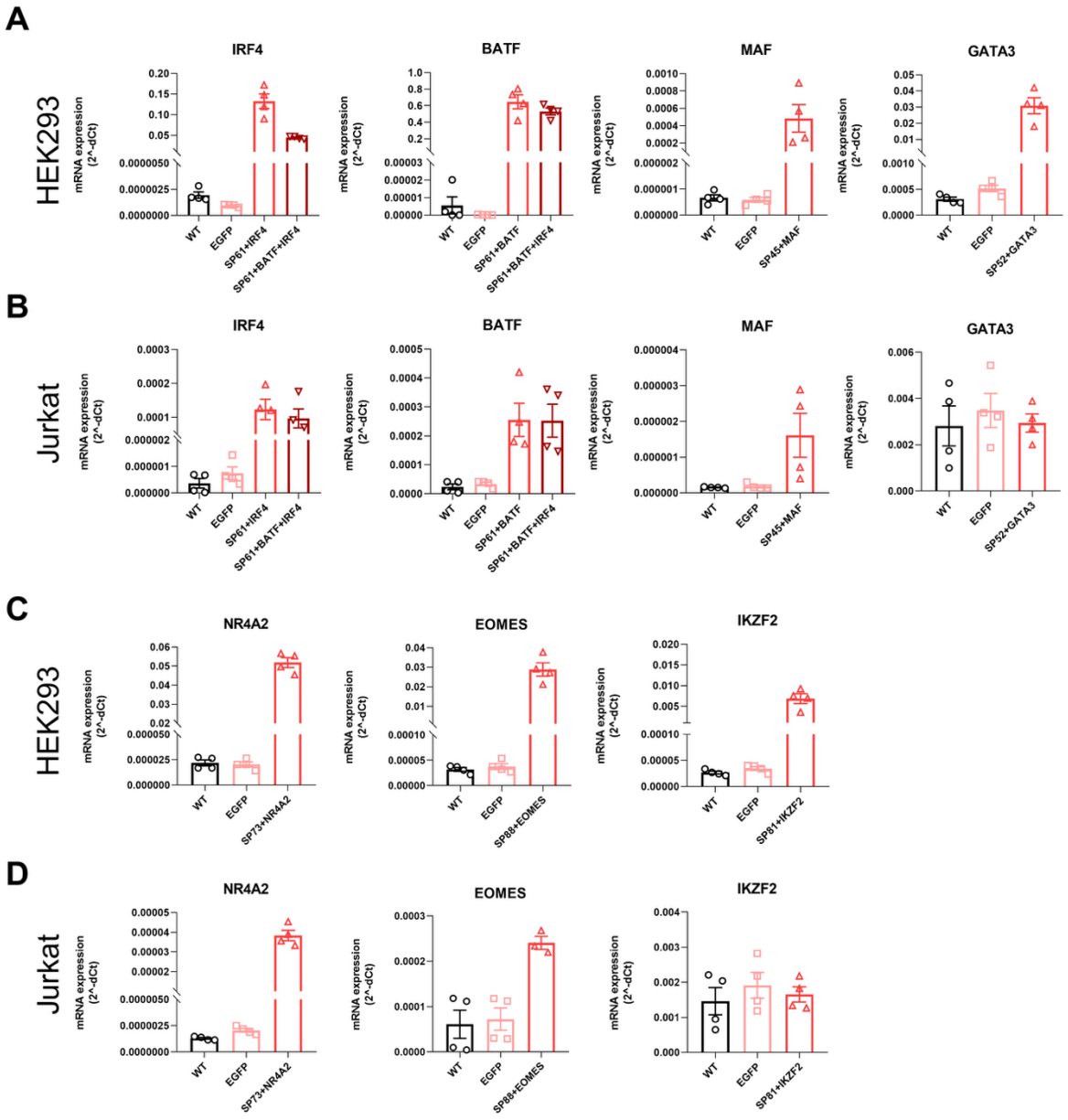


Supplementary Figure 33. NR4A2 binding sites are required for activation of the synthetic promoter in vivo (**A-B**) Correlation analysis between mCherry reporter expression driven by a control SP lacking the NR4A2 BS and NR4A2 protein levels in tumor-derived engineered CD8$^+$ T cells isolated from a melanoma mouse model. NR4A2 expression is increased at high tumor burden (day 14) and decreases upon tumor regression (day 28), while mCherry expression remains low throughout, indicating lack of synthetic promoter activation in the absence of NR4A2 binding sites. (**C-D**) Correlation analysis between mCherry reporter expression and the frequency of terminally exhausted (Ttex) cells within the CD8$^+$ TCR$^+$ T cells population. Ttex cells were identified using the same sequential gating strategy described in Figure 7H and Materials and Methods section. The correlation analysis show low mCherry expression despite detectable expression of

activation and dysfunction-associated markers further supporting the requirement of NR4A2 binding sites for reporter activation. (E) mCherry expression in tumor- and spleen-derived T cells 14- and 28-days post-transfer. Day 0 indicates mCherry expression levels prior to adoptive transfer. Reported R^2 values are: (A) 0,31; (B) 0,42; (C) 0,50 (D) 0,27. (**E**) mCherry expression driven by SP01 in tumor and spleen derived T cells at Day 14 and Day 28. Day 0 represents expression of mCherry before injection in mice. Data are shown as mean ± SEM. *n* = 6 mice per harvesting. Source data are provided as a Source Data file.

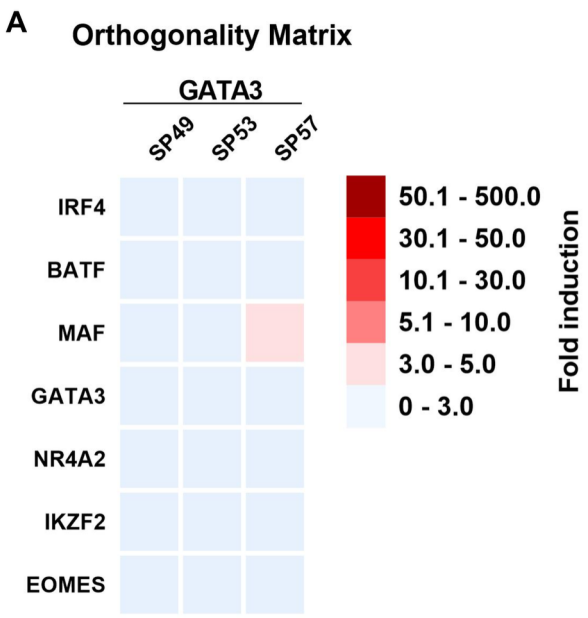


Supplementary Figure 34. Spleen-derived T cells display activation-associated marker expression during tumor growth. Markers associated with T cell activation and dysfunction were assessed in engineered T cells prior to adoptive transfer (day 0) and following isolation from the spleen at day 14 and day 28 post-transfer in a melanoma mouse model. Similar to the patterns observed in tumor-derived T cells, expression of most markers is increased at day 14 during active tumor growth and reduced at day 28 following tumor regression. Data are represented as mean ± SEM of *n* = 5 individual mice at day 14 and *n* = 6 at day 28. Statistical significance was assessed using the Kruskal–Wallis test with Dunn's post hoc test. *P* value: **P* < 0.05, ** *P* < 0.01, *** *P* < 0.001. Source data are provided as a Source Data file.

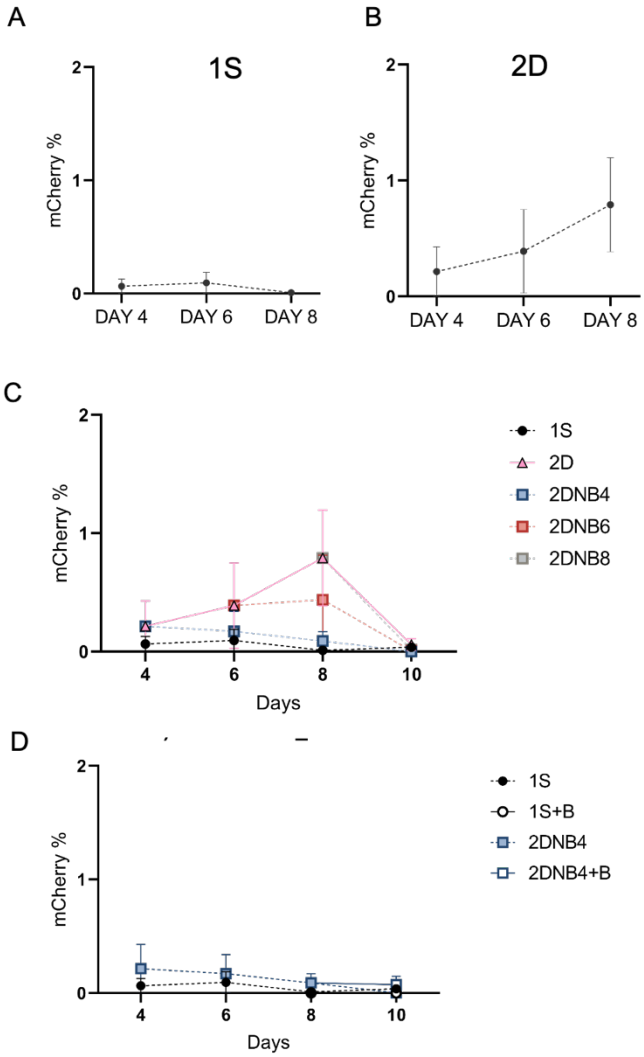


Supplementary Figure 35. Quantification of TFs expression upon transient delivery in HEK293 and Jurkat cells. Transcription factors (TFs) (A-B) IRF4, BATF, MAF, GATA3 and (C-D) NR4A2, EOMES and IKZF2 encoding plasmids were transiently delivered to HEK293 and Jurkat cells, and TF mRNA levels were quantified 48 h post-transfection or electroporation by qPCR ($2^{\wedge-dCt}$). Expression levels were compared across non-transfected wild-type cells (WT), cells expressing EGFP alone, and cells expressing the full inducible system comprising both the synthetic promoter (SP) and the cognate TF. TF transcript levels were substantially lower in Jurkat cells than in HEK293 cells, in some cases by approximately two to three orders of magnitude, consistent with the reduced promoter activation observed in Jurkat cells during transient screening. Notably, GATA3 and IKZF2 transcripts were not detectably expressed in Jurkat cells under these conditions. Data are represented as mean ± SEM of n = 4 independent biological replicates. Primer sequences used for this figure are reported in Supplementary Methods. Source data are provided as a

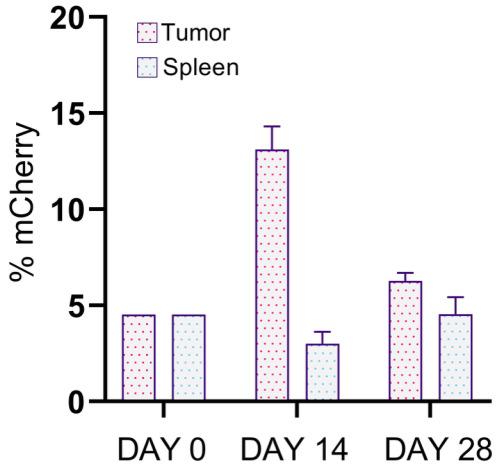
Source Data file.



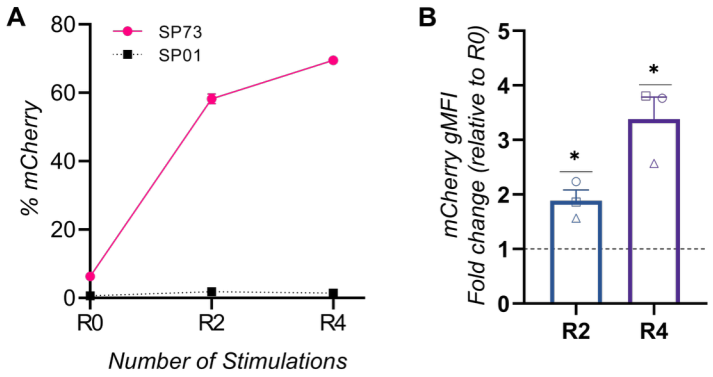
Supplementary Figure 36. Orthogonality analysis of GATA3-responsive synthetic promoters across non-cognate transcription factors. Orthogonality matrix showing activation of GATA3-responsive synthetic promoters when co-expressed with the panel of transcription factors tested in this study. Promoter activity was assessed by reporter expression under identical experimental conditions. Data are shown as mean of *n* = 3 independent biological replicates. Source data are provided as a Source Data file.



Supplementary Figure 37. SP01 lacking NR4A2 binding sites does not respond to stimulation or reversibility regimens. (**A-B**) Primary human CD8⁺ T cells were transduced with SP01, a synthetic promoter lacking NR4A2 binding sites, and subjected to the same stimulation protocols described in Figure 5. mCherry expression was quantified following single acute stimulation (1S), repeated stimulation every 48 h (2D). (**C-D)** mCherry expression by SP01 upon stimulation withdrawal/re-stimulation conditions (2DNB4, 2DNB6, 2DNB8). Across all conditions, SP01 exhibited minimal reporter activation, including during chronic stimulation and after re-exposure to anti-CD3/CD28 dynabeads, confirming that promoter activation observed with NR4A2-responsive SPs is dependent on the presence of NR4A2 binding sites. Data are shown as mean ± SEM from *n* = 3 independent donors. Source data are provided as a Source Data file.

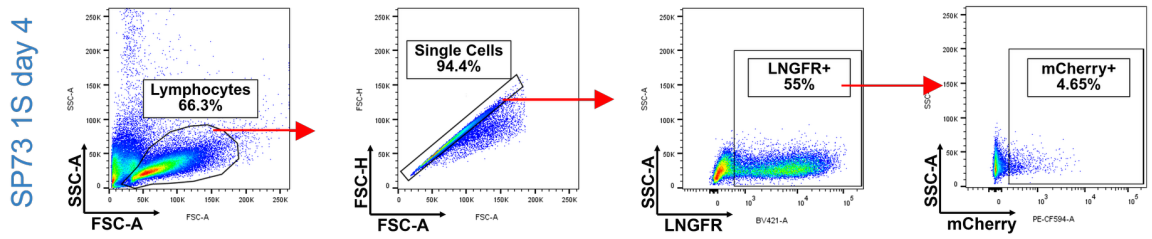


Supplementary Figure 38. Comparison of mCherry expression in spleen- and tumor-derived T cells during tumor progression and regression. Primary human CD8⁺ T cells engineered with NY-ESO-1 TCR and NR4A2-responsive SP73 were adoptively transferred into tumor-bearing NSG mice. Spleen- and tumor-derived T cells were isolated at day 14 and day 28 post-transfer and analyzed by flow cytometry. *n* = 6 mice per harvest. Data are shown as mean ± SEM. Source data are provided as a Source Data file.

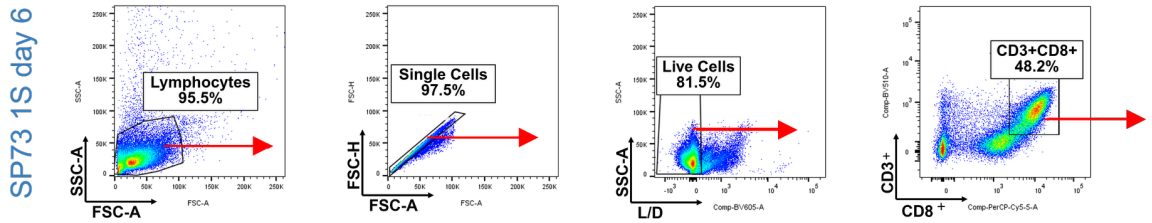


Supplementary Figure 39. NR4A2 binding sites are required for activation of the synthetic promoter SP73 in engineered CD8⁺ T cells co-cultured with melanoma cells in vitro. (**A**) Quantification of mCherry⁺ T cells over time during co-culture. SP73-engineered T cells show reporter induction whereas cells transduced with the SP01 control promoter (lacking NR4A2 binding sites) remain at background levels. (**B**) Fold change of SP73-responsive synthetic promoter calculated from mCherry gMFI values relative to the corresponding R0 condition. Data are shown as mean ± SEM of *n* = 3 independent donors. **P* < 0.05 vs. hypothetical value of 1 by One-sample *t*-test. Source data are provided as a Source Data file.

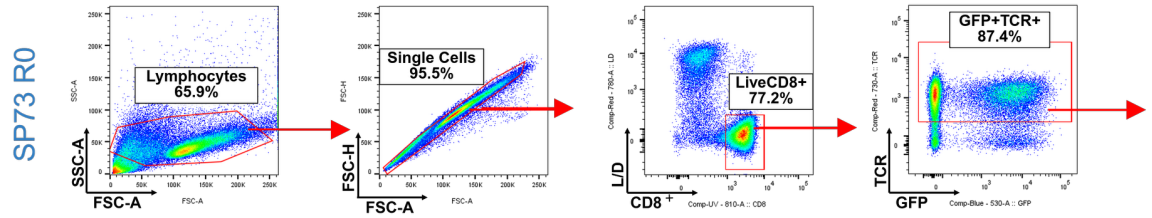
A - Gate Strategy for mCherry expression - in vitro (Dynabeads) - CD8+ T Cells



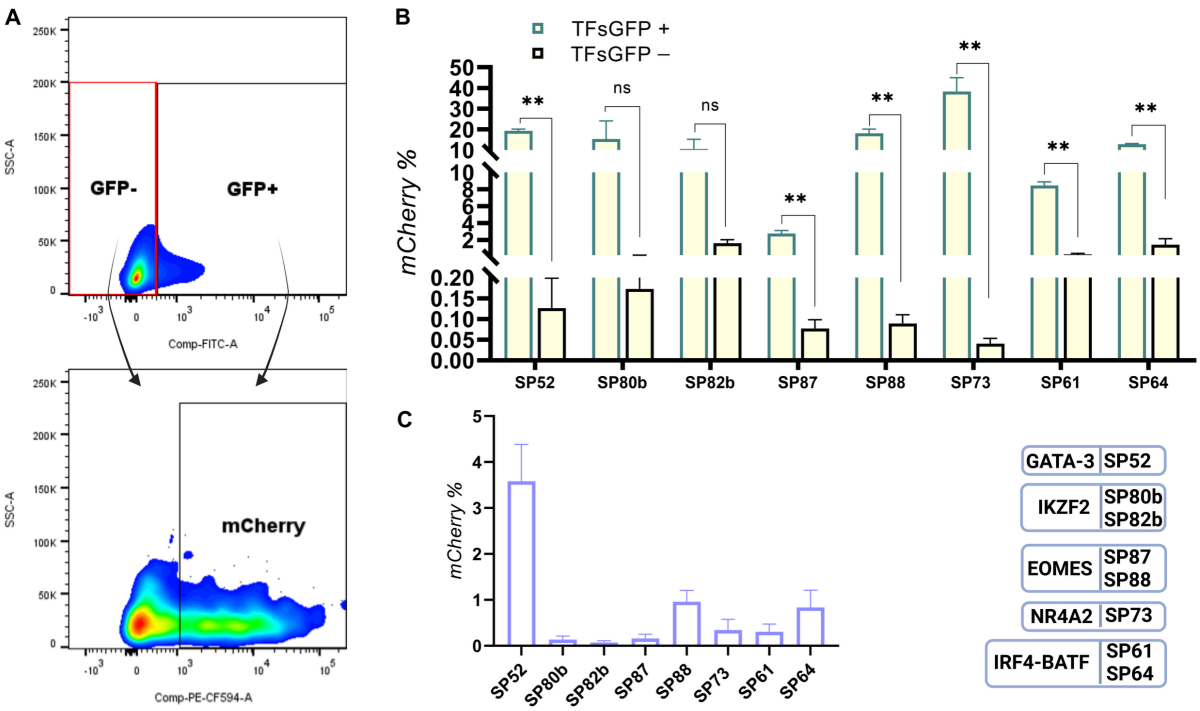
B - Gate Strategy for Activation/Dysfunction Markers - in vitro (Dynabeads) - CD8+ T Cells



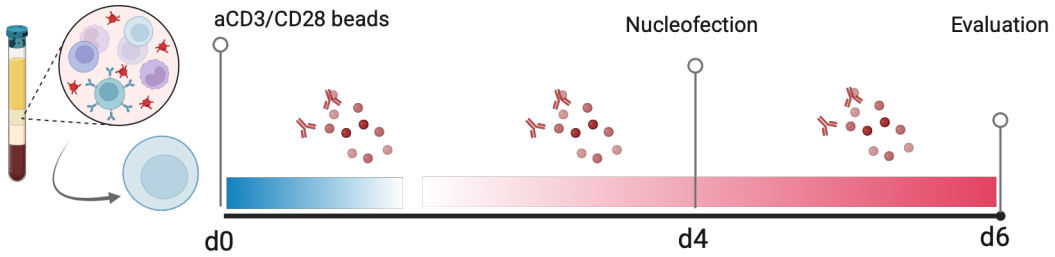
C - Gate Strategy for mCherry expression and T-cell state - co-culture/in vivo - CD8+ T Cells



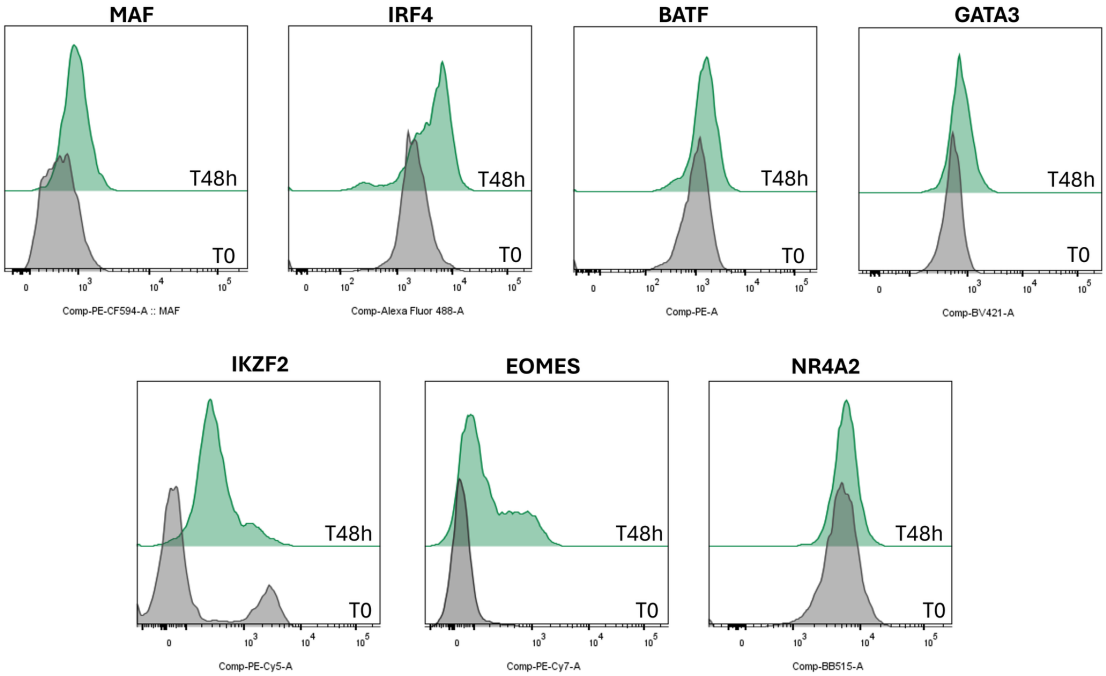
Supplementary Figure 40. Flow cytometry gating strategy in engineered T cells experiments. (A) Gating hierarchy used to evaluate mCherry expression applied to transduced CD8⁺ T cells. Cells were gated on the LNGFR+ population (reporter for transduction) to assess mCherry activity upon Dynabeads™ stimulation. (**B**) Gating hierarchy used to evaluate the expression of markers associated with both T-cell activation and chronic stimulation with anti CD3/CD28 Dynabeads™. (**C**) Gating hierarchy used to evaluate mCherry expression and T cell state in complex environments (co-culture and in vivo). Within this parent gate, terminally exhausted (Ttex) CD8⁺ T cells were further identified by sequential gating on LAG-3⁺ TIGIT ⁺, CD39⁺ PD-1⁺ and NR4A2⁺ TIM-3⁺ populations, as detailed in Figure 7 and in the Methods section.



Supplementary Figure 41. Synthetic promoter activation in Jurkat cells depends on successful transcription factor expression. (**A**) Representative gating strategy used to identify GFP-positive and GFP-negative Jurkat cell populations following co-electroporation of synthetic promoter (SP)–mCherry constructs with TF–P2A–GFP reporter plasmids. For IRF4-BATF condition, a dual GFP/BFP-positive gate strategy was applied. mCherry expression was subsequently quantified within each gated population. (**B**) Percentage of mCherry-positive cells among GFP-positive and GFP-negative populations for the indicated synthetic promoters. Promoter activation was consistently enriched within the TF–GFP-positive population, demonstrating that SP activation depends on successful expression of the cognate transcription factor. (**C**) Basal activity (OFF-state leakiness) of representative synthetic promoters co-electroporated with a GFP control plasmid in the absence of TF overexpression. Under these conditions, mCherry expression remained low across most promoters. Data represent $n = 3$ independent biological replicates. Statistical significance was assessed using a paired *t*-test comparing GFP-positive and GFP-negative populations for each promoter. p value: $^{**}P < 0.005$. Source data are provided as a Source Data file.



Supplementary Figure 42. Experimental workflow for evaluating representative synthetic promoter (SP) activity under endogenous transcription factor regulation in primary human CD8⁺ T cells. Primary human CD8⁺ T cells were isolated at day 0 and stimulated with anti-CD3/CD28 Dynabeads. Following two rounds of stimulation, cells were nucleofected with selected SP–mCherry reporter plasmids on day 4. Synthetic promoter activity and endogenous transcription factor expression were assessed by flow cytometry 48 h after nucleofection (day 6). Created in BioRender. Olandesi, R. (2026) https://BioRender.com/5hvrour.



Supplementary Figure 43. Endogenous transcription factor expression in primary human CD8⁺ T cells during repeated stimulation and transient nucleofection. Representative flow cytometry histograms from one representative donor showing endogenous expression of MAF, IRF4, BATF, GATA3, IKZF2, EOMES and NR4A2 following intracellular immunostaining. T0 corresponds to cells immediately before nucleofection, after two rounds of anti-CD3/CD28 Dynabead stimulation, whereas T48h corresponds to cells collected 48 h after nucleofection, during which an additional round of CD3/CD28 stimulation was applied. Histograms illustrate the changes in endogenous TF abundance associated with continued stimulation during the transient synthetic promoter assay.

Supplementary Methods

Quantitative Real-Time PCR (qPCR)

Total RNA was extracted from transfected (HEK293) or electroporated (Jurkat) cells using standard purification methods, with concentration, purity, and integrity validated by spectrophotometric analysis. Target-specific primers were designed using online tools to span exon-exon junctions, and their specificities were confirmed in silico via BLAST and empirically verified through melt-curve analysis. The following primers pairs were used in this study: IRF4 forward 5'-AACAAACTGGAGAGAGACCAGACC-3', reverse 5'-CCTCTCCAAAGCATAGAGTCACC-3'; BATF forward 5'-CTGCACCTGGAGAGCGAAGA-3', reverse 5'-ATGAGGTTGGTGGAATGCGT-3'; MAF forward 5'-GAGACCGACCGCATCATCAG-3', reverse 5'-GGTAGCCGGTCATCCAGTAG-3'; GATA3 forward 5'-TCCTGTGCGAACTGTCAGAC-3', reverse 5'-TCGGTTTCTGGTCTGGATGC-3'; NR4A2 forward 5'-GAGTCTGATCAGTGCCCTCG-3', reverse 5'-TGGAGCCAGTCAGGAGATCA-3'; EOMES forward 5'-GTGGCAAAGCCGACAATAACA-3', reverse 5'-CAATATGCAGTCGGGGTTGG-3'; IKZF2 forward 5'-ATGGGGAAAAGCTCATGCGA-3', reverse 5'-ATAACTGGGGCCACTTCAGC-3'; 18S forward 5'-TGTGCCGCTAGAGGTGAAATT-3', reverse 5'-TGGCAAATGCTTTCGCTTT-3'.

Real-time PCR amplifications were performed on a QuantStudio 3 Real-Time PCR System (Thermo Fisher Scientific). Reactions were assembled in a total volume of 20 µL containing Applied Biosystems™ SYBR™ Green Universal Master Mix (2×) (Thermo Fisher Scientific), forward and reverse primers at a final concentration of 300 nM each, and diluted cDNA template. Nuclease-free water was used to bring the reaction to its final volume. Non-template controls (NTC) were included for every primer pair on each plate to monitor for contamination or primer-dimer artifacts. All biological samples and controls were run in technical triplicates.

The thermal cycling conditions were configured as follows: Initial Denaturation/Polymerase Activation: 95°C for 10 minutes. PCR Cycling (40 cycles): Denaturation at 95°C for 15 seconds, followed by Annealing/Extension at 60°C for 60 seconds (with fluorescence data collection). Melt Curve Stage: A continuous temperature ramp from 60°C to 95°C to verify product specificity. Cycle threshold (Ct) values were determined using the QuantStudio Design & Analysis Software. Relative gene expression levels were calculated using the delta Ct method, expressed as 2^(-delta Ct), where delta Ct = Ct (target gene) - Ct (reference gene). The 18S rRNA gene was utilized as the internal reference gene to normalize target gene expression. The resulting 2^(-delta Ct) values represent the relative abundance of the target gene normalized to the reference gene.

Supplementary Note 1

Transfer Learning and Promoter Ranking

We explored whether promoter ranking could be partially inferred from externally trained models in situations where the dataset of interest is too small to support de novo model training. Due to the cardinality of the experimental dataset (n=82), we used a transfer learning strategy, a technique where a model pre-trained on one task is adapted to a different but related task (Supplementary Figure 18A-C). In specific, we wanted to examine whether features learnt from one promoter characterization metric were predictive of another. To apply transfer learning to our SPs task we use a dataset of approximately 25,000 human promoters from the Eukaryotic Promoter Database (EPD)[1]. The EPD dataset contains characterization information for each promoter in the form of the number of ChIP-seq tags matching the 250bp region to each side of the transcription start site. The ~80-sequence T-cell promoter dataset was treated strictly as an independent test set, and no fine-tuning was performed using these data. Our approach consisted in training a deep neural network taking as input the EPD promoter sequences and predicting as output the expression in the form of ChIP-seq tag count. We then applied the network trained on EPD to the SPs, and ranked the SPs based on the fluorescence of the reporter gene. The fluorescence-based absolute mean induction of these promoters was Box-Cox transformed, and the model produced a predicted ranking of the sequences based on their expected strength. In each case, we used the mean absolute induction when the promoter is at its strongest expression to build the ranking. The correlation between the predicted and actual rankings of the SPs was 0.49, indicating a moderate but significant agreement (Supplementary Figure 18D). This result does not support strong generalization between the EPD prediction and the SP ranking. Some of the captured features may however help in generalizable predictions. Finally, we performed a preliminary exploration of model interpretability by examining how the first convolutional filter weighted positions within the strongest and weakest promoters, identifying sequence regions that contributed most to prediction (Supplementary Figure 23). Overall, the analysis shows that transfer learning can be applied in common scenarios where promoter datasets are too small to train a standalone model. Reporting a rank correlation in such cases facilitates a metric that can be applied regardless on the methods used for promoter characterization.

For the purposes of modelling exhaustion synthetic promoter 'strength' we utilized two datasets in a transfer learning strategy. Our primary dataset of 82 exhaustion SPs is not numerous enough to directly apply a deep learning approach. Transfer learning is a technique where a pre-trained model on one task is adapted to a different but related task. To apply transfer learning to the exhaustion SPs task we use a dataset of approximately 25,000 human promoters from the Eukaryotic Promoter Database (EPD)[1].

Our secondary EPD dataset contains characterization information for each promoter in the form of the number of ChIP-seq tags matching the region 250bp to each side of the transcription start site. Our approach was to begin by training a deep neural network taking as input the EPD promoter sequences and predicting as output the expression in the form of ChIP-seq tag count.

Having trained a network on the EPD promoter task we must now approach how to use this network on the exhaustion SPs task. Our difficulty lies in the different characterization metrics used to quantify promoter 'strength' in each dataset. To apply the EPD network to

the exhaustion SPs we ranked order the SPs based on fluorescence. We attempted to recapitulate the CAR-T 'strength' ranking using the EPD network.

Preprocessing:

Before use in training, all EPD promoter sequences were represented as 200 bp regions and one-hot encoded into binary matrices of shape (200, 4). To account for the wide range of expression levels, we applied a Box-Cox transformation. The Box-Cox transformation is defined as:

$$y_i^{(\lambda)} = \begin{cases} \frac{y_i^{\lambda} - 1}{\lambda} & \text{if } \lambda \neq 0, \\ \ln y_i & \text{if } \lambda = 0, \end{cases}$$

where y_i is the original data and λ is the transformation parameter. Intuitively a Box-Cox transform attempts to move the data closer to a normal distribution. We follow the lead of Zrimec et al.[2] in this approach. Training on normally distributed data confers the benefit of maintaining balanced weight updates during backpropagation leading to faster convergence of the network.

To augment the EPD dataset, we applied a sliding window approach with a 5 bp stride over each 200 bp sequence. For each stride position, we added the original sequence, its complement, and its reverse complement. This approach has the benefit of providing us more training and test instances and attempting to train for translational invariance, allowing the network to identify predictive motifs even when their position on the strand is different.

Finally, we split the augmented EPD dataset for training and testing with an 80/20 split. After pre-processing and augmentation, we have 351,540 training instances and 87,885 test instances.

Our EPDNet architecture comprises a hybrid approach combining convolutional layers and gated recurrent units (Supplementary Figure 18A-C). Convolutional layers slide a weighted window across the encoded sequence. Crucially convolutional layers allow the network to decide for itself which motifs are most predictive of promoter 'strength'. It is necessary to take this approach given the lack of knowledge about the structure of the promoter 'grammar'.

After each convolutional layer we apply batch normalization. This technique normalizes the input of each layer by adjusting and scaling the activations. It helps stabilize and accelerate training by reducing internal covariate shift, making the network less sensitive to initialization and learning rate. Max pooling is used to down-sample the content of each convolution window extracting only the most highly weighted feature. This helps to reduce computation and control overfitting.

Gated Recurrent Units (GRUs) are designed to efficiently capture sequential dependencies in data. They use gating mechanisms to control the flow of information, allowing the network to maintain and update relevant context over longer sequences. By combining convolutions and GRUs our network can leverage both short and long range motifs in its predictions. Finally, we process the input through two dense layers to a final ReLU output which gives the predicted promoter 'strength'.

We train our network to minimize the Huber loss function using the Adam optimizer on an Nvidea L4 GPU. The Huber loss function is defined as:

$$L_\delta(y,\hat{y}) = \begin{cases} \frac{1}{2}\cdot(y-\hat{y})^2 & \text{if } |y-\hat{y}| \le \delta, \\ \delta\cdot\left(|y-\hat{y}| - \frac{1}{2}\delta\right) & \text{if } |y-\hat{y}| > \delta, \end{cases}$$

where y is the true value, $\hat{y}$ is the predicted value, and δ is a threshold parameter. Intuitively, the Huber loss attempts to be sensitive to outliers by switching between mean squared error for small errors and mean absolute error for large errors. We follow Kotopka et al.[3] in the use of the Huber loss function in a sequence-to-function network.

In Supplementary Figure 18A-C, we can see the progression during training of the EPD network. Training proceeds smoothly with no signs of over-fitting. Finally, we apply the trained network to the test data achieving a test loss of 0.03 and a mean absolute error or 0.25.

Real vs Predicted Rank

After training on the EPD dataset, the neural network was applied to the 82 Synthetic Promoters. The fluorescence-based absolute mean induction of these promoters was Box-Cox transformed, and the model produced a predicted ranking of the sequences based on their expected 'strength'. In each case we use the mean absolute induction when the promoter is at its strongest expression to build the ranking. The correlation between the predicted and actual rankings of the Synthetic Promoters was 0.49, indicating a moderate but significant agreement. Supplementary Figure 18D shows the exact differences in the correct and predicted ranking for each individual promoter. This result demonstrates that the transfer learning approach, leveraging a large dataset of human promoters, effectively captures relevant features for predicting gene expression in a distinct Synthetic Promoter context despite the limited dataset size.

Supplementary References

1. Dreos, R., Ambrosini, G., Groux, R., Cavin Périer, R. & Bucher, P. The eukaryotic promoter database in its 30th year: focus on non-vertebrate organisms. *Nucleic Acids Res.* **45**, D51–D55 (2017).

2. Zrimec, J. *et al.* Deep learning suggests that gene expression is encoded in all parts of a co-evolving interacting gene regulatory structure. *Nature Communications* **11**, 6141 (2020).

3. Kotopka, B. J. & Smolke, C. D. Model-driven generation of artificial yeast promoters. *Nature Communications* **11**, 2113 (2020).

4. Starkey, F. et al. A modular synthetic promoter toolkit for programmable gene control in activated and chronically stimulated T cells, GitHub, DOI 10.5281/zenodo.22079494 (2026).