

Opposing NF- κ B-associated transcriptional changes accompany high CAR density and signaling rest in CAR-T cells

Qian Wang

UCLA

Zhoulin Chen

UCLA

Haohan Yang

UCLA

Youcheng Yang

UCLA

Shubing Wang

brianbrian@ucla.edu

UCLA

Research Article

Keywords:

Posted Date: October 15th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11157821/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

The intensity and persistence of chimeric antigen receptor (CAR) signaling shape T cell state, but the transcriptional relationship between receptor density and signaling rest remains unclear. Here we combined literature-guided hypothesis generation with analysis of public CAR-T cell datasets to identify a shared response across these contexts. High CAR density and four-day signaling rest were associated with opposing enrichment of activation-related programs, including NF- κ B-associated transcription. The opposing transcriptional response included the NF- κ B feedback regulator NFKBIA and was accompanied by broad chromatin remodeling, with AP-1-related motifs enriched at regions that became less accessible after signaling rest. Virtual NFKBIA knockout prioritized TCF7 as a recurrent and robust network-response candidate across donor-specific regulatory models, supported by consistent rankings across subsampling runs. These findings define a shared transcriptional program across distinct CAR-T signaling contexts and nominate candidate regulators for experimental investigation of CAR-T cell plasticity.

Introduction

Chimeric antigen receptor (CAR)-T cell state is shaped by the intensity and persistence of engineered receptor signaling. Memory-associated features of infusion products have been linked to therapeutic response, whereas exhaustion-associated transcriptional programs have been associated with poorer outcomes^{1,2}. Consistent with this relationship, antigen-independent tonic CAR signaling has been shown to drive progressive T-cell dysfunction and exhaustion-associated states³. More directly, analyses of BCMA-targeting CAR-T cells have associated high receptor density with increased tonic signaling, enhanced acute cytotoxicity, greater differentiation and reduced persistence⁴. These observations suggest that receptor abundance is not simply a measure of transgene expression, but a variable associated with the balance between immediate effector activity and longer-term cellular state.

Interrupting CAR signaling provides a complementary view of this relationship. Transient receptor downregulation or pharmacological inhibition can restore function and remodel chromatin in dysfunctional CAR-T cells⁵. However, transcriptional reinvigoration does not necessarily imply complete epigenetic resetting, and the extent of plasticity can depend on cellular context^{6,7}. Whether transcriptional programs associated with high CAR density change in the opposite direction after signaling rest therefore remains unclear. NF- κ B-associated transcription provides a candidate axis for this comparison. NFKBIA, which encodes the inhibitor I κ B α , is itself induced by NF- κ B as part of a negative-feedback circuit, such that increased NFKBIA expression can accompany pathway engagement rather than reduced signaling^{8,9}. NF- κ B signaling can also support CAR-T cell survival in specific costimulatory contexts³, arguing against interpreting pathway activity as uniformly detrimental. We therefore focused on coordinated NF- κ B-associated transcription rather than assigning functional meaning to an individual feedback regulator.

AI-assisted literature analysis can help organize heterogeneous evidence into explicit, testable hypotheses^{10–13}. We used BioPathfinder to trace source observations through curated claims, hypothesis generation and repeated evidence appraisal^{14,15}. Public transcriptomic and chromatin-accessibility datasets then allowed a focused evaluation of the resulting biological question. This separation makes the provenance of a hypothesis accessible while leaving its biological support to the subsequent analyses. Guided by this framework, we compared CAR-density and signaling-rest datasets within their respective study designs. We identified a shared NF-κB-associated transcriptional program with opposing changes across the two contexts, examined chromatin accessibility after rest and used virtual NFKBIA knockout to nominate candidate network responses. Together, these analyses define a transcriptional relationship across distinct signaling contexts and motivate direct tests of its connection to CAR-T cell plasticity.

Results

BioPathfinder connected literature evidence to a hypothesis about CAR density and the reversibility of CAR-T cell dysfunction (Fig. 1a). Three input papers yielded 266 source windows, of which 200 were selected; 15 curated claims were reduced to 11 for hypothesis generation (Fig. 1b). Ten hypotheses underwent three model reviews, and each was scored across seven dimensions of scientific rationale and evidential support. Hypotheses were reviewed in descending order of overall score for expert assessment of *in silico* validation potential, leading to the selection of H08, a top-three-ranked hypothesis with high feasibility and novelty (Fig. 1c). Its evidence trace linked density-associated tonic signaling and dysfunction to molecular profiling in BCMA-targeting CAR-T cells⁴, motivating the original hypothesis that CAR-density-associated chromatin states might distinguish reversible from persistent dysfunction (Fig. 1d). H08 centered on the persistence of a dysfunctional state after stimulation ceased, distinguishing its biological question from hypotheses focused on marrow entry, expansion or microenvironmental suppression. This focus linked receptor abundance to cellular plasticity and provided a tractable starting point for molecular comparison. We therefore asked whether programs elevated with CAR density were attenuated during signaling rest, using each study's internally controlled contrast. The available datasets enabled focused evaluation of the transcriptional programs underlying this hypothesis.

High CAR density and signaling rest were associated with opposing transcriptional programs. Single-cell profiles from the BCMA CAR-density study resolved five transcriptional states across 25,767 CAR-T cells from three donors (Fig. 2a,b), with a fixed density-associated expression score varying across the embedding (Fig. 2c). Memory-like cells combined TCF7 and CCR7 expression with low cytotoxic-marker expression; effector-like cells showed stronger NKG7, GZMB and PRF1 expression. Cycling cells expressed MKI67 and TOP2A, whereas the dysfunction-associated state showed elevated PDCD1; mixed-state cells combined less clearly separated programs (Fig. 2b). Cycling cells had the highest median density score, linking the density-associated signature to a proliferative transcriptional state (Fig. 2c). Thus, the signature captured variation in proliferative and activation-associated transcription

across the annotated states. Donor-aware analyses of high versus low density and an independent four-day rest experiment identified opposing NFKBIA expression changes (Fig. 2d)^{4,5}. NFKBIA increased by 1.53 log₂ units with high density and decreased by 1.74 log₂ units after rest (FDR = 3.79×10^{-8} and 0.0248, respectively). Six specified programs, including TNFA signaling via NF- κ B, MTORC1, MYC targets, the unfolded protein response and an exhaustion-related signature, showed opposing enrichment and joint false discovery rate (FDR) < 0.05 in both contrasts (Fig. 2e). TNFA–NF- κ B enrichment was positive for density and negative for rest (Fig. 2f). Among the six displayed programs, MYC targets V1 showed the strongest density enrichment (NES = 2.55), whereas TNFA–NF- κ B showed the strongest opposing rest enrichment (NES = - 2.42), closely followed by MYC targets V2 (NES = - 2.40). Paired expression profiles supported increased NFKBIA with high density and decreased expression after rest (Fig. 2g). Of 51 genes shared by the two pathway leading edges, 23 had significant positive density effects and negative rest effects (Fig. 2h). The core included CSF1 and CSF2, EGR2 and EGR3, and NFKBIA, NFKBIE, NFKB2 and RELB, linking cytokine-associated transcription with inducible regulators and NF- κ B pathway components. This shared core links the two signaling contexts at the transcriptional level.

Signaling rest was also accompanied by broad directional changes in chromatin accessibility. Among 86,212 peaks jointly testable in ON – OFF and four-day rest – ON comparisons, 9,254 had significant opposing effects (FDR $\leq$ 0.05 in both): 5,328 opened with signaling and closed after rest, whereas 3,926 followed the converse pattern (Fig. 3a). AP-1-related motifs, including BATF::JUN, BATF and BATF3, were enriched in the first class against a background stratified by sequence composition, peak length and baseline accessibility (Fig. 3b). Principal component analysis summarized the 12 samples from three donors (**Supplementary Fig. 1a**), and 97.8–99.5% of peaks in the selected classes supported both expected directions within individual donors (**Supplementary Fig. 1b**). Their opposing contrast effects are shown in **Supplementary Fig. 1c**. The two return classes separated regions associated with ongoing signaling from regions that gained accessibility during rest. Enrichment of several related AP-1 motifs in the signaling-open class concentrated the regulatory association within a coherent motif family. Together, these results identify a reproducible chromatin-accessibility program associated with signaling and its reversal during rest, with AP-1-related regulatory elements prominently enriched among signaling-responsive regions.

Virtual NFKBIA knockout prioritized TCF7 as a donor-dependent network-response candidate. We perturbed separate donor networks inferred from a shared set of 1,000 genes in 25,767 CAR-T cells (Fig. 4a)¹⁶. Nineteen non-target genes met model FDR < 0.05 under both subsampling seeds in at least one donor (Fig. 4b). Candidate responses differed markedly among donors: D10 included S100A10, S100A11 and IFITM2, D14 included CD74 and multiple HLA class II genes, and D18 included CCR7, CD27 and TXNIP. TCF7 was the sole candidate meeting the two-seed criterion in two donors, ranking second in both D14 runs and fourth and third in D18 (Fig. 4c). The displayed local networks summarized donor-specific relationships among candidates (Fig. 4d). Perturbation ranks were concordant between seeds within each donor (Spearman ρ = 0.869–0.947; **Supplementary Fig. 2a**). Within-donor rank stability reinforced the reproducibility of each donor's response profile across network subsampling. This

convergence supports prioritizing TCF7 when examining how perturbations reshape the inferred CAR-T regulatory state. These results identify a reproducible network response and support prioritizing TCF7 as a candidate network response to NFKBIA perturbation.

Discussion

Our analysis identifies a shared transcriptional axis linking CAR receptor abundance with signaling persistence across distinct CAR-T experimental systems. High CAR density and signaling rest shifted activation-associated programs in opposite directions, including TNFA–NF- κ B, MYC, MTORC1 and unfolded-protein-response signaling. A 23-gene NF- κ B-associated core further defined a compact molecular signature connecting these contexts. NFKBIA is particularly informative because its induction can accompany NF- κ B pathway engagement through negative feedback rather than simply reflecting pathway inhibition. These findings support a model in which receptor abundance and signaling interruption engage reciprocal components of a common CAR-T transcriptional state.

Chromatin accessibility analysis extended this relationship to the regulatory level. Thousands of regions changed in opposing directions with signaling and rest, with strong consistency across donors. AP-1-related motifs, including BATF::JUN, BATF and BATF3, were enriched among signaling-responsive regions, implicating AP-1-associated regulatory programs in signaling-dependent chromatin remodeling. Virtual NFKBIA perturbation further prioritized TCF7 across donor-specific networks, nominating a potential connection between NF- κ B-associated feedback and transcriptional programs associated with less differentiated T-cell states. Together, these analyses define complementary transcriptional, chromatin and network features for studying CAR-T state transitions.

This study extends an emerging paradigm in which AI-guided hypothesis generation is coupled with expert assessment and data-driven validation. BioPathfinder was developed to transform heterogeneous literature evidence into traceable, mechanistically grounded hypotheses for CAR-T engineering¹⁵. A related application subsequently identified C1Q-associated adaptive myeloid remodeling during early response to BCMA CAR-T therapy through hypothesis generation followed by expert-guided analysis of patient single-cell data¹⁷. Together with the present study, these examples suggest that AI-assisted hypothesis generation can provide a reproducible starting point for biological discovery when paired with expert evaluation of novelty, feasibility and appropriate validation strategies. This framework may be extended beyond conventional CAR-T biology to broader immune-cell engineering. Biomaterial-based cancer immunotherapy increasingly integrates material properties, molecular cues and delivery systems to regulate immune-cell activation, differentiation, metabolism and persistence¹⁸. The same framework may also generalize unconventional T cells^{19,20}. Such systems create a substantially larger intervention space in which evidence-guided AI approaches could connect engineering variables to molecular and functional cell states.

Several limitations motivate direct experimental validation. CAR density and signaling rest were examined in distinct experimental systems, and virtual perturbation nominates network responses rather

than establishing direct regulatory relationships. Future studies should therefore manipulate CAR density and signaling duration within matched donor and CAR contexts, integrate longitudinal transcriptomic and chromatin profiling with functional rechallenge, and test NFKBIA perturbation with rescue. Such experiments will determine whether the regulatory programs identified here directly contribute to reversible CAR-T cell-state transitions.

Methods

Data sources and study design

Public RNA-seq and ATAC-seq data were obtained from two CAR-T studies (Table 1). GSE197843 and GSE197848 comprise bulk and single-cell profiles of BCMA CAR-T cells carrying an ARI-0002h-derived 4-1BB/CD3 ζ CAR. Bulk comparisons paired high- and low-density cells within donor and CD4/CD8 subset. GSE164949 and GSE164946 comprise RNA-seq and ATAC-seq profiles of HA.28 ζ .FKBP CAR-T cells subjected to transient signaling rest⁵. The primary comparison contrasted four-day rest with continuous signaling at day 15. Analyses accounted for donor pairing within each dataset. Sample inclusion and treatment metadata are provided in the Supplementary Data.

Table 1
Overview of the public datasets and experimental designs used in this study.

GEO accession	Assay	Included donors and samples	Design / conditions	Analysis in this study
GSE197843	Bulk RNA-seq	6 donors; 24 samples	High versus low CAR density, paired within donor and CD4/CD8 subset	Density-associated differential expression and pathway analysis
GSE197848	Single-cell RNA-seq	3 donors; 3 samples; 25,767 cells	BCMA CAR-T cells from D10, D14 and D18	Cell-state annotation, density-associated scoring and virtual knockout
GSE164949	Bulk RNA-seq	3 donors; 6 primary-comparison samples	Four-day rest versus continuous signaling at day 15	Rest-associated differential expression and pathway analysis
GSE164946	ATAC-seq	3 donors; 12 samples	OFF, ON, four-day rest and eight-day rest at day 15	Chromatin accessibility, return-peak classes and motif analysis

GSE197843 and GSE197848 originate from the same CAR-density study (Rodriguez-Marquez et al., 2022), using an ARI-0002h-derived BCMA CAR with 4-1BB/CD3 ζ signaling domains. GSE164949 and GSE164946 originate from the same signaling-rest study (Weber et al., 2021), using HA.28 ζ .FKBP CAR-T cells; the ATAC-seq profiles are from CD8-positive cells. RNA-seq counts for GSE164949 refer to the

primary four-day-rest versus ON comparison; 21 samples contributed across primary and secondary comparisons. Sample identifiers and inclusion details are provided in the Supplementary Data.

Bulk transcriptomic analysis

Bulk RNA-seq data from the CAR-density study were analyzed using edgeR with TMM normalization and robust quasi-likelihood modeling. Donor and CD4/CD8 subset were included in the model, and the primary analysis estimated the overall high- versus low-CAR-density effect across both subsets. For the signaling-rest dataset, deposited log2-transformed TPM values were analyzed using limma with donor pairing incorporated into the model. Lowly expressed genes were excluded before differential-expression testing, and multiple-testing correction was performed using the Benjamini–Hochberg method. Robustness of the density-associated effects was further assessed by repeating the analysis after sequentially omitting individual donors.

Single-cell analysis

Single-cell profiles were retained for cells with at least 1,000 UMIs, 500–6,000 detected genes, mitochondrial counts $\leq 10\%$ and detectable CD3D or CD3E. Doublets were screened with Scrublet. Donor-aware scVI models were trained on highly variable genes using a negative-binomial likelihood. UMAP and Leiden clustering were applied to the latent representation, with alternative clustering resolutions and model seeds used to assess stability.

Cell states were assigned from predefined marker programs using cluster-level expression, detection frequency and separation between program scores. Marker expression was library-size normalized and log transformed; state-level summaries were weighted by cell abundance. The density signature comprised genes with $FDR \leq 0.05$ and concordant effects in at least five of six estimable donor-omission fits. For each cell, expression values were converted to percentile ranks with average ranks for ties. The density score was the mean percentile rank of the upregulated signature minus that of the downregulated signature, requiring at least 70% coverage of each set. Marker sets, signature membership and model settings are provided in the Supplementary Data.

Gene set enrichment and shared core

Gene set enrichment analysis was performed to compare pathway-level transcriptional changes associated with CAR density and signaling rest. We tested Hallmark pathways together with exhaustion-related C7 immunologic signatures using preranked GSEA. Enrichment significance was corrected for multiple testing using the Benjamini–Hochberg method. To identify a shared NF- κ B-associated transcriptional program, we compared the leading-edge genes of the TNFA_SIGNALING_VIA_NFKB pathway across the two analyses and retained genes that were significantly increased with high CAR

density and decreased following signaling rest ($\text{FDR} \leq 0.05$ in both datasets). Full gene-set definitions and enrichment results are provided in the Supplementary Data.

ATAC and motif analysis

ATAC-seq data were analyzed to identify regions whose accessibility changed with CAR signaling and reversed after four days of signaling rest. Differential accessibility was assessed within donors using edgeR, with TMM normalization and donor-adjusted quasi-likelihood models. We defined “return” peaks as regions that changed significantly between OFF and ON conditions and then changed in the opposite direction after four days of rest, with $\text{FDR} \leq 0.05$ in both comparisons. This fixed set of peaks was used for downstream visualization and donor-level consistency analyses. To identify regulatory programs associated with these accessibility changes, we tested for transcription factor motif enrichment within each return-peak class. JASPAR 2024 vertebrate motifs were scanned using motifmatchr, and enrichment was evaluated relative to jointly testable background peaks while accounting for GC content, peak length and baseline accessibility. Multiple testing was controlled using the Benjamini–Hochberg procedure. Coordinate handling and additional sensitivity analyses are described in the Supplementary Methods.

Promoter mapping and cross-modal comparison

Promoters were defined from GENCODE v19 transcription start sites as strand-aware intervals extending 2 kb upstream and 500 bp downstream. Overlapping peaks were assigned to the nearest transcription start site by midpoint distance, with gene and transcript identifiers resolving ties. Each gene was represented by its nearest assigned testable peak; remaining ties were resolved by peak coordinates. Gene symbols linked rest-associated RNA and accessibility effects. Cross-modal agreement was summarized using Spearman correlation and effect-direction concordance among genes covered in both modalities.

Virtual knockout and sensitivity analyses

Virtual knockout analyses used scTenifoldKnn and scTenifoldNet to infer donor-specific networks from a shared gene set. Eligible genes were detected in at least 5% of cells and had positive variance in every donor; mitochondrial and ribosomal genes were excluded. Genes were ranked by their mean variance rank across donors after library-size normalization and log transformation. The top 1,000 genes were selected, with NFKBIA appended when required. Selected counts were normalized to CPM using full-library totals. Networks were inferred from repeated cell subsamples, with seeds 42 and 1729 varying subsampling while tensor initialization remained fixed. Full network parameters are provided in the Supplementary Data.

Virtual knockout set NFKBIA outgoing weights to zero and quantified changes in manifold position between the original and perturbed networks. Candidates required model FDR < 0.05 under both subsampling seeds in at least one donor. Perturbation ranks were compared over the shared gene space. Local network displays retained edges with concordant signs across seeds and ranked them by mean absolute weight.

BioPathfinder evidence processing and hypothesis review

BioPathfinder processed three source studies^{4,21,22}. Literature passages were curated into claims and selected for hypothesis generation using field-scoped evidence scoring. The planner used gpt-5.6-luna as its LLM backbone to generate traceable hypotheses linking source-supported premises to proposed biological relationships. Structured outputs underwent schema and provenance checks, with model-derived priors labeled separately. Ten hypotheses were evaluated in three repeated reviews using gpt-5.6-sol. Primary merit was the mean score across question fit, claim clarity, evidence-bridge support, novelty, mechanistic depth, decision value and risk balance. H08 was selected for computational follow-up. Source metadata, selection records, prompt policies, hypotheses and review scores are provided in the Supplementary Data. Biomni was used to support the bioinformatic analyses associated with BioPathfinder-generated hypotheses¹¹.

Declarations

Ethical Approval and Accordance

Not applicable. This study involved secondary analysis of publicly available datasets and did not involve participant recruitment or the collection of new human samples.

Consent to participate

Not applicable.

Consent to publish

Not applicable.

Data availability

The public sequencing datasets analyzed in this study are available in the NCBI Gene Expression Omnibus (GEO): GSE197843 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE197843>) contains bulk RNA-seq data from the CAR-density study; GSE197848 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE197848>) contains single-cell RNA-seq data from the same study; GSE164949 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE164949>) contains RNA-seq data from the signaling-rest study; and GSE164946

(<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE164946>) contains ATAC-seq data from the signaling-rest study.

Reference resources used in the analyses include the JASPAR 2024 CORE vertebrate non-redundant motif collection

(https://jaspar.elixir.no/download/data/2024/CORE/JASPAR2024_CORE Vertebrates_non-redundant_pfms_jaspar.txt), GENCODE human release 19

(https://www.gencodegenes.org/human/release_19.html), and the Hallmark (<https://www.gsea-msigdb.org/gsea/msigdb/human/collections.jsp#H>) and C7 immunologic signature (<https://www.gsea-msigdb.org/gsea/msigdb/human/collections.jsp#C7>) collections of the Molecular Signatures Database (MSigDB). The gene-set definitions used in this study are provided in the Supplementary Data. Sample annotations, processed analysis results, virtual-knockout outputs, and hypothesis-review scores supporting this study are provided in the Supplementary Data.

Code availability

BioPathfinder source code and analysis scripts are available from the BioPathfinder GitHub repository at <https://github.com/naiQAQian/BioPathfinder>¹⁵. Analysis scripts used in this study are also provided in the Supplementary Data.

Author contributions

S.W., Q.W. and Z.C. designed the study and analyses. Q.W. performed the AI experiments. Q.W., Z.C. and H.Y. performed the bioinformatics analyses. S.W. and Y.Y. supervised the study.

Funding

The authors received no funding for this study.

Competing interests

The authors declare no competing interests.

References

1. Fraietta, J. A. *et al.* Determinants of response and resistance to CD19 chimeric antigen receptor (CAR) T cell therapy of chronic lymphocytic leukemia. *Nature Medicine* 2018 24:5 **24**, 563–571 (2018).
2. Deng, Q. *et al.* Characteristics of anti-CD19 CAR T cell infusion products associated with efficacy and toxicity in patients with large B cell lymphomas. *Nat. Med.* **26**, 1878–1887 (2020).
3. Long, A. H. *et al.* 4-1BB costimulation ameliorates T cell exhaustion induced by tonic signaling of chimeric antigen receptors. *Nature Medicine* 2015 21:6 **21**, 581–590 (2015).

4. Rodriguez-Marquez, P. et al. CAR density influences antitumoral efficacy of BCMA CAR T cells and correlates with clinical outcome. *Sci. Adv.* **8**, 514 (2022).
5. Weber, E. W. et al. Transient rest restores functionality in exhausted CAR-T cells through epigenetic remodeling. *Science* (1979). **372**, (2021).
6. Philip, M. et al. Chromatin states define tumour-specific T cell dysfunction and reprogramming. *Nature* 2017 545:7655 **545**, 452–456 (2017).
7. Pauken, K. E. et al. Epigenetic stability of exhausted T cells limits durability of reinvigoration by PD-1 blockade. *Science* (1979). **354**, 1160–1165 (2016).
8. Hoffmann, A., Levchenko, A., Scott, M. L. & Baltimore, D. The I κ B-NF- κ B signaling module: Temporal control and selective gene activation. *Science* (1979). **298**, 1241–1245 (2002).
9. Sun, S.-C., Ganchi, P. A., Ballard, D. W. & Greene, W. C. NF- κ B Controls Expression of Inhibitor I κ B α : Evidence for an Inducible Autoregulatory Pathway. *Science* (1979). **259**, 1912–1915 (1993).
10. Ghareeb, A. E. et al. Robin: A multi-agent system for automating scientific discovery. *Nature* 2026 1–3 (2025) doi:<https://doi.org/10.1038/s41586-026-10652-y>.
11. Huang, K. et al. Autonomous biomedical research with an artificial intelligence agent. *Science* (1979). **393**, (2026).
12. Ning, J. et al. UniMedVL: Unifying Medical Multimodal Understanding and Generation through Observation-Knowledge-Analysis. <https://arxiv.org/pdf/2510.15710> (2025).
13. Deng, Z. et al. Project Imaging-X: A Survey of 1000+ Open-Access Medical Imaging Datasets for Foundation Model Development. *Chinese Academy of Sciences* **11**, (2026).
14. Nan, H., Shen, X., Yang, Y., Wang, S. & Li, Y. R. Large language model-guided CAR-T in silico platform for cytokine optimization in liver cancer with low antigen density. *Journal of Hematology & Oncology* 2026 19:1 **19**, 49- (2026).
15. Wang, S. et al. BioPathfinder: Evidence-guided multi-agent platform enables hypothesis discovery for CAR-T engineering. *bioRxiv* 2026.07.15.738646 (2026) doi:10.64898/2026.07.15.738646.
16. Osorio, D. et al. scTenifoldKnk: An efficient virtual knockout tool for gene function predictions via single-cell gene regulatory network perturbation. *Patterns* **3**, 100434 (2022).
17. Wang, S., Wang, Q., Li, Y.-R. & Li, S. C1Q-associated adaptive myeloid remodelling accompanies early response to BCMA CAR-T therapy in multiple myeloma. *bioRxiv* 2026.08.23.746487 (2026) doi:10.64898/2026.08.23.746487.
18. Liu, Z. et al. Engineering biomaterials for next-generation cancer immunotherapy: From cell manufacturing to in vivo modulation. *Cell Biomaterials* **2**, 100380 (2026).
19. Li, Y. R., Nan, H., Yang, Y., Wang, S. & Yang, L. In situ programming of unconventional T cells. *Trends Biotechnol.* **0**, (2026).
20. Li, Y.-R. et al. Stem cell mitochondrial transfer rejuvenates CAR-NKT cell metabolism and antitumor activity. <https://doi.org/10.21203/RS.3.RS-8953703/V1> (2026) doi:10.21203/RS.3.RS-8953703/V1.

21. Wang, Y. *et al.* Stressed target cancer cells drive nongenetic reprogramming of CAR T cells and solid tumor microenvironment. *Nature Communications* 2023 14:1 **14**, 5727- (2023).

22. Yan, Z. X. *et al.* Cholesterol efflux from C1QB-expressing macrophages is associated with resistance to chimeric antigen receptor T cell therapy in primary refractory diffuse large B cell lymphoma. *Nature Communications* 2024 15:1 **15**, 5183- (2024).

Figures

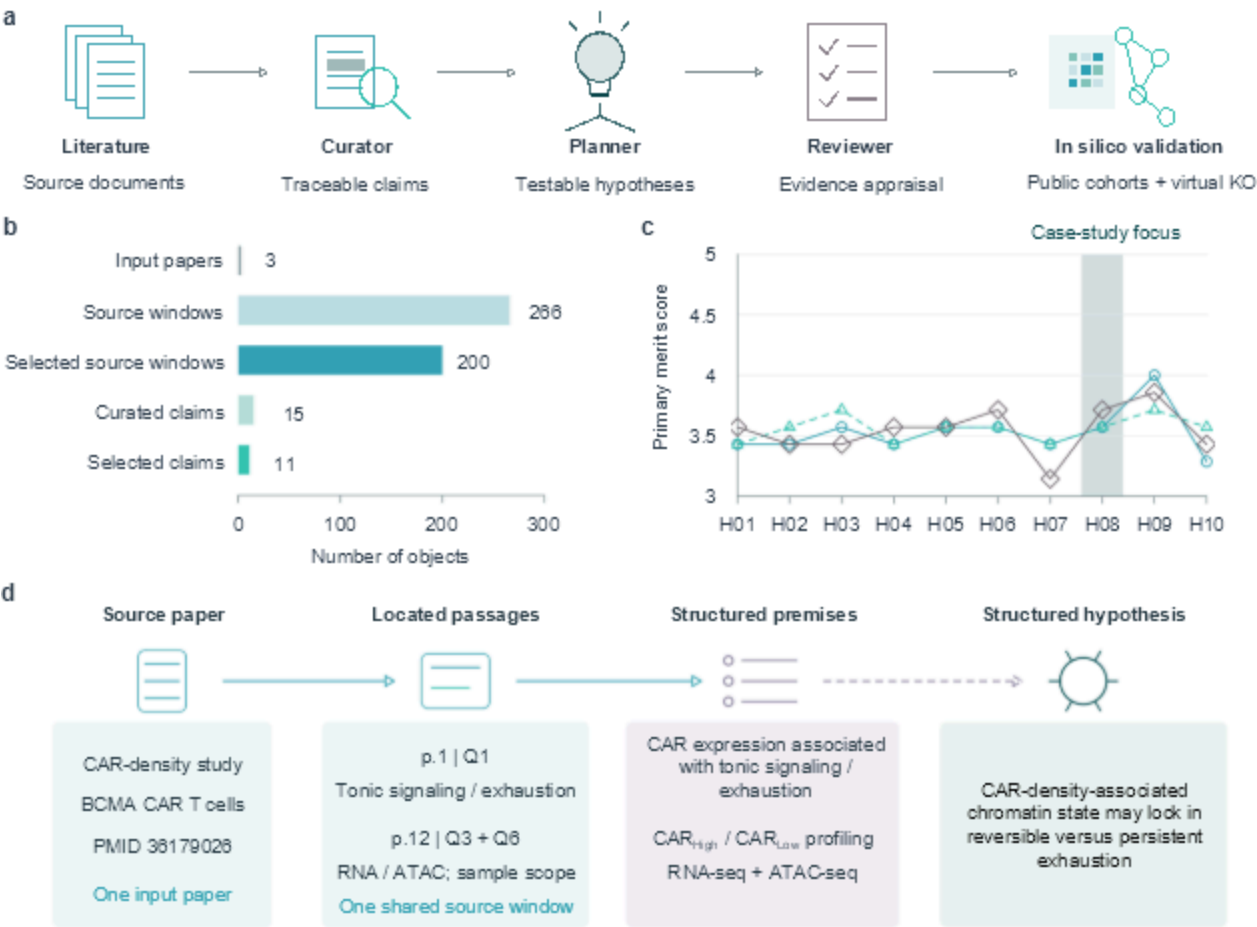


Figure 1

Literature-guided hypothesis generation and evidence tracing with BioPathfinder.

a, Overview of the BioPathfinder workflow, from literature curation and hypothesis generation to evidence review and in silico validation. **b**, Summary of evidence selection from three input studies, including source windows and curated claims retained for hypothesis generation. **c**, Merit scores of ten generated hypotheses across three independent model reviews. H08, highlighted as the case-study focus, ranked among the top candidates and was selected after expert assessment of novelty and feasibility for in silico validation. **d**,Evidence trace underlying H08, linking observations from a BCMA CAR-T density study to the hypothesis that CAR-density-associated chromatin states may contribute to reversible or persistent dysfunction.

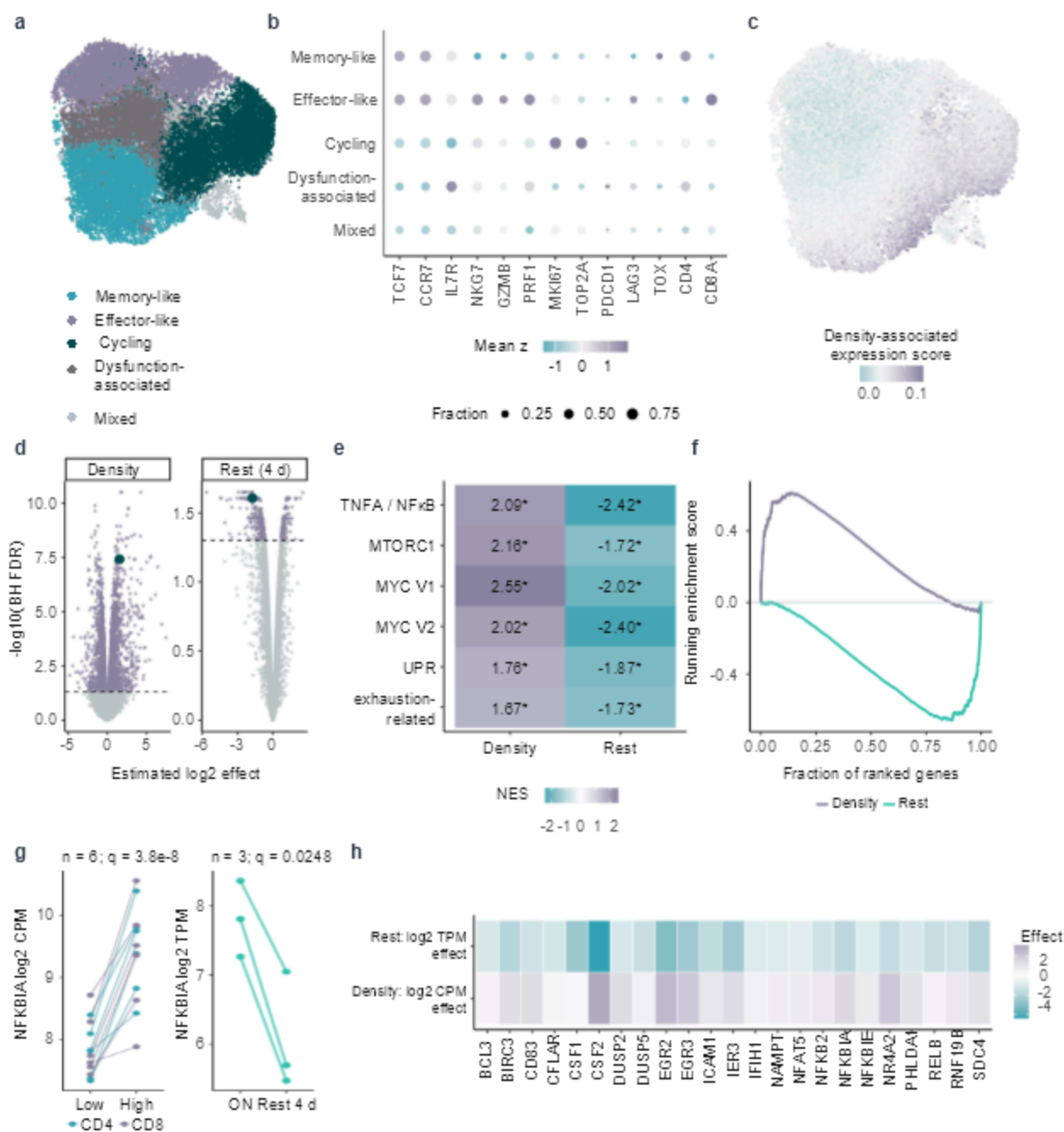


Figure 2

High CAR density and signaling rest are associated with opposing transcriptional programs.

a, UMAP of 25,767 CAR-T cells from three donors, colored by five marker-defined transcriptional states. **b**, Expression of representative state-associated genes across the five transcriptional states. Dot size indicates the fraction of expressing cells and color indicates scaled mean expression. **c**, Distribution of a fixed CAR-density-associated expression score across the single-cell embedding. **d**, Differential expression associated with high versus low CAR density and with four-day signaling rest versus continuous signaling, highlighting NFKB1A. **e**, Gene-set enrichment analysis showing opposing regulation of TNFA–NF-κB, MTORC1, MYC, unfolded-protein-response and exhaustion-related programs across the

two conditions. **f**, Representative enrichment profiles for TNFA signaling via NF-κB. **g**, Paired NFKBIA expression across donors, showing increased expression with high CAR density and decreased expression after signaling rest. **h**, Shared NF-κB leading-edge genes with significant positive effects in the CAR-density comparison and negative effects after signaling rest. Asterisks indicate FDR < 0.05.

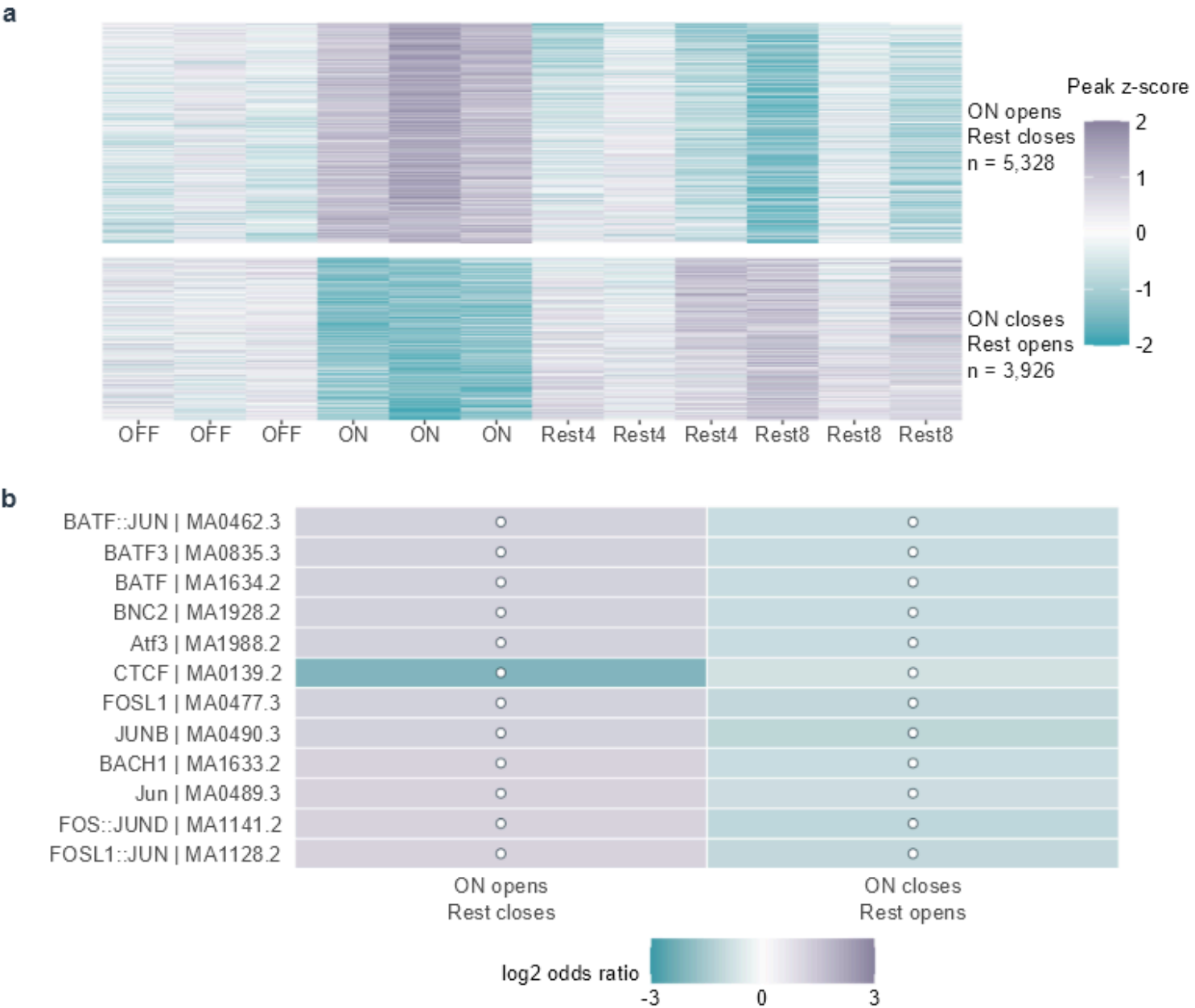


Figure 3

Rest-associated accessibility changes define return-peak classes with distinct motif composition.

a, Heatmap of 9,254 chromatin regions showing significant opposing accessibility changes between signaling activation and four-day rest in three donors. Of these, 5,328 regions opened with signaling and closed after rest, whereas 3,926 showed the opposite pattern. Eight-day rest samples are shown for comparison. **b**, Transcription-factor motif enrichment within the two return-peak classes. AP-1-related

motifs, including BATF::JUN, BATF and BATF3, were preferentially enriched among regions that opened with signaling and closed after rest. Colors indicate motif enrichment strength, and dots denote significant enrichment after multiple-testing correction.

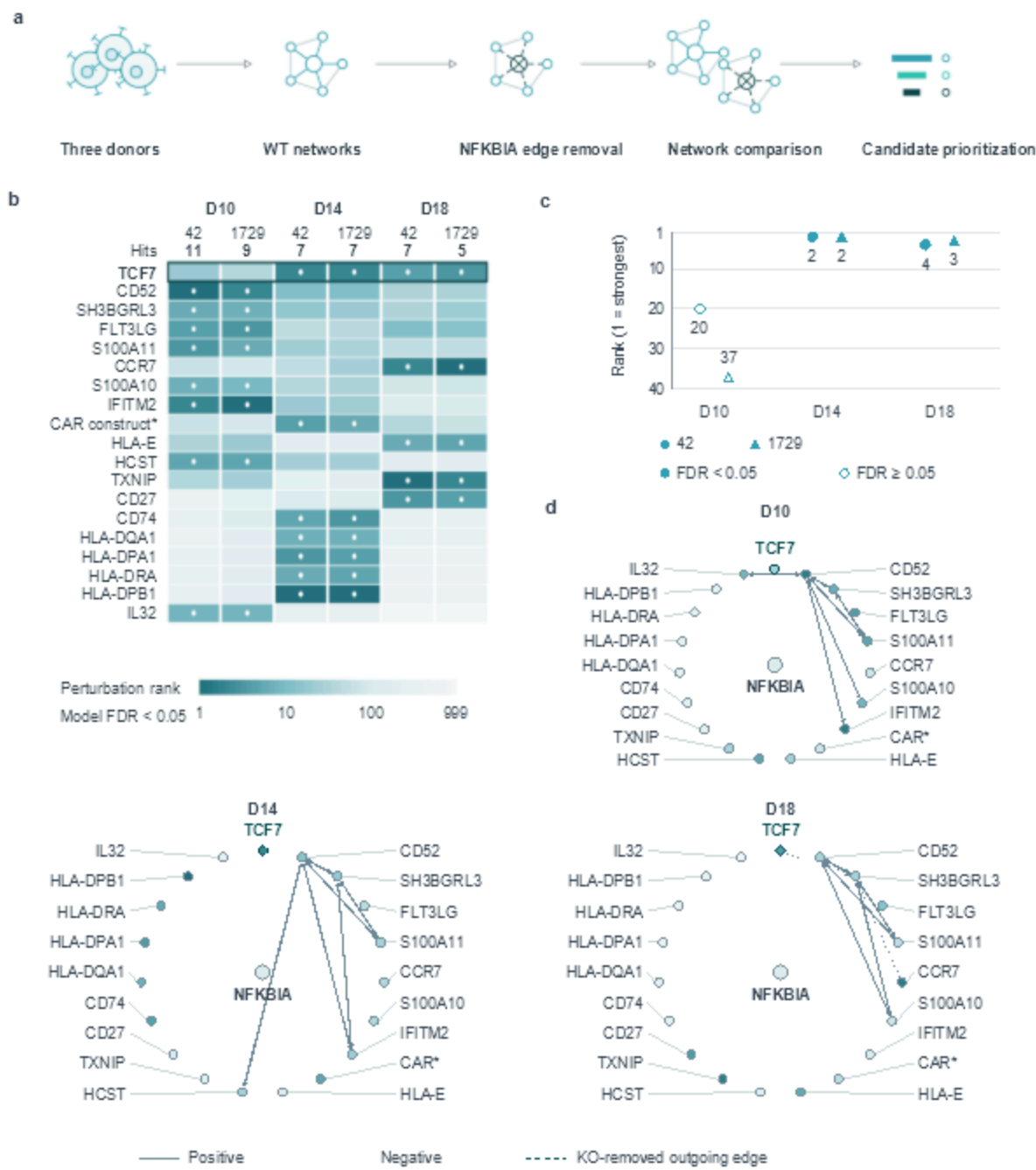


Figure 4

Virtual NFKBIA knockout nominates TCF7 as a context-dependent network-response candidate.

Donor-specific regulatory networks were inferred from 25,767 CAR-T cells from three donors and subjected to virtual NFKBIA knockout. **a**, Overview of the perturbation workflow. **b**, Perturbation ranks of reproducible candidate genes identified across donors and subsampling runs. **c**, TCF7 was among the strongest responses in D14 and D18, but not D10. **d**, Local donor-specific networks connecting NFKBIA

with prioritized response candidates. Edge style denotes inferred interaction sign, and node shading indicates perturbation rank. Together, these analyses nominate TCF7 as a recurrent, context-dependent response to NFKBIA perturbation.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [v1supplementarymaterials.docx](#)
- [SupplementaryData.zip](#)