

Refactored intracellular domains enhance CAR expression and function

Liam Bai, Aakriti Kumar, Hetvi Desai, Jiahe Li,
Dinesh de Alwis, Daria Hazuda, Gevorg Grigoryan, Wujie Wang*, Jason Yi*

Generate Biomedicines, 101 South St, Somerville, MA 02143

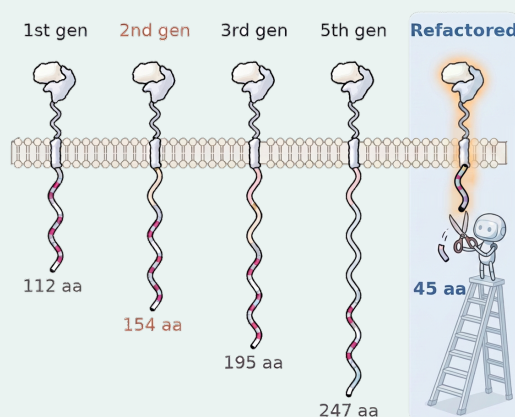
*Correspondence:

wwang@generatebiomedicines.com,

jyi@generatebiomedicines.com

Abstract

Chimeric antigen receptor (CAR) intracellular domains (ICDs) couple antigen recognition to T-cell activation. Here we worked with AI agents to refactor CAR ICDs, proposing designs and refining them as measurements accumulated over three rounds. This approach yielded **krgr- ζ 1-40**, a 45-amino-acid ICD less than a third the length of the clinical **4-1BB-CD3 ζ** (BB ζ) ICD. Compared with BB ζ , **krgr- ζ 1-40** increased CAR surface expression and antigen-driven IFN- γ release, without detectable antigen-independent IFN- γ release. **krgr- ζ 1-40** extended CAR expression and target-cell killing after transient mRNA delivery, supported the expression of four CARs from a shared mRNA dose, and enhanced B-cell depletion after lentiviral and lipid-nanoparticle delivery. These results show that iterative agent-guided design with experimental feedback can yield compact ICDs that enhance CAR expression and sustain CAR T-cell function.



Introduction

Chimeric antigen receptors (CARs) redirect T cells to recognize cell-surface antigens independently of major histocompatibility complexes (MHCs). The concept originated with chimeric T-cell receptors whose variable regions were replaced by those of an antibody, which conferred antibody-type specificity on T cells [1]. In parallel, Irving and Weiss showed that the cytoplasmic domain of the T-cell receptor (TCR) ζ chain is sufficient to couple a chimeric receptor to TCR signal transduction pathways [2]. Combining these ideas, Eshhar and colleagues fused a single-chain antibody fragment to the ζ or γ chain, producing single-chain chimeric receptors that triggered cytokine secretion and target-cell lysis upon antigen binding [3]. Adding a costimulatory domain from CD28 [4,5] or 4-1BB [6–8] to CD3 ζ produced the second-generation CARs used in approved therapies, and the two differ in signaling, metabolism and persistence [9,10].

Engineering these receptors requires balancing antigen-induced activity with behavior in the absence of antigen. Antigen-independent (tonic) CAR signaling can promote exhaustion, and 4-1BB and CD28 costimulation have different effects on exhaustion in tonic-signaling models [11]. Activation strength also depends on the number of immunoreceptor tyrosine-based activation motifs (ITAMs) in CD3 ζ . Mutating two of the three CD3 ζ ITAMs lowers the activation potential of a CD28-based CAR, which preserves memory differentiation and limits exhaustion [12]. Conversely, additional ITAMs increase T-cell proliferation [13,14] and the efficiency of receptor triggering, likely by bringing multiple ZAP-70 molecules into proximity [15], and the larger number of ITAMs in the TCR-CD3 complex is thought to contribute to the greater antigen sensitivity of the TCR compared with CARs [15,16]. These studies support the prevailing view that more ITAMs produce stronger activation, so that removing ITAMs trades acute activity for persistence. They also suggest that the composition of the CAR intracellular domain (ICD) can be tuned beyond the two clinical designs, CD28-CD3 ζ and 4-1BB-CD3 ζ .

Efforts to expand this signaling vocabulary have screened libraries of signaling domains for T-cell function, from randomly assembled domains to pooled and combinatorial designs [17–24]. Pooled screens have tested dozens of intracellular domains from diverse immune receptors in primary T cells [19–21], shuffled costimulatory domains with single-cell readouts [22–24] and recombined short signaling motifs to decode their effects on CAR T-cell phenotype [17]. Others have built CARs from CD3 subunits beyond CD3 ζ [25,26], from the ITAM adaptor DAP12 [27] and from proximal signaling molecules [28,29]. Together, these studies show that the signaling vocabulary of CARs extends beyond CD28 and 4-1BB. However, they explored which domains to combine, exchange or reorder rather than how much of the ICD could be removed. The closest test, truncating CD3 ζ to a single membrane-proximal ITAM in a CD28-based CAR, preserved acute effector function and improved tumor control, but was designed to examine ITAM position and dosage and retained the full CD28 domain [12]. Because shortening the signaling domain is otherwise expected to reduce function, how far the ICD as a whole can be compacted, and whether compaction can improve function, remain unknown.

Refactoring in software reorganizes internal structure while preserving observable behavior [30]. In biology, refactoring has simplified systems shaped by natural selection, from the overlapping genetic elements of bacteriophage T7 to G protein-coupled receptor signaling in yeast, so that they can be engineered rationally [31,32]. Applied to CAR ICDs, refactoring would reorganize signaling motifs into compact domains that preserve the core function of coupling antigen recognition to T-cell activation. This requires choosing which binding partners to recruit, how to combine and place their motifs, and which sequences to remove. Unlike the folded proteins of structure-based design [33–35], signaling domains such as the CD3 ζ cytoplasmic domain are largely disordered and act through short motifs that recruit partners inside cells [17,36,37]. The functional consequences of these design choices must be established experimentally. AI agents that combine literature search, data analysis and hypothesis generation to propose and revise experiments [38,39] offer a way to reason over the design space as measurements accumulate.

A compact ICD would also remove most of this disordered sequence, which two mechanistic models implicate in receptor signaling, with opposite predictions for a compact ICD. In the first, signaling components assemble into membrane condensates that promote signal transduction [40,41]. In T cells, these condensates have been attributed mainly to multivalent crosslinking of LAT [40,42], as in other multivalent signaling systems [43]. However, TCR–ZAP-70 and LAT occupy distinct domains within microclusters, with the TCR–ZAP-70 domain assembling first [44], and LAT condenses only after a delay following TCR engagement [44,45], so LAT crosslinking does not explain how the receptor itself clusters. In the broader condensate field, phase separation is often driven by intrinsically disordered regions (IDRs) themselves, in which interacting stickers, such as aromatic and charged residues, are separated by solvating spacers [46–48], and in CARs, an engineered CD3 ϵ motif that condenses through cation– π interactions improved antigen sensitivity and persistence [49]. If the receptor’s disordered tails drive condensation and condensation drives activation, removing disordered sequence should impair activation. In the second model, basic-rich stretches (BRS) of the CD3 ϵ and CD3 ζ tails bind acidic phospholipids of the inner leaflet [50–57], inserting ITAM tyrosines into the bilayer and restraining their phosphorylation until antigen engagement releases them [50,51,55], although disrupting the CD3 ζ BRS has also been found to weaken TCR signaling or synapse accumulation [52,54]. Removing the BRS should instead release tonic signaling. A compact ICD therefore risks either loss of activation or loss of antigen dependence, and changing disorder, basic residues or motif spacing can test these competing expectations.

Compactness may also change how many receptors reach the cell surface. CAR function depends on both receptor and antigen density [58–61], so a compact ICD that raises surface expression could improve function even without greater signaling per receptor. Transient mRNA delivery, which can generate functional CAR T cells [62,63], provides a setting in which this relationship can be followed as receptors are lost from the cell surface. Expression from limited mRNA also matters for in vivo CAR T-cell generation, in which targeted lipid nanoparticles carrying CAR mRNA [64–66] or targeted lentiviral vectors [67] modify T cells inside the body. With lipid nanoparticles, an ICD that yields more receptor from less mRNA could lower the dose needed for function.

Here we addressed these questions by refactoring CAR ICDs through successive rounds of design, in which scientists and AI agents proposed changes to motif composition and sequence context and the resulting measurements guided the next designs. We asked how far the ICD could be compacted and whether removing its disordered sequence would cost activation or antigen dependence, as the condensate and lipid-binding models predict. We also asked whether a compact ICD that yields more surface receptor would improve function after mRNA and lipid-nanoparticle delivery, and whether experimental feedback could steer design ideas across rounds. Successive rounds produced shorter ICDs with higher expression and function, without the losses that either model predicts. Earlier-round measurements also shifted large language model (LLM) predictions toward the design ideas that worked, although measurement still revealed gains the models missed. These results establish iterative design by scientists and AI agents as a route to compact CAR ICDs with higher expression and more durable function.

Results

Agent-guided design and screening identify **brs- ζ 1-40**, a compact CAR ICD

CAR intracellular domains are typically assembled from intact cytoplasmic domains of CD3 ζ and a costimulatory receptor such as 4-1BB or CD28, yet T cells use a much larger vocabulary of signaling modules [17,19,20,22,25,68–71]. To make this vocabulary explicit and to frame the design of compact ICDs, we compiled published evidence into a signaling map (Fig. 1a). The map links 34 ICD modules, drawn from CD3 chains and other ITAM receptors, co-receptors, the CD28 family, LAT, the TNF-receptor family and cytokine receptors, through binding partners and relays to signaling pathways and reported T-cell phenotypes. It draws on 150 interactions from 253 primary studies, each graded by the strength of its evidence [68,72–75]. Our iterative exploration across rounds 1–3 covered most modules on the map (Fig. 1a).

The map summarizes established rules of CAR signaling. CD3 ζ ITAM number and identity calibrate activation strength, ligand sensitivity and T-cell fate [12,15,76]. 4-1BB signals through TRAF1–3 and noncanonical NF- κ B [77,78] and recruits THEMIS–SHP-1, which attenuates Lck-dependent priming [79]. CD28 and 4-1BB differ in the kinetics, magnitude and metabolic consequences of signaling [9,10], and point mutations in CD28 motifs change persistence and function [80,81]. Membrane-proximal basic residues bind acidic inner-leaflet lipids [51,52]; CD3 ϵ motifs recruit Lck, PI3K and Csk [82–84]; and a JAK–STAT module can add a cytokine signal [85]. ICD size, rather than costimulatory identity, determines constitutive CD3 ζ phosphorylation [86], and pooled and combinatorial libraries show that non-canonical domains and module combinations can change CAR T-cell phenotype [17,19,20,22]. Grading the evidence behind each rule family by experimental system showed that most support comes from T cells in vitro and from mouse models, whereas clinical evidence is limited to comparisons of CD28 and 4-1BB costimulation.

We built an agent system that combines LLM reasoning with literature search and computational tools. Scientists specified design objectives and reference sequences, and tool-using agents searched the literature and proposed sequence changes. Scientists then reviewed the proposed designs, steered the agents and compared candidate sequences. After several cycles of review we selected 186 candidates for the round 1 library screen. Representative designs ranged from 36 to 109 amino acids (aa), compared with 154 for the second-generation 4-1BB–CD3 ζ reference ICD (BB ζ ; Fig. 1b).

As an initial exploration, we screened the 186 designs as CARs with the D-domain BCMA binder (ddBCMA) [87] in Jurkat NFAT reporter cells against BCMA-positive and -negative target cells. Of the 186 designs, 38 matched or exceeded the BCMA-driven signal of BB ζ , and 32 of these also gave lower signal with BCMA-negative targets (Fig. 1c). Grouping the designs by native ITAM count and length, we found that 25 of the 130 designs retaining only one ITAM supported antigen-driven signaling at or above the level of BB ζ (Fig. 1d).

brs- ζ 1-40 is a 54-aa ICD joining the membrane-proximal basic clusters and first ITAM of CD3 ζ to a CD40 segment required for TRAF6 binding in native CD40 [88,89]. CD40-derived signaling domains have previously enhanced CAR T-cell proliferation, survival and persistence [23,90,91]. However, a 19-residue CD40 motif spanning the same TRAF6-binding site was among the most inhibitory motifs for CAR T-cell proliferation in a pooled screen of CARs with full-length CD3 ζ [17], so whether this segment can act as a costimulatory domain within this scaffold was unclear. Unexpectedly, **brs- ζ 1-40** matched the NFAT signal of BB ζ with BCMA-positive targets, whereas with BCMA-negative targets its signal was about one-sixteenth that of BB ζ (Fig. 1c).

To test whether the designs transferred to another binder and to primary T cells, we re-screened 78 designs, selected by ICD length and NFAT activity, as bb2121 BCMA scFv CARs [92] in primary T cells (Fig. 1e). Strikingly, nearly all compact designs increased surface CAR: 73 of the 78 exceeded BB ζ (median, 2.7-fold), and 30 also released more antigen-driven IFN- γ . In this screen, **brs- ζ 1-40** displayed 4.1-fold the surface CAR of BB ζ and the second-highest IFN- γ of the 78 designs. Without target cells, its IFN- γ remained below the detection limit. Jurkat NFAT and primary T-cell IFN- γ ranked the re-screened designs similarly (Spearman $\rho = 0.69$; Fig. 1f), even though the two screens differed in binder, target cells and measured functional response. Together, these results identify **brs- ζ 1-40** as a compact ICD that is highly expressed and antigen responsive in primary T cells.

brs- ζ 1-40 increases CAR expression across BCMA binders and retains antigen-dependent function

We selected ten compact ICDs, including **brs- ζ 1-40**, to test in primary T cells across a range of effector-to-target (E:T) ratios (Fig. 2a–c). All ten compact ICDs displayed more surface CAR than BB ζ , from 2.3- to 17-fold (Fig. 2a).

We next asked whether this higher expression was accompanied by cytokine release and target-cell killing. CAR T cells expressing the compact designs were co-cultured with BCMA-positive H929 myeloma cells across an E:T titration. **brs- ζ 1-40** released the most IFN- γ in the panel, 2.0-fold the peak of BB ζ (Figs. 2b and S1e), and ranked highest on both replicate plates (Fig. S1d). Several designs killed H929 cells as effectively as BB ζ , whereas the design with a TRAF motif inserted before ITAM1 (**brs-traf- ζ 1**) killed only about half of the H929 cells even at

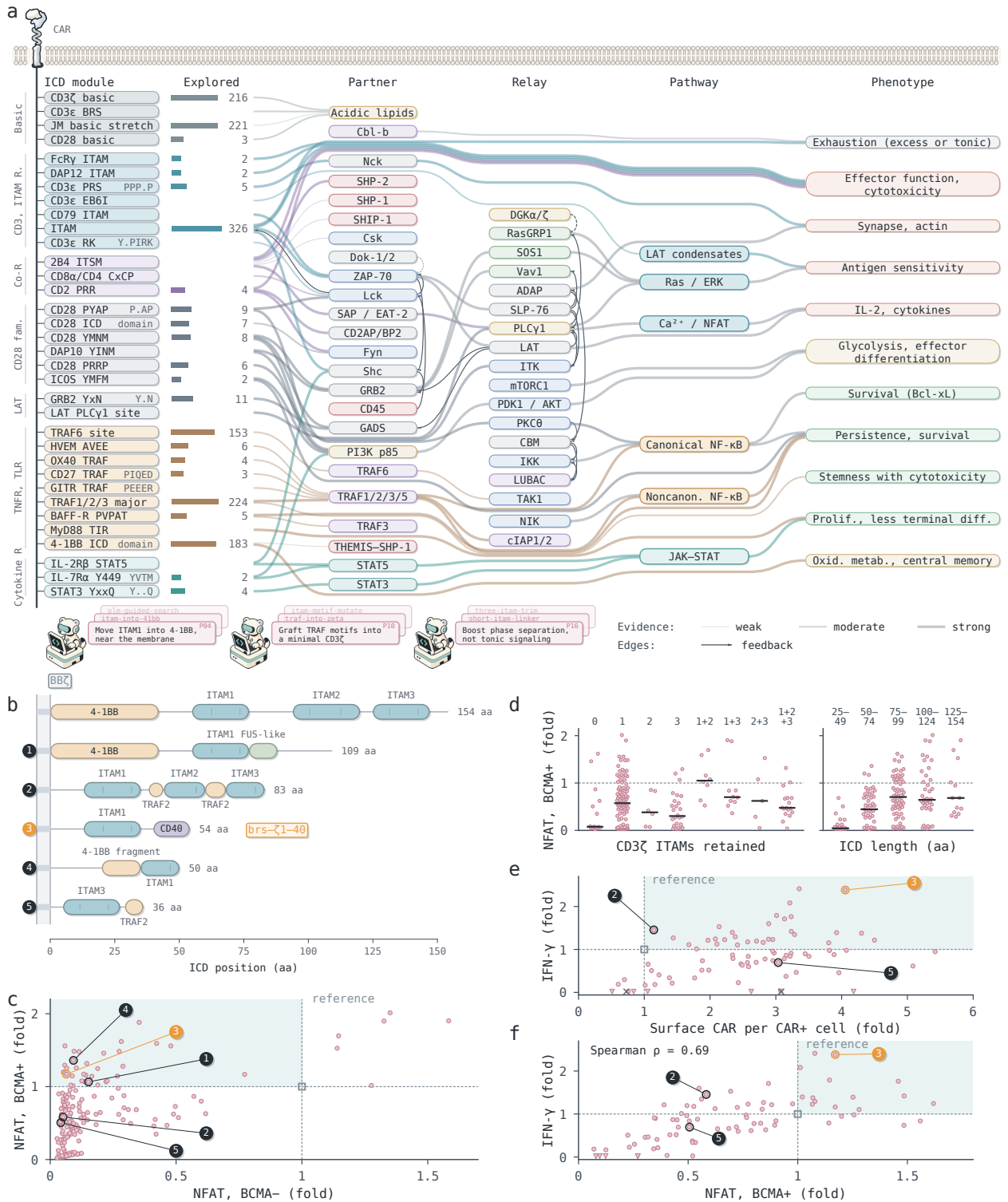


Figure 1: Agent-guided design and screening of compact ICDs. **a**, Signaling map of ICD modules. Designs compose sequence motifs that recruit partner proteins, which relay to signaling pathways. Lines show the strong activating interactions, plus the best-supported interaction of any otherwise unconnected node. Explored, number of the 339 ICDs tested in rounds 1–3 that carry each module’s sequence motif (bar length on a log scale); blank, none. **b**, Representative designs proposed by scientists and agents, drawn to residue scale. **c**, Many compact designs show higher on-target signaling and lower background signaling than BBζ (4-1BB-CD3ζ). Y axis shows on-target activation in BCMA-positive cells; X axis shows background activation in BCMA-negative cells; expressed as fold over reference (open square). Shading marks higher on-target and lower background signal than BBζ. **d**, A single ITAM can support strong signaling. Jurkat NFAT signal expressed as fold of BBζ, grouped by native CD3ζ ITAMs retained and by ICD length; bars, medians. **e**, Designs re-tested in primary T cells: most showed higher surface expression than BBζ, and some released more IFN-γ. Surface CAR density and antigen-driven IFN-γ, expressed as fold of BBζ; ×, ITAM Y→F and no-mRNA negative controls; shading marks higher surface CAR and IFN-γ than BBζ. **f**, Correlation between Jurkat NFAT and primary T-cell IFN-γ for the 78 re-tested ICDs. Each axis is relative to its assay’s BBζ reference (open square; dashed lines); shading marks higher NFAT and IFN-γ than BBζ. Spearman $\rho = 0.69$ across all 78 designs.

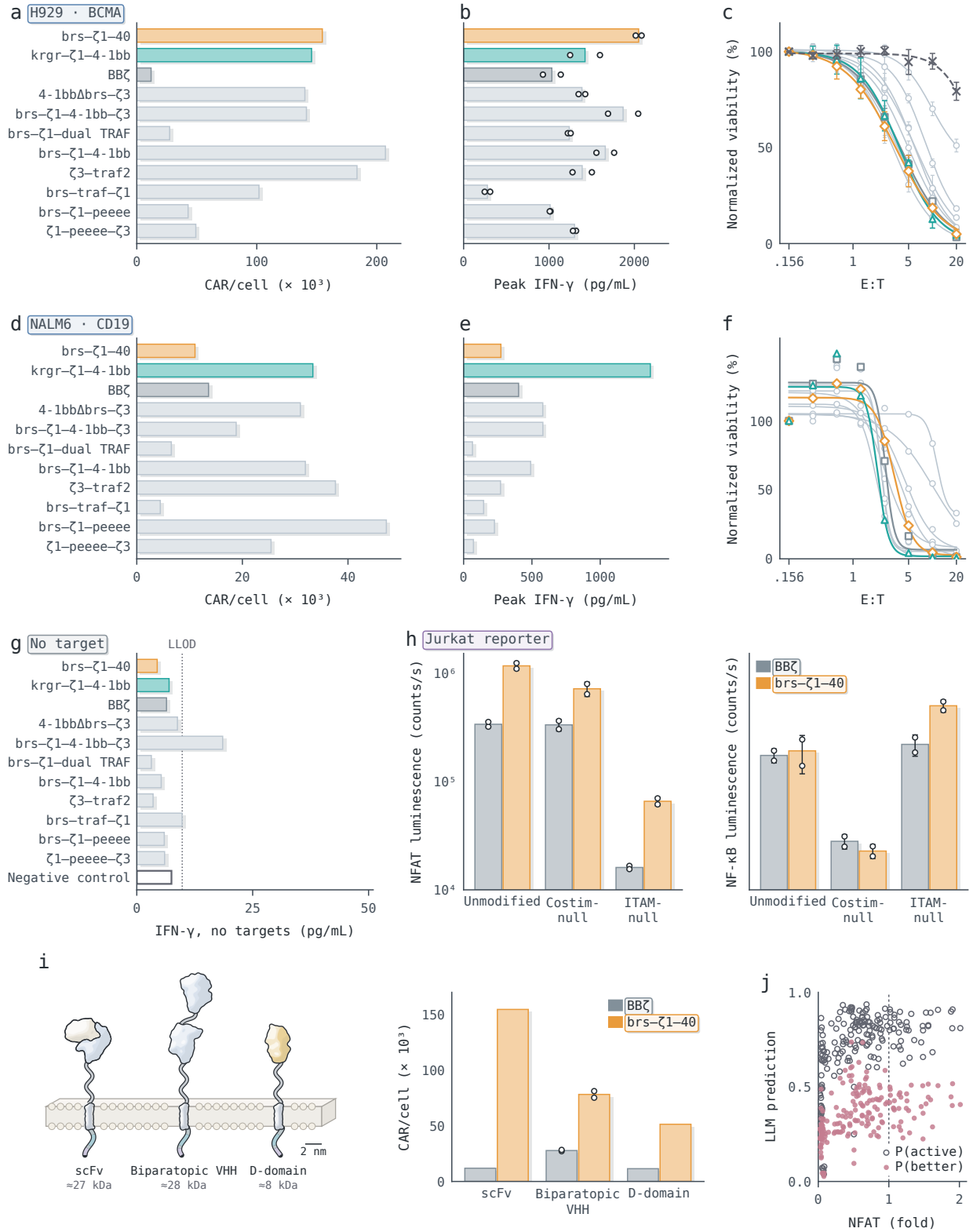


Figure 2: Compact ICD expression, killing and signaling across CAR binders. **a–c**, Surface CAR density (**a**), peak IFN- γ across eight E:T ratios (**b**) and killing of H929 myeloma cells (**c**) by primary T cells expressing bb2121 BCMA CARs. **d–f**, The same ICDs in FMC63 CD19 CARs: surface CAR density (**d**), peak IFN- γ across eight E:T ratios (**e**) and killing of NALM6 leukemia cells (**f**). Rows in **b** and **e** follow **a** and **d**. Killing is shown as residual target-cell luminescence normalized to the lowest E:T ratio for each construct and replicate; lower values indicate more killing. **g**, Peak IFN- γ from BCMA CAR T cells without targets across eight effector inputs; dotted line (LLOD), 9.8 pg/mL lower limit of detection. **h**, Jurkat NFAT and NF- κ B responses to costimulation- or ITAM-disrupting mutations in bsr-ζ1-40 and BBζ (4-1BB-CD3ζ). In bsr-ζ1-40, E42A/E45A substitutions alter the CD40-derived TRAF6-binding segment (QEPQE→QAPQA), and Y21F/Y32F substitutions replace both ITAM1 tyrosines. The corresponding BBζ mutants disrupt 4-1BB costimulation or replace all six tyrosines across the three CD3ζ ITAMs with phenylalanine. **i**, Three binder formats, and bsr-ζ1-40 vs. BBζ expression across them. **j**, Zero-shot LLM predictions for the round 1 library against NFAT relative to BBζ (Fig. 1c); mean of five models.

the highest E:T (Fig. 2c). **brs-ζ1-40** and **BBζ** killed H929 cells with similar potency. These results show that **brs-ζ1-40** releases more IFN-γ than **BBζ** while killing BCMA-positive target cells as effectively.

To determine whether these properties depend on the binder, we tested the same eleven ICDs with the FMC63 anti-CD19 scFv [93] against NALM6 leukemia cells (Fig. 2d-f). Seven of the ten compact designs again displayed more surface CAR than **BBζ**, by up to 3.5-fold (Fig. 2d). In contrast, **brs-ζ1-40** displayed slightly less CAR than **BBζ** with FMC63, whereas **krgr-ζ1-4-1bb** displayed 2.5-fold more. **brs-ζ1-40** released less NALM6-induced IFN-γ than **BBζ**, whereas **krgr-ζ1-4-1bb** released the most in the panel, 3.4-fold **BBζ** (Figs. 2e and S1f). **brs-ζ1-40** and **BBζ** killed NALM6 cells with similar potency (Fig. 2f), and **krgr-ζ1-4-1bb** was among the most potent. Thus, the expression and cytokine advantage of **brs-ζ1-40** seen with bb2121 did not extend to FMC63, whereas **krgr-ζ1-4-1bb** was highly expressed and active with both binders.

Because tonic signaling can promote exhaustion [11] and may rise with receptor density [59,94], we next measured IFN-γ released by the BCMA CAR T cells in the absence of target cells (Fig. 2g). Across effector inputs, every construct except **brs-ζ1-4-1bb-ζ3** remained below the 9.8 pg/mL detection limit, including **brs-ζ1-40** despite its 13-fold higher receptor density. These results indicate that the high surface density of **brs-ζ1-40** does not lead to detectable tonic IFN-γ release.

To determine which segments of **brs-ζ1-40** drive its signaling, we compared signaling mutants of **brs-ζ1-40** and **BBζ** in Jurkat NFAT and NF-κB reporter cells stimulated with BCMA-positive target cells (Fig. 2h). Costimulation-null variants carried alanine substitutions in the CD40-derived QEPQE motif of **brs-ζ1-40** or the TRAF-binding motif of 4-1BB, and ITAM-null variants carried Y→F substitutions in every ITAM. As expected, the costimulation-null variant of **BBζ** kept its NFAT signal but lost more than 80% of its NF-κB signal, whereas the ITAM-null variant lost about 95% of its NFAT signal. **brs-ζ1-40** followed a similar pattern: its costimulation-null variant lost more than 80% of NF-κB signal but, unlike that of **BBζ**, also about 40% of NFAT signal, and its ITAM-null variant lost about 95% of NFAT signal. In this experiment, unmodified **brs-ζ1-40** gave 3.5-fold higher NFAT and similar NF-κB signal compared with **BBζ**, more than the matched NFAT signal of the library screen (Fig. 1c). These results show that, as in **BBζ**, NFAT signaling by **brs-ζ1-40** depends on its ITAM tyrosines and NF-κB signaling on its CD40-derived QEPQE motif.

We then asked whether the expression advantage of **brs-ζ1-40** extends to other BCMA binder formats represented among clinical CAR T cells (Fig. 2i). With the biparatopic VHH binder of ciltacabtagene autoleucel, **brs-ζ1-40** displayed 2.8-fold more CAR per cell than **BBζ** at similar CAR-positive fractions. With ddBCMA, it displayed 4.5-fold more, compared with 13-fold with the bb2121 scFv (Fig. 2a). Together, these results show that **brs-ζ1-40** increases surface CAR expression across scFv, VHH and miniprotein BCMA binders while retaining antigen-dependent killing and signaling that depends on its ITAM and its CD40-derived QEPQE motif.

We also asked whether such designs could have been anticipated from sequence alone. Given each library sequence and a description of the assay, five general-purpose LLMs predicted NFAT activity poorly. Their predicted probability that a design would exceed **BBζ** in NFAT stayed below 0.75 and hardly separated the 38 designs that did from the rest (mean 0.40 versus 0.34; Fig. 2j).

Experiment-informed agent redesign yields **krgr-ζ1-40**, a 45-aa ICD with higher expression and IFN-γ

To determine which features of the round 1 hits could be combined in a shorter ICD, we worked with Claude Code to design 15 round 2 ICDs from four hits characterized in primary T cells (Fig. 3a). The designs tested six groups of changes: ablations, costimulatory motif swaps, ITAM swaps, two TRAF-binding motifs, KRGR anchor transfers and minimal domains. Ablations tested whether individual segments were required, while motif swaps and anchor transfers tested whether features of different hits could be combined; **krgr-ζ1-40** was an anchor transfer. **brs-ζ1-40** had released the most IFN-γ of the four hits, and earlier cross-binder cytokine measurements motivated transferring the KRGR anchor from **krgr-ζ1-4-1bb** (Fig. 3a). The Jurkat screen had also shown that KRGR was required for the activity of **krgr-ζ1-4-1bb** (Fig. S1b). These observations led us to replace the first 13 residues of the basic-rich segment of **brs-ζ1-40** with the four-residue KRGR anchor while retaining its ITAM and CD40-derived segment (Fig. 3c). The resulting **krgr-ζ1-40** was 45 aa long, shorter than either parent (54 and 56 aa) and less than a third the length of **BBζ**.

To determine whether the KRGR anchor preserved the properties of **brs-ζ1-40**, we tested the 15 round 2 designs alongside **brs-ζ1-40** and **BBζ** in a primary T-cell screen (Fig. 3b). Twelve of the 15 round 2 designs displayed more surface CAR than **BBζ**, and six released more antigen-driven IFN-γ. **krgr-ζ1-40** improved on the surface expression of **brs-ζ1-40** by 1.5-fold and **BBζ** by 6.2-fold (Fig. 3b, left). Its IFN-γ was the highest of the 15 round 2 designs (Fig. 3b, middle). Without target cells, all round 2 designs and **brs-ζ1-40** remained below the detection limit (Fig. 3b, right). Thus, replacing the 13-residue basic-rich segment of **brs-ζ1-40** with the four-residue KRGR anchor increased both expression and antigen-driven IFN-γ without increasing tonic IFN-γ.

To compare killing, we titrated **krgr-ζ1-40**, **brs-ζ1-40** and **BBζ** CAR T cells against H929 cells across a range of E:T ratios (Fig. 3d). Killing curves for **krgr-ζ1-40** and **brs-ζ1-40** were left-shifted relative to that of **BBζ**, at 2.0- and 1.6-fold lower IC50 E:T, respectively, so fewer CAR T cells were needed per target cell for the same killing.

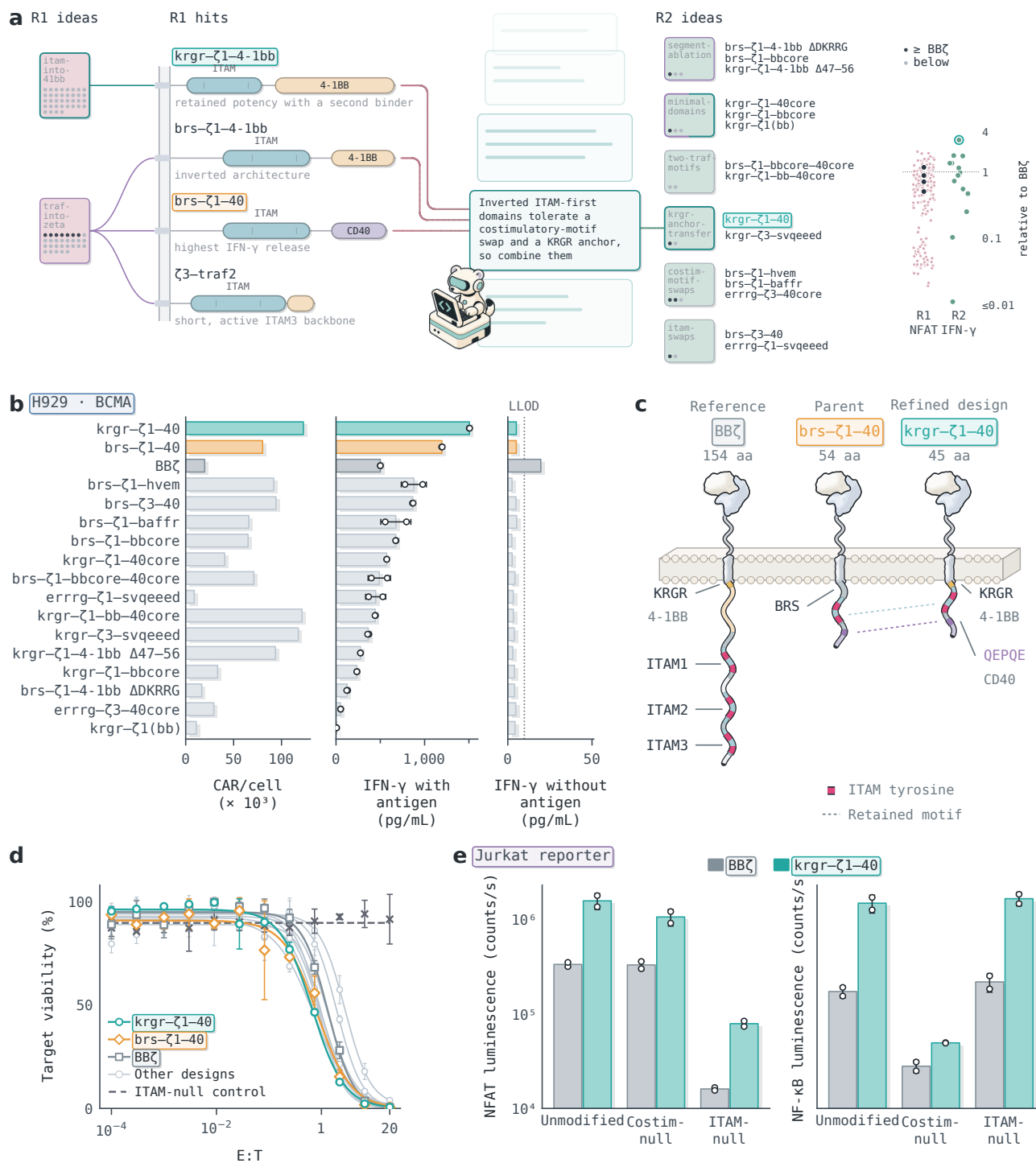


Figure 3: *krgr-ζ1-40* refinement and functional characterization. **a**, Evolution of design ideas from round 1 to round 2. Four round 1 hits were the parents of the 15 round 2 designs, which recombine sequence features from these hits. Center, a simplified reasoning step by which the agent drew on three round 1 hits to combine the KRGR anchor of *krgr-ζ1-4-1bb* with the ITAM and CD40-derived segment of *brs-ζ1-40*. Right, round 1 and round 2 activation readouts relative to BBζ (4-1BB-CD3ζ), in Jurkat and primary T cells respectively. **b**, Round 2 results in H929 cells with the bb2121 binder: CAR density, IFN-γ and IFN-γ without antigen. **c**, Reference BBζ (154 aa), parent *brs-ζ1-40* (54 aa) and refined *krgr-ζ1-40* (45 aa). **d**, Killing of H929 myeloma cells by round 2 designs, including *krgr-ζ1-40*, *brs-ζ1-40* and BBζ across E:T ratios. **e**, Jurkat NFAT and NF-κB responses to costimulation- or ITAM-disrupting mutations in *krgr-ζ1-40*. In *krgr-ζ1-40*, E33A/E36A substitutions alter the CD40-derived TRAF6-binding segment (QEPQE→QAPQA), and Y12F/Y23F substitutions replace both ITAM1 tyrosines.

To determine whether *krgr-ζ1-40* signals through the same segments as *brs-ζ1-40*, we tested the signaling mutants of Fig. 2h (Fig. 3e). Unmodified *krgr-ζ1-40* gave 4.7-fold higher NFAT and 8.6-fold higher NF-κB signal than *BBζ*, in proportion to its higher surface CAR (Fig. S2). In contrast, *brs-ζ1-40* gave about a quarter of the NF-κB signal per surface CAR of *BBζ*, which may reflect suboptimal costimulatory signaling (Fig. S2). As with *brs-ζ1-40*, NFAT signaling depended on the ITAM tyrosines, with about 95% lost in the ITAM-null variant, and NF-κB signaling on the CD40-derived motif, with more than 95% lost in the costimulation-null variant, compared with about 85% for *BBζ* (Fig. 3e). Together, these results show that the KRGR anchor preserves the ITAM- and CD40-dependent signaling of *brs-ζ1-40* while further increasing expression and antigen-driven IFN-γ. We therefore carried *krgr-ζ1-40* forward as the lead compact ICD.

***krgr-ζ1-40* extends CAR expression and function across mRNA, lentiviral and lipid-nanoparticle delivery**

We next measured how *krgr-ζ1-40* expression scales with the amount of mRNA delivered and how long expression and function last after transient delivery (Fig. 4a-e). Surface CAR density increased with the mRNA dose for all three ICDs, but *krgr-ζ1-40* gave the highest density at every dose, 8.4-fold that of *BBζ* and 2.4-fold that of *brs-ζ1-40* at the highest dose (Figs. 4a and S3). *krgr-ζ1-40* matched the density that *BBζ* reached at the highest dose with one-eighth of the mRNA.

To follow expression over time, we transfected primary T cells with *krgr-ζ1-40* or *BBζ* CAR mRNA and measured surface CAR daily for 17 days (Fig. 4b,c). On day 1, *krgr-ζ1-40* T cells displayed 12-fold more CAR than *BBζ* T cells, and, as expected after transient delivery, expression of both CARs then declined at a similar rate. Because it started higher, *krgr-ζ1-40* remained above 2,000 CAR/cell until day 13, compared with day 4 for *BBζ*, and above the 1,000 CAR/cell LLOD through day 17 (Fig. 4b). Thus, after transient delivery, the higher starting expression of *krgr-ζ1-40* kept surface CAR detectable for longer than with *BBζ*, without slowing its decline.

We next asked whether the longer expression of *krgr-ζ1-40* extends its function. Every second day, we co-cultured the CAR T cells with fresh H929 cells and measured IFN-γ and target viability (Fig. 4d,e). *krgr-ζ1-40* T cells released 3.1-fold more IFN-γ than *BBζ* T cells on day 1, and IFN-γ declined over time for both CARs (Figs. 4d and S4a). However, *krgr-ζ1-40* T cells still released IFN-γ on day 13, whereas IFN-γ from *BBζ* T cells fell below the detection limit from day 7. Similarly, at the highest E:T, *krgr-ζ1-40* T cells killed more than 90% of H929 cells through day 13, whereas *BBζ* T cells had lost most of their killing by day 9 (Fig. 4e). From day 1 to day 7, the killing IC₅₀ rose 11-fold for *BBζ* T cells but only 2.5-fold for *krgr-ζ1-40* T cells, which still killed half of the H929 cells through day 13 (Fig. S4b). Together, these results show that *krgr-ζ1-40* extends both CAR expression and function after transient mRNA delivery.

Because *krgr-ζ1-40* yielded more receptor per unit of mRNA (Fig. 4a), we next asked whether it supports the co-expression of several CARs from a shared mRNA dose (Fig. 4f,g). We electroporated T cells with a cocktail of four CAR mRNAs carrying BCMA, CD19, HER2 or MUC16 binders on either *krgr-ζ1-40* or *BBζ*, each at half the single-CAR dose, and detected each CAR with its own antigen in a separate well. With *krgr-ζ1-40*, more than 80% of cells were positive for each CAR, compared with at most a third with *BBζ* (Fig. 4f), and CAR density was 2.6- to 12.9-fold higher (Fig. 4g). Each CAR delivered alone at the full dose showed a similar advantage, with 4.5- to 10.1-fold higher density (Fig. S5). These results show that *krgr-ζ1-40* supports the co-expression of four CARs from a single electroporation at half the per-CAR mRNA dose.

To determine whether these advantages extend to persistent function, we transduced T cells with lentiviral vectors encoding FMC63 anti-CD19 CARs (Fig. 4h,i). *krgr-ζ1-40* T cells displayed 6.8-fold more CAR than *BBζ* T cells (*n* = 2 donors; Fig. 4h). We then co-cultured the T cells with Raji lymphoma multilayer 3D aggregates at low E:T ratios and measured IFN-γ on days 1, 3, 5 and 7 (Fig. 4i). IFN-γ from both CARs peaked on day 3 and then declined steadily through day 7. *krgr-ζ1-40* T cells released more IFN-γ than *BBζ* T cells in 30 of 32 matched E:T and day conditions. By day 7, when both curves had reached the same plateau, *krgr-ζ1-40* T cells reached half-maximal IFN-γ at a 2.9-fold lower E:T than *BBζ* T cells. Because each T cell must engage more tumor cells in succession at a lower E:T, this shift suggests that *krgr-ζ1-40* T cells stayed active through 2.9-fold more consecutive encounters with antigen-positive tumor cells.

We next asked whether lentivirally transduced *krgr-ζ1-40* T cells also kill target cells more efficiently. In a separate experiment, *krgr-ζ1-40* T cells killed primary human B cells at 2.4-fold lower IC₅₀ E:T than *BBζ* T cells (Fig. 4j), and NALM6 leukemia cells at 1.3-fold lower IC₅₀ E:T (Fig. S4c). Thus, *krgr-ζ1-40* T cells killed both normal and leukemic CD19-positive cells at lower E:T.

Finally, we asked whether *krgr-ζ1-40* enhances CAR T-cell function when the CAR is delivered directly to unselected blood cells [66]. We added CD3-targeted lipid nanoparticles (tLNPs) carrying CAR mRNA to whole PBMC cultures from five donors, so that CAR T cells were generated in situ alongside autologous B cells (Fig. 4k,l). One day later, *krgr-ζ1-40* gave more CAR-positive T cells than *BBζ* at every dose from 12 ng mRNA per 10⁶ cells upward (Fig. 4k). B-cell depletion increased with dose for both CARs, but *krgr-ζ1-40* reached 50% depletion at a 4.6-fold lower dose than *BBζ* (Fig. 4l). Together, these results show that *krgr-ζ1-40* increases CAR expression and enhances CAR T-cell function after mRNA, lentiviral and lipid-nanoparticle delivery.

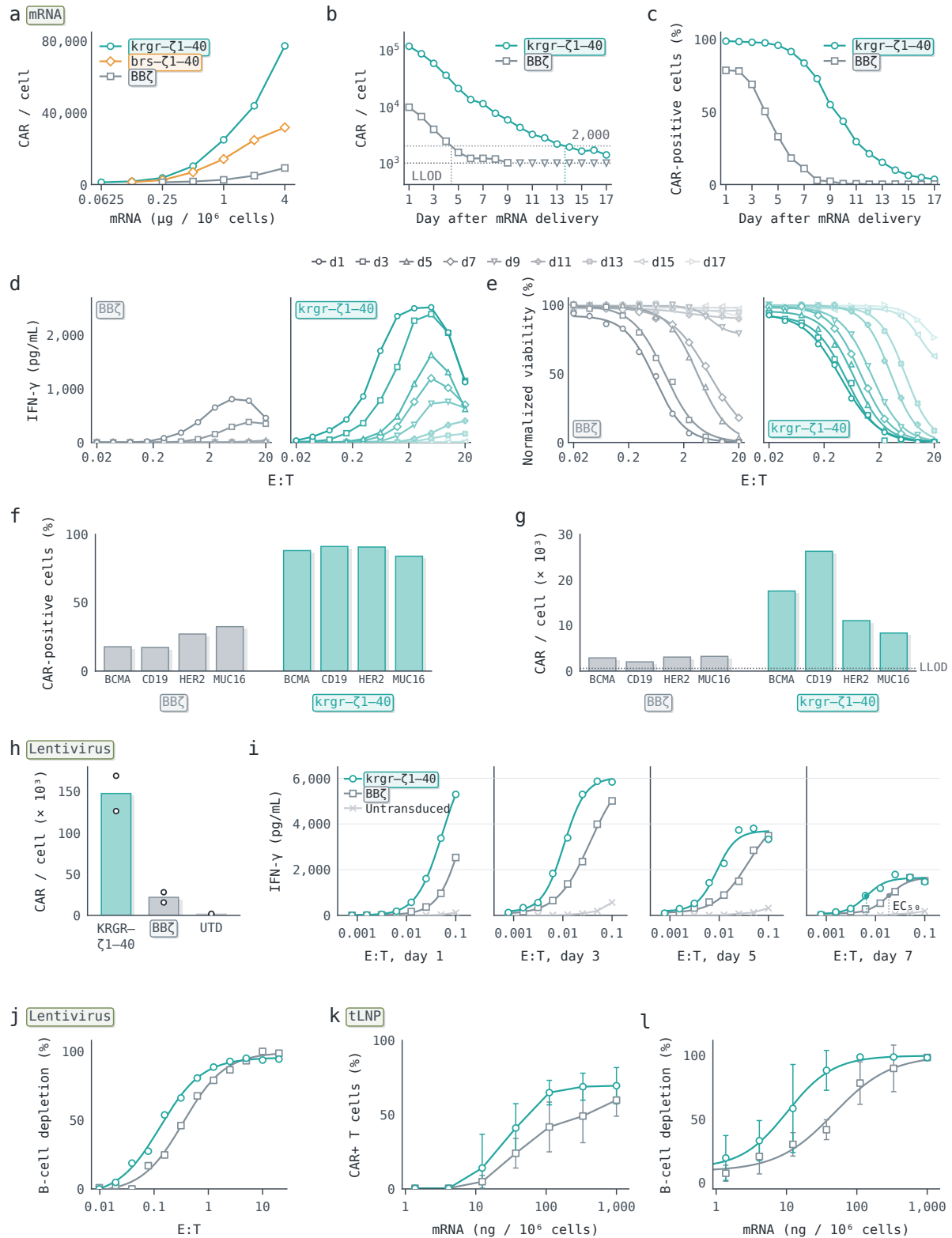


Figure 4: krg- ζ 1-40 expression, persistence and function after mRNA electroporation, lentiviral and tLNP delivery. **a**, Surface CAR density 24 h after electroporation with increasing CAR mRNA doses. **b**, Surface CAR density over 17 days after mRNA delivery. **c**, CAR-positive cells over the same period. **d**, IFN- γ across E:T ratios on alternate days after mRNA delivery; later days are paler. **e**, H929 viability across E:T ratios, as in **d**. **f**, CAR-positive cells 24 h after electroporation with a cocktail of four CAR mRNAs (BCMA, CD19, HER2 and MUC16 binders), each CAR detected with its own antigen. **g**, CAR density in the same cells. **h**, Surface density of lentivirally transduced FMC63 CD19 CARs; circles, two donors; UTD, untransduced T cells. **i**, IFN- γ during Raji 3D-aggregate co-culture. **j**, Depletion of CD19-positive B cells in PBMCs across E:T ratios. **k**, CAR-positive T cells after CD3-targeted lipid nanoparticles (tLNPs) carrying FMC63 CAR mRNA were added to PBMCs from five donors; mean $\pm$ s.d. **l**, B-cell depletion in the same cultures.

Iterative experimental feedback guides CAR ICD refactoring

Having established **krgr- ζ 1-40** as the lead compact ICD, we asked whether it could be refined further. In round 3, we worked with AI agents to design 164 variants of **krgr- ζ 1-40** and screened them in primary T cells (Fig. S6a,b). The variants tested three groups of changes: 113 rational edits, 30 single and combined point mutations, and 21 synonymous codon recodes. Rational edits changed the juxtamembrane (JM) segment, the costimulatory motif, spacers, charge and length; point mutations tested local sequence changes (Fig. 5a); and codon recodes tested whether expression could be raised through the mRNA sequence alone (Fig. S6e). Surface CAR varied widely among the 21 recodes, which encode the same protein, so codon choice alone changed CAR expression. Of the 122 ICD variants that passed the expression threshold, 17 released at least as much IFN- γ as **krgr- ζ 1-40** on the same plate (Fig. S6b).

Compact ICDs are difficult to design from structure or from training data. Signaling ICDs such as the CD3 ζ cytoplasmic domain are intrinsically disordered [36], so apart from short motifs bound to their partners, such as a phosphorylated ITAM on the ZAP-70 tandem SH2 domains (PDB 2OQ1) [95] and a CD40 peptide on TRAF6 (PDB 1LB6) [89], they offer little structure to model. Their function is measured mainly in cell-based assays, and published CAR libraries offer little training data for compact ICDs [17–24]. Machine learning offers several routes to interpretable sequence design, including libraries of CAR signaling motifs [17], sparse features of protein language models that map to biological concepts [96,97], and grammars that constrain sequence generation [98]. However, these approaches rely on training data or established rules that remain scarce for ICDs. Design therefore relies on reasoning about experimental data, which can be biased toward the designer’s prior beliefs when many choices are defensible [99–103].

To design in this space, we worked with many AI agents in parallel [38,39], which searched the literature and analyzed our assay data. In each round, we and the agents reviewed the preceding measurements together, proposed ideas and developed designs with matched controls. Within each idea, designs changed one feature at a time against matched controls, such as JM charge at fixed length or stepwise trims of ICD length (Fig. S7). The agents wrote and ran the analysis code, and each proposal came with a written rationale [103]. Each design was assigned to an idea when it was proposed, and we traced these ideas across rounds 1–3 (Fig. 5a). Round 1 comprised 186 designs in 14 ideas, and **krgr- ζ 1-40** became the parent of the 164 round 3 variants.

Current models did not reliably identify which compact ICDs would fail (Fig. S8). Neither Boltz-2 [104] co-fold confidence nor protein language-model likelihoods (ESM-2 [105] and ProtGPT2 [106]) reliably separated active from inactive ICDs. We also asked five general-purpose LLMs, three queried in a single call and two run as coding agents, given the ICD sequences and a description of the assays, for two probabilities per design: that it was active, $P(\text{active})$, and that it would exceed the round’s comparator, $P(\text{better})$. Each model answered zero-shot and again with the earlier rounds’ results in the prompt (in context). The LLMs called most designs active, including most inactive round 3 variants: the specificity of their calls was 0.15–0.34 zero-shot and 0.18–0.51 in context (Fig. S8a), consistent with the poor calibration of LLM predictions of experimental outcomes [107].

As a retrospective proxy for how much signal the earlier rounds’ data carried about each idea, we measured how far the models’ predictions moved when those data were added to the prompt, a belief change analogous to Bayesian surprise [108,109] (Fig. 5b). Each model was run three times per condition. We removed each model’s overall tendency to call designs active or better, so that ideas were compared by relative belief. Given round 1, the models moved **krgr- ζ 1-40** from sixth to third among the 15 round 2 designs for $P(\text{active})$ and from fifth to second for $P(\text{better})$; in the primary T-cell screen, which tested the 15 round 2 designs alongside 78 round 1 designs, it released the most IFN- γ of the 93 ICDs (Fig. 3b). Across ideas, the in-context shift followed the outcome. It correlated with the gap between the observed fraction and the zero-shot prediction (Spearman ρ , 0.52 for $P(\text{active})$ and 0.35 for $P(\text{better})$), and of the ideas whose prediction moved by 0.03 or more, 12 of 16 moved toward the outcome for $P(\text{active})$ and 11 of 14 for $P(\text{better})$. For 3 ideas in each task, however, the outcome lay outside the in-context range: the models overrated **native-brs-jm**, **foreign-itam** and **traf6-super** for $P(\text{active})$, overrated **plm-ranked-mutant** and **recruit-pick** for $P(\text{better})$ and underrated **jm-lysine-charge**.

To ask what the LLMs learned in context, we compared both predictions with one readout, IFN- γ relative to each round’s comparator, for the 125 round 2 and 3 designs scored in both tasks (Fig. 5c). The earlier rounds’ data raised the correlation of $P(\text{better})$ with the readout for single designs (Spearman ρ , from 0.33 to 0.51) and for the 21 ideas (from 0.40 to 0.63). For $P(\text{active})$, they raised it for ideas (from 0.37 to 0.54) but not for single designs (from 0.40 to 0.37), and in context $P(\text{better})$ tracked single designs more closely than $P(\text{active})$ did (difference in ρ , +0.14; 95% interval +0.01 to +0.28). The size of each design’s update also followed the outcome for $P(\text{better})$ (Spearman ρ , 0.46 across 125 designs; 0.66 in round 2 and 0.44 in round 3): the models raised $P(\text{better})$ for 17 of the 21 designs that exceeded the comparator and for 43 of the 104 that did not. For $P(\text{active})$, the update hardly followed the outcome (ρ , 0.16; Fig. S9b). In both tasks, most of this signal lay between ideas; within ideas, ρ was 0.24 for $P(\text{better})$ and 0.06 for $P(\text{active})$. Thus, the earlier rounds’ data improved the models’ predictions for ideas in both tasks, and for single designs only for $P(\text{better})$.

krgr- ζ 1-40 gave the clearest idea-level update that matched the outcome. It combined two round 1 design ideas: the KRGR anchor of **krgr- ζ 1-4-1bb** and the ITAM1 and CD40 core of **brs- ζ 1-40** (Fig. 3a). Given round 1, the

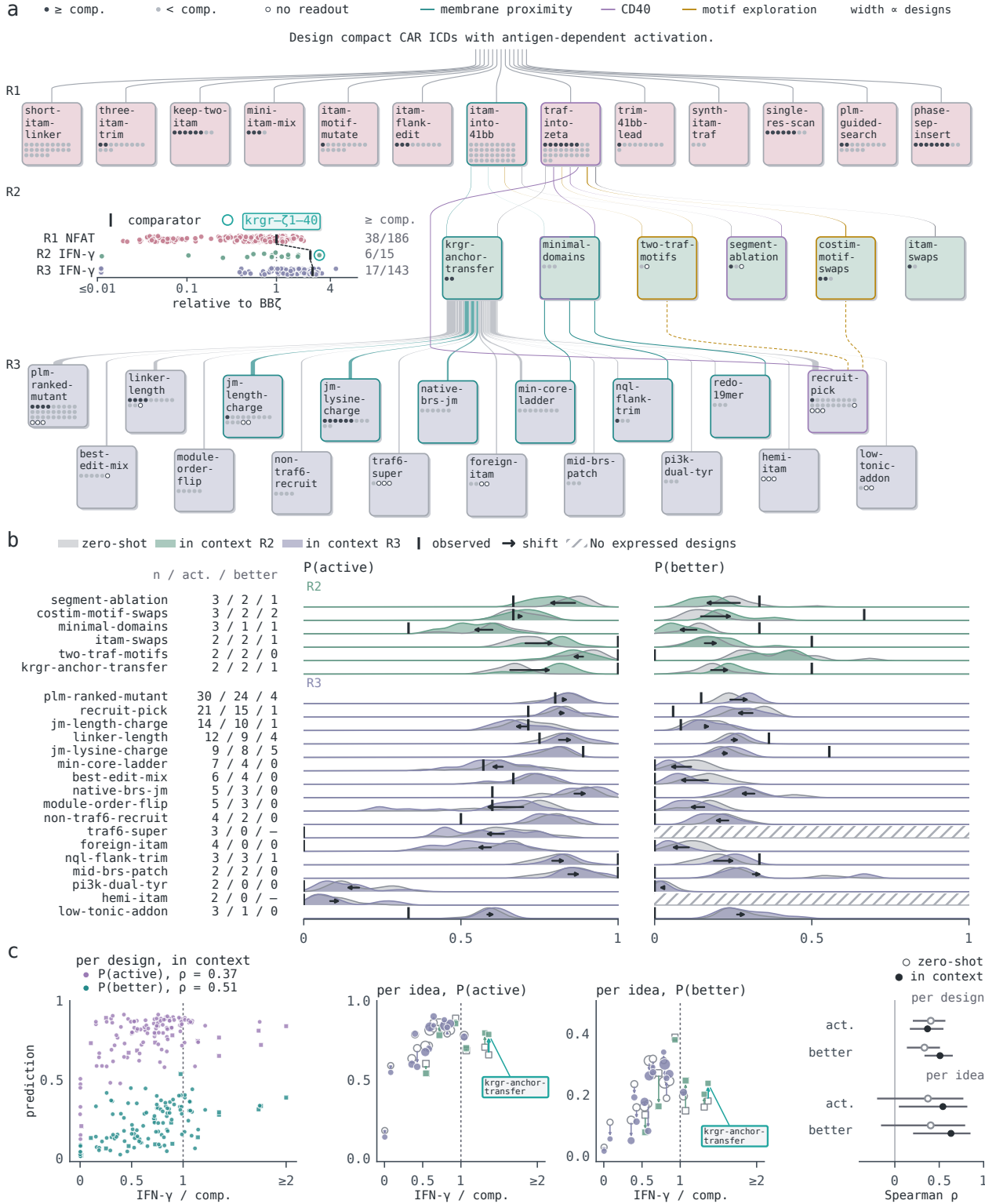


Figure 5: Iterative experimental feedback guides the design of compact ICDs. **a**, Ideas across the three design rounds (left) and each round's readouts (right). Each card is one idea, labeled by its code name (ideas and code names were generated by AI agents; Supplementary Table 1); dots, its designs (dark, at or above the comparator: BB ζ (4-1BB-CD3 ζ) in round 1, the same-plate parent in round 2 and round 3). Ribbons trace lineage; colored borders mark two threads: teal, the distance of the first ITAM tyrosine (Y1) from the membrane; lavender, the CD40 core. Right, every design relative to BB ζ on its plate, colored by round; black bars, key comparators. **b**, For each idea, the mean predicted probability over its designs that a design is active, P(active), or exceeds the round's comparator, P(better), from five LLMs (three runs each), zero-shot (gray) or given the earlier rounds' results (round 2, green; round 3, purple), offsets removed; ridges, distribution over models and runs; arrow, shift of the mean from zero-shot to in context; tick, observed fraction. Left, designs, active designs and designs above the comparator. **c**, Both predictions against IFN- γ relative to the round's comparator (round 2, BB ζ ; round 3, krgr- ζ 1-40; dashed line, equal to the comparator) for the 125 designs scored in both tasks; ensemble of the five models with equal weight, each averaged over three runs. Left, per design, in context; squares, round 2; circles, round 3. Middle, per idea, arrow from the zero-shot mean (open) to the in-context mean (filled); size, designs per idea. Right, Spearman correlation with the readout, zero-shot (open) and in context (filled), with bootstrap 95% intervals.

models raised the predicted activity of this idea more than that of any other round 2 idea (Fig. 5c). In round 3, a single JM edit, R2G (KRGR to KGGR), gave 1.6-fold more surface CAR and 1.2-fold more IFN- γ than **krgr- ζ 1-40** in the screen (Fig. S6b) and a 1.5-fold higher IFN- γ peak across E:T ratios (Fig. S6d). This result suggests that further changes to the membrane-proximal JM segment could improve compact ICDs.

We next examined designs for which the earlier rounds gave a clear signal. Round 1 showed that ICDs with Y1 at 8 or less were inactive and that moving Y1 from 5 to 9 with KRGR restored signaling (Fig. S9a). Given the assay descriptions, the models already rated a round 2 design with Y1 at 5 unlikely to be active (mean P(active), 0.16 zero-shot), lowered it further given round 1 (0.10), and it was inactive (Fig. S9b). Round 3 moved this boundary closer to the membrane: a basic JM spacer kept a design active at Y1 = 6, where a neutral spacer did not (Fig. S9a). Basic JM residues contribute to membrane association and signaling [51,110]; in our compact ICDs, JM sequence also influenced how close to the membrane an ITAM could remain functional. Natural ITAM receptors place Y1 at 13–33 residues [111], so with a basic JM segment, compact ICDs can place Y1 closer to the membrane than natural receptors do. Among designs of the same idea, the models’ in-context predictions tracked activity when the designs varied JM length and charge (Spearman ρ , 0.75) and in part when they varied linker length (0.38), but not among the 30 point mutants or the 21 motif swaps (0.01 and -0.19).

The models could also converge on wrong hypotheses. They rated all five designs that restored the native basic-rich JM segment of CD3 ζ at about 0.9, yet two were inactive. Although native CD3 ζ BRS motifs support membrane association and TCR signaling [52,54], **krgr- ζ 1-40** retained antigen-dependent function after most of these segments were removed (Fig. 3b,d,e). Every in-context run called the same 21 of the 43 inactive round 3 designs active, consistent with the correlated errors that LLMs share [112]. Designs that replaced the second YxxL motif of ITAM1 with the CD28 YNMN motif, proposed in several rounds, expressed but did not signal, as ZAP-70 binding requires both YxxL motifs [95,113]. Conversely, the CD40 core, which binds TRAF6 in native CD40 through a motif distinct from those of 4-1BB [88,89], raised IFN- γ in each round: the lead design with the CD40 core released more IFN- γ than the same ICD with the 4-1BB segment in its place (up to 2.4-fold; Fig. S9c), a gain the models did not anticipate (the CD40 design was never predicted more than 0.07 above its counterpart). Asked whether round 3 designs would exceed **krgr- ζ 1-40**, the models also overrated the point mutants and motif swaps, of which 4 of 27 and 1 of 17 did, and underrated the JM lysine-charge designs, of which 5 of 9 did (Fig. 5b). These results show that the models followed directions that the earlier data made clear but still missed the gain from the CD40 core and overestimated the promise of several design ideas.

Discussion

Here we sought to determine how compact a functional CAR ICD can be. We found that most of a clinical second-generation ICD could be removed: **krgr- ζ 1-40** retains a single ITAM and a CD40-derived segment in 45 aa. Our results establish three properties of compact ICDs. First, they increase surface CAR expression, across designs, binders and delivery methods. Second, they retain antigen-dependent signaling through defined segments, NFAT through the single ITAM and NF- κ B through the CD40-derived motif, with antigen-independent IFN- γ below detection despite higher receptor density. Third, higher expression translated into more durable and broader function after mRNA, lentiviral and lipid-nanoparticle delivery.

Why do compact ICDs express so highly? Because the effect was shared across designs, it may reflect a property of short ICDs more than of a particular sequence. One possibility is that shorter receptors are synthesized, folded or delivered to the cell surface more efficiently. Alternatively, they may be removed from the surface more slowly, because constitutive CD3 ζ phosphorylation depends on ICD size [86] and a shorter ICD carries fewer lysines, whose ubiquitination drives CAR downmodulation after antigen engagement [114]. The effect also depended on the binder, as **brs- ζ 1-40** did not increase the expression of the FMC63 CAR whereas **krgr- ζ 1-4-1bb** did (Fig. 2d). Measuring CAR mRNA, protein synthesis and surface half-life across ICDs and binders would distinguish faster production from slower removal.

Do compact ICDs signal more per receptor, or do they place more receptors on the cell surface? Higher expression contributes: **brs- ζ 1-40** lost its IFN- γ advantage with FMC63, where it also lost its expression advantage (Fig. 2d,e), and the NFAT and NF- κ B signals of **krgr- ζ 1-40** rose in proportion to its surface CAR (Figs. 3e and S2). Signaling per receptor also bears on antigen sensitivity. Because additional ITAMs raise the efficiency with which ligand binding is converted into a signal [15], and CARs are already far less sensitive to antigen than TCRs [16,115], a single-ITAM ICD might need more engaged receptors for the same response. However, when we titrated antigen density [60] at a matched mRNA dose, **krgr- ζ 1-40** T cells needed about 1.7-fold more BCMA per target cell than **BB ζ** T cells for half-maximal antigen-specific IFN- γ , and released more antigen-specific IFN- γ from a lower antigen-independent floor (Fig. S10). Because **krgr- ζ 1-40** displayed 8.4-fold more CAR at this dose (Fig. 4a), its higher surface density likely offsets a lower antigen sensitivity per receptor. Measuring ITAM and ZAP-70 phosphorylation in T cells matched for surface CAR would test this directly.

Our results also bear on the two models that implicate the BRS of the disordered CD3 ζ tail in receptor signaling. **krgr- ζ 1-40** removed most BRS regions from **BB ζ** yet retained antigen-dependent activation (Fig. 3b,d,e). Its antigen-independent IFN- γ also remained below detection at 6.2-fold the surface CAR of **BB ζ** (Fig. 3b). Losing

most of the BRS therefore impaired neither activation nor antigen dependence, and neither model’s prediction held. For condensation, LAT crosslinking [40,42,116] does not address the result, because LAT condenses after and apart from the TCR–ZAP-70 domain [44,45], and CARs can form microclusters without LAT [117]. Instead, the result runs counter to extending the stickers-and-spacers model of IDR phase separation [46–48] to the ICD, in which losing charged stickers such as the BRS should impair activation. One possibility is that antigen binding and the multivalent partners of the ITAM and CD40 motif, ZAP-70 [95] and TRAF6 [89], cluster the receptor without its disordered tail. For membrane sequestration, in which the BRS hold ITAM tyrosines at the membrane until antigen engagement [50–56], the KRGR anchor and the basic residues after the ITAM may suffice to hold a single ITAM there (Fig. S11). Alternatively, antigen-independent CAR signaling may arise mainly from the 4-1BB domain [118] and additional ITAMs that compact ICDs lack, or from charged patches on the binder [119].

Although the native BRS was dispensable, a short basic segment near the membrane was required. With a basic JM segment, a single ITAM remained functional closer to the membrane than in natural ITAM receptors or BBζ (Figs. S9a and S12), and JM segments without lysine or arginine lowered surface CAR. Natural spacing may instead serve regulation, such as holding ITAMs at the membrane until antigen engagement [120]. KRGR, the first four residues of 4-1BB, supplies such a segment. A conserved basic juxtamembrane motif anchors the OX40 tail at the membrane by binding anionic lipids, with similar evidence for 4-1BB and CD40 [110] and for CD28 [121], but to our knowledge KRGR had not been used before as a standalone anchor. Because we measured neither condensates nor basal ITAM phosphorylation, these results argue against a requirement for the BRS rather than against condensation or membrane sequestration. Imaging CAR and LAT clusters separately and measuring ITAM phosphorylation across ICDs that differ in BRS content and juxtamembrane spacing, at matched surface density, would distinguish these possibilities.

How did biological reasoning guide the refactoring? The round 2 designs illustrate the process: measurements from earlier hits motivated combining a short membrane-proximal anchor with an ITAM and a CD40-derived segment. Agents helped turn these observations into sequence proposals, and subsequent experiments determined which combinations worked. Round 3 then tested which changes to the lead preserved or impaired function, allowing us to trace ideas across designs and rounds (Fig. 5a). Because current models did not reliably predict which designs would work (Figs. S8 and S11a–e), we relied on the screens to choose among designs. Experimental feedback guides design in a similar way in directed evolution, antibody design and agent-guided experimentation [122–126]. Given the earlier rounds’ results, the LLMs’ predictions tracked which ideas worked, and tracked which designs worked only when the models were asked whether a design would exceed the comparator (Fig. 5c), consistent with LLMs judging candidates more reliably against one another than on an absolute scale [127]. Whether agent-guided design yields better designs per round than expert or model-based design, and whether it reduces the number of rounds needed, remain to be tested against a baseline matched for experimental budget and data [128], which could also separate the contributions of literature retrieval, biological reasoning and earlier measurements.

Several aspects of our work merit further investigation. The Y1 and JM constraints identified here require prospective tests on new designs. Several experiments used T cells from few donors. Our successive-day challenges measured sustained function rather than serial killing by the same cells, and we did not measure exhaustion, which tonic signaling promotes [11] and rest can reverse [129]. Whether the CD40-derived segment recruits TRAF6 in a CAR, how many cells co-express all four CARs and how compact-ICD CAR T cells perform in vivo remain to be determined.

In summary, successive rounds of agent-guided rational design and experimental feedback identified a 45-aa CAR ICD that increases expression without detectable antigen-independent IFN-γ, extends function after transient mRNA delivery and enhances function after lentiviral and lipid-nanoparticle delivery. The screens also exposed sequence constraints of compact ICDs, such as how close to the membrane the first ITAM tyrosine can sit. Because *krgr-ζ1-40* yielded more receptor per unit of mRNA and occupies less of the payload, compact ICDs may lower the dose needed to generate CAR T cells in vivo [64–67] and enable more complex CAR therapies, such as CAR cocktails.

Methods

CAR constructs and mRNA synthesis

Each CAR comprised the CD8α signal peptide, a FLAG (DYKDDDDK) tag, the antigen-binding domain, the CD8α hinge and transmembrane domain, and the ICD. The BCMA binders were the D-domain binder ddBCMA [87] in the Jurkat screen and Fig. 2i, the bb2121 scFv [92] in primary T cells, and the biparatopic VHH of ciltacabtagene autoleucel in Fig. 2i. CD19, HER2 and MUC16 CARs carried the FMC63 scFv [93], a trastuzumab-derived scFv and an anti-MUC16 scFv from Generate Biomedicines, respectively. ICD coding sequences were codon-optimized for human expression with CodonTransformer [130] and DNA Chisel [131]. DNA templates were synthesized as linear double-stranded DNA by Integrated DNA Technologies, as eBlocks for the compact-ICD CARs with the ddBCMA binder and as gBlocks for all other constructs, and were transcribed without linearization. The constructs carried proprietary in-house 5′ and 3′ UTRs.

mRNA was transcribed in vitro with the HiScribe T7 ARCA mRNA Kit (with tailing) (New England Biolabs, E2060S), which yields a Cap0 structure, and polyadenylated for 1 h with *E. coli* poly(A) polymerase (New England Biolabs, M0276L). Reactions were incubated in a Mastercycler X50 thermocycler (Eppendorf, 6313000018). mRNA was purified with the Total RNA Purification 96-Well Kit (Norgen Biotek, 25100), quantified on a Lunatic spectrophotometer (Unchained Labs) and stored at -80°C until use.

Cell lines

Jurkat cells carrying firefly luciferase under an NFAT- or NF- κ B-responsive promoter (BPS Bioscience, 60621 and 60651, respectively) were cultured in RPMI 1640 Medium, GlutaMAX Supplement, HEPES (Gibco, 72400047) with 10% fetal bovine serum (FBS; Gibco, A5670502) and 1% penicillin–streptomycin (Gibco, 15140122). HEK293T cells (ATCC, CRL-3216) were cultured in DMEM, high glucose (Gibco, 11965118) with 10% FBS and 1% penicillin–streptomycin. NCI-H929 multiple myeloma cells expressing GFP and firefly luciferase (ATCC, CRL-3580-GFP-LUC2) were cultured in RPMI-1640 with 10% FBS, 0.05 mM 2-mercaptoethanol and 1% penicillin–streptomycin. NALM6 leukemia cells (ATCC, CRL-3273), which we transduced with firefly luciferase lentivirus (BPS Bioscience, 79692-P), and Raji lymphoma cells (ATCC, CCL-86) were cultured in RPMI-1640 with 10% FBS and 1% penicillin–streptomycin. All cells were maintained at 37°C with 5% CO_2 and tested negative for mycoplasma by h-IMPACT 1 PCR testing (IDEXX BioAnalytics, 41-00093).

Primary T cells

Cryopreserved peripheral blood mononuclear cells (PBMCs) from healthy donors were purchased from STEMCELL Technologies (70025). PBMCs were thawed into LymphoONE T-Cell Expansion Xeno-Free Medium (Takara Bio, WK552S) supplemented with 10% human AB serum (MP Biomedicals, 092938249), 10 ng/mL IL-7 (STEMCELL Technologies, 78196.1) and 5 ng/mL IL-15 (STEMCELL Technologies, 78218.1), seeded at 1×10^6 cells/mL and activated with T Cell TransAct (Miltenyi Biotec, 130-111-160) at 1:100. Cells were counted with acridine orange/propidium iodide (AO/PI) staining (Revvity, CS2-0106) on a Cellometer K2 (Revvity), and cells with post-thaw viability below 75% were not used. T cells were expanded for three to five days before receptor delivery.

mRNA electroporation

T cells were washed twice in Opti-MEM Reduced Serum Medium (Gibco, 31985070) and resuspended in Opti-MEM at 1×10^6 cells per 20 μL well of the 96-well Nucleocuvette plates from the P3 Primary Cell 4D-Nucleofector X Kit (96-well) (Lonza, V4SP-3096). mRNA was added at 4 μg per 10^6 cells unless stated, and cells were electroporated on a 4D-Nucleofector with the 96-well Shuttle (Lonza) using pulse code DS-137. Cells were recovered immediately into prewarmed medium with 10% human serum at 1×10^6 cells/mL and rested overnight. Jurkat reporter cells were electroporated with the SF Cell Line 4D-Nucleofector X Kit (96-well) (Lonza, V4SC-2096) using pulse code CL-120, with 5×10^5 cells and 8 μg of mRNA per well.

Lentiviral transduction

Lentiviral particles encoding FMC63 CARs with the BB ζ or *krgr*- ζ 1–40 ICD were custom-produced by Vector-Builder. PBMCs were plated at 1×10^5 cells per well in Falcon 96-well clear flat-bottom TC-treated plates (Corning, 353072) and activated with TransAct. After 24 h, lentiviral particles were added directly to the cells at a multiplicity of infection of 10, in T-cell medium without human serum and without a transduction enhancer or spinoculation. On day 4, cells were pooled into G-Rex 6-well plates (Wilson Wolf, 80240M) and expanded until 10 days after transduction, when CAR density was measured and the co-culture assays were set up. Untransduced T cells from the same donor served as controls.

CD3-targeted lipid nanoparticles

CD3-targeted lipid nanoparticles (tLNPs) encapsulating FMC63 CAR mRNA with the BB ζ or *krgr*- ζ 1–40 ICD were produced by GenScript (lots U236FPBQG0-1B and U9797RPHG0-5A, respectively), with an anti-CD3 antibody (clone OKT3) as the targeting ligand and SM-102 as the ionizable lipid. The encapsulated mRNA carried GenScript UTRs, a Cap1 structure, N1-methyl-pseudouridine and a 100-nt poly(A) tail. PBMCs were thawed without activation into LymphoONE with 10% human serum and cultured at 25×10^6 cells per 30 mL in G-Rex 6-well plates. One day after thaw, tLNPs were added directly to the PBMC cultures in an eight-point, 3-fold dilution series starting at 1 μg mRNA per 10^6 cells. Cells were analyzed by flow cytometry one day later. Because tLNPs were added to whole PBMCs, CAR T cells were generated in the same culture as the autologous B cells they targeted.

Flow cytometry and CAR quantification

Cells were stained in BSA Stain Buffer (BD Biosciences, 554657) with a PE-conjugated anti-DYKDDDDK antibody, REAfinity (Miltenyi Biotec, 130-128-816), at 1:100 for 1 h at 4°C in Falcon 96-well non-treated V-bottom plates (Corning, 353263). FMC63 CARs, including those in Fig. 2d, were detected with the same antibody. Cells were

acquired on an ID7000 Spectral Cell Analyzer (Sony Biotechnology) together with QuantiBRITE PE beads (BD Biosciences, 340495) run in the same session at the same settings. CAR molecules per cell were interpolated from the bead standard curve of PE molecules per bead, assuming one PE molecule per antibody. Lymphocytes were gated on forward against side scatter, and CAR-positive cells on PE fluorescence above a CAR-negative reference (Fig. S3). We report the CAR-positive percentage and CAR molecules per CAR-positive cell. The lower limit of detection was 1,000 CAR molecules per cell. Flow data were analyzed in a gating viewer that we built in-house as part of our electronic lab notebook.

Jurkat reporter assays

To make BCMA-positive targets, we transfected HEK293T cells at 100,000 cells per well with 2.4 μ g of linear DNA encoding full-length BCMA using Lipofectamine 3000 Transfection Reagent (Invitrogen, L3000001) and plated them at 60,000 cells per well in Nunc F96 MicroWell white polystyrene plates (Thermo Scientific, 236105). BCMA expression was confirmed by flow cytometry about 24 h after transfection with a PE-conjugated anti-CD269 (BCMA) antibody, REAfinity (clone REA315; Miltenyi Biotec, 130-118-970), and untransfected HEK293T cells served as BCMA-negative targets. CAR-expressing reporter cells were added 18 h after electroporation at a 1:1 ratio of CAR-positive cells to target cells and co-cultured for 4 h at 37 °C. Luminescence was developed in the same wells with 35 μ L of ONE-Glo Luciferase Assay System (Promega, E6120) and read on an EnVision Multimode Plate Reader (Revvity).

The library screen (Fig. 1c,d) comprised two independent experiments. For each design, we divided its mean luminescence across the two experiments by the mean of the BB ζ wells in the same condition. In the KRGR comparison (Fig. S1b), each BCMA-stimulated well was divided by the same-plate BB ζ well. For the signaling mutants (Figs. 2h and 3e), NFAT and NF- κ B reporter cells were electroporated with the same constructs in parallel and stimulated with BCMA-positive HEK293T cells. We report raw luminescence from two technical readings and, in Fig. S2, luminescence divided by the CAR density measured in the same reporter line.

Co-culture assays with primary T cells

Luciferase-expressing target cells were plated at 10,000 cells per well, and effector T cells were added at the stated E:T ratios in 100 μ L of LymphoONE with 10% human serum and without cytokines, in 96-well solid white flat-bottom TC-treated plates (Corning, 3917). E:T ratios count CAR-positive T cells per target cell unless stated. After 24 h at 37 °C, we collected 50 μ L of supernatant for IFN- γ and read residual target-cell luminescence with ONE-Glo in white plates on the EnVision reader. Lower luminescence indicates more killing. Tonic IFN- γ was measured from wells on the same plate that held T cells without target cells.

IFN- γ was measured with the BD CBA Human IFN- γ Flex Set (Bead E7) (BD Biosciences, 558269), with BSA Stain Buffer (BD Biosciences, 554657) as the assay buffer. Supernatant (50 μ L) was incubated with 50 μ L of capture bead mix and 50 μ L of detection reagent for 3 h at room temperature on a plate shaker. Samples were then washed once, acquired on the ID7000 and quantified against a standard curve in our electronic lab notebook. The lower limit of detection was 9.8 pg/mL. Samples above the standard range were diluted.

Primary T-cell screens

In the primary T-cell re-screen (Figs. 1e and 3b), the 78 round 1 designs, the 15 round 2 designs, brs- ζ 1-40, BB ζ and negative controls were expressed as bb2121 CARs in T cells. CAR density was measured 24 h after electroporation, and T cells were co-cultured with H929 cells at 5:1 for 24 h. IFN- γ was measured on two replicate plates (Fig. S1c), and values were divided by those of the same-plate BB ζ . In the round 3 screen (Fig. S6a,b), each of the 164 krgr- ζ 1-40 variants was electroporated once into T cells and co-cultured with H929 cells at 10:1 in two wells, and IFN- γ was averaged across the two wells. Each of the two plates carried krgr- ζ 1-40, BB ζ , ITAM-null and no-mRNA controls, and variant values were divided by those of the same-plate krgr- ζ 1-40. Variants with fewer than 30% CAR-positive cells were excluded.

E:T titrations

For the characterization panel (Fig. 2a-g), ten compact ICDs, BB ζ and a negative control were expressed as bb2121 or FMC63 CARs and co-cultured with H929 or NALM6 cells at eight 2-fold E:T ratios from 20:1 to 0.16:1. BCMA co-cultures had two replicate wells and CD19 co-cultures had one well per condition. For Fig. 3d, round 2 designs including krgr- ζ 1-40, brs- ζ 1-40 and BB ζ were co-cultured with H929 cells at twelve 3-fold E:T ratios from 20:1 to 0.0001:1, in two wells per condition. For Fig. S6c,d, krgr- ζ 1-40 and four round 3 hits were co-cultured with H929 cells at twelve 2-fold E:T ratios from 10:1 to 0.0049:1, in two wells per condition. Killing was normalized to each construct's lowest E:T in Fig. 2c,f, to each construct's highest well in Fig. 3d and to each construct's low-E:T plateau in Fig. S6c. For Fig. 2i, brs- ζ 1-40 and BB ζ CARs with the VHH or ddBCMA binders were measured about 24 h after electroporation, in T cells from a different donor for each binder.

Expression and function after transient mRNA delivery

For the mRNA dose titration (Fig. 4a), T cells were electroporated with twelve 2-fold doses of CAR mRNA from 4 to 0.002 μg per 10^6 cells, and CAR density was measured 24 h later. Cells electroporated with the lowest dose of krgr- ζ 1-40 mRNA served as the CAR-negative reference for the 4 μg plots in Fig. S3.

To follow expression and function over time (Figs. 4b–e and S4), we electroporated T cells with 4 μg of krgr- ζ 1-40 or BB ζ CAR mRNA and cultured them in LymphoONE with 10% human serum, IL-7 and IL-15, refreshing the medium every other day. CAR density and the CAR-positive fraction were measured daily from day 1 to day 17. On every second day from day 1 to day 17, T cells were co-cultured with fresh H929 cells (5,000 per well) for about 24 h, at eleven 2-fold E:T ratios from 20:1 that count total T cells, with one well per condition. H929 viability was normalized to the highest luminescence of each construct and day.

Multi-CAR co-expression

T cells were electroporated on day 4 after activation with a cocktail of four CAR mRNAs carrying BCMA (bb2121), CD19 (FMC63), HER2 (trastuzumab) and MUC16 binders on the same ICD, 2 μg of each (8 μg in total), or with each CAR mRNA alone at 4 μg (Fig. S5). After 24 h, each CAR was detected in a separate well with its PE-conjugated antigen (ACROBiosystems, BCA-HP2H7 for BCMA, CD9-HP2H5 for CD19 and HE2-HP2E3 for HER2), the MUC16 CAR with biotinylated human MUC16 (residues 13810–14451, Fc, Avitag; ACROBiosystems, CA5-H82F8) and streptavidin-PE (BD Pharmingen, 554061), and all with QuantiBRITE calibration. Untransfected T cells served as the CAR-negative control.

Antigen-density titration

HEK293T cells (ATCC, CRL-3216) that we transduced with firefly luciferase lentivirus (BPS Bioscience, 79692-P) were transfected at 100,000 cells per well with eight 2-fold amounts of linear DNA encoding full-length BCMA, from 2.4 μg , using Lipofectamine 3000 to give eight BCMA densities, which we measured with the anti-BCMA antibody above and QuantiBRITE PE beads at co-culture setup. T cells electroporated with 4 μg of krgr- ζ 1-40 or BB ζ CAR mRNA were co-cultured with these targets at 5:1 for 18 h, with one well per condition, and IFN- γ was measured as above (Fig. S10).

Assays with lentivirally transduced T cells

The CAR density of lentivirally transduced T cells (Fig. 4h) was measured by anti-FLAG staining with QuantiBRITE calibration 10 days after transduction. For the 3D aggregate co-culture (Fig. 4i), Raji cells were seeded at 100,000 cells per well in round-bottom ultra-low-attachment spheroid microplates (Corning, 4520), centrifuged at $200 \times g$ for 10 min and left to aggregate for 72 h. Transduced or untransduced T cells were then added at eight 2-fold E:T ratios from 0.1:1 to 0.00078:1, counting CAR-positive T cells against the seeded Raji cells, with one well per condition. On days 1, 3, 5 and 7, 100 μL of supernatant was collected for IFN- γ and replaced with 100 μL of fresh LymphoONE with 10% human serum and 100 IU/mL IL-2 (Miltenyi Biotec, 130-097-743).

For B-cell depletion (Fig. 4j), autologous PBMCs were seeded to give 10,000 CD3⁺CD19⁺ B cells per well, based on the CD19⁺ fraction measured by flow cytometry before seeding. Transduced T cells, activated and transduced separately from the same PBMCs, were added back at twelve 2-fold ratios of CAR-positive T cells to B cells, from 20:1 to 0.0098:1, in 100 μL of cytokine-free LymphoONE with 10% human serum, with one well per condition. After 24 h at 37 °C, cells were stained with anti-CD3-BV510 (clone UCHT1; BD Biosciences, 563109) and anti-CD19-FITC (clone HIB19; BD Biosciences, 555412), and B cells were counted as CD3⁺CD19⁺ events by flow cytometry, without a viability dye. Each count was expressed as a percentage of the highest count for the same construct, and depletion as 100 minus this percentage. In the same experiment, NALM6-luc cells were plated at 10,000 cells per well and co-cultured with the transduced or mock T cells for 24 h at twelve 2-fold E:T ratios from 20:1. Residual luminescence was read as in the other co-culture assays, and killing was calculated relative to mock T cells at each ratio (Fig. S4c).

B-cell depletion after tLNP delivery

PBMCs from five donors were dosed with tLNPs as above (Fig. 4k,l). One day later, cells were stained with anti-DYKDDDDK-PE, anti-CD3 and anti-CD19 as above. CAR-positive T cells are reported as a percentage of CD3⁺ cells. B-cell counts were expressed as a percentage of the highest count in the same plate row, and depletion as 100 minus this percentage. Four donors had two replicate rows per ICD, averaged after normalization, and one donor had one.

Dose-response fitting and replicate agreement

Data were analyzed in Python with NumPy, SciPy and Matplotlib. Killing and B-cell depletion curves were fitted with a decreasing four-parameter logistic, $y = B + A/[1 + (x/C)^h]$, by unweighted least squares over all normalized readings. The bottom B was bounded between 0 and 100%, the amplitude A between 0 and 200 percentage points,

the Hill slope h between 0.1 and 8, and the midpoint C to within two decades of the tested range. We kept the best of 12 starting points. IC50 is the E:T or dose at which the fitted curve crosses 50% viability. IFN- γ against E:T (Fig. 4i) or antigen density (Fig. S10) was fitted with an increasing four-parameter logistic with all four parameters free. Agreement between replicate plates (Fig. S1c,d) is the squared Pearson correlation of $\log_{10}$ values for wells with both replicates above the detection limit. Error bars show the mean $\pm$ s.d. of replicate wells.

Agent-guided sequence design

We used AI agents with access to computational tools to propose compact CAR ICDs from scientist-defined objectives and reference sequences. In round 1, we used Claude Opus 4.6, Claude Sonnet 4.6, GPT-5.2, Kimi K2.5 and MiniMax M2.5. `brs- $\zeta$ 1-40` was proposed by Claude Opus 4.6 in round 1. We prompted the agents to shorten CD3 ζ , vary ITAM number and identity, incorporate costimulatory motifs, or jointly redesign activation and costimulatory elements (Supplementary materials, “Round 1 design prompts”).

The agents had access to literature search, protein annotation, sequence analysis, Python execution, structure prediction and visual assessment of predicted structures. The computational tools ran as serverless functions on Modal.

In rounds 2 and 3, we worked with Claude Code (Opus 4.7 and Opus 4.8, respectively) to interpret experimental measurements and refine the designs. We used literature evidence and results from earlier rounds to define design objectives, revise candidate sequences and develop matched controls.

We organized the designs of each round by idea. An idea was a design question with a written rationale, for example whether the KRGR-KQGQ JM segment acts through charge or through spacing. Each design was assigned to one idea when it was proposed, and the agents gave each idea a short code name. Because no screen could cover all the sequences that an idea allows, the designs within an idea were chosen as contrasts that one plate could interpret. Each design changed one feature relative to the same-plate comparator, and the set included charge-matched pairs at fixed length, residue-by-residue trims, re-tests of earlier designs beside matched controls and tyrosine-null calibrators. For some designs, the rationale also stated the expected direction of the result. Fig. S7 shows six ideas organized in this way, including one whose readouts did not separate its designs.

Evaluation of agent-generated design hypotheses

We evaluated the agents’ design hypotheses retrospectively against the screens (Figs. 5b,c, S8 and S9).

Design data. The benchmark comprised 339 unique ICDs with activity calls: 186 from round 1 (Jurkat NFAT; active at 0.25 or more of BB ζ), 15 from round 2, screened in primary T cells alongside the 78 re-screened round 1 designs, which thus also received round 2 calls (IFN- γ , active at 0.5 or more of the same-plate BB ζ), and 138 new round 3 ICDs (H929 killing, active at 0.5 or more of `krgr- $\zeta$ 1-40`). Constructs encoding the same protein took the majority call; plate controls were excluded, as was `krgr- $\zeta$ 1-40` from round 3, where it served as the reference. For P(better), a design exceeded the comparator when its IFN- γ was at least that of the comparator on the same plate (round 2, BB ζ , 6 of 15 designs; round 3, `krgr- $\zeta$ 1-40`, 16 of the 117 designs with IFN- γ).

Predictions. Five LLMs, each run three times on the same task text (three queried directly and two run as coding agents), gave the probability that each ICD was active, P(active), and, in a separate task, that its readout exceeded that of the round’s comparator, P(better), zero-shot and in context, that is, given the measured ICDs of the earlier rounds (round 1 for round 2; rounds 1 and 2 for round 3). To compare ideas by relative rather than absolute belief, we removed each model’s offset: for every model, round and condition, one shift on the log-odds scale set the model’s mean prediction over the round’s test ICDs to the active fraction of the earlier rounds’ labeled ICDs (those given in the in-context prompt), so that no test label was used. Because one shift is applied to all of a model’s predictions in a round and condition, it leaves each model’s ranking of the designs, and so its AUROC, unchanged. It removes the models’ overall overprediction of activity (in round 3, a mean zero-shot prediction of 0.78 against an observed active fraction of 0.66), which would otherwise dominate the comparison of zero-shot and in-context predictions. For ensemble predictions, probabilities were averaged over runs within a model, and the five models were combined with equal weight.

Metrics. AUROC and specificity were computed per run against the measured calls, with 95% intervals from 1,000 bootstrap resamples of ICDs (Fig. S8). Specificity was computed for each model’s own calls and, independent of its calling threshold, at the highest threshold on its predicted probability that still called at least 90% of active ICDs active, with ties at the threshold called active. For each idea, we compared the mean prediction over its designs, zero-shot and in context, with the fraction of its designs observed active or above the comparator (Fig. 5b). An in-context prediction supported the outcome when its mean lay at least 0.05 closer to the observed fraction than the zero-shot mean. An idea was underrated or overrated when, treating its designs as independent trials with the in-context probabilities, a count as extreme as the one observed had a probability below 0.10, with the observed fraction above or below the in-context mean. The in-context shift of an idea was compared with the gap between its observed fraction and its zero-shot mean by Spearman correlation, and among ideas that moved by 0.03 or more, by the share that moved toward the outcome.

Finally, the Spearman correlation between the ensemble prediction and IFN- γ relative to each round's comparator was computed per design and per idea (the mean over its designs, with ratios capped at 2), for both predictions on the designs scored in both tasks, with intervals from 2,000 bootstrap resamples of designs or ideas (Fig. 5c). For ideas with at least four designs, the correlation of in-context P(active) with the activity readout was computed among the designs of one idea (Fig. S9b).

Acknowledgments

We thank Mike Nally, John Ingraham, Molly Gibson, and Kaggen Ausma for their feedback and support.

References

1. Gross G, Waks T, Eshhar Z. Expression of immunoglobulin-T-cell receptor chimeric molecules as functional receptors with antibody-type specificity. *Proceedings of the National Academy of Sciences of the United States of America*. 1989;86:10024–10028. doi:10.1073/pnas.86.24.10024.
2. Irving BA, Weiss A. The cytoplasmic domain of the T cell receptor ζ chain is sufficient to couple to receptor-associated signal transduction pathways. *Cell*. 1991;64:891–901. doi:10.1016/0092-8674(91)90314-O.
3. Eshhar Z, Waks T, Gross G, Schindler DG. Specific activation and targeting of cytotoxic lymphocytes through chimeric single chains consisting of antibody-binding domains and the γ or ζ subunits of the immunoglobulin and T-cell receptors. *Proceedings of the National Academy of Sciences of the United States of America*. 1993;90:720–724. doi:10.1073/pnas.90.2.720.
4. Finney HM, Lawson ADG, Bebbington CR, et al. Chimeric receptors providing both primary and costimulatory signaling in T cells from a single gene product. *The Journal of Immunology*. 1998;161:2791–2797. doi:10.4049/jimmunol.161.6.2791.
5. Maher J, Brentjens RJ, Gunset G, et al. Human T-lymphocyte cytotoxicity and proliferation directed by a single chimeric TCR ζ /CD28 receptor. *Nature Biotechnology*. 2002;20:70–75. doi:10.1038/nbt0102-70.
6. Finney HM, Akbar AN, Lawson ADG. Activation of resting human primary T cells with chimeric receptors: costimulation from CD28, inducible costimulator, CD134, and CD137 in series with signals from the TCR ζ chain. *The Journal of Immunology*. 2004;172:104–113. doi:10.4049/jimmunol.172.1.104.
7. Imai C, Mihara K, Andreansky M, et al. Chimeric receptors with 4-1BB signaling capacity provoke potent cytotoxicity against acute lymphoblastic leukemia. *Leukemia*. 2004;18:676–684. doi:10.1038/sj.leu.2403302.
8. Milone MC, Fish JD, Carpenito C, et al. Chimeric receptors containing CD137 signal transduction domains mediate enhanced survival of T cells and increased antileukemic efficacy in vivo. *Molecular Therapy*. 2009;17:1453–1464. doi:10.1038/mt.2009.83.
9. Salter AI, Ivey RG, Kennedy JJ, et al. Phosphoproteomic analysis of chimeric antigen receptor signaling reveals kinetic and quantitative differences that affect cell function. *Science Signaling*. 2018;11:eaat6753. doi:10.1126/scisignal.aat6753.
10. Kawalekar OU, O'Connor RS, Fraietta JA, et al. Distinct signaling of coreceptors regulates specific metabolism pathways and impacts memory development in CAR T cells. *Immunity*. 2016;44:380–390. doi:10.1016/j.immuni.2016.01.021.
11. Long AH, Haso WM, Shern JF, et al. 4-1BB costimulation ameliorates T cell exhaustion induced by tonic signaling of chimeric antigen receptors. *Nature Medicine*. 2015;21:581–590. doi:10.1038/nm.3838.
12. Feucht J, Sun J, Eyquem J, et al. Calibration of CAR activation potential directs alternative T cell fates and therapeutic potency. *Nature Medicine*. 2019;25:82–88. doi:10.1038/s41591-018-0290-5.
13. Holst J, Wang H, Eder KD, et al. Scalable signaling mediated by T cell antigen receptor-CD3 ITAMs ensures effective negative selection and prevents autoimmunity. *Nature Immunology*. 2008;9:658–666. doi:10.1038/ni.1611.
14. Guy CS, Vignali KM, Temirov J, et al. Distinct TCR signaling pathways drive proliferation and cytokine production in T cells. *Nature Immunology*. 2013;14:262–270. doi:10.1038/ni.2538.
15. James JR. Tuning ITAM multiplicity on T cell receptors can control potency and selectivity to ligand density. *Science Signaling*. 2018;11:eaan1088. doi:10.1126/scisignal.aan1088.
16. Gudipati V, Rydzek J, Doel-Perez I, et al. Inefficient CAR-proximal signaling blunts antigen sensitivity. *Nature Immunology*. 2020;21:848–856. doi:10.1038/s41590-020-0719-0.
17. Daniels KG, Wang S, Simic MS, et al. Decoding CAR T cell phenotype using combinatorial signaling motif libraries and machine learning. *Science*. 2022;378:1194–1200. doi:10.1126/science.abq0225.
18. Duong CPM, Westwood JA, Yong CSM, et al. Engineering T cell function using chimeric antigen receptors identified using a DNA library approach. *PLoS ONE*. 2013;8:e63037. doi:10.1371/journal.pone.0063037.
19. Goodman DB, Azimi CS, Kearns K, et al. Pooled screening of CAR T cells identifies diverse immune signaling domains for next-generation immunotherapies. *Science Translational Medicine*. 2022;14:eabm1463. doi:10.1126/scitranslmed.abm1463.
20. Gordon KS, Kyung T, Perez CR, et al. Screening for CD19-specific chimaeric antigen receptors with enhanced signalling via a barcoded library of intracellular domains. *Nature Biomedical Engineering*. 2022;6:855–866. doi:10.1038/s41551-022-00896-0.

21. Gordon KS, Perez CR, Garmilla A, et al. Pooled screening for CAR function identifies novel IL-13R α 2-targeted CARs for treatment of glioblastoma. *Journal for ImmunoTherapy of Cancer*. 2025;13:e009574. doi:10.1136/jitc-2024-009574.
22. Castellanos-Rueda R, Di Roberto RB, Bieberich F, et al. speedingCARs: accelerating the engineering of CAR T cells by signaling domain shuffling and single-cell sequencing. *Nature Communications*. 2022;13:6555. doi:10.1038/s41467-022-34141-8.
23. Castellanos-Rueda R, Wang KK, Forster JL, et al. Dissecting the role of CAR signaling architectures on T cell activation and persistence using pooled screens and single-cell sequencing. *Science Advances*. 2025;11:eadp4008. doi:10.1126/sciadv.adp4008.
24. Perez CR, Garmilla A, Nilsson A, et al. Library-based single-cell analysis of CAR signaling reveals drivers of in vivo persistence. *Cell Systems*. 2025;16:101260. doi:10.1016/j.cels.2025.101260.
25. Velasco Cárdenas RMH, Brandl SM, Meléndez AV, et al. Harnessing CD3 diversity to optimize CAR T cells. *Nature Immunology*. 2023;24:2135–2149. doi:10.1038/s41590-023-01658-z.
26. Baeuerle PA, Ding J, Patel E, et al. Synthetic TRuC receptors engaging the complete T cell receptor for potent anti-tumor response. *Nature Communications*. 2019;10:2087. doi:10.1038/s41467-019-10097-0.
27. Wang E, Wang L-C, Tsai C-Y, et al. Generation of potent T-cell immunotherapy for cancer using DAP12-based, multichain, chimeric immunoreceptors. *Cancer Immunology Research*. 2015;3:815–826. doi:10.1158/2326-6066.CIR-15-0054.
28. Tousley AM, Rotiroti MC, Labanieh L, et al. Co-opting signalling molecules enables logic-gated control of CAR T cells. *Nature*. 2023;615:507–516. doi:10.1038/s41586-023-05778-2.
29. Balagopalan L, Moreno T, Qin H, et al. Generation of antitumor chimeric antigen receptors incorporating T cell signaling motifs. *Science Signaling*. 2024;17:eadp8569. doi:10.1126/scisignal.adp8569.
30. Fowler M. Definition of refactoring. 1 September 2004. martinofowler.com. Accessed 25 October 2026.
31. Chan LY, Kosuri S, Endy D. Refactoring bacteriophage T7. *Molecular Systems Biology*. 2005;1:2005.0018. doi:10.1038/msb4100025.
32. Shaw WM, Yamauchi H, Mead J, et al. Engineering a model cell for rational tuning of GPCR signaling. *Cell*. 2019;177:782–796.e27. doi:10.1016/j.cell.2019.02.023.
33. Watson JL, Juergens D, Bennett NR, et al. De novo design of protein structure and function with RFdiffusion. *Nature*. 2023;620:1089–1100. doi:10.1038/s41586-023-06415-8.
34. Ingraham JB, Baranov M, Costello Z, et al. Illuminating protein space with a programmable generative model. *Nature*. 2023;623:1070–1078. doi:10.1038/s41586-023-06728-8.
35. Stark H, Faltings F, Choi M, et al. BoltzGen: toward universal binder design. *bioRxiv*. 2025. doi:10.1101/2025.11.20.689494.
36. Sigalov A, Aivazian D, Stern L. Homooligomerization of the cytoplasmic domain of the T cell receptor ζ chain and of other proteins containing the immunoreceptor tyrosine-based activation motif. *Biochemistry*. 2004;43:2049–2061. doi:10.1021/bi035900h.
37. Sigalov AB, Aivazian DA, Uversky VN, Stern LJ. Lipid-binding activity of intrinsically unstructured cytoplasmic domains of multichain immune recognition receptor signaling subunits. *Biochemistry*. 2006;45:15731–15739. doi:10.1021/bi061108f.
38. Gottweis J, Weng W-H, Daryin A, et al. Accelerating scientific discovery with Co-Scientist. *Nature*. 2026;655:487–496. doi:10.1038/s41586-026-10644-y.
39. Ghareeb AE, Chang B, Mitchener L, et al. A multi-agent system for automating scientific discovery. *Nature*. 2026;655:497–505. doi:10.1038/s41586-026-10652-y.
40. Su X, Ditlev JA, Hui E, et al. Phase separation of signaling molecules promotes T cell receptor signal transduction. *Science*. 2016;352:595–599. doi:10.1126/science.aad9964.
41. Chen H, Xu X, Hu W, et al. Self-programmed dynamics of T cell receptor condensation. *Proceedings of the National Academy of Sciences of the United States of America*. 2023;120:e2217301120. doi:10.1073/pnas.2217301120.
42. Zeng L, Palaia I, Šarić A, Su X. PLC γ 1 promotes phase separation of T cell signaling components. *The Journal of Cell Biology*. 2021;220:e202009154. doi:10.1083/jcb.202009154.
43. Li P, Banjade S, Cheng HC, et al. Phase transitions in the assembly of multivalent signalling proteins. *Nature*. 2012;483:336–340. doi:10.1038/nature10879.
44. Yi J, Balagopalan L, Nguyen T, McIntire KM, Samelson LE. TCR microclusters form spatially segregated domains and sequentially assemble in calcium-dependent kinetic steps. *Nature Communications*. 2019;10:277. doi:10.1038/s41467-018-08064-2.
45. McAfee DB, O’Dair MK, Lin JJ, et al. Discrete LAT condensates encode antigen information from single pMHC:TCR binding events. *Nature Communications*. 2022;13:7446. doi:10.1038/s41467-022-35093-9.
46. Wang J, Choi JM, Holehouse AS, et al. A molecular grammar governing the driving forces for phase separation of prion-like RNA binding proteins. *Cell*. 2018;174:688–699.e16. doi:10.1016/j.cell.2018.06.006.
47. Choi JM, Holehouse AS, Pappu RV. Physical principles underlying the complex biology of intracellular phase transitions. *Annual Review of Biophysics*. 2020;49:107–133. doi:10.1146/annurev-biophys-121219-081629.

48. Borchers W, Bremer A, Borgia MB, Mittag T. How do intrinsically disordered protein regions encode a driving force for liquid–liquid phase separation? *Current Opinion in Structural Biology*. 2021;67:41–50. doi:10.1016/j.sbi.2020.09.004.
49. Xu X, Chen H, Ren Z, et al. Phase separation of chimeric antigen receptor promotes immunological synapse maturation and persistent cytotoxicity. *Immunity*. 2024;57:2755–2771.e8. doi:10.1016/j.immuni.2024.11.005.
50. Aivazian D, Stern LJ. Phosphorylation of T cell receptor ζ is regulated by a lipid dependent folding transition. *Nature Structural Biology*. 2000;7:1023–1026. doi:10.1038/80930.
51. Xu C, Gagnon E, Call ME, et al. Regulation of T cell receptor activation by dynamic membrane binding of the CD3 ϵ cytoplasmic tyrosine-based motif. *Cell*. 2008;135:702–713. doi:10.1016/j.cell.2008.09.044.
52. DeFord-Watts LM, Dougall DS, Belkaya S, et al. The CD3 ζ subunit contains a phosphoinositide-binding motif that is required for the stable accumulation of TCR–CD3 complex at the immunological synapse. *The Journal of Immunology*. 2011;186:6839–6847. doi:10.4049/jimmunol.1002721.
53. DeFord-Watts LM, Tassin TC, Becker AM, et al. The cytoplasmic tail of the T cell receptor CD3 ϵ subunit contains a phospholipid-binding motif that regulates T cell functions. *The Journal of Immunology*. 2009;183:1055–1064. doi:10.4049/jimmunol.0900404.
54. Zhang H, Cordoba SP, Dushek O, et al. Basic residues in the T-cell receptor ζ cytoplasmic domain mediate membrane association and modulate signaling. *Proceedings of the National Academy of Sciences of the United States of America*. 2011;108:19323–19328. doi:10.1073/pnas.1108052108.
55. Shi X, Bi Y, Yang W, et al. Ca²⁺ regulates T-cell receptor activation by modulating the charge property of lipids. *Nature*. 2013;493:111–115. doi:10.1038/nature11699.
56. Li H, Meng L, Chu C, et al. Lipid-regulated phosphorylation hierarchy of the T cell receptor tyrosine motifs. *Molecular Cell*. 2025;85:3694–3710.e8. doi:10.1016/j.molcel.2025.09.005.
57. Gagnon E, Schubert DA, Gordo S, et al. Local changes in lipid environment of TCR microclusters regulate membrane binding by the CD3 ϵ cytoplasmic domain. *Journal of Experimental Medicine*. 2012;209:2423–2439. doi:10.1084/jem.20120790.
58. Walker AJ, Majzner RG, Zhang L, et al. Tumor antigen and receptor densities regulate efficacy of a chimeric antigen receptor targeting anaplastic lymphoma kinase. *Molecular Therapy*. 2017;25:2189–2201. doi:10.1016/j.ymthe.2017.06.008.
59. Eyquem J, Mansilla-Soto J, Giavridis T, et al. Targeting a CAR to the TRAC locus with CRISPR/Cas9 enhances tumour rejection. *Nature*. 2017;543:113–117. doi:10.1038/nature21405.
60. Majzner RG, Rietberg SP, Sotillo E, et al. Tuning the antigen density requirement for CAR T-cell activity. *Cancer Discovery*. 2020;10:702–723. doi:10.1158/2159-8290.CD-19-0945.
61. Rodriguez-Marquez P, Calleja-Cervantes ME, Serrano G, et al. CAR density influences antitumoral efficacy of BCMA CAR T cells and correlates with clinical outcome. *Science Advances*. 2022;8:eabo0514. doi:10.1126/sciadv.abo0514.
62. Zhao Y, Moon E, Carpenito C, et al. Multiple injections of electroporated autologous T cells expressing a chimeric antigen receptor mediate regression of human disseminated tumor. *Cancer Research*. 2010;70:9053–9061. doi:10.1158/0008-5472.can-10-2880.
63. Granit V, Benatar M, Kurtoglu M, et al. Safety and clinical activity of autologous RNA chimeric antigen receptor T-cell therapy in myasthenia gravis (MG-001): a prospective, multicentre, open-label, non-randomised phase 1b/2a study. *The Lancet Neurology*. 2023;22:578–590. doi:10.1016/S1474-4422(23)00194-1.
64. Rurik JG, Tombácz I, Yadegari A, et al. CAR T cells produced in vivo to treat cardiac injury. *Science*. 2022;375:91–96. doi:10.1126/science.abm0594.
65. Hunter TL, Bao Y, Zhang Y, et al. In vivo CAR T cell generation to treat cancer and autoimmune disease. *Science*. 2025;388:1311–1317. doi:10.1126/science.ads8473.
66. Wang Q, Xiao ZX, Zheng X, et al. In vivo CD19 CAR T-cell therapy for refractory systemic lupus erythematosus. *The New England Journal of Medicine*. 2025;393:1542–1544. doi:10.1056/NEJMc2509522.
67. Xu J, Liu L, Parone P, et al. In-vivo B-cell maturation antigen CAR T-cell therapy for relapsed or refractory multiple myeloma. *The Lancet*. 2025;406:228–231. doi:10.1016/S0140-6736(25)01030-X.
68. Chen L, Flies DB. Molecular mechanisms of T cell co-stimulation and co-inhibition. *Nature Reviews Immunology*. 2013;13:227–242. doi:10.1038/nri3405.
69. Roex G, Gordon KS, Lion E, et al. Expanding the CAR toolbox with high throughput screening strategies for CAR domain exploration: a comprehensive review. *Journal for ImmunoTherapy of Cancer*. 2025;13:e010658. doi:10.1136/jitc-2024-010658.
70. Butler SE, Hartman CJ, Huang YH, et al. Toward high-throughput engineering techniques for improving CAR intracellular signaling domains. *Frontiers in Bioengineering and Biotechnology*. 2023;11:1101122. doi:10.3389/fbioe.2023.1101122.
71. Wang P, Wang Y, Zhao X, et al. Chimeric antigen receptor with novel intracellular modules improves anti-tumor performance of T cells. *Signal Transduction and Targeted Therapy*. 2025;10:20. doi:10.1038/s41392-024-02096-5.
72. Courtney AH, Lo W-L, Weiss A. TCR signaling: mechanisms of initiation and propagation. *Trends in Biochemical Sciences*. 2018;43:108–123. doi:10.1016/j.tibs.2017.11.008.

73. Gaud G, Lesourne R, Love PE. Regulatory mechanisms in T cell receptor signalling. *Nature Reviews Immunology*. 2018;18:485–497. doi:10.1038/s41577-018-0020-8.
74. Watts TH. TNF/TNFR family members in costimulation of T cell responses. *Annual Review of Immunology*. 2005;23:23–68. doi:10.1146/annurev.immunol.23.021704.115839.
75. Guedan S, Calderon H, Posey AD, Maus MV. Engineering and design of chimeric antigen receptors. *Molecular Therapy Methods & Clinical Development*. 2019;12:145–156. doi:10.1016/j.omtm.2018.12.009.
76. Majumdar S, Echelibe H, Bettini M, et al. The impact of CD3 ζ ITAM multiplicity and sequence on CAR T-cell survival and function. *Frontiers in Immunology*. 2025;15:1509980. doi:10.3389/fimmu.2024.1509980.
77. Li G, Boucher JC, Kotani H, et al. 4-1BB enhancement of CAR T function requires NF- κ B and TRAFs. *JCI Insight*. 2018;3:e121322. doi:10.1172/jci.insight.121322.
78. Philipson BI, O'Connor RS, May MJ, et al. 4-1BB costimulation promotes CAR T cell survival through noncanonical NF- κ B signaling. *Science Signaling*. 2020;13:eaay8248. doi:10.1126/scisignal.aay8248.
79. Sun C, Shou P, Du H, et al. THEMIS-SHP1 recruitment by 4-1BB tunes LCK-mediated priming of chimeric antigen receptor-redirected T cells. *Cancer Cell*. 2020;37:216–225.e6. doi:10.1016/j.ccell.2019.12.014.
80. Guedan S, Madar A, Casado-Medrano V, et al. Single residue in CD28-costimulated CAR-T cells limits long-term persistence and antitumor durability. *Journal of Clinical Investigation*. 2020;130:3087–3097. doi:10.1172/jci133215.
81. Boucher JC, Li G, Kotani H, et al. CD28 costimulatory domain-targeted mutations enhance chimeric antigen receptor T-cell function. *Cancer Immunology Research*. 2021;9:62–74. doi:10.1158/2326-6066.cir-20-0253.
82. Wu W, Zhou Q, Masubuchi T, et al. Multiple signaling roles of CD3 ϵ and its application in CAR-T cell therapy. *Cell*. 2020;182:855–871.e23. doi:10.1016/j.cell.2020.07.018.
83. Hartl FA, Beck-Garcia E, Woessner NM, et al. Noncanonical binding of Lck to CD3 ϵ promotes TCR signaling and CAR function. *Nature Immunology*. 2020;21:902–913. doi:10.1038/s41590-020-0732-3.
84. Li L, Guo X, Shi X, et al. Ionic CD3–Lck interaction regulates the initiation of T-cell receptor signaling. *Proceedings of the National Academy of Sciences of the United States of America*. 2017;114:E5891–E5899. doi:10.1073/pnas.1701990114.
85. Kagoya Y, Tanaka S, Guo T, et al. A novel chimeric antigen receptor containing a JAK–STAT signaling domain mediates superior antitumor effects. *Nature Medicine*. 2018;24:352–359. doi:10.1038/nm.4478.
86. Ramello MC, Benzaïd I, Kuenzi BM, et al. An immunoproteomic approach to characterize the CAR interactome and signalosome. *Science Signaling*. 2019;12:eaap9777. doi:10.1126/scisignal.aap9777.
87. Buonato JM, Edwards JP, Zaritskaya L, et al. Preclinical efficacy of BCMA-directed CAR T cells incorporating a novel D domain antigen recognition domain. *Molecular Cancer Therapeutics*. 2022;21:1171–1183. doi:10.1158/1535-7163.MCT-21-0552.
88. Ishida T, Mizushima S, Azuma S, et al. Identification of TRAF6, a novel tumor necrosis factor receptor-associated factor protein that mediates signaling from an amino-terminal domain of the CD40 cytoplasmic region. *The Journal of Biological Chemistry*. 1996;271:28745–28748. doi:10.1074/jbc.271.46.28745.
89. Ye H, Arron JR, Lamothe B, et al. Distinct molecular mechanism for initiating TRAF6 signalling. *Nature*. 2002;418:443–447. doi:10.1038/nature00888.
90. Julamanee J, Terakura S, Umemura K, et al. Composite CD79A/CD40 co-stimulatory endodomain enhances CD19CAR-T cell proliferation and survival. *Molecular Therapy*. 2021;29:2677–2690. doi:10.1016/j.ymthe.2021.04.038.
91. Maynard LH, Cavanaugh EJ, Zhu H, et al. Pooled CAR-T screening in pig-tailed macaques identifies designs with enhanced proliferation, trafficking, and persistence. *Blood*. 2025;146:2531–2543. doi:10.1182/blood.2025028683.
92. Raje N, Berdeja J, Lin Y, et al. Anti-BCMA CAR T-cell therapy bb2121 in relapsed or refractory multiple myeloma. *The New England Journal of Medicine*. 2019;380:1726–1737. doi:10.1056/NEJMoa1817226.
93. Nicholson IC, Lenton KA, Little DJ, et al. Construction and characterisation of a functional CD19 specific single chain Fv fragment for immunotherapy of B lineage leukaemia and lymphoma. *Molecular Immunology*. 1997;34:1157–1165. doi:10.1016/S0161-5890(97)00144-2.
94. Frigault MJ, Lee J, Basil MC, et al. Identification of chimeric antigen receptors that mediate constitutive or inducible proliferation of T cells. *Cancer Immunology Research*. 2015;3:356–367. doi:10.1158/2326-6066.CIR-14-0186.
95. Hatada MH, Lu X, Laird ER, et al. Molecular basis for interaction of the protein tyrosine kinase ZAP-70 with the T-cell receptor. *Nature*. 1995;377:32–38. doi:10.1038/377032a0.
96. Adams E, Bai L, Lee M, Yu Y, AlQuraishi M. From mechanistic interpretability to mechanistic biology: training, evaluating, and interpreting sparse autoencoders on protein language models. In: *Proceedings of the 42nd International Conference on Machine Learning (ICML 2025)*. PMLR 267; 2025:460–476. proceedings.mlr.press.
97. Simon E, Zou J. InterPLM: discovering interpretable features in protein language models via sparse autoencoders. *Nature Methods*. 2025;22:2107–2117. doi:10.1038/s41592-025-02836-7.
98. Kusner MJ, Paige B, Hernández-Lobato JM. Grammar variational autoencoder. In: *Proceedings of the 34th International Conference on Machine Learning (ICML 2017)*. PMLR 70; 2017:1945–1954. proceedings.mlr.press.

99. Gelman A, Loken E. The statistical crisis in science. *American Scientist*. 2014;102:460–465. doi:10.1511/2014.111.460.
100. Simmons JP, Nelson LD, Simonsohn U. False-positive psychology: undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*. 2011;22:1359–1366. doi:10.1177/0956797611417632.
101. Silberzahn R, Uhlmann EL, Martin DP, et al. Many analysts, one data set: making transparent how variations in analytic choices affect results. *Advances in Methods and Practices in Psychological Science*. 2018;1:337–356. doi:10.1177/2515245917747646.
102. Borjas GJ, Breznau N. Ideological bias in the production of research findings. *Science Advances*. 2026;12:eadz7173. doi:10.1126/sciadv.adz7173.
103. Miao J, Pritchard JK, Zou J. The agentic garden of forking paths. *arXiv*. 2026. doi:10.48550/arXiv.2607.01507.
104. Passaro S, Corso G, Wohlwend J, et al. Boltz-2: towards accurate and efficient binding affinity prediction. *bioRxiv*. 2025. doi:10.1101/2025.06.14.659707.
105. Lin Z, Akin H, Rao R, et al. Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science*. 2023;379:1123–1130. doi:10.1126/science.ade2574.
106. Ferruz N, Schmidt S, Höcker B. ProtGPT2 is a deep unsupervised language model for protein design. *Nature Communications*. 2022;13:4348. doi:10.1038/s41467-022-32007-7.
107. Schwag UM, Lau E, Ehsani Oskouie H, et al. SciPredict: can LLMs predict the outcomes of scientific experiments in natural sciences? *arXiv*. 2026. doi:10.48550/arXiv.2604.10718.
108. Itti L, Baldi P. Bayesian surprise attracts human attention. In: *Advances in Neural Information Processing Systems 18 (NIPS 2005)*; 2005. proceedings.neurips.cc.
109. Agarwal D, Majumder BP, Adamson R, et al. AutoDiscovery: open-ended scientific discovery via Bayesian surprise. In: *Advances in Neural Information Processing Systems 38 (NeurIPS 2025)*; 2025:28504–28542. doi:10.52202/085713-0847.
110. Zhu J, Xu X, Liu P, et al. A conserved basic motif mediates membrane anchoring and TRAF2 recruitment across costimulatory TNFRs. *The EMBO Journal*. 2026;45:5827–5850. doi:10.1038/s44318-026-00868-z.
111. UniProt Consortium. UniProt: the Universal Protein Knowledgebase in 2025. *Nucleic Acids Research*. 2025;53:D609–D617. doi:10.1093/nar/gkae1010.
112. Kim EM, Garg A, Peng K, Garg N. Correlated errors in large language models. In: *Proceedings of the 42nd International Conference on Machine Learning (ICML 2025)*. PMLR 267; 2025:30038–30066. proceedings.mlr.press.
113. Iwashima M, Irving BA, van Oers NSC, et al. Sequential interactions of the TCR with two distinct cytoplasmic tyrosine kinases. *Science*. 1994;263:1136–1139. doi:10.1126/science.7509083.
114. Li W, Qiu S, Chen J, et al. Chimeric antigen receptor designed to prevent ubiquitination and downregulation showed durable antitumor efficacy. *Immunity*. 2020;53:456–470.e6. doi:10.1016/j.immuni.2020.07.011.
115. Salter AI, Rajan A, Kennedy JJ, et al. Comparative analysis of TCR and CAR signaling informs CAR designs with superior antigen sensitivity and in vivo function. *Science Signaling*. 2021;14:eabe2606. doi:10.1126/scisignal.abe2606.
116. Zhang W, Sloan-Lancaster J, Kitchen J, et al. LAT: the ZAP-70 tyrosine kinase substrate that links T cell receptor to cellular activation. *Cell*. 1998;92:83–92. doi:10.1016/s0092-8674(00)80901-0.
117. Dong R, Libby KA, Blaeschke F, et al. Rewired signaling network in T cells expressing the chimeric antigen receptor (CAR). *The EMBO Journal*. 2020;39:e104730. doi:10.15252/embj.2020104730.
118. Gomes-Silva D, Mukherjee M, Srinivasan M, et al. Tonic 4-1BB costimulation in chimeric antigen receptors impedes T cell survival and is vector-dependent. *Cell Reports*. 2017;21:17–26. doi:10.1016/j.celrep.2017.09.015.
119. Chen J, Qiu S, Li W, et al. Tuning charge density of chimeric antigen receptor optimizes tonic signaling and CAR-T cell fitness. *Cell Research*. 2023;33:341–354. doi:10.1038/s41422-023-00789-0.
120. Lee MS, Glassman CR, Deshpande NR, et al. A mechanical switch couples T cell receptor triggering to the cytoplasmic juxtamembrane regions of CD3ζ. *Immunity*. 2015;43:227–239. doi:10.1016/j.immuni.2015.06.018.
121. Dobbins J, Gagnon E, Godec J, et al. Binding of the cytoplasmic domain of CD28 to the plasma membrane inhibits Lck recruitment and signaling. *Science Signaling*. 2016;9:ra75. doi:10.1126/scisignal.aaf0626.
122. Yang J, Lal RG, Bowden JC, et al. Active learning-assisted directed evolution. *Nature Communications*. 2025;16:714. doi:10.1038/s41467-025-55987-8.
123. Frey NC, Hötzel I, Stanton SD, et al. Lab-in-the-loop therapeutic antibody design with deep learning. *bioRxiv*. 2025. doi:10.1101/2025.02.19.639050.
124. Yang C, Wang X, Lu Y, et al. Large language models as optimizers. In: *The Twelfth International Conference on Learning Representations (ICLR 2024)*; 2024. doi:10.48550/arXiv.2309.03409.
125. Boiko DA, MacKnight R, Kline B, Gomes G. Autonomous chemical research with large language models. *Nature*. 2023;624:570–578. doi:10.1038/s41586-023-06792-0.
126. Cutts Z, Collier J, Chen R, et al. Aptamers as a sandbox for lab-in-the-loop agentic molecular discovery. *arXiv*. 2026. doi:10.48550/arXiv.2610.24806.
127. Liusie A, Manakul P, Gales MJF. LLM comparative assessment: zero-shot NLG evaluation through pairwise comparisons using large language models. In: *Proceedings of the 18th Conference of the European Chapter*

- of the Association for Computational Linguistics (Volume 1: Long Papers). Association for Computational Linguistics; 2024:139–151. doi:10.18653/v1/2024.eacl-long.8.
128. Kim J, Romero PA. Benchmarking and behavioral characterization of LLM agents for protein design. *bioRxiv*. 2026. doi:10.64898/2026.05.06.723381.
 129. Weber EW, Parker KR, Sotillo E, et al. Transient rest restores functionality in exhausted CAR-T cells through epigenetic remodeling. *Science*. 2021;372:eaba1786. doi:10.1126/science.aba1786.
 130. Fallahpour A, Gureghian V, Filion GJ, Lindner AB, Pandi A. CodonTransformer: a multispecies codon optimizer using context-aware neural networks. *Nature Communications*. 2025;16:3205. doi:10.1038/s41467-025-58588-7.
 131. Zulkower V, Rosser S. DNA Chisel, a versatile sequence optimizer. *Bioinformatics*. 2020;36:4508–4509. doi:10.1093/bioinformatics/btaa558.
 132. Kumar M, Possemato AP, Zee BM, et al. Expanding the global map of protein post-translational modifications with immunoaffinity enrichment and nDIA analysis on the Orbitrap Astral mass spectrometer. *Molecular & Cellular Proteomics*. 2026;25:101562. doi:10.1016/j.mcpro.2026.101562.
 133. Callahan A, Trychanh SS, Ro T, et al. Phosphotyrosine proteomics reveals novel Zap70 and Itk pathway targets downstream of TCR and CAR in Jurkat T cells. *Scientific Reports*. 2026;16:16482. doi:10.1038/s41598-026-47234-x.
 134. Yaron-Barir TM, Joughin BA, Huntsman EM, et al. The intrinsic substrate specificity of the human tyrosine kinome. *Nature*. 2024;629:1174–1181. doi:10.1038/s41586-024-07407-y.
 135. Ginell GM, Emenecker RJ, Lotthammer JM, et al. Sequence-based prediction of intermolecular interactions driven by disordered regions. *Science*. 2025;388:eadq8381. doi:10.1126/science.adq8381.
 136. Shah NH, Löbel M, Weiss A, Kuriyan J. Fine-tuning of substrate preferences of the Src-family kinase Lck revealed through a high-throughput specificity screen. *eLife*. 2018;7:e35190. doi:10.7554/eLife.35190.
 137. Bauch A, Adamczyk I, Buczek P, et al. openBIS: a flexible framework for managing and analyzing complex data in biology research. *BMC Bioinformatics*. 2011;12:468. doi:10.1186/1471-2105-12-468.
 138. Vrana J, de Lange O, Yang Y, et al. Aquarium: open-source laboratory software for design, execution and data management. *Synthetic Biology*. 2021;6:ysab006. doi:10.1093/synbio/ysab006.

Supplementary materials

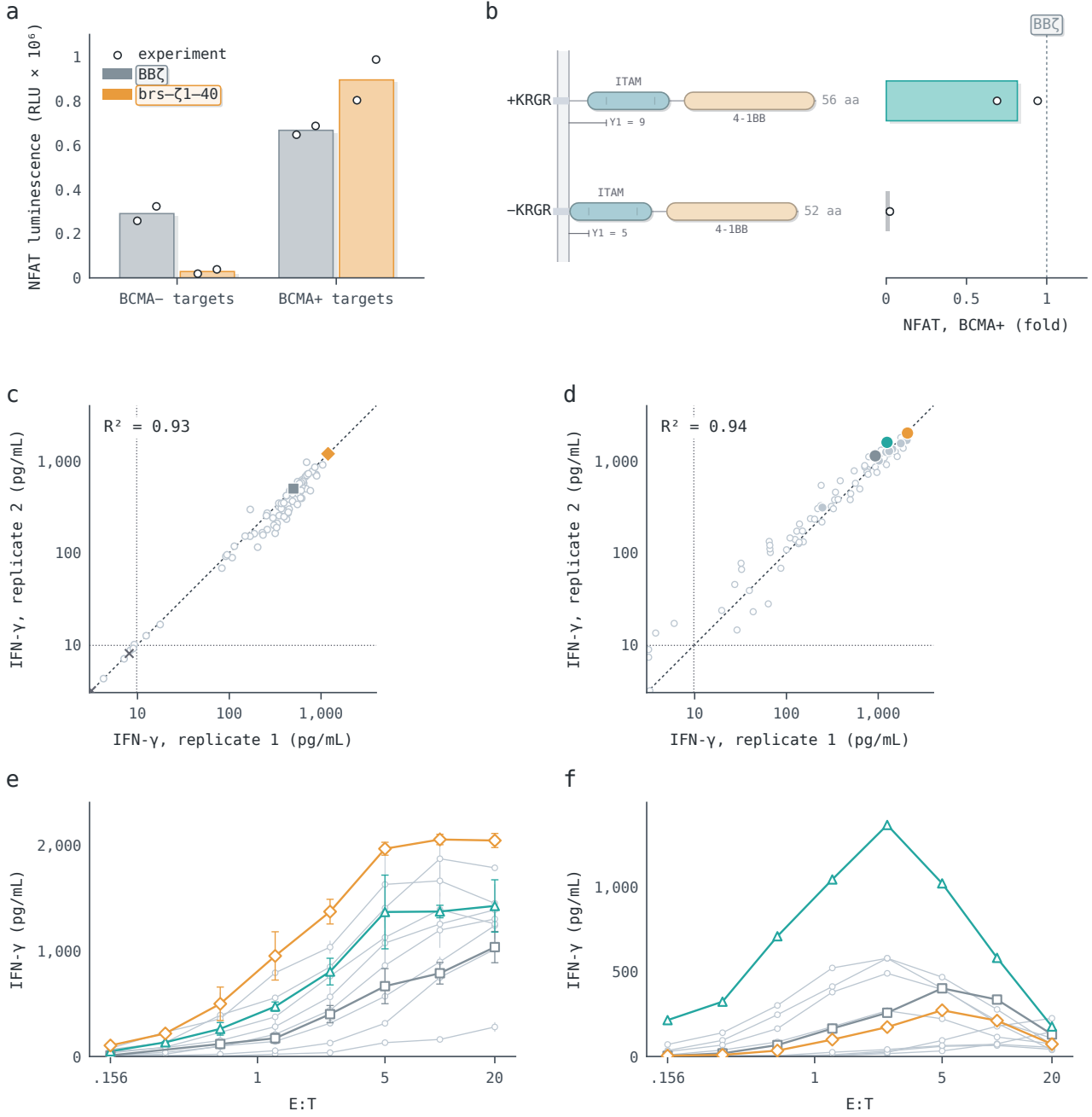


Figure S1: Individual-experiment Jurkat NFAT responses, the KRGR comparison, agreement between replicate IFN- γ measurements and IFN- γ across E:T ratios. **a**, NFAT luminescence of brs- $\zeta 1-40$ and BB ζ (4-1BB-CD3 ζ) without antigen and with BCMA-positive targets in each of two independent experiments. **b**, BCMA-stimulated NFAT for krgr- $\zeta 1-4-1bb$ with (+KRGR) and without (−KRGR) its four N-terminal residues, relative to the same-plate BB ζ (dashed line). Without KRGR, the first ITAM tyrosine (Y1) sits four residues closer to the membrane and the design did not function. **c,d**, H929-induced IFN- γ in replicate 1 against replicate 2, measured on separate plates; dashed line, identity; dotted lines, lower limit of detection; R^2 , squared Pearson correlation of $\log_{10}$ values for wells above that limit (c, 74 wells; d, 73 wells). **c**, The primary T-cell re-screen of Fig. 1e, which plots replicate 1: library designs (light gray circles), brs- $\zeta 1-40$ (orange diamond), BB ζ (gray square) and negative controls (×). **d**, The BCMA CARs of Fig. 2b: small open light-gray circles, non-peak E:T wells; filled circles, each design's peak, as ranked in Fig. 2b: brs- $\zeta 1-40$ (orange), krgr- $\zeta 1-4-1bb$ (teal), BB ζ (gray) and other designs (small, light gray). **e,f**, IFN- γ across the eight E:T ratios behind the peak values of Fig. 2b (e; H929 targets, BCMA CARs; mean $\pm$ s.d. of the two replicate plates, or the single reading where only one exists) and Fig. 2e (f; NALM6 targets, CD19 CARs; one well per E:T), with points joined without fitting: brs- $\zeta 1-40$ (orange diamonds), krgr- $\zeta 1-4-1bb$ (teal triangles), BB ζ (gray squares) and other designs (light gray circles).

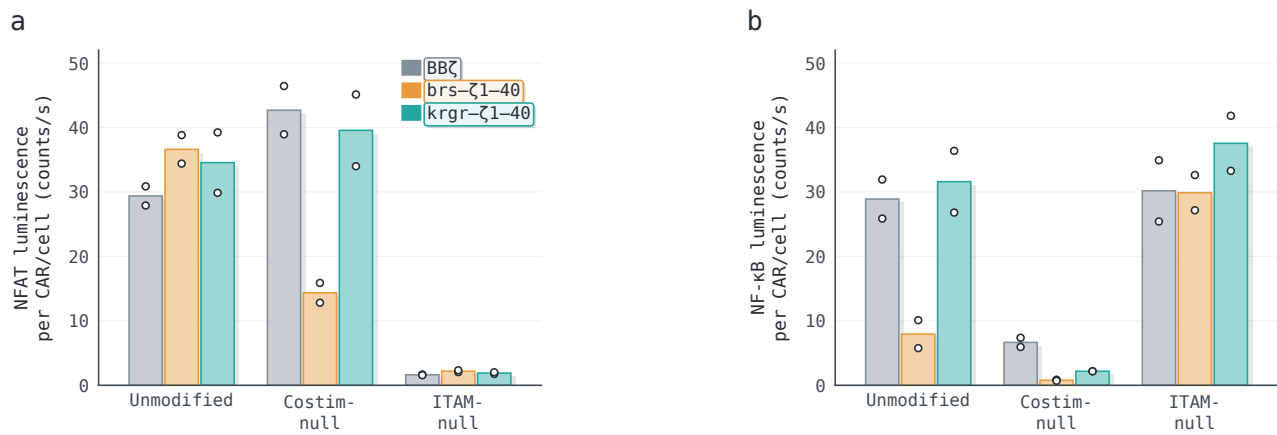


Figure S2: Reporter signal per surface CAR for the signaling mutants of Figs. 2h and 3e. a,b, Jurkat NFAT (a) and NF-κB (b) luminescence divided by surface CAR density for the unmodified, costimulation-null and ITAM-null forms of BBζ (4-1BB-CD3ζ), brs-ζ1-40 and krgr-ζ1-40. Bars, means; circles, individual readings.

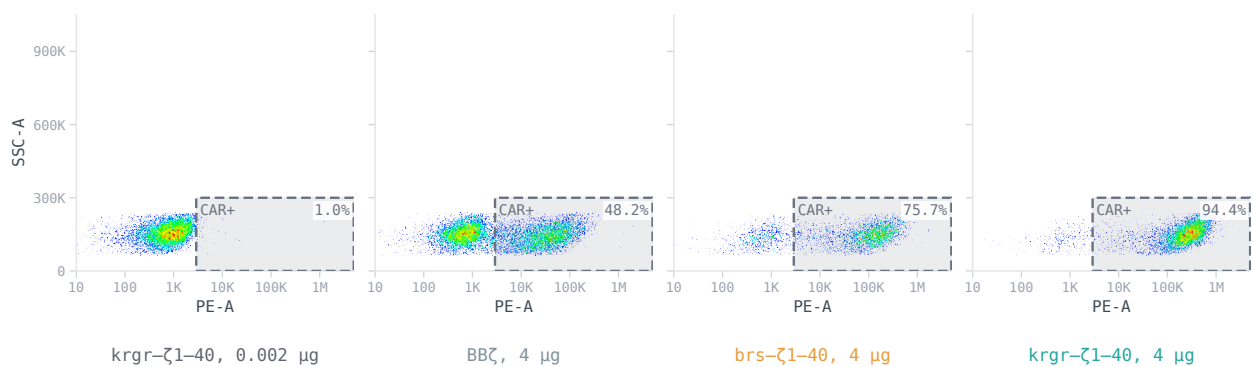


Figure S3: Surface CAR staining at the highest mRNA dose of Fig. 4a. Side scatter against anti-CAR staining of primary T cells expressing BBζ (4-1BB-CD3ζ), brs-ζ1-40 or krgr-ζ1-40 CARs, beside a CAR-negative reference (left). Color shows event density; dashed rectangles, the CAR-positive gate, with the percentage of CAR-positive cells.

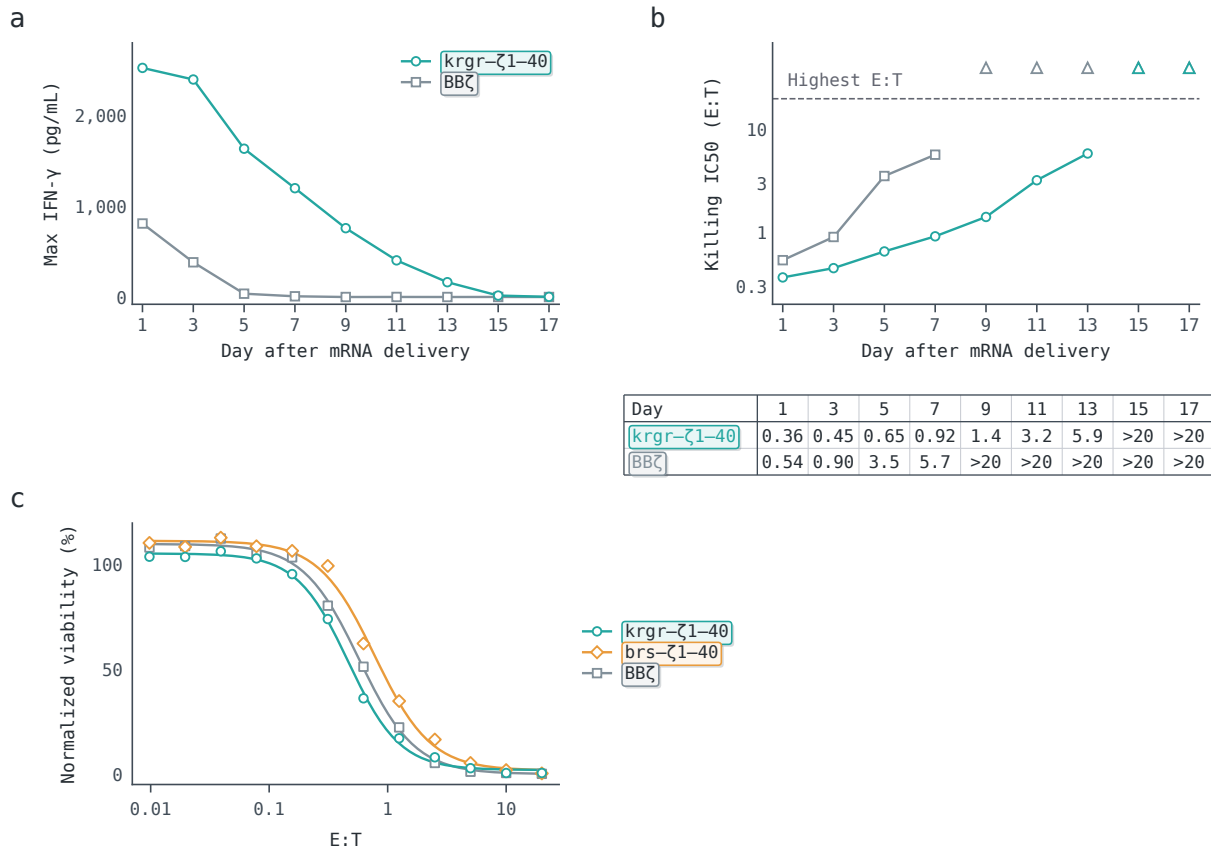


Figure S4: Loss of IFN- γ release and killing after mRNA electroporation, and NALM6 killing by lentiviral CAR T cells. **a**, Highest IFN- γ across the E:T ratios of Fig. 4d on each day. **b**, Killing IC₅₀ on each day, the E:T at which the fitted H929 viability curves of Fig. 4e cross 50%; triangles, days whose fits stay above 50% at the highest E:T. The values under b list the same IC₅₀s (>20, above 50% at the highest E:T). **c**, Viability of NALM6 leukemia cells after co-culture with the lentivirally transduced CAR T cells of Fig. 4j, and with brs- ζ 1-40 CAR T cells from the same experiment, across twelve 2-fold E:T ratios, as 100 minus killing relative to mock T cells; curves, descriptive four-parameter fits as in Fig. 4j.

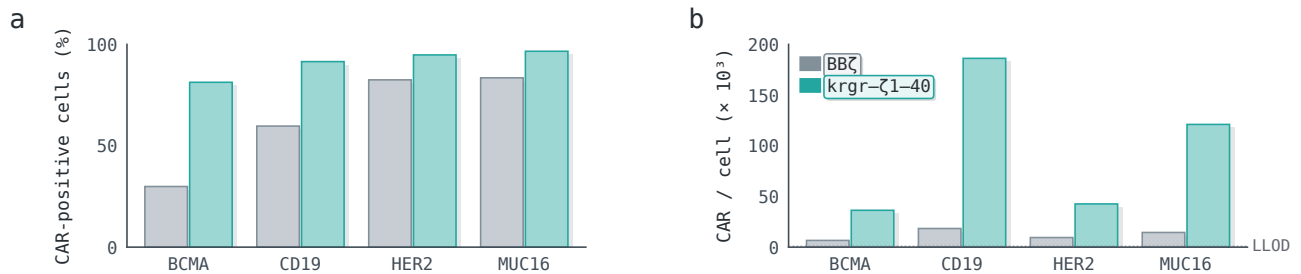


Figure S5: Single-CAR expression after mRNA electroporation. **a,b**, Each of the four CARs of Fig. 4f,g (BCMA, CD19, HER2 and MUC16 binders) electroporated alone on BBζ (4-1BB-CD3ζ) or krggr- ζ 1-40: CAR-positive cells (a) and CAR molecules per cell (b) at 24 h, each CAR detected with its own antigen. Dotted line, lower limit of detection.

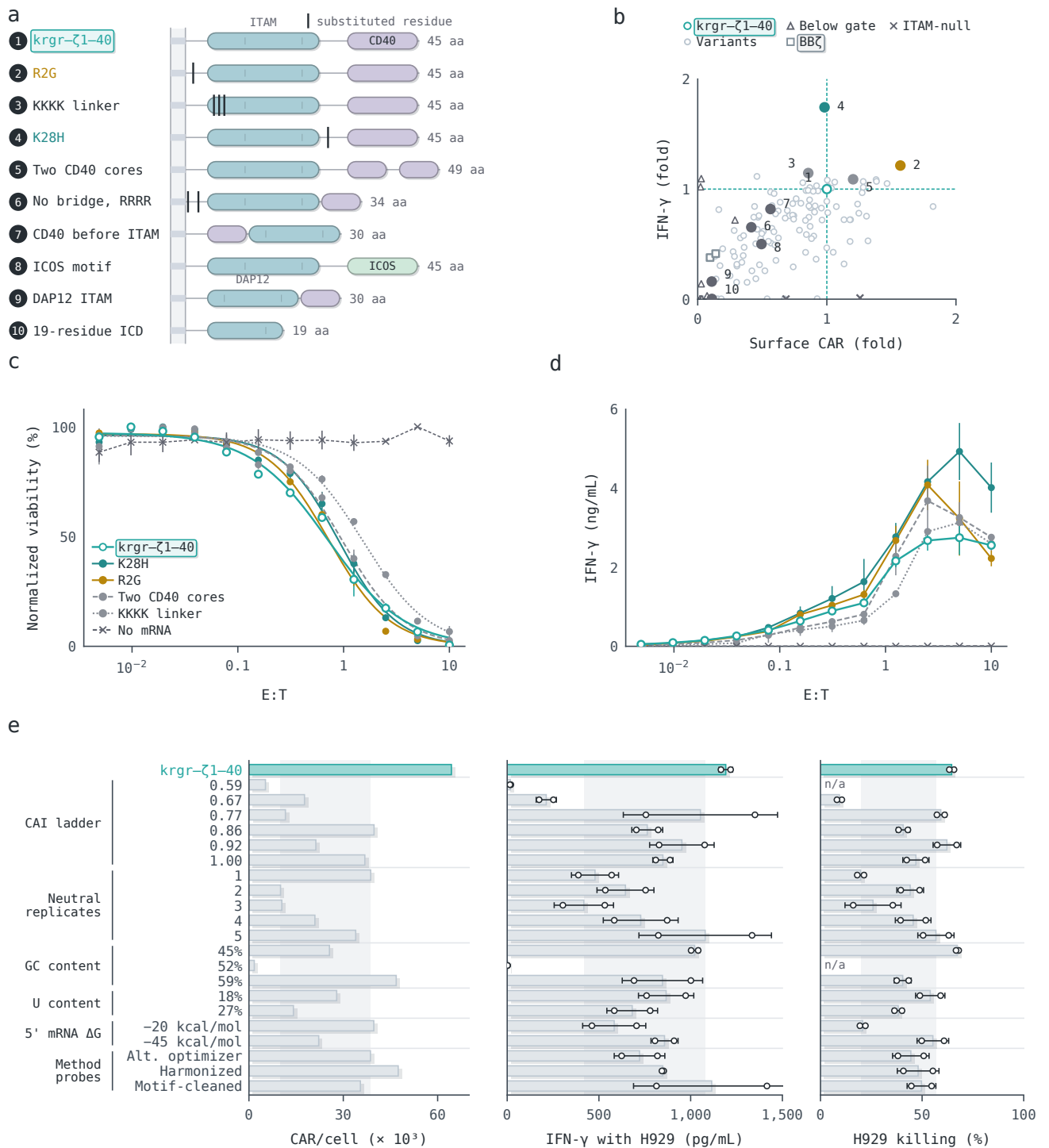


Figure S6: Round 3 variants of *krgr-ζ1-40*. **a**, Representative round 3 designs drawn to residue scale with segments as in Fig. 1b and numbered as in b; ticks mark substituted residues. **b**, Surface CAR density and H929-induced IFN- γ of each variant, relative to the same-plate *krgr-ζ1-40*; open circles, variants that passed the expression gate; triangles, variants below the gate with a measured CAR density; filled circles, the designs of a. **c,d**, H929 viability (c) and IFN- γ (d) across E:T ratios for *krgr-ζ1-40* and four screen hits. **e**, Synonymous codon recodes of the *krgr-ζ1-40* CAR, grouped by codon-design family, with their CAR/cell, H929-induced IFN- γ and killing below *krgr-ζ1-40* with its validated codons; pale band, the range of five neutral recodes; n/a, failed the expression gate.



Figure S7: Design ideas built as controlled contrasts. Each card is one idea of Fig. 5a, labeled by its code name and framed in its round's color (green, round 2; indigo, round 3). Left, the idea's designs drawn to residue scale from the transmembrane end, with segments as in Fig. 1b; right, IFN-γ and surface CAR relative to the round's comparator on the same plate (dashed line): BBζ (4-1BB-CD3ζ) in round 2 and krgr-ζ1-40 in round 3, which is drawn as an open teal circle. **a**, Transfer of the KRGR anchor onto two round 1 hits, brs-ζ1-40 (orange diamonds) and ζ3-traf2. **b**, JM length and charge, with charge-matched pairs at fixed length. **c**, Shorter ICDs that keep ITAM1 and the CD40 core. **d**, The inactive 19-aa design re-tested beside a matched-length design with the native ITAM1 spacer and a tyrosine-null calibrator. **e**, Stepwise trimming of the three residues before the first ITAM1 tyrosine. **f**, An inconclusive idea: the length of the linkers before (N-linker) and after (C-linker) ITAM1. One well per round 3 design, each relative to krgr-ζ1-40 on its own plate.

LLM evaluations of designs

Data. The benchmark uses 339 unique ICD proteins with activity calls from three rounds: round 1, the library screen (186 ICDs; Jurkat NFAT, active at 0.25 or more of BB ζ); round 2, the 15 new designs and 78 re-screened round 1 designs in primary T cells (93 ICDs; IFN- γ , active at 0.5 or more of the BB ζ CAR on the plate); and round 3, the krgr- ζ 1-40 variants of Figs. 5 and S6 (138 ICDs not measured before; H929 killing, active at 0.5 or more of krgr- ζ 1-40). Constructs encoding the same protein take the majority call; plate controls and krgr- ζ 1-40 are excluded.

Settings. All five models received one task text (T1). It describes the CAR, the three assays and the files: test.csv, the ICDs to predict, and in context train.csv, the labeled ICDs of the earlier rounds. The single-call LLMs received the same text with the files inline and answered in a JSON schema; the coding agents worked in an isolated sandbox that held the files and wrote predictions.csv (T2). Each prediction comprises a short reasoning, a probability and a label, and each model was run three times per condition. Zero-shot, a model saw no labels; in context, it saw the labeled ICDs of the earlier rounds (round 1, 186 ICDs, for the 15 round 2 designs; rounds 1 and 2, 279 labeled measurements of 201 unique ICDs, for the 138 round 3 ICDs). Every run returned a prediction for every ICD.

	Single-call LLMs	Coding agents
Models	GPT-6.1 (openai/gpt-6.1-sol), Gemini 4 (google/gemini-4-argon), Claude Opus 5 (anthropic/claude-opus-5)	Claude Code 2.1.309 running Claude Opus 5.5; Codex CLI 0.163.0 running gpt-6-astra
Access	OpenRouter; strict JSON output schema; no temperature setting; provider-default reasoning	isolated sandbox with the task files, a shell, Python (numpy, pandas, scipy, scikit-learn) and network access for the model API only; reasoning effort high
Query	all test ICDs of a round in one query, files inline (T1, T2)	TASK.md, test.csv and in context train.csv in the working directory (T1-T4)

The baselines are the Y1 rule, Boltz-2 co-fold confidence and ESM-2 likelihood.

Comparator task. In a second task (C1), the same five models estimated the probability that each ICD’s readout exceeded that of the round’s comparator on the same plate, P(better): BB ζ for the 15 round 2 designs and krgr- ζ 1-40 for the 117 round 3 designs with an IFN- γ measurement. The task names the comparator and gives its sequence; the round 3 readout is IFN- γ , as in the round 3 screen. In context, train.csv gave each earlier measurement’s round, assay, ratio to BB ζ and label (higher, a ratio of 1 or more, or lower): the round 1 library in Jurkat NFAT cells and the 78 library designs re-screened in primary T cells (264 measurements) for round 2, with the round 2 plate added (282 measurements) for round 3. The single-call LLMs ran with reasoning effort high, Claude Opus 5 through the Claude API with adaptive thinking; each model ran three times per condition.

Scoring. AUROC and specificity were computed per run against the measured calls and averaged over runs, with 95% intervals from 1,000 bootstrap resamples of ICDs. Specificity was computed for each model’s own calls and at the highest threshold on its predicted probability that still called at least 90% of active ICDs active; the latter, like AUROC, depends only on the order of the predicted probabilities, not on where the model places its call threshold.

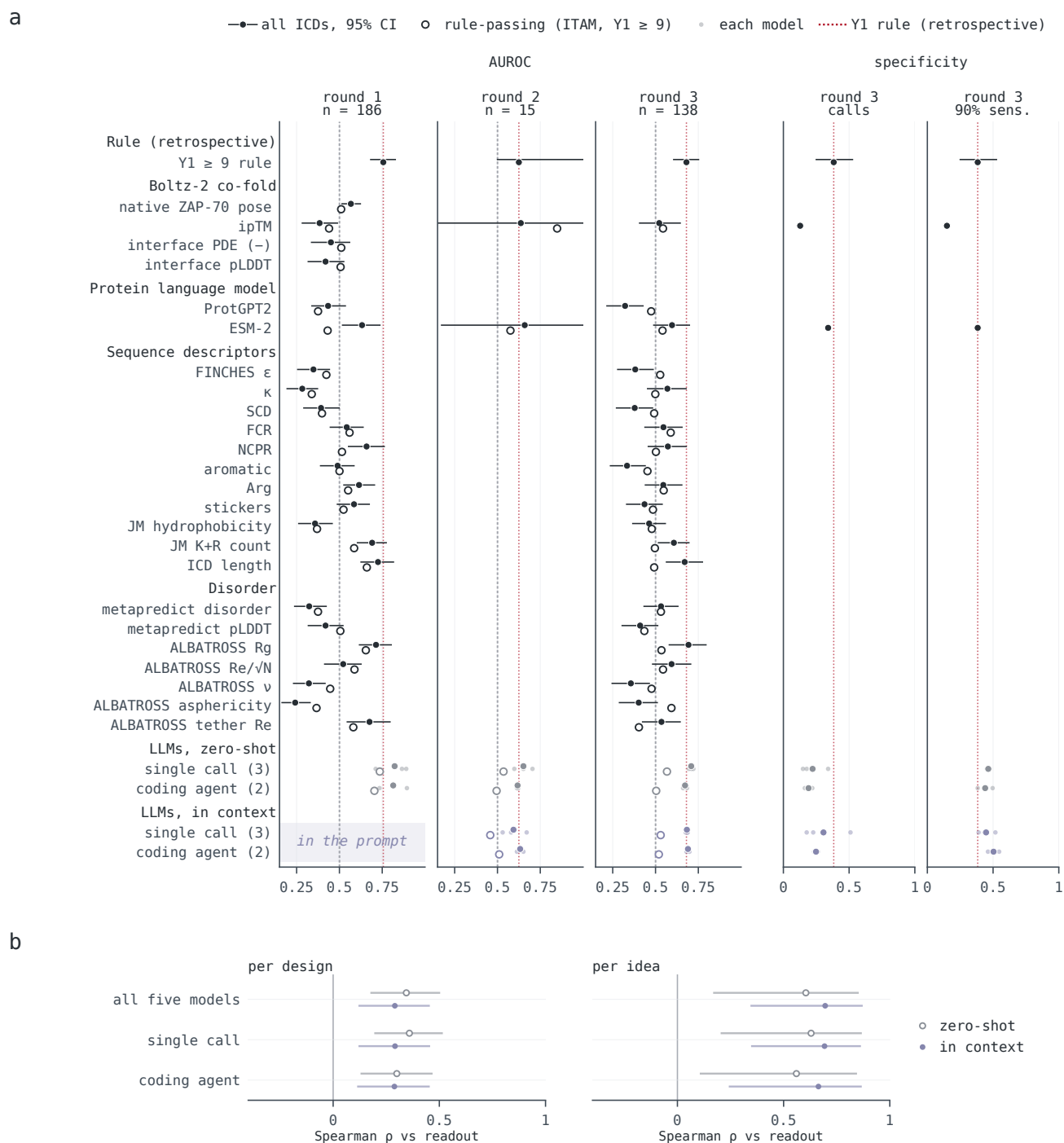


Figure S8: Benchmarks of the model predictions. **a**, Every method on the same ICDs: the 186 library ICDs of round 1, the 15 new designs of round 2 and the 138 new ICDs of round 3. In round 2, only the Y1 rule, Boltz-2 ipTM, ESM-2 and the language models were scored. AUROC for active versus inactive ICDs, for all ICDs (filled, with 95% bootstrap interval) and for ICDs that pass the Y1 rule (open), and round 3 specificity of each method's own calls and at the threshold that keeps 90% of active ICDs (90% sens.), shown for the methods that give calls or a stored score. Scores include the Y1 rule, Boltz-2 co-fold scores, protein language-model likelihoods, and condensate, charge, composition and disorder descriptors. Language models, the mean over the single-call LLMs (GPT-6.1, Gemini 4, Claude Opus 5) or the coding agents (Claude Code, Codex), with each model as a small dot; in context, the prompt held the labeled ICDs of rounds 1 and 2 (shaded, round 1 in the prompt). The Y1 rule was derived retrospectively, after all screens were measured, and is shown as a reference rather than a prediction. Dashed line, chance; dotted line, the Y1 rule. **b**, Spearman correlation of P(active) with the activity readout (round 2, IFN- γ relative to BB ζ (4-1BB-CD3 ζ); round 3, killing relative to krgr- ζ 1-40), per design and per idea, zero-shot (open) and in context (filled), for all five models, the single-call LLMs and the coding agents; bars, 95% bootstrap intervals.

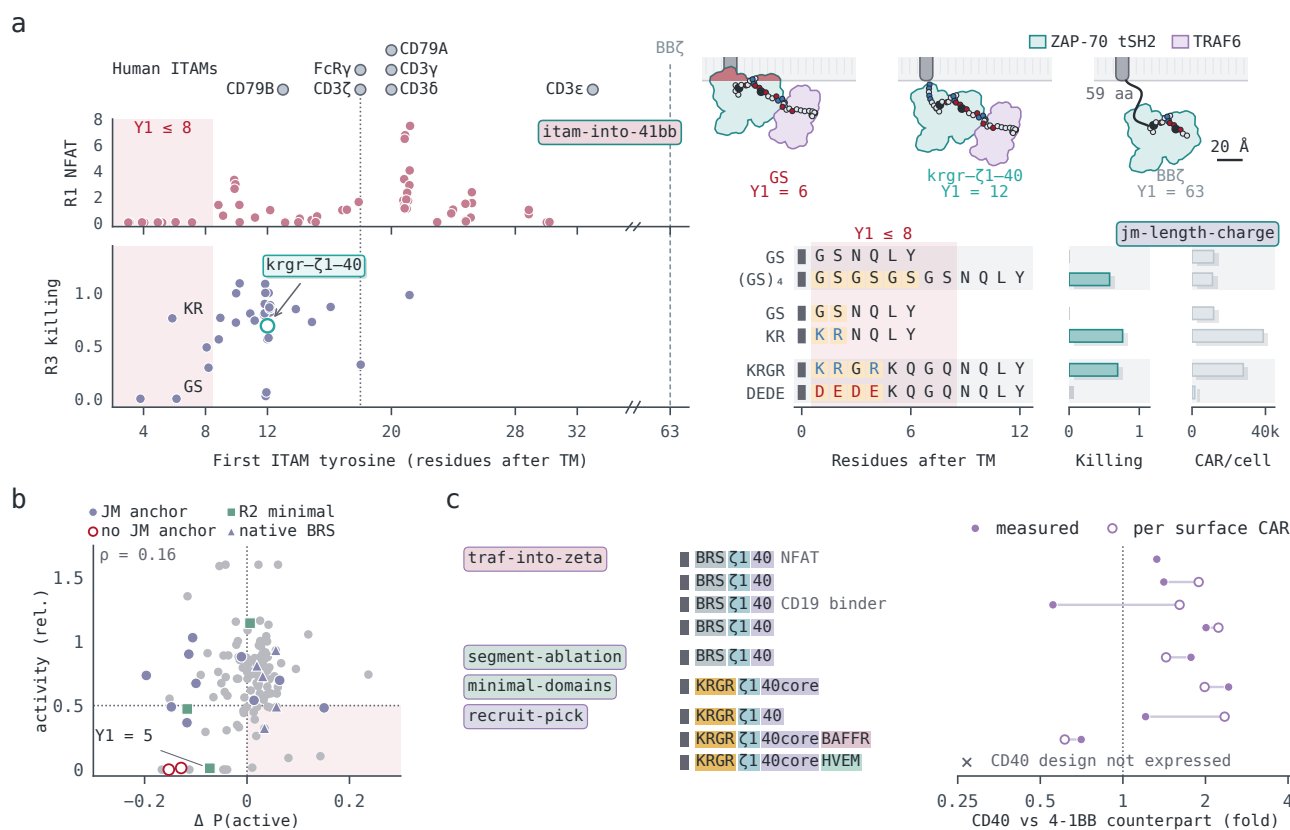


Figure S9: Membrane proximity, model predictions and the CD40 core. **a**, Membrane proximity: Y1 of human ITAM-bearing chains, round 1 NFAT and round 3 killing; shading, Y1 of 8 or less. Right, JM contrasts with killing and surface CAR/cell. Top right, Boltz-2 models of ZAP-70 and TRAF6 docked at Y1 = 6, 12 and 63. **b**, Whether the models' update tracked the outcome. Change in the ensemble's $P(\text{active})$ from zero-shot to in context (the five models as one ensemble, as in Fig. 5c) against the activity readout relative to the reference (round 2, IFN- γ ; round 3, killing relative to *krgr-ζ1-40*), for the 125 designs scored in both tasks (gray), with the cases of Fig. 5b on top: shortened round 3 designs with (filled circles) or without (open circles) a JM segment before the first module, round 2 minimal domains (squares) and round 3 native-BRS JM designs (triangles). Dotted lines, zero change and the 0.5 activity threshold; shading, raised in context but inactive. Values above 1.6 are drawn at 1.6. **c**, The CD40-core thread: each CD40 design relative to its 4-1BB counterpart, measured (filled) and per surface CAR (open). One measurement per round 3 construct.

Prompts.

Prompt	Use
T1	task, P(active); bracketed text in context only; {answer}, T2
T2	answer section: coding agents (with the web sentence of T3), or single-call LLMs (files follow inline)
T3	the web sentence of T2
T4	coding-agent command line
C1	comparator task, P(better), round 3 against krgr- ζ 1-40; for round 2 the reference is BB ζ and train.csv lists only the round 1 measurements; answer section as T2 with p_higher and label higher or lower

T1

```
# Task
You will predict the outcome of a cell assay for chimeric antigen receptor (CAR) designs. All CARs in a round share the same signal peptide, FLAG tag, antigen-binding domain, CD8alpha hinge and CD8alpha transmembrane domain; only the intracellular domain (ICD) differs. Each sequence is the ICD in one-letter amino-acid code, from the first residue after the CD8alpha transmembrane segment (...IYIWAPLAGTCGVLLLSLVITLYC) to the C-terminus.
For each ICD in test.csv, decide whether the CAR is active in the assay used for it (below). Reason from the ICD itself: which signaling motifs or domains it contains and what they imply for CAR signaling[, using the labelled ICDs in train.csv as evidence].

# Assays
[train.csv names the assay of each labelled ICD in its assay column.]
The ICDs in test.csv were measured in assay {A, B or C}.

A. Jurkat NFAT reporter (round 1).
- CAR: anti-BCMA D-domain binder (ddBCMA).
- Cells: Jurkat NFAT-luciferase reporter cells expressing the CAR.
- Co-culture: 4 h with BCMA-positive or BCMA-negative HEK293T targets, 1:1 CAR-positive effectors to targets.
- Readout: luminescence, two independent runs.
- Score: (luminescence with BCMA-positive targets - with BCMA-negative targets), divided by the same quantity for a second-generation reference CAR (4-1BB-CD3zeta ICD) on the same plate, averaged over the two runs.
- Active: score of 0.25 or more.

B. Primary T-cell IFN-gamma (round 2).
- CAR: bb2121 anti-BCMA scFv.
- Cells: human T cells activated with TransAct, expanded 3-4 days, then electroporated with CAR mRNA and rested overnight.
- Co-culture: 24 h with H929 myeloma cells (BCMA-positive), 5:1 CAR-positive effectors to targets.
- Readout: IFN-gamma by cytometric bead array, one co-culture.
- Score: IFN-gamma divided by that of a second-generation reference CAR (4-1BB-CD3zeta ICD) on the same plate.
- Active: score of 0.5 or more.

C. Primary T-cell killing (round 3).
- CAR: bb2121 anti-BCMA scFv.
- Cells: human T cells activated with TransAct, electroporated with CAR mRNA (4 ug per million cells) on day 4, rested about 24 h.
- Co-culture: 24 h with luciferase-expressing H929 cells, 10:1 CAR-positive effectors to targets, two wells.
- Readout: remaining target luminescence L.
- Score: (L_no mRNA - L) / (L_no mRNA - L_reference), averaged over the two wells. The reference is a CAR whose ICD carries a single CD3zeta ITAM.
- Active: score of 0.5 or more.

# Data
- test.csv (id,sequence): {n_test} ICDs to predict.
[- train.csv (assay,label,sequence): {n_train} ICDs measured earlier, labelled active or inactive in their own assay.]

{answer}
```

T2

Coding agents

```
# Tools and answer
You may inspect the files and write and run Python (numpy, pandas, scipy, scikit-learn are installed).
{web sentence, T3}
You have one attempt; do not ask questions.
Write predictions.csv with header id,p_active,label,reasoning and one row for every id in test.csv (quote fields as CSV requires):
- reasoning: two to four sentences for this ICD;
- p_active: your probability (0 to 1) that the CAR is active; use the whole range, not only 0 or 1;
- label: active or inactive.
Then reply with one short paragraph describing your method.
```

Single-call LLMs

Answer

The files are given below. You have one attempt; do not ask questions.

Answer in the required JSON format, with one prediction for every id in test.csv:

- reasoning: two to four sentences for this ICD;
- p_active: your probability (0 to 1) that the CAR is active; use the whole range, not only 0 or 1;
- label: active or inactive.

T3

Do not use the internet: no web search, no web pages, no package downloads.

T4

Read TASK.md in the current directory and complete the task it describes.

C1

Task

You will predict the outcome of a cell assay for chimeric antigen receptor (CAR) designs. All CARs in a round share the same signal peptide, FLAG tag, antigen-binding domain, CD8alpha hinge and CD8alpha transmembrane domain; only the intracellular domain (ICD) differs. Each sequence is the ICD in one-letter amino-acid code, from the first residue after the CD8alpha transmembrane segment (...IYIWAPLAGTCGVLLLSLVITLYC) to the C-terminus.

For each ICD in test.csv, estimate the probability that its CAR gives a higher readout than the reference CAR measured on the same plate in the assay used for it (below). Reason from the ICD itself and from how it differs from the reference: which signaling motifs or domains it contains and what they imply for CAR signaling, using the measured ICDs in train.csv as evidence.

Reference

KRGR- ζ 1-40: a compact intracellular domain designed in an earlier round, from which the test ICDs were derived. ICD sequence:

<sequence>

Assays

train.csv gives each measured ICD's design round and assay in its round and assay columns.

The ICDs in test.csv were measured in assay C, against the reference above.

A. Jurkat NFAT reporter (round 1).

- CAR: an anti-BCMA binder.
- Cells: Jurkat NFAT-luciferase reporter cells expressing the CAR.
- Co-culture: 4 h with BCMA-positive HEK293T targets, 1:1 CAR-positive effectors to targets.
- Readout: luminescence, two independent runs.
- Ratio: luminescence of the CAR divided by that of the 4-1BB-CD3zeta reference CAR on the same plate, averaged over the two runs.

B. Primary T-cell IFN-gamma (round 1 re-screen of 78 library designs, and round 2).

- CAR: an anti-BCMA binder (bb2121, from bluebird bio).
- Cells: human T cells activated with TransAct, expanded 3-4 days, then electroporated with CAR mRNA and rested overnight.
- Co-culture: 24 h with H929 myeloma cells (BCMA-positive), 5:1 CAR-positive effectors to targets.
- Readout: IFN-gamma by cytometric bead array, one co-culture.
- Ratio: IFN-gamma of the CAR divided by that of the 4-1BB-CD3zeta reference CAR on the same plate.

C. Primary T-cell IFN-gamma (round 3).

- CAR: an anti-BCMA binder (bb2121, from bluebird bio).
- Cells: human T cells activated with TransAct, electroporated with CAR mRNA (4 ug per million cells) on day 4, rested about 24 h.
- Co-culture: 24 h with luciferase-expressing H929 cells, 10:1 CAR-positive effectors to targets.
- Readout: IFN-gamma in the co-culture supernatant.
- Ratio: IFN-gamma of the CAR divided by that of the reference CAR on the same plate.

Data

- test.csv (id,sequence): {n_test} ICDs to predict.
 - train.csv (round,assay,sequence,ratio_to_4-1BB-CD3zeta,label): 282 measurements made earlier:
 - round 1, assay A: the round 1 library in Jurkat NFAT reporter cells;
 - round 1, assay B: 78 of those library designs re-screened in primary T cells (IFN-gamma);
 - round 2, assay B: the round 2 designs and three re-tested round 1 designs in primary T cells (IFN-gamma).
- All are anti-BCMA CARs, but the two assays differ in cell type and readout, so a design can rank differently in them. An ICD measured in both assays has a row for each. ratio is the readout divided by that of 4-1BB-CD3zeta on the same plate; label is higher (ratio 1 or more) or lower.

Tools and answer

You may inspect the files and write and run Python (numpy, pandas, scipy, scikit-learn are installed).

Do not use the internet: no web search, no web pages, no package downloads.

You have one attempt; do not ask questions.

Write predictions.csv with header id,p_higher,label,reasoning and one row for every id in test.csv (quote fields as CSV requires):

- reasoning: two to four sentences for this ICD;
- p_higher: your probability (0 to 1) that the CAR's readout is higher than the reference's; use the whole range, not only 0 or 1;
- label: higher or lower.

Then reply with one short paragraph describing your method.

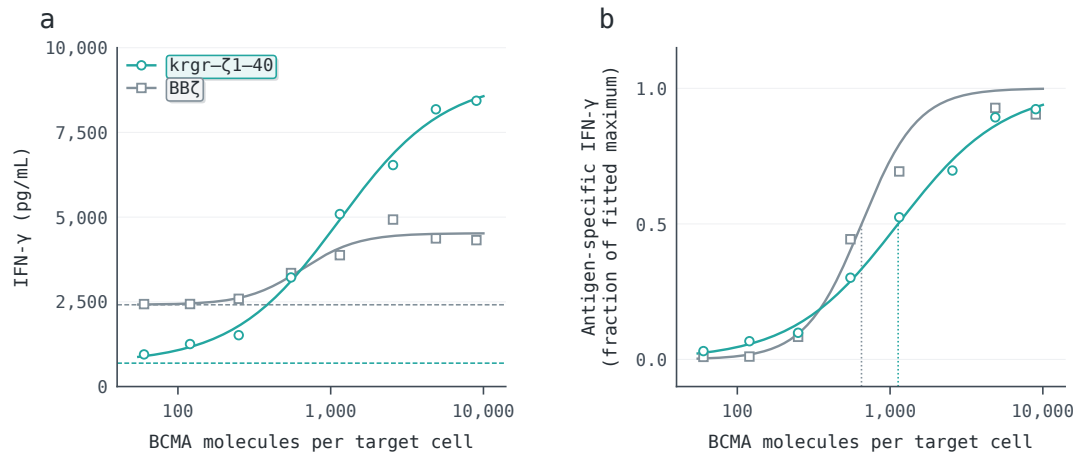


Figure S10: Antigen sensitivity of IFN-γ release at a matched CAR mRNA dose. **a**, IFN-γ from BBζ (4-1BB-CD3ζ) and *krgr-ζ1-40* CAR T cells after co-culture with target cells carrying eight densities of BCMA, at the highest mRNA dose of Fig. 4a; dashed lines, each construct's fitted antigen-independent floor. **b**, The same readings with the fitted floor subtracted, as a fraction of each fitted span; dotted lines, EC50s of 650 (BBζ) and 1,130 (*krgr-ζ1-40*) BCMA molecules per target cell.

Structural and sequence analyses of CAR ICDs

To examine the structural context of ICD activity, we modeled adaptor binding and membrane interactions. Predicted complexes reproduced native ZAP-70-ITAM and TRAF6-CD40 binding poses for active and inactive ICDs alike, so binding geometry did not explain activity (Fig. S11a-e).

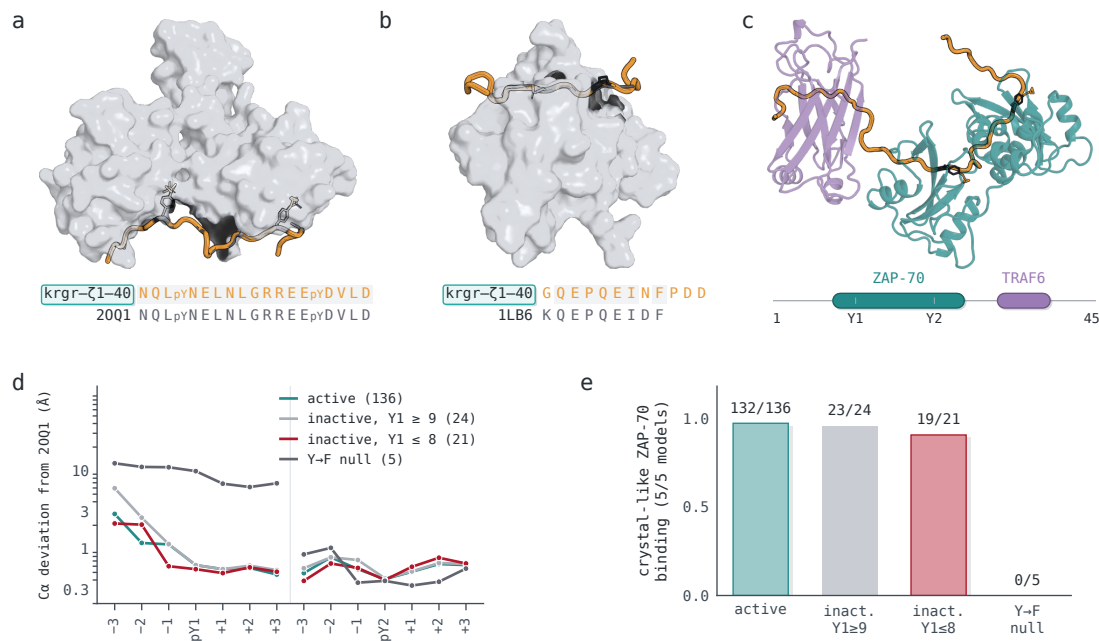


Figure S11: Structure models of *krgr-ζ1-40* and related ICDs. **a**, Boltz-2 co-fold of *krgr-ζ1-40* with the ZAP-70 tandem SH2 domains, superposed on PDB 20Q1. **b**, Co-fold of the CD40-derived segment with the TRAF6 TRAF-C domain, superposed on PDB 1LB6. **c**, Ternary co-fold of the ICD with ZAP-70 (teal) and TRAF6 (mauve), which bind separate segments. **d**, Deviation of each ITAM half-site from 20Q1 for active, inactive and Y→F null ICDs; only the nulls are displaced. **e**, Fraction of library ICDs whose first ITAM is crystal-like in all five models, by group.

We next asked whether natural sequences and physical models could explain the spacing constraint. Membrane-proximal tyrosines in human single-pass proteins were less frequently annotated or detected as phosphorylated (Fig. S12a-c). Lck substrate scores (Fig. S12f) and a steric model of ZAP-70 access (Fig. S12k) did not account for the sharp activity boundary observed in round 1.

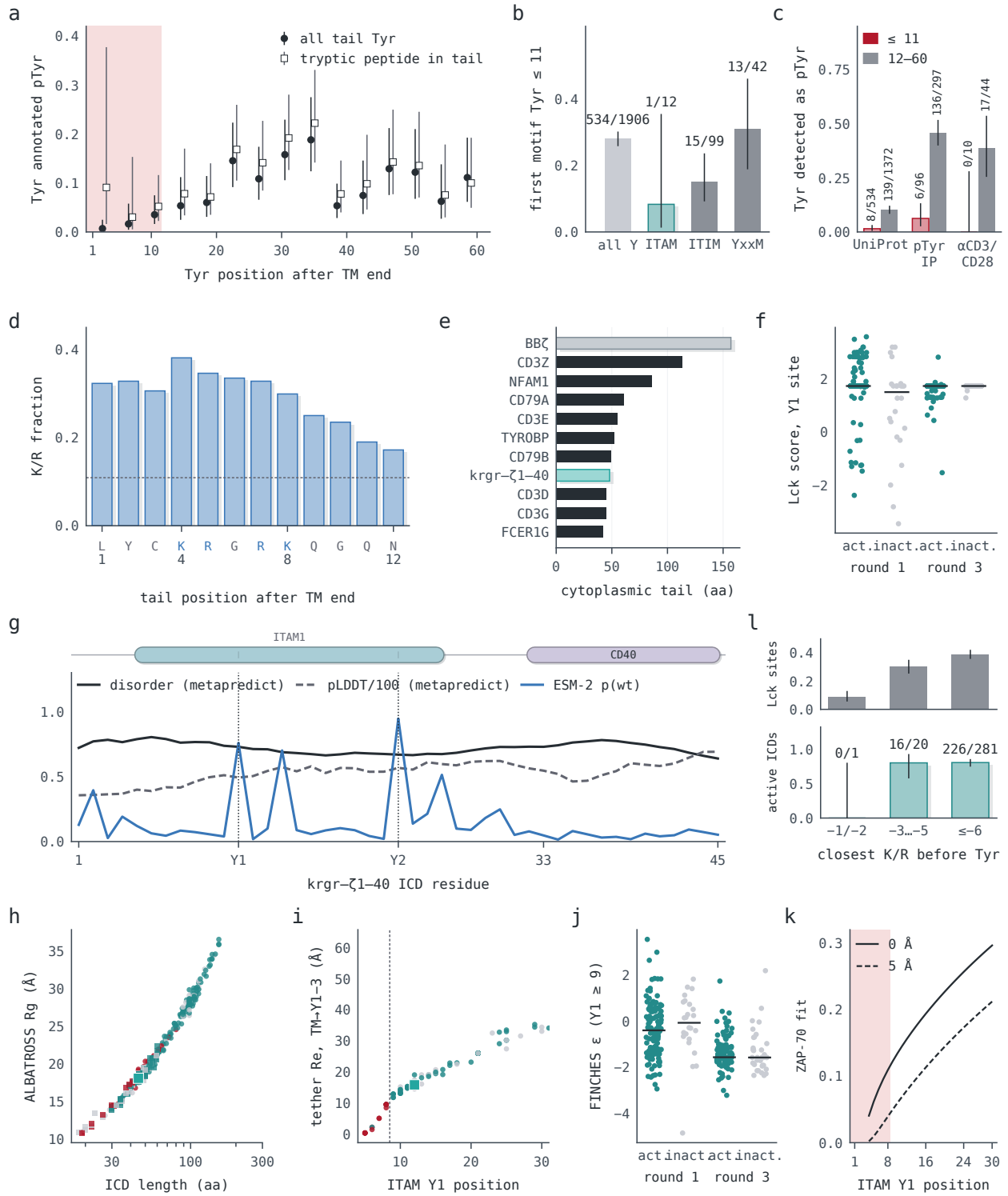


Figure S12: Sequence, proteome and physical-model context for the juxtamembrane spacing. **a**, Fraction of cytoplasmic tyrosines annotated as phosphotyrosine in human single-pass proteins by distance from the transmembrane end, for all tail tyrosines (filled circles) and those whose tryptic peptide lies wholly in the tail (open squares). **b**, Fraction of motif tyrosines within 11 residues of the membrane for all tyrosines, ITAMs, ITIMs and YxxM motifs. **c**, Phosphotyrosine detection within 11 residues (crimson) and at 12–60 (gray) in UniProt and in two Jurkat phosphotyrosine datasets [132,133]. **d**, K/R fraction at each tail position. **e**, Cytoplasmic tail lengths of natural ITAM chains, *krgr-ζ1-40* and BBζ (4-1BB-CD3ζ). **f**, Lck substrate score of the Y1 site [134] for active and inactive ICDs with Y1 of 9 or more. **g**, Per-residue disorder, predicted pLDDT and ESM-2 wild-type probability along *krgr-ζ1-40*. **h**, Radius of gyration against ICD length (teal, active; crimson, Y1 of 8 or less; gray, other inactive; square, *krgr-ζ1-40*). **i**, End-to-end distance from residue 1 to Y1 – 3 against Y1. **j**, Homotypic interaction strength (FINCHES ϵ [135]) for ICDs with Y1 of 9 or more. **k**, Probability that the 2OQ1 complex fits below the membrane in a steric model, against Y1; the model rises gradually rather than stepping at Y1 = 9 as in Fig. S9a. **l**, Efficient Lck phosphorylation of library substrate sites (top) and the fraction of active ICDs (bottom) by the position of the nearest K/R before the tyrosine [136]. Positions in a–e count from the UniProt transmembrane end, three residues earlier than in Fig. S9a.

Agentic digital infrastructure

We built an electronic lab notebook system to store all our sequences, protocols, and experimental data. Claude Code serves as the primary interface, with dedicated graphical interfaces for tasks that demand consistent visual interaction, such as plate map visualization and flow cytometry gating.

Routine procedures were written as shared agent skills instead of fixed application code. For example, coding agents performed codon optimization from a shared skill that defined the procedure and adapted it to individual sequences and constraints, without each edge case being encoded in advance [137,138].

We use minimal data models to store only core information and leverage agents to infer many relationships and details at runtime (Fig. S13). For example, after each experiment, users upload raw data to a shared filesystem, where agents use plate maps to link measurements to designed variants and experimental conditions for analysis and plotting. Typically, we discuss results and make decisions with comprehensive analyses and plots in hand within minutes of instrument readout.

The PostgreSQL database and file storage ran on Supabase.

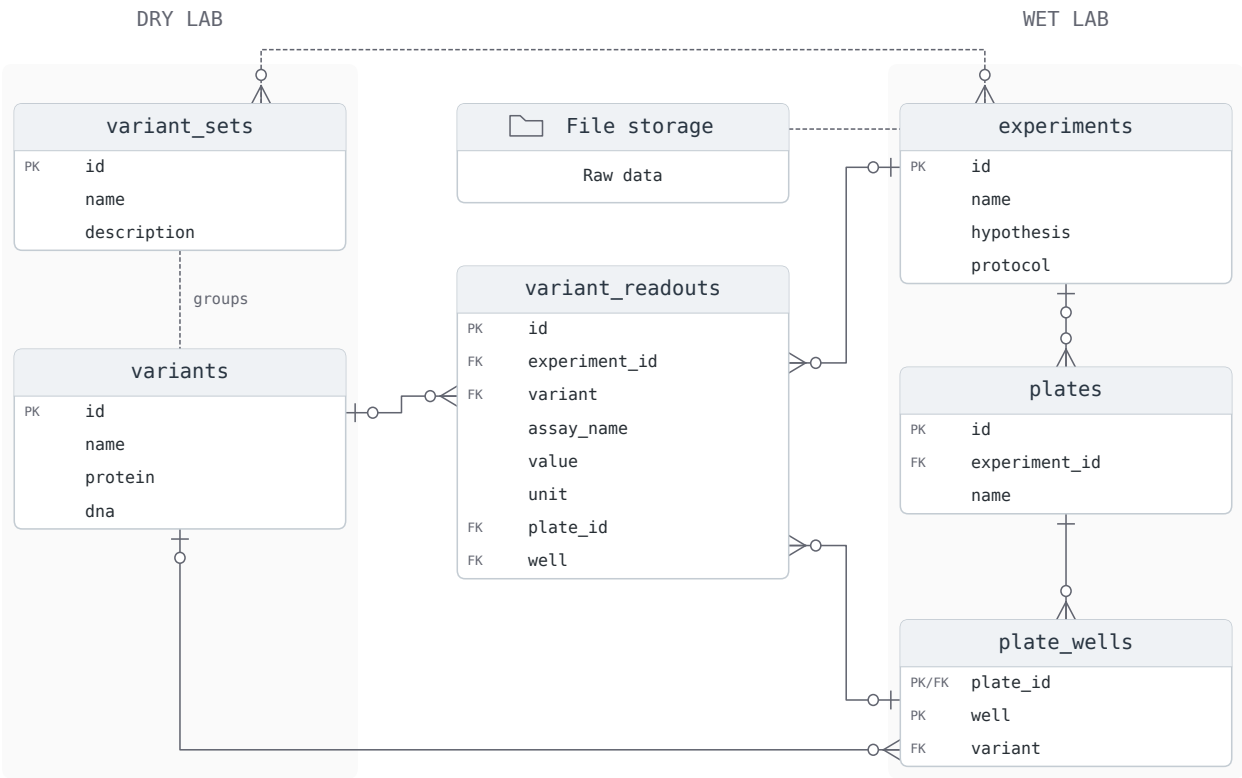


Figure S13: Electronic lab notebook data model. Selected tables and fields link experiments, variants, plates, and readouts. Computational scientists typically populate **variants** after sequence design; wet-lab scientists create **experiments** and **plates** during experiment planning. Agents read and write all tables. **variant_readouts**, populated by agents, links variants to experimental measurements for analysis. PK, primary key; FK, foreign key.

Supplementary Table 1. The 38 ideas of Fig. 5a. The ideas, their code names and descriptions, and this table were generated by AI agents. The full table (CSV; Markdown) gives, for each idea, its code name, round, plain-language description, short recorded intention, designs, designs with a readout and designs at or above the comparator, its parent ideas with the evidence type of each link, and the Fig. 5a thread it is on (ITAM distance from membrane, teal; CD40 core, lavender; both; or none). The short version below gives each code name with a brief description and its round. Designs are listed with, in parentheses, those with a readout and those at or above the comparator; comparators are as in Fig. 5a. Parent ideas are linked by a parent design, a KRGR-anchor donor, a citation in the design rationale or an inferred shared edit, and every round 3 idea also descends from **krgr- ζ 1-40** in **krgr-anchor-transfer**. Threads are named as in the Fig. 5a key; ideas off the drawn threads are assigned from their description and shown in italics (ITAM composition, motif exploration, sequence scan and reference ICDs are themes, not Fig. 5a threads).

Code name	Description	Round	Designs (read / $\geq$)	Parent idea(s)	Thread
itam-into-4lbb	A CD3 ζ ITAM placed inside or next to the 4-1BB domain	1	36 (36 / 0)	–	membrane proximity
traf-into-zeta	A 4-1BB TRAF-binding site added to a shortened CD3 ζ	1	29 (29 / 7)	–	CD40
short-itam-linker	Shorter linkers and scaffold between the CD3 ζ ITAMs	1	23 (23 / 0)	–	<i>ITAM composition</i>
plm-guided-search	Variants proposed with a protein language model	1	14 (14 / 2)	–	<i>sequence scan</i>
itam-motif-mutate	ITAM motif choice and point substitutions	1	14 (14 / 1)	–	<i>ITAM composition</i>
three-itam-trim	All three ITAMs kept, other residues trimmed	1	12 (12 / 2)	–	<i>ITAM composition</i>
trim-4lbb-lead	A shortened 4-1BB ahead of one or two CD3 ζ ITAMs	1	10 (10 / 1)	–	<i>motif exploration</i>
itam-flank-edit	Edits to the residues that flank the ITAMs	1	9 (9 / 3)	–	<i>ITAM composition</i>
phase-sep-insert	Inserts intended to promote phase separation	1	9 (9 / 7)	–	<i>motif exploration</i>
keep-two-itam	Two CD3 ζ ITAMs kept	1	8 (8 / 6)	–	<i>ITAM composition</i>
single-res-scan	Single substitutions in a 4-1BB–ITAM1 scaffold	1	8 (8 / 6)	–	<i>sequence scan</i>
ref-icd-set	Published reference domains	1	7 (7 / 0)	–	<i>reference ICDs</i>
mini-itam-mix	Combinations of reduced-ITAM elements	1	4 (4 / 3)	–	<i>ITAM composition</i>
synth-itam-traf	A synthetic ITAM after a TRAF-binding peptide	1	3 (3 / 0)	–	<i>motif exploration</i>
segment-ablation	Delete one parent segment at a time	2	3 (2 / 1)	itam-into-4lbb; traf-into-zeta	CD40
costim-motif-swaps	Replace the costimulatory motif with another	2	3 (3 / 2)	traf-into-zeta	motif exploration
minimal-domains	Shorten the ICD to its smallest working form	2	3 (3 / 0)	itam-into-4lbb; traf-into-zeta	membrane proximity; CD40
itam-swaps	Exchange ITAM1 and ITAM3	2	2 (2 / 1)	traf-into-zeta	<i>ITAM composition</i>
two-traf-motifs	Stack two TRAF-binding motifs	2	2 (1 / 0)	itam-into-4lbb; traf-into-zeta	motif exploration
krgr-anchor-transfer	Move the KRGR anchor onto other ICDs	2	2 (2 / 2)	itam-into-4lbb; traf-into-zeta	membrane proximity
plm-ranked-mutant	Point mutants (1–4 substitutions) ranked by a protein language-model regressor	3	30 (27 / 4)	krgr-anchor-transfer	<i>sequence scan</i>
recruit-pick	Other costimulatory motifs, or stacks of them, after ITAM1	3	21 (17 / 1)	costim-motif-swaps; krgr-anchor-transfer; traf-into-zeta; two-traf-motifs	CD40
jm-length-charge	Length and charge of the segment between membrane and ITAM1	3	14 (12 / 1)	itam-into-4lbb; krgr-anchor-transfer	membrane proximity
linker-length	Length of the linkers before and after ITAM1	3	12 (11 / 4)	krgr-anchor-transfer; segment-ablation	<i>membrane proximity</i>
jm-lysine-charge	Single-lysine and C-flank charge substitutions	3	11 (11 / 6)	krgr-anchor-transfer	membrane proximity
min-core-ladder	Shorter ICDs that keep ITAM1 and the CD40 core	3	8 (8 / 0)	itam-swaps; krgr-anchor-transfer; minimal-domains; segment-ablation	membrane proximity
best-edit-mix	Combinations of single edits from this round	3	6 (5 / 0)	krgr-anchor-transfer	<i>sequence scan</i>
native-brs-jm	The native CD3 ζ basic stretch in place of KRGR	3	6 (6 / 0)	costim-motif-swaps; krgr-anchor-transfer	membrane proximity
module-order-flip	CD40 core placed before ITAM1	3	5 (5 / 0)	krgr-anchor-transfer	<i>CD40</i>

Code name	Description	Round	Designs (read / $\geq$)	Parent idea(s)	Thread
non-traf6-recruit	Recruiters outside the TRAF6 class	3	4 (4 / 0)	costim-motif-swaps; krgr-anchor-transfer	<i>motif exploration</i>
traf6-super	One high-affinity TRAF6 motif in place of the CD40 segment	3	4 (1 / 0)	krgr-anchor-transfer; minimal-domains; traf-into-zeta	<i>CD40</i>
foreign-itam	An FcR γ or DAP12 ITAM in place of CD3 ζ ITAM1	3	4 (2 / 0)	itam-motif-mutate; krgr-anchor-transfer	<i>ITAM composition</i>
nql-flank-trim	The three residues before the first ITAM tyrosine trimmed	3	3 (3 / 1)	krgr-anchor-transfer; minimal-domains	membrane proximity
mid-brs-patch	The CD3 ζ basic patch restored between ITAM1 and CD40	3	3 (3 / 0)	krgr-anchor-transfer	<i>CD40</i>
pi3k-dual-tyr	Second ITAM1 tyrosine recoded to a PI3K motif	3	3 (3 / 0)	krgr-anchor-transfer; ref-icd-set	<i>motif exploration</i>
hemi-itam	A single-tyrosine signal in place of the ITAM	3	3 (0 / 0)	itam-motif-mutate; krgr-anchor-transfer	<i>ITAM composition</i>
redo-19mer	The failed 19-residue design re-tested with controls	3	3 (3 / 0)	krgr-anchor-transfer; minimal-domains	membrane proximity
low-tonic-addon	Costimulatory add-ons chosen to lower target-free signaling	3	3 (1 / 0)	krgr-anchor-transfer; round 1 library	<i>motif exploration</i>

Round 1 design prompts

We recovered 17 distinct prompts from 34 archived round 1 agent runs.

Model assignments. Selected model assignments are listed below.

Prompt	Models
P02	GPT-5.2, Claude Sonnet 4.6
P03	Claude Opus 4.6
P04	MiniMax M2.5, Claude Sonnet 4.6, Claude Opus 4.6, Kimi K2.5
P05	GPT-5.2, Kimi K2.5, Claude Opus 4.6
P06	Claude Opus 4.6, Claude Sonnet 4.6
P07	GPT-5.2
P08	Claude Sonnet 4.6
P09	Claude Sonnet 4.6
P10	GPT-5.2
P11	GPT-5.2, Claude Opus 4.6
P12	Claude Opus 4.6
P13	Claude Opus 4.6
P14	Claude Opus 4.6
P15	Claude Sonnet 4.6
P16	Claude Opus 4.6, GPT-5.2
P17	Claude Opus 4.6

Prompts.

The task-specific text below was embedded in the system message. Each run also received the following instruction, followed by an automatically generated execution plan:

Please complete the task. Start by downloading any required data files.

P01

The goal is to minimize the intracellular domains of CAR constructs. Design minimized CD3ζ variants with grafted TRAF-binding motifs (15-20 amino acids) at biochemically optimal positions.

P02

Our goal is to minimize the intracellular domains of CAR constructs. Find an ideal location within 41BB for ITAM1 of CD3zeta. lace ITAM1 close to the membrane where it naturally is and embedded among non-functional IDR sequences within 41BB could be optimal way to save sequence budget.

P03

Our goal is to minimize the intracellular domains of CAR constructs. Design 5-10 minimized CD3ζ variants with grafted TRAF-binding motifs at optimal positions. Each design should be grounded in thoroughly-researched biochemical reasoning.

P04

Our goal is to minimize the intracellular domains of CAR constructs. Design 5-10 minimized sequences where we replace CD3ζ entirely by moving its ITAM1 into 41BB. Placing ITAM1 close to the membrane where it naturally is and embedding it among non-functional IDR sequences within 41BB could be a good way to save sequence budget. Each design must be grounded in thoroughly-researched biochemical reasoning.

P05

Design 5-10 variants of the given CAR 4-1BB costim + minimized CD3ζ sequences for optimal signaling. Each design must be grounded in thoroughly-researched biochemical reasoning.

P06

Our goal is to minimize the intracellular domains of CAR constructs. Design 5-10 minimized variants by modifying the given 4-1BB and CD3 ζ domains. Each design must be grounded in thoroughly-researched biochemical reasoning.

P07

Our goal is to minimize the intracellular domains of CAR constructs. Design 5-10 minimized variants by modifying the given CD3 ζ domain. Each design must be grounded in thoroughly-researched biochemical reasoning.

P08

Our goal is to minimize the intracellular domains of CAR constructs. Design 5-10 minimized variants by modifying the given CD3 ζ domain. Focus on shortening the non ITAM regions while maintaining optimal signaling. Each design must be grounded in thoroughly-researched biochemical reasoning.

P09

Our goal is to minimize the intracellular domains of CAR constructs. Design 5-10 minimized variants by removing one ITAM from the given CD3 ζ domain along with other appropriate adjustments. Each design must be grounded in thoroughly-researched biochemical reasoning.

P10

Our goal is to minimize the intracellular domains of CAR constructs. Design minimized CD3 ζ variants with grafted TRAF-binding motifs at optimal positions. Each design must be grounded in thoroughly-researched biochemical reasoning.

P11

Our objective is to minimize the intracellular signaling domains of CAR constructs while preserving or enhancing functional signaling. Propose engineered variants of CD3 ζ with reduced sequence length, incorporating innovative modifications to both the 4-1BB co-stimulatory region and CD3 ζ signaling motifs. Designs should be mechanistically justified, biochemically plausible, and prioritize synergistic integration of signaling elements rather than simple truncation. Highlight novel strategies for domain fusion, motif optimization, or modular redesign that maintain signaling potency despite minimized size.

P12

Imagine next-generation ultra-compact CAR intracellular domains. How might CD3 ζ and 4-1BB signaling elements be redesigned, fused, or re-encoded into minimal yet fully functional architectures? Propose inventive, biochemically realistic constructs that compress signaling information into shorter sequences while retaining robust activation. Prioritize unconventional but mechanistically sound ideas that create synergy between co-stimulatory and activation motifs.

P13

Our goal is to minimize the intracellular domains of CAR constructs. Design 5-10 minimized variants keeping only 1 ITAM from the given CD3 ζ domain along with other appropriate adjustments that ensure optimal signaling. Each design must be grounded in thoroughly-researched biochemical reasoning, such as the kinetic proofreading theory of intracellular signaling.

P14

Design 5-10 variants of the given CAR 4-1BB costim + minimized ITAM-3 only CD3 ζ sequences for optimal signaling. Each design must be grounded in thoroughly-researched biochemical reasoning, such as the kinetic proofreading model of intracellular signaling.

P15

Our goal is to minimize the intracellular domains of CAR constructs. Design 5-10 minimized variants keeping only ITAM2 from the given CD3 ζ domain along with other appropriate adjustments that ensure optimal signaling. Each design must be grounded in thoroughly-researched biochemical reasoning, such as the kinetic proofreading theory of intracellular signaling.

P16

Starting from the provided ITAM1-only CAR sequence, design a variant with enhanced phase separation potential which can drive a more robust biocondensate formation. Do this without triggering antigen-independent aggregation which can lead to high tonic CAR activation.

P17

Starting from the provided ITAM1-only CAR sequence, design a set of 5-10 variants with enhanced phase separation potential which can drive a more robust biocondensate formation. Do this without triggering antigen-independent aggregation which can lead to high tonic CAR activation.