nature methods

ARTICLES

https://doi.org/10.1038/s41592-022-01436-z

Check for updates

# Antigen identification and high-throughput interaction mapping by reprogramming viral entry

Connor S. Dobson[1,2], Anna N. Reich[1,2], Stephanie Gaglione[2,3], Blake E. Smith[2,4], Ellen J. Kim[1,2], Jiayi Dong[1,2], Larance Ronsard[5], Vintus Okonkwo[5], Daniel Lingwood[5], Michael Dougan[4,6], Stephanie K. Dougan[4,7] and Michael E. Birnbaum[1,2,5,8] ✉

**Deciphering immune recognition is critical for understanding a broad range of diseases and for the development of effective vaccines and immunotherapies. Efforts to do so are limited by a lack of technologies capable of simultaneously capturing the complexity of adaptive immunoreceptor repertoires and the landscape of potential antigens. To address this, we present receptor–antigen pairing by targeted retroviruses, which combines viral pseudotyping and molecular engineering approaches to enable one-pot library-on-library interaction screens by displaying antigens on the surface of lentiviruses and encoding their identity in the viral genome. Antigen-specific viral infection of cell lines expressing human T or B cell receptors allows readout of both antigen and receptor identities via single-cell sequencing. The resulting system is modular, scalable and compatible with any cell type. These techniques provide a suite of tools for targeted viral entry, molecular engineering and interaction screens with broad potential applications.**

The adaptive immune system recognizes and responds to the diverse pathogens that humans encounter throughout their lives. B cells recognize extracellular antigens via the B cell receptor (BCR), leading to the production of secreted antibodies to inhibit and eliminate pathogens. T cells use T cell receptors (TCRs) to recognize short peptide fragments presented by major histocompatibility complex (MHC) proteins (pMHCs), enabling antigen-specific coordination of the immune response and elimination of malignant or infected cells. This remarkable ability to sense, respond to and remember threats is key to any successful adaptive immune response and accordingly is the cornerstone of successful vaccines and immunotherapies.

The importance of T cell-mediated immunity has led to the development of a series of approaches dedicated to identifying antigens or pMHC–TCR pairs[1]. Conventional T cell assays, such as ELISPOT or intracellular cytokine staining[2], provide direct readouts of T cell function. While these approaches can be highly multiplexed, they do not readily lend themselves to TCR sequencing and are limited in their ability to identify single reactive antigens. More recently, T-Scan and similar approaches have broadened the antigenic scope of functional assays to genome scale[3,4], but cannot provide paired receptor information without panning of predetermined TCRs or iterated steps of target antigen identification followed by sorting of reactive cells for TCR sequencing. Other recent cell-based reporter assays convert pMHCs into the recognition domain of immune signaling complexes[5,6] or leverage trogocytosis[7] or interleukin (IL)-2 capture[8] to identify the pMHC targets of a given TCR. These approaches can identify successful TCR–pMHC interactions for single TCRs from antigen libraries on the scale of $10^3$ to $10^4$, but require substantial library redundancy, are limited in their ability to multiplex TCRs and may require multiple rounds of screening. Requiring a discrete experiment per TCR imposes a significant scalability constraint; each individual has a TCR repertoire of as many as ~$10^{12}$ unique clones[1,9,10] and very little overlap is expected between individual repertoires even for people who share common MHC alleles.

Recombinant protein-based screening can broaden either the number of T cell clones or the number of antigens that are feasible to screen. Baculoviral display libraries have enabled screening of ~$10^5$ pMHCs by panning of infected Sf9 cells with fluorescently labeled recombinant TCRs[11,12]. Yeast display of pMHCs enables screening of ~$10^8$ unique pMHC antigens[13,14], but can only examine a limited number of recombinantly expressed TCRs at a time and can require significant optimization for each MHC allele. Conversely, recent advances in barcoded pMHC multimers enable screens on the order of $10^3$ antigens in bulk[15,16] or hundreds while maintaining receptor–antigen pairing[17,18]. While such analyses can be performed on polyclonal T cells, they are inherently bottlenecked by several technical limitations: (1) the need to manually assemble individual barcoded multimers; (2) the ability to correctly identify interactions in large pools of multimers; and (3) the relatively small set of MHC molecules that have been recombinantly expressed successfully.

Identifying antigenic targets of B cells poses similar limitations. Recent approaches have shown that antigen–BCR pairs can be identified via the oligonucleotide tagging of recombinantly expressed proteins[19,20], which is inherently limited in scale. Thus, our understanding of antigen recognition by both arms of the adaptive immune system is currently constrained by significant hurdles to experimental scale.

Deciphering the full complexity of immune recognition requires the ability to screen for interaction pairs while incorporating diversity of both antigen receptors and their targets at the same time.

[1]Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA. [2]Koch Institute for Integrative Cancer Research, Cambridge, MA, USA. [3]Department of Chemical Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA. [4]Program in Immunology, Harvard Medical School, Boston, MA, USA. [5]Ragon Institute of MGH, MIT and Harvard, Cambridge, MA, USA. [6]Department of Medicine, Division of Gastroenterology, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA. [7]Department of Cancer Immunology and Virology, Dana-Farber Cancer Institute, Boston, MA, USA. [8]Singapore-MIT Alliance for Research and Technology Centre, Singapore, Singapore. ✉e-mail: mbirnb@mit.edu

If only the target antigens are known, it is difficult to understand the cellular factors regulating a successful response. On the other hand, if antigen receptors are characterized without knowledge of their specificity, it is difficult to understand which target antigens are important for preventing or eliminating disease. This represents a broader experimental challenge for interaction screens commonly known as 'library-versus-library screening'. The most well-established approach to this problem is yeast two-hybrid[21], in which intracellular protein pairs are used to drive expression of a reporter gene. Other previously established approaches include automated individual enzyme-linked immunosorbent assay (ELISA) screens of recombinantly expressed protein pairs[22–24], as well as mass-spectrometry-based identification of interacting pairs[25,26]. There have been several recent efforts at designing new systems, including those based on yeast mating or spatial colocalization of DNA-barcoded molecules[27,28]. However, these approaches can be labor-intensive, are often not suited for complex, extracellular protein complexes such as immune receptors and may be inefficient at low (micromolar) affinities. Together, these limitations have thus far rendered such approaches unsuitable for applications such as identifying TCR–pMHC pairs.

To overcome the above limitations, we have developed a technique that combines lentiviral surface display[29–31] with a versatile pseudotyping strategy and viral genome engineering to enable one-pot library-versus-library screening (Fig. 1). We demonstrate that our pseudotyping strategy, an engineered fusogen termed VSVGmut coexpressed with a targeting moiety, is general and versatile for both receptor and ligand usage. We leverage these abilities to present 'receptor–antigen pairing by targeted retroviruses' (RAPTR), which matches receptors with their cognate antigens based on specific infection of receptor-expressing cells by antigen-displaying viruses. Putative hits can be identified by bulk or single-cell sequencing, enabling screens of single or polyclonal receptors. We demonstrate the feasibility of this approach for both TCRs and BCRs, including a library-on-library screen consisting of 96 pMHC antigens and receptors enriched from a library of >450,000 TCRs (thousands of potential interactions).

## Results

**VSVGmut-pseudotyped lentiviruses enable modular tropism.** Developing a scalable pipeline for antigen–receptor screening presents several requirements: (1) a straightforward ability to track both receptor and antigen sequences; (2) mammalian expression systems to maximize the ability to express complex, multimeric proteins; and (3) the ability to generate selection reagents without the need to recombinantly express and characterize each protein for every experiment. To address these needs, we turned to lentiviruses, which have decades of precedent for facile molecular manipulations. They can be created at large scale, are already known to infect T cells and, upon successful infection, leave a permanent record of infection due to their integration into the host genome.

Lentiviruses have been pseudotyped via a number of different strategies to enable their use as biotechnology tools and gene therapy vectors[32]. Due to its robustness and efficient infection of many cell types, vesicular stomatitis virus G protein (VSVG, referred to as VSVGwt in this manuscript) is the most common pseudotype for laboratory studies and cell manufacturing for clinical applications[33]. More recently, approaches to enable cell type-specific targeting, via coexpressing receptor-blinded versions of Sindbis virus[34–36] or paramyxovirus envelope proteins[37–42] with targeting moieties have been described. While these approaches show promise, they have been reported to have strict limitations on targeting ligand- and receptor-binding topology due to their mechanism of entry.

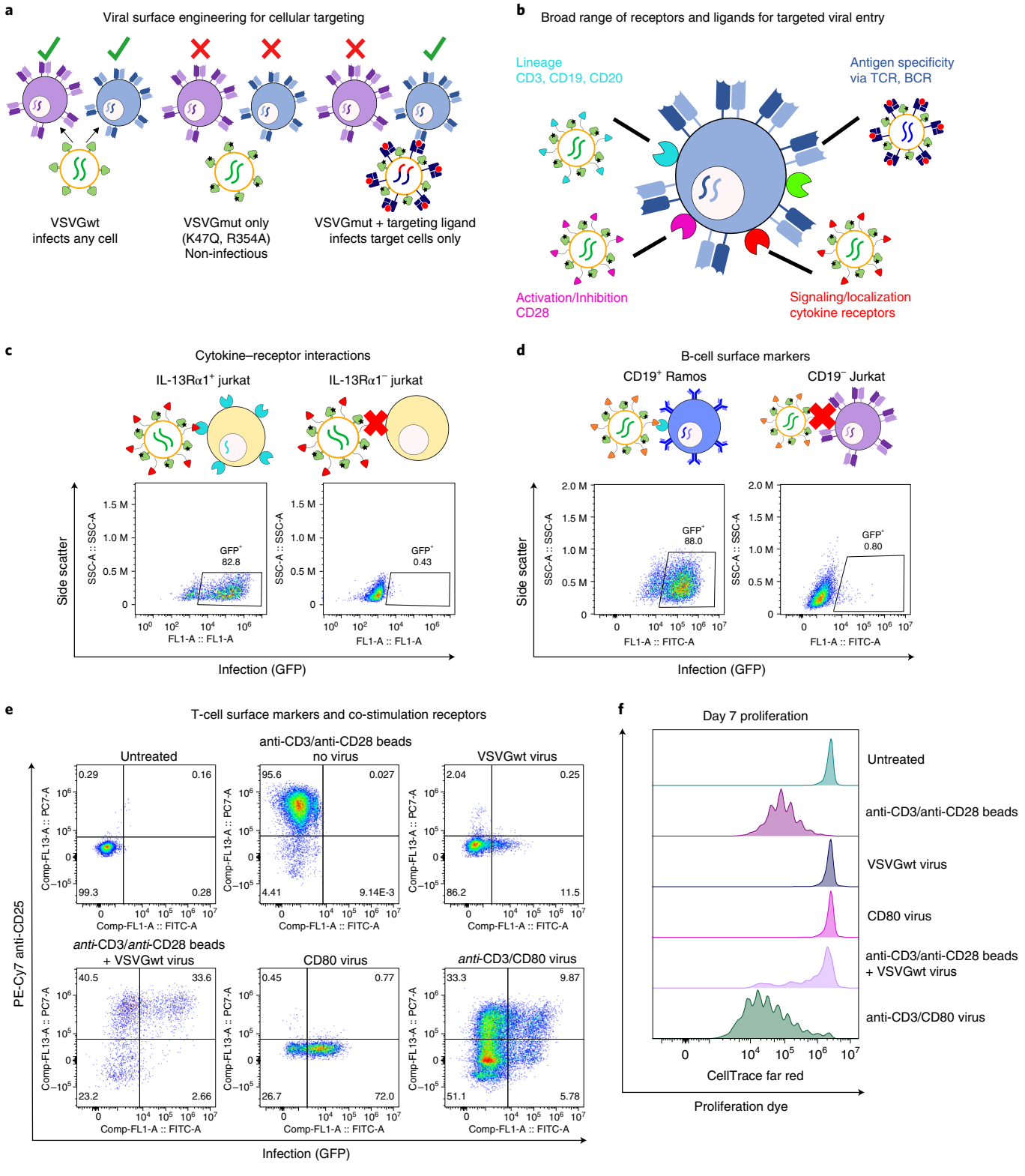
As these factors constrain the generality of an interaction screening system, we developed an alternative strategy based upon VSVG. We used the recently described crystal structure of VSVG in complex with its native receptor[43], the low-density lipoprotein receptor (LDLR), to engineer VSVGmut, which incorporates the K47Q and R354A mutations reported to ablate affinity for the LDLR family of receptors (Fig. 1a). To retarget VSVGmut-pseudotyped viruses, we coexpressed a variety of surface-bound molecules during viral production (Fig. 1b). To benchmark against established systems[38], we used IL-13 as a viral targeting ligand. After testing several surface architectures (Extended Data Fig. 1), we found that while many distinct constructs conferred specific infection of IL-13Rα1-expressing cells, the dimerized, surface-tethered IL-13 yielded the most efficient infection while retaining excellent specificity (Fig. 1c). We also found that VSVGmut-pseudotyped viruses displaying an anti-CD19 single-chain antibody fragment (scFv) efficiently infected CD19$^+$ B cell lines, but not CD19$^-$ T cell lines (Fig. 1d).

Next, we sought to exploit the unique modularity of the VSVGmut system to develop synergistic targeting strategies. While display of the anti-CD3 Fab UCHT1 yielded only modest infection of Jurkat T cell lines, CD80 mediated robust infection (Extended Data Fig. 2a). Although the infection rate for the synergistic strategy was not better than for CD80 alone, we reasoned that the ability to incorporate multiple signals during infection could be useful for engineering primary cells by providing user-defined phenotypic inputs. Therefore, we used these viruses to infect primary CD8$^+$ T cells and measured infection and CD25 upregulation as a marker of activation (Fig. 1e). We observed that the anti-CD3/CD80 combination viruses both infected and activated cells, but viruses displaying CD80 alone infected but did not activate the cells. We also observed that the anti-CD3/CD80 combination viruses induced robust T cell proliferation, while viruses displaying CD80 alone did not cause proliferation (Fig. 1f). This work presents a viral targeting system capable of simultaneously delivering multiple synergistic signals while infecting cells.

**Antigen-specific T cell targeting by engineered lentiviruses.** Following the observation that anti-CD3 viruses yielded infection of T cells via a component of the TCR complex, we next sought to determine whether we could achieve antigen-specific cell entry via the TCR itself. One potential challenge to repurposing the TCR–pMHC interaction for viral entry is its affinity, which is typically 1–50 μM[44]. We therefore generated cell lines expressing a range of previously characterized affinity variants for the 1G4 TCR[45,46], which recognizes a peptide derived from the NY-ESO-1 cancer testis

**Fig. 1 | The VSVGmut pseudotyping system is modular and utilizes a broad range of receptors and ligands. a**, Schematic of the VSVGmut (VSVG K47Q and R354A) pseudotyping system enables specific targeting by coexpression of receptor-blinded VSVG and a modular targeting ligand. **b**, VSVGmut is compatible with a broad range of receptors and targeting ligands, including immune cell surface markers, signaling and antigen receptors. **c**, Specific infection of IL-13Rα1-expressing Jurkat cells (left) as compared to parental cell line (IL-13Rα1$^-$, right) via VSVGmut lentiviruses displaying surface-tethered IL-13; data are representative of two biologically independent experiments. **d**, Specific infection of CD19$^+$ Ramos cells (left), but not CD19$^-$ Jurkat cells (right) by VSVGmut-pseudotyped viruses displaying an anti-CD19 scFv; data are representative of two biologically independent experiments. **e**, Infection of primary CD8$^+$ T cells (multiplicity of infection of 1) via lentiviruses displaying anti-CD3 Fab and/or the co-stimulatory receptor CD80 or VSVGwt viruses with or without anti-CD3/anti-CD28 magnetic beads; data were collected at day 4 after infection and are representative of three biological replicates. **f**, Proliferation of primary CD8$^+$ T cells following viral infection described in **e**, as measured by dilution of cell tracking dye added at day 0 before the addition of virus. Data are for day 7 after infection.

antigen (SLLMWITQV) presented in the context of HLA-A*02:01, with reported affinities ranging from picomolar to micromolar (Fig. 2a). To display these molecules on the virus surface, we expressed them as single-chain trimers[47], which consist of covalently linked peptide, β-2-microglobulin and MHC. We found that pMHC-displaying lentiviruses are able to efficiently infect T cells in an antigen-dependent manner. We observed similar infection efficiency across the tested affinity range, with only a modest reduction for the lowest-affinity variant (Fig. 2b). To ensure that TCR-mediated infection was generalizable, we displayed several individual pMHCs as single-chain trimers alongside VSVGmut and used them to infect either Jurkat cells expressing its parental TCR (off-target) or on-target J76 cell lines expressing their cognate TCRs (Fig. 2c). We observed specific infection across three different TCR–pMHC pairs with minimal background infection. As VSVG enables cell entry via an endocytic route[48] and because endocytosis is one of the earliest downstream events following T cell activation[49–51], we next sought to determine whether viral infection was



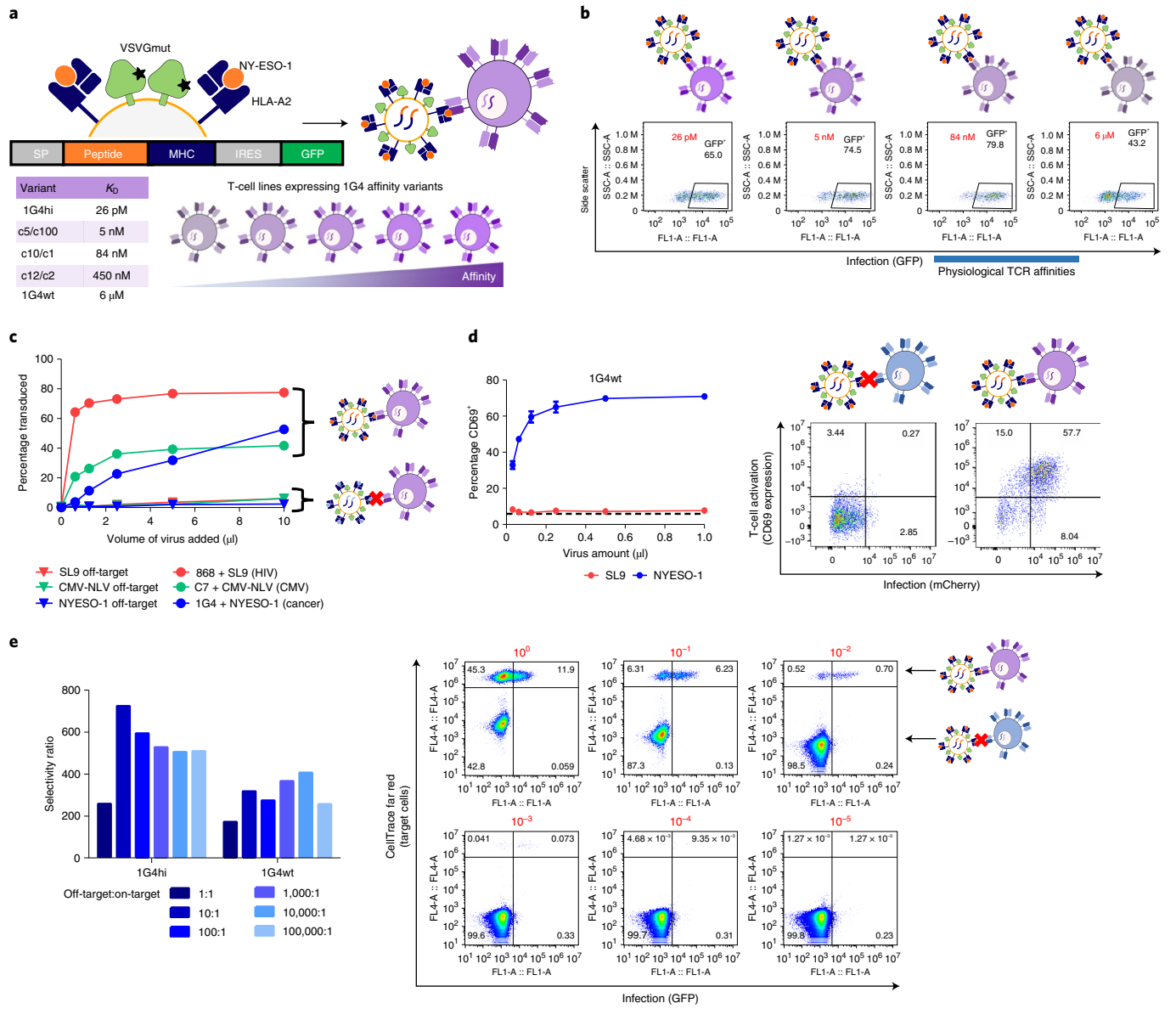



**Fig. 2 | TCR-mediated infection is sensitive, specific and induces signaling. a**, Schematic of the 1G4 TCR-HLA-A2/NY-ESO-1 system used for TCR–pMHC characterization. **b**, Representative infection data for various 1G4 TCR affinity variants by HLA-A2/NY-ESO-1 displaying VSVGmut lentiviruses. **c**, infection of J76 cells via specific TCR–pMHC interactions via pMHC-displaying VSVGmut lentiviruses for three independent TCR–pMHC pairs (HLA-A2/SL9-displaying viruses infecting 868 TCR-expressing J76 cells; HLA-A2/NLV-displaying viruses infecting C7 TCR-expressing J76 cells; HLA-A2-NY-ESO-1-displaying viruses infecting 1G4 TCR-expressing J76 cells). **d**, Upregulation of CD69 on J76 cells transduced with the 1G4wt TCR during viral entry; data shown represent mean + s.d. across three biological replicates, dashed line represents CD69 expression in untreated J76-1G4wt cells. **e**, Selectivity ratio of on-target to off-target infection of J76 cells expressing 1G4 TCR variants mixed at indicated ratios with off-target Jurkat cells; representative of three independent experiments, flow plots correspond to 1G4wt TCR-expressing J76 cells with target cell frequencies indicated in red. The selectivity ratio was calculated as the transduction rate of on-target cells divided by the transduction rate of off-target cells.

accompanied by TCR signaling. We measured CD69 expression following viral infection (Fig. 2d) and observed robust CD69 upregulation during TCR-mediated viral entry, but not during off-target combinations or VSVGwt infection (Extended Data Fig. 2c). Moreover, TCR-mediated entry was inhibited by dasatinib, a TCR signaling inhibitor[52], whereas TCR-independent infection via VSVGwt was unaffected (Extended Data Fig. 2b). Thus, we concluded that pMHC-targeted viruses integrate both binding and signaling as a means of infection. Notably, pMHC-based targeting was more efficient than infection by viruses displaying an anti-CD3 Fab. While it is unclear why this is the case, it is possible that the pMHC-based approach is more efficient at inducing TCR signaling when engaged as monomeric proteins or that the orientation of the pMHC is more favorable for viral entry than the CD3 epitope recognized by the anti-CD3 Fab.

We next sought to determine the limitations of our system by characterizing its sensitivity and specificity using the 1G4-NY-ESO-1 system. To determine the ability of our approach to discern on-target interactions in complex mixtures, we labeled 1G4-expressing cells with a cell tracking dye, mixed them at varying ratios with unlabeled Jurkat cells and infected the mixture with NY-ESO-1/HLA-A2-displaying viruses (Fig. 2e). We calculated a selectivity ratio of on-target infection relative to off-target infection by dividing the transduction rate of on-target cells by the

transduction rate of off-target cells and observed on-target selectivity greater than 200:1, down to target cell frequencies of $1 \times 10^{-5}$. Thus, we concluded that our system can detect on-target interactions in the micromolar affinity range, even in complex mixtures containing hundreds of thousands of non-target T cells.

**A viral packaging system to maintain protein-barcode linkage.** To fully enable interaction screening at library scale, we next turned to the challenge of producing lentiviral libraries. While lentiviral screens have been widely used for functional genomics, standard lentiviral packaging techniques pose inherent limitations that are particularly relevant to interaction screens. The key challenge stems from the fact that at least hundreds of plasmids containing library elements are mixed in each transfected packaging cell (Fig. 3a). In functional genomics, this leads to well-described intermolecular recombination of library elements[53,54], which is a significant source of noise in combinatorial experiments. To date, solutions to this problem have included plasmid dilution (accompanied by a 100× reduction in viral titer)[54] or simply restricting libraries to sizes suitable for arrayed screens[55–57]. This problem is compounded in our approach, as multiple plasmids entering the same packaging cell would cause multiple different targeting molecules to be expressed on each virus surface, compromising the link between viral genotype and surface phenotype.

To obviate this issue, we exploited a detail of HIV-1 replication that results in the copying of sequences between the polypurine tract and the 3′ long terminal repeat (LTR) to the 5′ end of the genome during reverse transcription and integration[58]. There have been several recent reports exploiting this phenomenon to enable genomic screens via copying of CRISPR guide RNA cassettes[59,60]. Here, we insert a promoter to enable repackaging of the viral genome following infection of a cell[61] (Fig. 3b). Conventional third-generation lentiviral packaging approaches are inherently self-inactivating in part due to an inability to transcribe packageable viral genomes following infection; the LTRs, which are weak promoters, are truncated and there is no promoter upstream of the viral genome following successful integration. Our approach, which we term 'lentiviruses activated by promoter shuffling' (LeAPS), enables generation of lentiviruses from previously transduced cells upon the re-introduction of helper plasmids by placing a strong promoter 5′ of the integrated viral genome, allowing the integrated vector to produce additional viral genomes at a high copy number. As a result, we can use lentiviral transduction to provide a single library member per packaging cell without incorporating additional viral elements into the genome. We first validated that viruses produced with the LeAPS strategy yielded high-titer virus ($>10^6$ TU ml$^{-1}$ unconcentrated), whereas conventional constructs (without a LeAPS promoter) yielded very little following transduction (Fig. 3c). By stably expressing pMHC cassettes in LeAPS-transduced packaging cells, we verified that viruses produced in this manner maintain antigen-specific targeting capability (Fig. 3d).

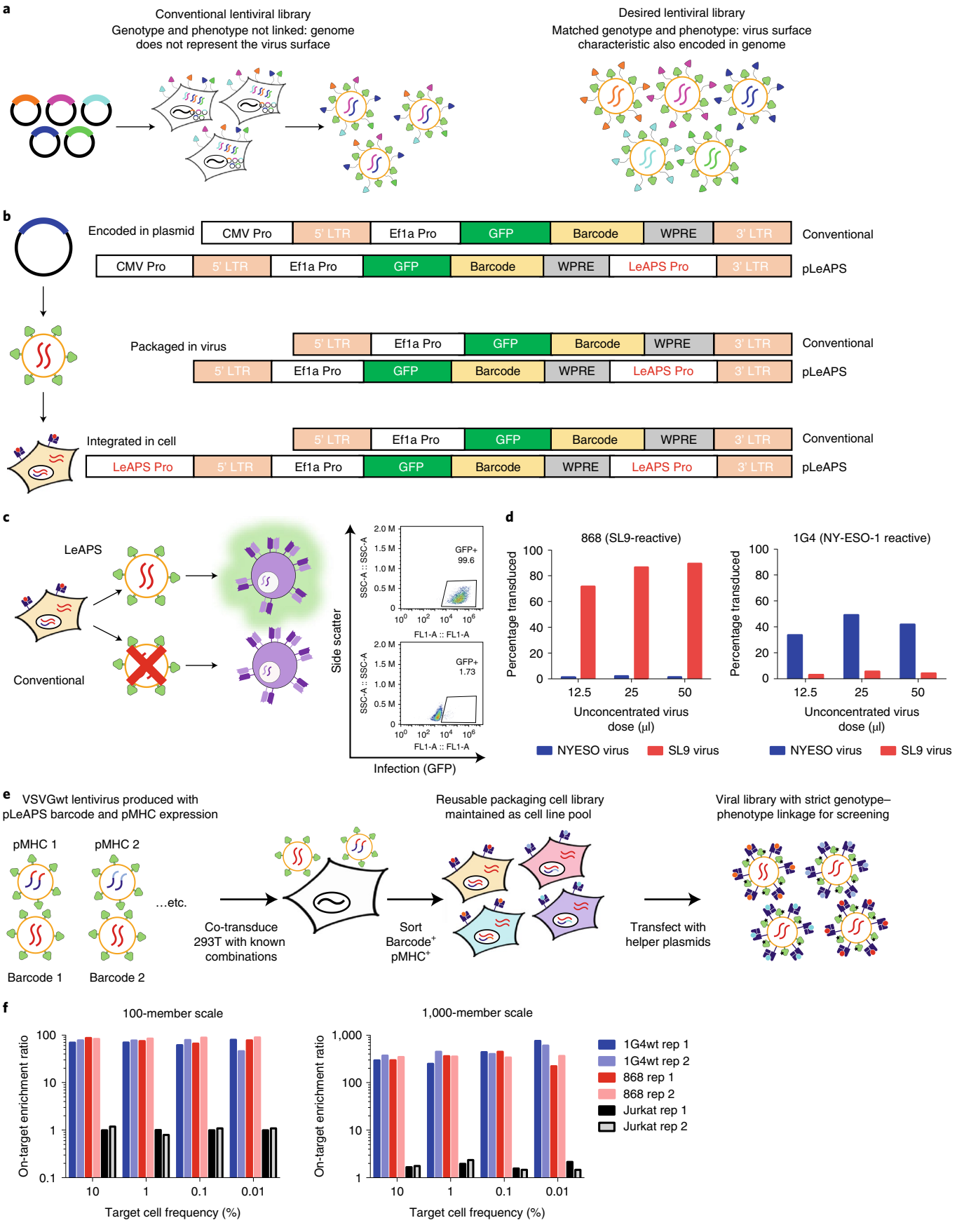
To develop a scalable library assembly strategy, we co-transduced 293T cells with two VSVGwt-pseudotyped viruses: one that drives surface expression of a known pMHC using a conventional lentivirus and another that uses the LeAPS system to deliver a repackageable genome driving expression of a fluorescent protein and a defined barcode (Fig. 3e). Following initial transduction, packaging cells for each library member were pooled and sorted for expression of pMHC and barcode in a single sorting step. This yields a library packaging line that can be used to generate the same lentiviral library indefinitely via single transfections with only helper plasmids. By contrast, conventional approaches require arrayed transfection to produce each library member for each experiment[62], which poses a substantial impediment to experimental throughput.

To validate the utility of our approach, we employed a two-component system of the HIV SL9 antigen (SLYNTVATL) paired with a LeAPS–mCherry cassette and the NY-ESO-1 antigen paired with a LeAPS–green fluorescent protein (GFP) cassette. By mixing the packaging cells for each at different ratios, we assessed the feasibility of screening pMHC libraries at 100-member and 1,000-member scales. To assess the feasibility of a 1,000-member library, we mixed the SL9 and NY-ESO-1 LeAPS packaging cell lines at a ratio of 1:1,000 by cell number, produced the virus in pools via transfection with helper plasmids and used the resulting viral pools for target infections of TCR knock-in T cell lines (Extended Data Fig. 3). We then used flow cytometry to quantify signal to noise across different frequencies of on-target cells (Fig. 3f). Across both library sizes and extending from target cell frequencies of 10% down to 0.01%, we observed robust enrichment of on-target interactions, but minimal enrichment of off-target interactions (no enrichment of either component on Jurkat cells expressing an irrelevant TCR). These results gave us confidence that the screening step would be feasible at scales of at least 1,000 pMHC variants using the LeAPS library generation approach.

**Receptor–antigen pairing by targeted retroviruses.** With all the necessary tools for RAPTR validated, we assembled a library of 96 pMHC constructs consisting of known seroprevalent viral or vaccine antigens from the Immune Epitope Database (IEDB)[63] and further filtered for binding to HLA-A2 using NetMHC4.1 (ref. [64]; Supplementary Tables 1 and 2). We co-transduced each into 293T cells alongside a LeAPS vector encoding a barcode, pooled the resulting cells proportionally and sorted for cells expressing both barcode (via mCherry) and pMHC (via GFP). We verified that antigen and barcode expression in our packaging cell library was stable after at least nine passages after sorting (Extended Data Fig. 4a). We next used the library to infect a cell line expressing the C7 TCR[65], which is known to recognize the cytomegalovirus (CMV) peptide NLVPMVATV (known as NLV). Without any cell-sorting step required, we sequenced genomic DNA from transduced cells to enumerate barcode frequencies (Fig. 4a). We observed a large enrichment of NLV barcode in transduced cells relative to the packaging cell line (Fig. 4b), with NLV representing 15% of all reads in transduced cells and no notable enrichment of other sequences. We further validated this approach in C7 replicates (Fig. 4c and Extended Data Fig. 4b) and in a separate screen of the JM22 TCR[65], which recognizes the GL9 peptide from influenza A (Fig. 4d and Extended Data Fig. 4c). In all cases, we observed robust enrichment of on-target, but not off-target, barcodes.

**RAPTR can be adapted to B cell receptors.** Many of the same constraints that limit TCR–pMHC interaction screens are also shared by efforts to identify B cell antigens at scale. Most approaches rely on barcoded multimers for very few targets and the largest antigen library reported to date is nine antigens[20]. As BCRs are also rapidly endocytosed upon antigen engagement[66,67], we reasoned that we could use

**Fig. 3 | LeAPS-based lentiviral packaging enables scalable viral library construction. a**, Schematic overview of the limitation of conventional lentiviral packaging strategies for surface displayed lentiviral libraries. **b**, Schematic overview of the mechanism of promoter translocation used in the LeAPS system. **c**, Packaging cells transduced with viruses containing LeAPS-encoding genomes and then transfected with lentiviral helper plasmids (top) and transduced with viruses containing standard genomes (bottom). **d**, LeAPS-produced viruses maintain TCR specificity in two independent TCR–pMHC systems, **e**, Schematic overview of the implementation of LeAPS for viral library assembly. **f**, On-target enrichment of libraries using defined mixtures of A2-NY-ESO-1 and A2-SL9 packaging cells to represent 100- or 1,000-variant libraries.


a
Conventional lentiviral library
Genotype and phenotype not linked: genome does not represent the virus surface
Desired lentiviral library
Matched genotype and phenotype: virus surface characteristic also encoded in genome
b
Encoded in plasmid
Packaged in virus
Integrated in cell
CMV Pro
5′ LTR
Ef1a Pro
GFP
Barcode
WPRE
3′ LTR
LeAPS Pro
Conventional
pLeAPS
c
LeAPS
Conventional
Side scatter
Infection (GFP)
SSC-A :: SSC-A
FL1-A :: FL1-A
GFP+ 99.6
GFP+ 1.73
d
868 (SL9-reactive)
1G4 (NY-ESO-1 reactive)
Percentage transduced
Unconcentrated virus dose (µl)
NYESO virus
SL9 virus
e
VSVGwt lentivirus produced with pLeAPS barcode and pMHC expression
pMHC 1
pMHC 2
…etc.
Barcode 1
Barcode 2
Co-transduce 293T with known combinations
Reusable packaging cell library maintained as cell line pool
Sort Barcode+ pMHC+
Transfect with helper plasmids
Viral library with strict genotype–phenotype linkage for screening
f
100-member scale
1,000-member scale
On-target enrichment ratio
Target cell frequency (%)
1G4wt rep 1
1G4wt rep 2
868 rep 1
868 rep 2
Jurkat rep 1
Jurkat rep 2

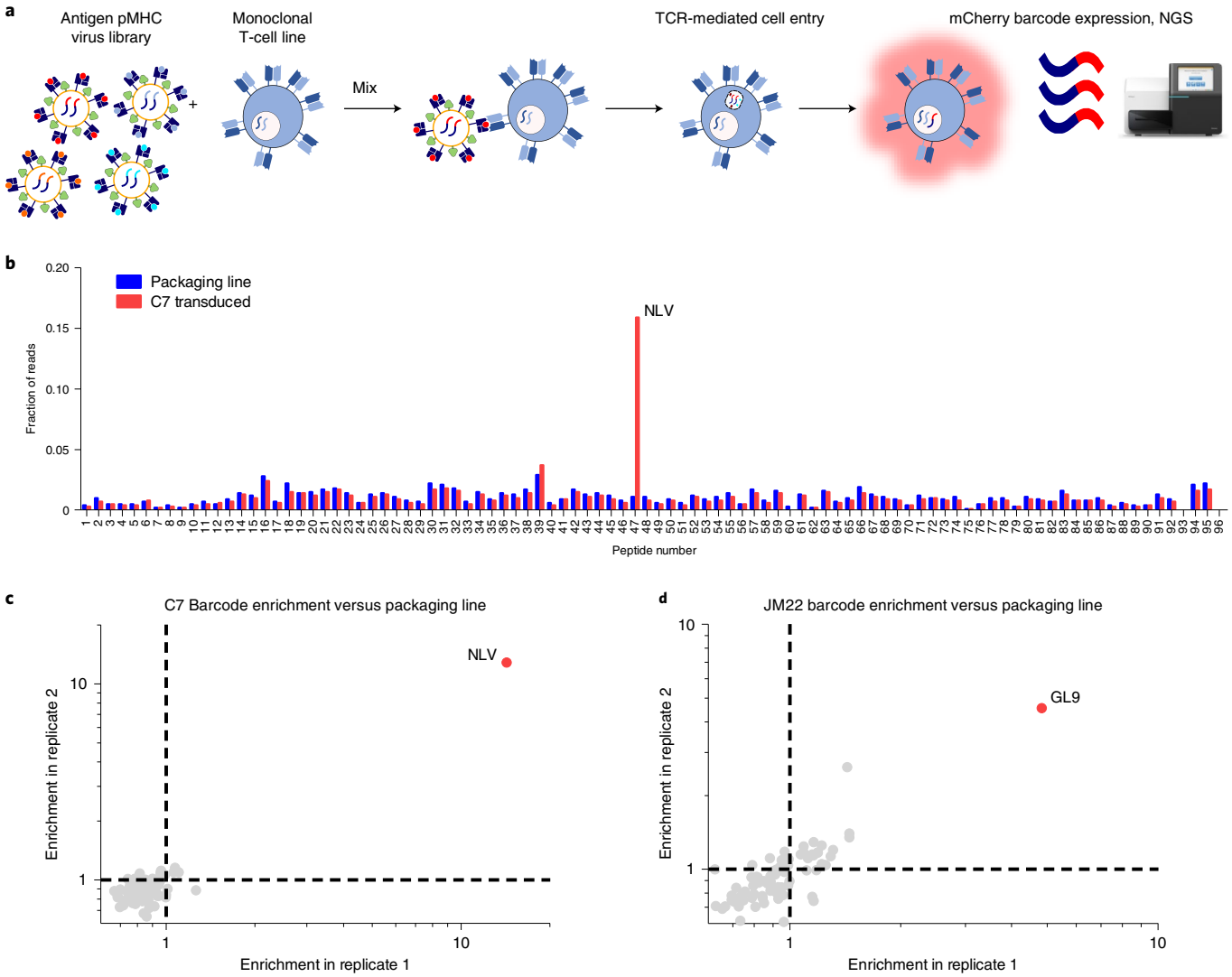



**Fig. 4 | RAPTR enables facile antigen identification for TCRs. a**, Workflow for RAPTR on monoclonal cell lines. The 96-member pMHC-displaying lentiviral library was mixed with J76 cells expressing a known TCR. After 2 d of infection, genomic DNA from unsorted transduced cells was collected and analyzed by next-generation sequencing (NGS). **b**, Read fraction of barcodes following infection of C7 TCR-transduced J76 cells. Blue bars represent the read fractions of barcodes from the 293T packaging cell line used to produce the viral library and provides the unenriched (null) distribution-absent TCR selection. The red bars indicate the read fraction in transduced C7-expressing J76 cells, revealing the selectivity of the TCR based on increasing read fraction of the barcode for the cognate ligand (NLV). **c**, Comparison of antigen enrichment relative to the packaging line upon library infection of C7 TCR-transduced J76 cells across two additional replicates. **d**, Comparison of antigen enrichment relative to the packaging line upon library infection of JM22 TCR-transduced J76 cells across two replicates.

stabilized viral surface antigens or known immunogens as targeting ligands for viral infection. We thus generated a 'hybrid' pseudotype coexpressing either the surface-tethered receptor-binding domain (RBD) or the full-length, prefusion-stabilized spike (S2P) protein from SARS-CoV-2 (ref. [68]) alongside VSVGmut (Fig. 5a). We used each of these to achieve efficient, antigen-specific infection of a Ramos cell line expressing CR3022, a BCR clone that is cross-reactive for the RBDs of SARS-CoV-1 and SARS-CoV-2 (refs. [69,70]; Fig. 5b). Similarly, hybrid pseudotypes incorporating HIV env CD4 binding site (CD4bs) constructs from diverse clades[71] each infected cell lines expressing the bNab VRC01 (ref. [72]) and a known affinity-reducing mutation (D368R) dramatically reduced, but did not fully abrogate, infection[73].

To determine the sensitivity for BCR-mediated cell entry, we mixed Ramos cells expressing either the CR3022 BCR ($IgM^+$) or no BCR ($IgM^-$) and tracked on-target cells via surface IgM expression (Fig. 5c and Extended Data Fig. 5). Using RBD-displaying viruses, we observed specific infection at target cell frequencies at least as rare as $1 \times 10^{-5}$, but could not calculate a meaningful selectivity ratio due to a lack of off-target infection. Next, to assess the potential feasible library size, we mixed on-target RBD-displaying viruses encoding GFP with off-target CD4bs-displaying viruses encoding mCherry and used them to infect CR3022 BCR-expressing Ramos cells (Fig. 5d and Extended Data Fig. 5c). Keeping the total amount of virus constant, we were able to robustly detect on-target interactions with minimal off-target infection in mixtures as dilute as 1:1,000 on-target viruses. We observed similar results when displaying SARS-CoV-2 S2P instead of RBD (Extended Data Fig. 5). These results gave us confidence in the feasibility of extracting a meaningful signal from interaction screens using lentiviral infection via the BCR.

Next, we sought to establish the capabilities of RAPTR for BCR profiling in a true library setting. We assembled a library

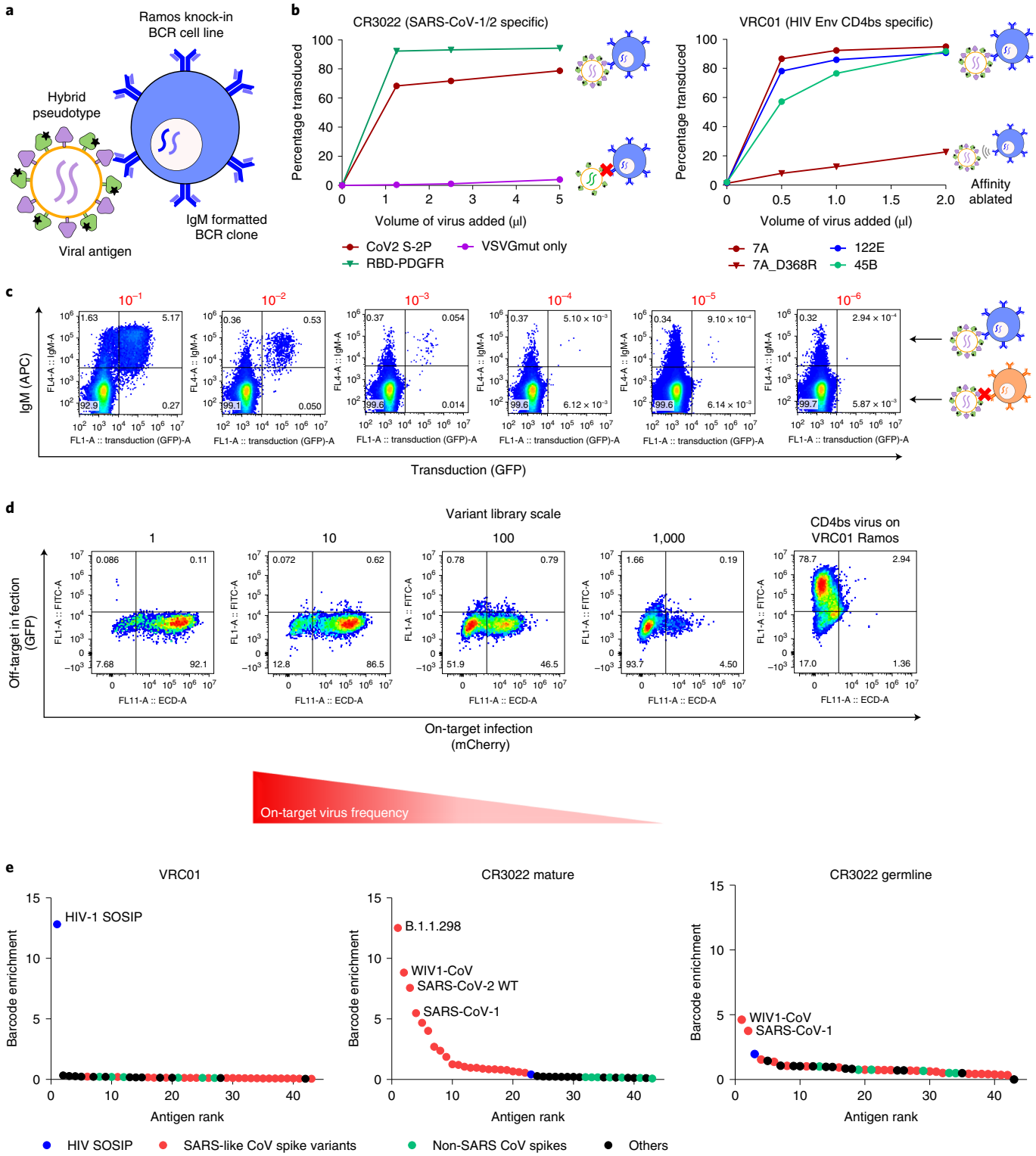
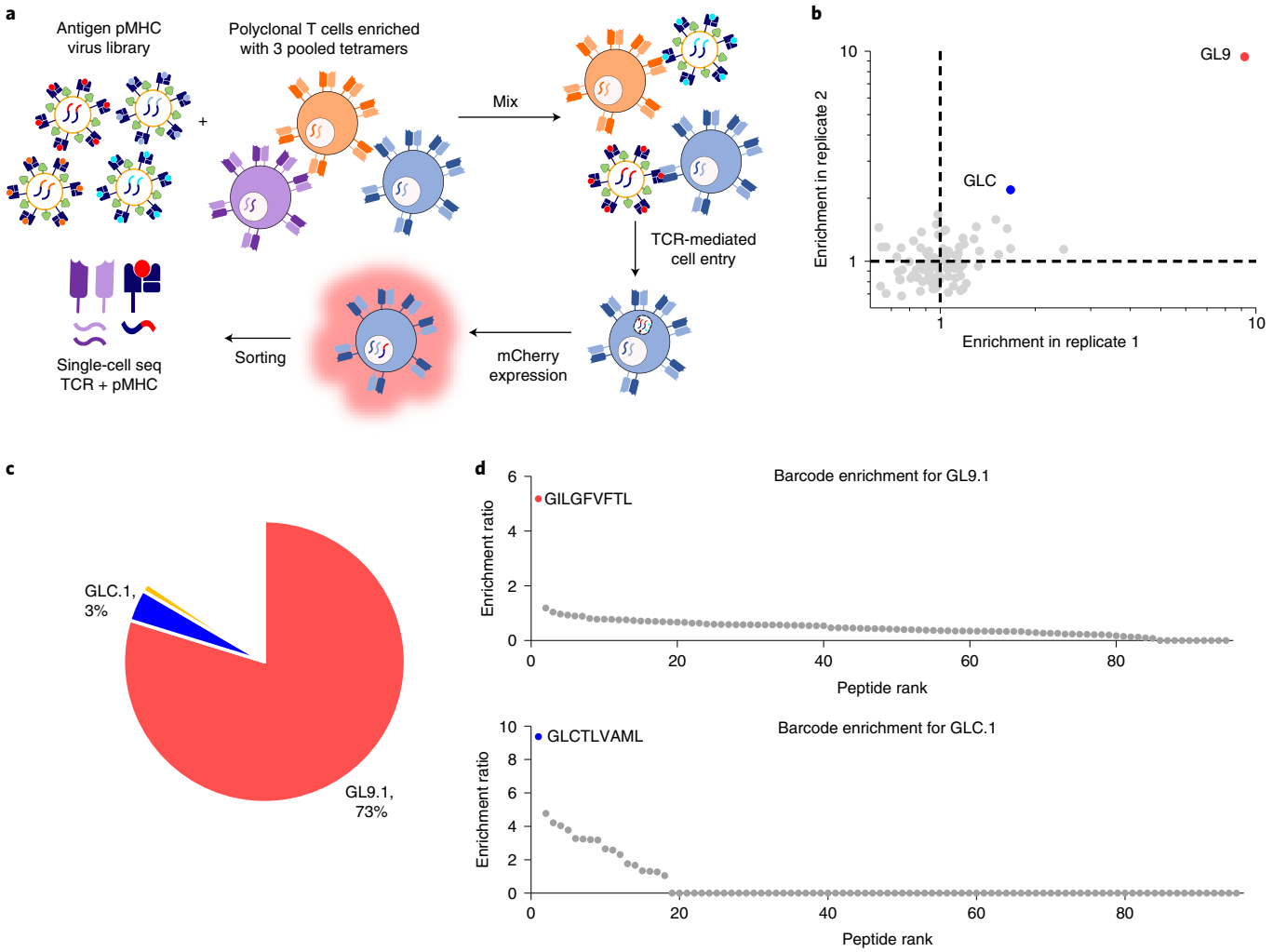



**Fig. 5 | RAPTR enables profiling of B cell reactivity via infection of BCR knock-in cells. a**, Overview of BCR-based targeting via the VSVGmut hybrid pseudotype in which a viral protein is coexpressed with VSVGmut. **b**, Targeting of Ramos cells expressing the SARS-CoV S protein-specific CR3022 BCR via VSVGmut lentiviruses displaying either SARS-CoV-2 spike (2P) or surface-bound SARS-CoV-2 S protein RBD or VRC01 BCR via env constructs representing multiple clades of HIV; representative data are from two independent experiments. **c**, SARS-CoV-2 RBD-displaying viral infection of CR3022 BCR-expressing Ramos cells (on-target, IgM$^+$) mixed into IgM$^-$ Ramos cells (off-target, IgM$^-$) at target cell frequencies indicated in red, demonstrating selectivity in a mixed population of cells. **d**, Infection of CR3022 BCR-expressing Ramos cells with mixtures of SARS-CoV-2 RBD (on-target, expressing mCherry) and HIV env (off-target, expressing GFP) viruses; variant library size indicates the ratio of off-target to on-target virus present. Infection of VRC01-expressing Ramos cells with HIV env hybrid pseudotyped viruses (right) demonstrates that the viral particles remain functional. **e**, Enrichment of antigen barcodes relative the viral packaging cell line following viral glycoprotein library infection of Ramos cells expressing BCRs for VRC01, mature CR3022 or the germline-reverted version CR3022, demonstrating selectivity even for antibody sequences before affinity maturation.




e

| | | | | Clone frequency at each pipeline stage | | |
|---|---|---|---|---|---|---|
| TCR ID | pMHC Target | CDR3β | No. of cells | Primary T cells | Starting library | Virus panning |
| GL9.1 | Flu M1 | CASSIRSSYEQYF | 1,068 | 0.02% | 0.02% | 73.25% |
| GLC.1 | BMLF1 | CASSTGQLSNTGELFF | 49 | 0.04% | 0.16% | 3.36% |

**Fig. 6 | RAPTR scales to polyclonal antigen identification in a single, scalable pipeline. a**, Schematic overview of the library-on-library RAPTR paradigm for TCR antigen identification. The 96-member viral library is mixed with polyclonal TCR-expressing J.RT3 cells; 2 d after infection, mCherry$^+$ cells were sorted and analyzed by single-cell RNA sequencing to identify expressed TCRs and enumerate integrated pMHC barcodes. **b**, Bulk sequencing of barcodes across all infected cells following infection of enriched CSS-930 TCR library with viral antigen library shows enrichment of immunodominant EBV and Influenza antigens in two independent infections. **c**, Clonal frequencies of enriched TCRs in cells analyzed by single-cell sequencing; white space represents TCRs found in fewer than two cells. **d**, Enrichment of antigen barcodes in the top two most abundant clones relative to the viral packaging cell line. **e**, Detailed tracking of TCR clones with matched antigens, including CDR3 sequence, number of cells analyzed and clonal frequencies at each stage of the pipeline.

composed of 21 prefusion-stabilized SARS-CoV-2 spike variants and 22 additional viral surface glycoproteins or immunogens from diverse sources as controls to enable proof-of-concept experiments (Supplementary Tables 3 and 4). When we used this library to infect cells expressing the VRC01 BCR, we observed clear enrichment of only HIV-1 SOSIP[74] (Fig. 5e), highlighting the potential for direct receptor de-orphanization. When we infected cells expressing the CR3022 BCR, we observed clear enrichment of SARS-CoV spikes, including multiple SARS-CoV-2 variants. Notably, a version of CR3022 that excluded the mutations acquired via somatic hypermutation yielded enrichment of only SARS-CoV-1 spike and the closely related WIV-1 spike, highlighting the ability of RAPTR to distinguish BCR cross-reactivity even for naive BCRs that have yet to undergo affinity maturation.

**RAPTR enables library-on-library screens.** To fully realize the potential of RAPTR for antigen identification, we sought to apply it to library-versus-library screening by using our 96-member

pMHC viral library to pan a previously reported library of >450,000 TCRs[75]. To simplify the first attempt, we pre-enriched for potentially reactive cells using tetramers for CMV, Epstein–Barr virus (EBV) and influenza (CEF) antigens presented by HLA-A2, resulting in a polyclonal pool of TCRs with greater prevalence of these specificities, similar to the routine process of pre-expanding cells (Fig. 6a). Following transduction and FACS to isolate transduced cells, we used bulk sequencing to identify which antigens were enriched in aggregate. In line with tetramer staining of the untransduced cells, we observed strong enrichment of influenza GL9 and more modest enrichment of EBV GLCTLVAML (GLC) (Fig. 6b and Extended Data Fig. 6).

With this validation in hand, we performed single-cell RNA sequencing using the 10X Genomics Chromium 5′ chemistry with V(D)J enrichment. We included a targeted primer at the reverse transcription step and a custom amplification protocol to enable recovery of our pMHC barcodes, akin to ECCITEseq and other similar methods[76,77]. In this experiment, we recovered 1,458 cells, with TCR clonotypes dominated by a few clones (Fig. 6c,e). We grouped cells of common clonotype together and analyzed pMHC barcode expression to determine enrichment relative to packaging library frequency. The most prevalent TCR clone, GL9.1, demonstrated strong enrichment of the GL9 pMHC signal (Fig. 6d). Upon literature search, we noted that GL9.1 represents a well-characterized public clonotype known to recognize the GL9 peptide presented by HLA-A2 (ref. [78]).

The next most abundant clone, GLC.1, was identified in a previous screen of this same TCR library by recombinant HLA-A2 GLC dextramer[75]. Our screen also identified GLC as the target pMHC (Fig. 6d). Notably, this was the only clone previously identified by GLC dextramer panning that was validated by monoclonal binding and functional tests. Our approach did not identify any of the previously reported false positives as hits, providing further evidence that RAPTR is an efficient means of identifying T cell antigens that integrates both binding and signaling. Finally, while T cells that recognize NLV are common in HLA-A2$^+$ CMV-seropositive individuals[79], we did not observe any NLV-reactive cells (Fig. 6b). However, this result is consistent with previously reported screening of the same library via orthogonal methods[75] and our tetramer staining data (Extended Data Fig. 6a).

## Discussion

RAPTR is a high-throughput platform for directly linking immune receptors with their cognate antigens. It is based on the combination of several conceptual and technological advances that enable (1) versatile and efficient targeted viral entry via the VSVGmut pseudotyping system; (2) a method for scalable, reproducible lentiviral library packaging (LeAPS); and (3) the use of viral entry as a means of screening for interactions. We demonstrated a system for efficient, antigen-specific infection via both the TCR and the BCR that integrates both binding and target cell activation. We then exploited this mode of infection to match receptors with target antigens in complex mixtures and in a library-versus-library format.

We demonstrated that the VSVGmut pseudotyping system is efficient, specific and uniquely modular by targeting a variety of cell surface proteins, including cytokine receptors, co-stimulatory molecules, lineage markers and both TCRs and BCRs. We report here as many different targeting approaches for VSVGmut as have been reported for any other single pseudotyping system[32], including antigen-specific infection of lymphocytes and a targeting approach that can simultaneously deliver multiple signals. In addition, the VSVGmut system produces viruses at higher average titers than other systems while maintaining a high degree of specificity. Whereas most pseudotyping systems have been limited to the use of small, stable, high-affinity targeting ligands (such as scFvs or DARPins), VSVGmut is versatile in ligand usage due to the modular design it permits. We report targeting even at micromolar affinities, whereas the efficiency of infection by paramyxoviral pseudotypes was reported to dramatically decrease in the low nanomolar range[80]. We observed efficient pMHC-based infection for both antiviral TCRs (C7, JM22 and the TCRs recovered shown in Fig. 6) and 1G4, which recognizes a self-antigen, suggesting that the full range of possible CD8$^+$ antigens should be able to be profiled.

**Applications of RAPTR.** We have demonstrated the utility of RAPTR for identifying antigens recognized by TCRs and BCRs. For T cell reactivity, our approach could be directly applied to rapidly deorphanize TCRs from existing single-cell studies, which routinely pair TCR sequences with gene expression for cancer, autoimmunity or infectious disease. Existing techniques, including yeast display, SABRs and T-Scan, require individual screens to pair each TCR with its cognate antigen, including at least one cell-sorting step per TCR. In contrast, RAPTR can be performed on pools of TCRs in a single step, with only one cell-sorting step required (and no sorting required for single-TCR experiments). While barcoded pMHC multimers can be built into single-cell workflows, large antigen pools are inherently difficult to produce and validate and require an equivalent effort for each new batch. These factors have largely limited their use to only a few specialized labs thus far. RAPTR library production is straightforward and far more scalable in comparison, requiring only a simple transfection per batch following initial library assembly. In principle, these reagents could be made and distributed at a large scale, either via sharing the packaging cell line or by leveraging current industrial infrastructure for production of lentiviruses.

For B cells, we envision applying pools of protein variants, such as the one presented here, to large libraries of BCRs. Such an approach would greatly accelerate isolation and optimization of cross-reactive monoclonal antibodies to a wide range of target classes. Current techniques require the recombinant expression of individual monoclonal antibodies followed by ELISA for each target of interest, which represents a major bottleneck to experimental scale and scope[81]. Using RAPTR with BCR knock-in cells, which could themselves be prepared as pooled lentiviral libraries[70], could alleviate this bottleneck. For example, even scales smaller than the one we presented would be sufficient to represent every subtype of influenza hemagglutinin and neuraminidase, multiple representatives of each of the major clades of HIV-1 and more. Larger libraries on the scale of thousands of variants, whose feasibility we demonstrated, could enable detailed antigenic site or epitope mapping for many monoclonal antibodies at once, a process currently limited to structural studies that are difficult to scale.

Finally, the demonstrated modularity of the system opens the door to many potential applications. As our approach only requires HEK cell surface expression of the targeting molecule, we hypothesize that RAPTR can be readily adapted to other MHC class I and class II alleles, as well as non-human systems for vaccine and immunotherapy development. This compares favorably to recombinant MHC expression or yeast display, each of which can require allele-specific optimizations and/or mutations to ensure proper folding[13,82,83]. However, our system is not restricted to immunological applications. In principle, it is applicable to any receptor–ligand system in which one binding partner can be displayed on a lentivirus and the other can be expressed as an endocytic receptor on cells. This may enable large-scale interactome and coevolution studies that are not readily achievable by existing techniques. Furthermore, the VSVGmut pseudotyping system may be useful as a new platform for the development of efficient, specific gene therapies.

**Limitations.** As with any technology, RAPTR has several limitations. Our proof-of-concept studies were performed on ~100-antigen scale and while our control experiments indicate that

this screening capability could be readily increased to thousands, these scales will still require preselection of antigen targets for analysis. Further streamlining library assembly and increasing transduction efficiency will be focuses for enabling larger-scale assays. At present, the LeAPS system involves one-time arrayed transfections and transductions at scales and complexity routinely performed for genomics screens[55–57], after which lentiviral library generation can be performed in a straightforward manner by standard transfections of the pooled packaging cell library. As a result, library generation is currently the largest bottleneck to throughput. Due to the one-time nature of the bottleneck, we believe that automation will readily enable library sizes of up to $10^3$, but adapting our approach for pooled library generation could aid in further scaling.

In addition, the single-cell sequencing step must be performed in sufficient depth to facilitate acceptable signal-to-noise determinations. While we chose to use a commercially available platform to ensure broad applicability, recent and future advances in the scale of single-cell analysis[84] can improve the utility of RAPTR regardless of platform. Like other genetically encoded techniques, RAPTR also requires previous selection of MHC haplotype(s) to generate libraries. The antigen identification studies we presented were on cell lines with receptors knocked in, rather than primary cells. While the decreasing cost and turnaround time of gene synthesis will make such resources increasingly available, future studies can apply RAPTR directly to primary cells. Moreover, immortalized TCR libraries are becoming increasingly common as library generation methods continue to improve[75,85–87] and can serve as renewable resources to conduct deep profiling of low-abundance or irreplaceable samples such as patient-derived tumor-infiltrating lymphocytes. Finally, for each of these cases, our data indicate that the ability to enrich for antigen-specific cells at library scale will be beneficial for efficient deorphanization, to ensure that target cells are present at sufficient frequencies for infection and analysis. Existing techniques, including stimulation via peptide pools, should readily achieve this but may require some optimization.

RAPTR is a versatile platform for high-throughput interaction screens that incorporates diversity of both receptors and ligands in a single assay. The resulting tools will enable detailed studies of antigen recognition for understanding and engineering the adaptive immune response, as well as broader interactome studies.

## Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41592-022-01436-z.

Received: 18 September 2021; Accepted: 1 March 2022;
Published online: 8 April 2022

## References

1. Joglekar, A. V. & Li, G. T cell antigen discovery. *Nat. Methods* **18**, 873–880 (2021).
2. McCutcheon, M. et al. A sensitive ELISPOT assay to detect low-frequency human T lymphocytes. *J. Immunol. Methods* **210**, 149–166 (1997).
3. Kula, T. et al. T-Scan: a genome-wide method for the systematic discovery of T cell epitopes. *Cell* **178**, 1016–1028 (2019).
4. Sharma, G., Rive, C. M. & Holt, R. A. Rapid selection and identification of functional CD8+ T cell epitopes from large peptide-coding libraries. *Nat. Commun.* **10**, 4553 (2019).
5. Joglekar, A. V. et al. T cell antigen discovery via signaling and antigen-presenting bifunctional receptors. *Nat. Methods* **16**, 191–198 (2019).
6. Kisielow, J., Obermair, F.-J. & Kopf, M. Deciphering CD4+ T cell specificity using novel MHC–TCR chimeric receptors. *Nat. Immunol.* **20**, 652–662 (2019).
7. Li, G. et al. T cell antigen discovery via trogocytosis. *Nat. Methods* **16**, 183–190 (2019).
8. Lee, M. N. & Meyerson, M. Antigen identification for HLA class I- and HLA class II-restricted T cell receptors using cytokine-capturing antigen-presenting cells. *Sci. Immunol.* **6**, eabf4001 (2021).
9. Robins, H. S. et al. Overlap and effective size of the human CD8+ T cell receptor repertoire. *Sci. Transl. Med.* **2**, 47ra64 (2010).
10. Arstila, T. P. et al. A direct estimate of the human αβ T cell receptor diversity. *Science* **286**, 958–961 (1999).
11. Wang, Y. et al. Using a baculovirus display library to identify MHC class I mimotopes. *Proc. Natl Acad. Sci. USA* **102**, 2476–2481 (2005).
12. Crawford, F., Huseby, E., White, J., Marrack, P. & Kappler, J. W. Mimotopes for alloreactive and conventional T cells in a peptide-MHC display library. *PLoS Biol.* **2**, E90 (2004).
13. Birnbaum, M. E. et al. Deconstructing the peptide-MHC specificity of T cell recognition. *Cell* **157**, 1073–1087 (2014).
14. Gee, M. H. et al. Antigen identification for orphan T cell receptors expressed on tumor-infiltrating lymphocytes. *Cell* **172**, 549–563 (2018).
15. Bentzen, A. K. et al. Large-scale detection of antigen-specific T cells using peptide-MHC-I multimers labeled with DNA barcodes. *Nat. Biotechnol.* **34**, 1037–1045 (2016).
16. Bentzen, A. K. et al. T cell receptor fingerprinting enables in-depth characterization of the interactions governing recognition of peptide-MHC complexes. *Nat. Biotechnol.* https://doi.org/10.1038/nbt.4303 (2018).
17. Overall, S. A. et al. High-throughput pMHC-I tetramer library production using chaperone-mediated peptide exchange. *Nat. Commun.* **11**, 1909 (2020).
18. Zhang, S.-Q. et al. High-throughput determination of the antigen specificities of T cell receptors in single cells. *Nat. Biotechnol.* **36**, 1156–1159 (2018).
19. Shiakolas, A. R. et al. Efficient discovery of potently neutralizing SARS-CoV-2 antibodies using LIBRA-seq with ligand blocking. *Nat. Biotechnol.* https://doi.org/10.1038/s41587-022-01232-2 (2022).
20. Setliff, I. et al. High-throughput mapping of B cell receptor sequences to antigen specificity. *Cell* **179**, 1636–1646 (2019).
21. Fields, S. & Song, O. A novel genetic system to detect protein–protein interactions. *Nature* **340**, 245–246 (1989).
22. Wojtowicz, W. M. et al. A human IgSF cell-surface interactome reveals a complex network of protein–protein interactions. *Cell* **182**, 1027–1043 (2020).
23. Verschueren, E. et al. The immunoglobulin superfamily receptome defines cancer-relevant networks associated with clinical outcome. *Cell* **182**, 329–344 (2020).
24. Özkan, E. et al. An extracellular interactome of immunoglobulin and LRR proteins reveals receptor–ligand networks. *Cell* **154**, 228–239 (2013).
25. Havugimana, P. C. et al. A census of human soluble protein complexes. *Cell* **150**, 1068–1081 (2012).
26. Babu, M. et al. Interaction landscape of membrane-protein complexes in *Saccharomyces cerevisiae*. *Nature* **489**, 585–589 (2012).
27. Younger, D., Berger, S., Baker, D. & Klavins, E. High-throughput characterization of protein–protein interactions by reprogramming yeast mating. *Proc. Natl Acad. Sci. USA* **114**, 12166–12171 (2017).
28. Gu, L. et al. Multiplex single-molecule interaction profiling of DNA-barcoded proteins. *Nature* **515**, 554–557 (2014).
29. Buchholz, C. J., Duerner, L. J., Funke, S. & Schneider, I. C. Retroviral display and high throughput screening. *Comb. Chem. High. Throughput Screen.* **11**, 99–110 (2008).
30. Buchholz, C. J. et al. In vivo selection of protease cleavage sites from retrovirus display libraries. *Nat. Biotechnol.* **16**, 951–954 (1998).
31. Schneider, R. M. et al. Directed evolution of retroviruses activatable by tumour-associated matrix metalloproteases. *Gene Ther.* **10**, 1370–1380 (2003).
32. Frank, A. M. & Buchholz, C. J. Surface-engineered lentiviral vectors for selective gene transfer into subtypes of lymphocytes. *Mol. Ther. Methods Clin. Dev.* **12**, 19–31 (2019).
33. Naldini, L. et al. In vivo gene delivery and stable transduction of nondividing cells by a lentiviral vector. *Science* **272**, 263–267 (1996).
34. Lei, Y., Joo, K.-I. & Wang, P. Engineering fusogenic molecules to achieve targeted transduction of enveloped lentiviral vectors. *J. Biol. Eng.* **3**, 8 (2009).
35. Yang, H., Joo, K.-I., Ziegler, L. & Wang, P. Cell type-specific targeting with surface-engineered lentiviral vectors co-displaying OKT3 antibody and fusogenic molecule. *Pharm. Res.* **26**, 1432–1445 (2009).
36. Yang, L., Bailey, L., Baltimore, D. & Wang, P. Targeting lentiviral vectors to specific cell types in vivo. *Proc. Natl Acad. Sci. USA* **103**, 11479–11484 (2006).
37. Frecha, C. et al. Stable transduction of quiescent T cells without induction of cycle progression by a novel lentiviral vector pseudotyped with measles virus glycoproteins. *Blood* **112**, 4843–4852 (2008).
38. Ou, W. et al. Specific targeting of human interleukin (IL)-13 receptor α2-positive cells with lentiviral vectors displaying IL-13. *Hum. Gene Ther. Methods* **23**, 137–147 (2012).
39. Funke, S. et al. Targeted cell entry of lentiviral vectors. *Mol. Ther.* **16**, 1427–1436 (2008).
40. Bender, R. R. et al. Receptor-targeted nipah virus glycoproteins improve cell-type selective gene delivery and reveal a preference for membrane-proximal cell attachment. *PLoS Pathog.* **12**, e1005641 (2016).

41. Pfeiffer, A. et al. In vivo generation of human CD19-CAR T cells results in B-cell depletion and signs of cytokine release syndrome. *EMBO Mol. Med.* **10**, e9158 (2018).
42. Agarwal, S. et al. In vivo generation of CAR T cells selectively in human CD4$^+$ lymphocytes. *Mol. Ther.* **28**, 1783–1794 (2020).
43. Nikolic, J. et al. Structural basis for the recognition of LDL-receptor family members by VSV glycoprotein. *Nat. Commun.* **9**, 1029 (2018).
44. Rossjohn, J. et al. T cell antigen receptor recognition of antigen-presenting molecules. *Annu. Rev. Immunol.* **33**, 169–200 (2015).
45. Zhao, Y. et al. High-affinity TCRs generated by phage display provide CD4$^+$ T cells with the ability to recognize and kill tumor cell lines. *J. Immunol.* **179**, 5845–5854 (2007).
46. Li, Y. et al. Directed evolution of human T-cell receptors with picomolar affinities by phage display. *Nat. Biotechnol.* **23**, 349–354 (2005).
47. Yu, Y. Y. L., Netuschil, N., Lybarger, L., Connolly, J. M. & Hansen, T. H. Cutting edge: single-chain trimers of MHC class I molecules form stable structures that potently stimulate antigen-specific T cells and B cells. *J. Immunol.* **168**, 3145–3149 (2002).
48. Matlin, K. S., Reggio, H., Helenius, A. & Simons, K. Pathway of vesicular stomatitis virus entry leading to infection. *J. Mol. Biol.* **156**, 609–631 (1982).
49. Alcover, A. & Alarcón, B. Internalization and intracellular fate of TCR–CD3 complexes. *Crit. Rev. Immunol.* **20**, 325–346 (2000).
50. Valitutti, S., Müller, S., Salio, M. & Lanzavecchia, A. Degradation of T cell receptor (TCR)-CD3-ζ complexes after antigenic stimulation. *J. Exp. Med.* **185**, 1859–1864 (1997).
51. Liu, H., Rhodes, M., Wiest, D. L. & Vignali, D. A. On the dynamics of TCR:CD3 complex cell surface expression and downmodulation. *Immunity* **13**, 665–675 (2000).
52. Blake, S., Hughes, T. P., Mayrhofer, G. & Lyons, A. B. The Src/ABL kinase inhibitor dasatinib (BMS-354825) inhibits function of normal human T-lymphocytes in vitro. *Clin. Immunol.* **127**, 330–339 (2008).
53. Xie, S., Cooley, A., Armendariz, D., Zhou, P. & Hon, G. C. Frequent sgRNA-barcode recombination in single-cell perturbation assays. *PLoS ONE* **13**, e0198635 (2018).
54. Feldman, D., Singh, A., Garrity, A. J. & Blainey, P. C. Lentiviral co-packaging mitigates the effects of intermolecular recombination and multiple integrations in pooled genetic screens. Preprint at *bioRxiv* https://doi.org/10.1101/262121 (2018).
55. Adamson, B. et al. A multiplexed single-cell CRISPR screening platform enables systematic dissection of the unfolded protein response. *Cell* **167**, 1867–1882 (2016).
56. Parnas, O. et al. A genome-wide CRISPR screen in primary immune cells to dissect regulatory networks. *Cell* **162**, 675–686 (2015).
57. Berger, A. H. et al. High-throughput phenotyping of lung cancer somatic mutations. *Cancer Cell* **30**, 214–228 (2016).
58. Hu, W.-S. & Hughes, S. H. HIV-1 reverse transcription. *Cold Spring Harb. Perspect. Med.* **2**, a006882 (2012).
59. OhAinle, M. et al. A virus-packageable CRISPR screen identifies host factors mediating interferon inhibition of HIV. *eLife* **7**, e39823 (2018).
60. Datlinger, P. et al. Pooled CRISPR screening with single-cell transcriptome readout. *Nat. Methods* **14**, 297–301 (2017).
61. Broussau, S. et al. Inducible packaging cells for large-scale production of lentiviral vectors in serum-free suspension culture. *Mol. Ther.* **16**, 500–507 (2008).
62. Hill, A. J. et al. On the design of CRISPR-based single-cell molecular screens. *Nat. Methods* **15**, 271–274 (2018).
63. Vita, R. et al. The Immune Epitope Database (IEDB): 2018 update. *Nucleic Acids Res.* **47**, D339–D343 (2019).
64. Jurtz, V. et al. NetMHCpan-4.0: Improved peptide-MHC class I interaction predictions integrating eluted ligand and peptide binding affinity data. *J. Immunol.* **199**, 3360–3368 (2017).
65. Yang, X. et al. Structural basis for clonal diversity of the public T cell response to a dominant human cytomegalovirus epitope. *J. Biol. Chem.* **290**, 29106–29119 (2015).
66. Malhotra, S., Kovats, S., Zhang, W. & Coggeshall, K. M. B cell antigen receptor endocytosis and antigen presentation to T cells require Vav and dynamin. *J. Biol. Chem.* **284**, 24088–24097 (2009).
67. Dougan, S. K. et al. Antigen-specific B-cell receptor sensitizes B cells to infection by influenza virus. *Nature* **503**, 406–409 (2013).
68. Hsieh, C.-L. et al. Structure-based design of prefusion-stabilized SARS-CoV-2 spikes. *Science* **369**, 1501–1505 (2020).
69. Yuan, M. et al. A highly conserved cryptic epitope in the receptor binding domains of SARS-CoV-2 and SARS-CoV. *Science* **368**, 630–633 (2020).
70. Weaver, G. C. et al. In vitro reconstitution of B cell receptor–antigen interactions to evaluate potential vaccine candidates. *Nat. Protoc.* **11**, 193–213 (2016).
71. Zhou, T. et al. Structural repertoire of HIV-1-neutralizing antibodies targeting the CD4 supersite in 14 donors. *Cell* **161**, 1280–1292 (2015).
72. Wu, X. et al. Rational design of envelope identifies broadly neutralizing human monoclonal antibodies to HIV-1. *Science* **329**, 856–861 (2010).
73. Li, Y. et al. HIV-1 neutralizing antibodies display dual recognition of the primary and coreceptor binding sites and preferential binding to fully cleaved envelope glycoproteins. *J. Virol.* **86**, 11231–11241 (2012).
74. Sliepen, K. et al. Structure and immunogenicity of a stabilized HIV-1 envelope trimer based on a group-M consensus sequence. *Nat. Commun.* **10**, 2355 (2019).
75. Spindler, M. J. et al. Massively parallel interrogation and mining of natively paired human TCRαβ repertoires. *Nat. Biotechnol.* **38**, 609–619 (2020).
76. Mimitou, E. P. et al. Multiplexed detection of proteins, transcriptomes, clonotypes and CRISPR perturbations in single cells. *Nat. Methods* **16**, 409–412 (2019).
77. Replogle, J. M. et al. Combinatorial single-cell CRISPR screens by direct guide RNA capture and targeted sequencing. *Nat. Biotechnol.* **38**, 954–961 (2020).
78. Sant, S. et al. Single-cell approach to influenza-specific CD8$^+$ T cell receptor repertoires across different age groups, tissues, and following influenza virus infection. *Front. Immunol.* **9**, 1453 (2018).
79. Wills, M. R. et al. The human cytotoxic T-lymphocyte (CTL) response to cytomegalovirus is dominated by structural protein pp65: frequency, specificity, and T-cell receptor usage of pp65-specific CTL. *J. Virol.* **70**, 7569–7579 (1996).
80. Hasegawa, K. et al. Affinity thresholds for membrane fusion triggering by viral glycoproteins. *J. Virol.* **81**, 13149–13157 (2007).
81. Utset, H. A., Guthmiller, J. J. & Wilson, P. C. Bridging the B cell gap: novel technologies to study antigen-specific human B cell responses. *Vaccines* **9**, 711 (2021).
82. Adams, J. J. et al. T cell receptor signaling is limited by docking geometry to peptide-major histocompatibility complex. *Immunity* **35**, 681–693 (2011).
83. Altman, J. D. & Davis, M. M. MHC-peptide tetramers to visualize antigen-specific T cells. *Curr. Protoc. Immunol.* **115**, 17.3.1–17.3.44 (2016).
84. Datlinger, P. et al. Ultra-high throughput single-cell RNA sequencing by combinatorial fluidic indexing. *Nat. Methods* **18**, 635–642 (2021).
85. Hu, Z. et al. A cloning and expression system to probe T-cell receptor specificity and assess functional avidity to neoantigens. *Blood* **132**, 1911–1921 (2018).
86. Oliveira, G. et al. Phenotype, specificity and avidity of antitumour CD8$^+$ T cells in melanoma. *Nature* **596**, 119–125 (2021).
87. Fahad, A. S. et al. Immortalization and functional screening of natively paired human T cell receptor repertoires. *Protein Eng. Des. Sel.* **35**, gzab034 (2022).

## Methods

**Ethics statement.** This work was reviewed and approved by the MIT Institutional Review Board (protocol no. 1801190804).

**Media and cells.** HEK293T cells (ATCC CRL-11268) were cultured in DMEM (ATCC) supplemented with 10% fetal bovine serum (FBS; Atlanta Biologics) and penicillin-streptomycin (pen/strep; Gibco).

Jurkat (ATCC TIB-152), J76 cells and Ramos cells (ATCC CRL-1596) were cultured in RPMI-1640 (ATCC) supplemented with 10% FBS and pen/strep. J76 cells[88] were a gift from M. Heemskerk and M. Davis.

CSS-930 TCR library cells were a gift from D. Johnson[75] and were cultured in RPMI-1640 (Thermo Fisher) supplemented with 10% FBS, pen/strep, 1× non-essential amino acids (Thermo Fisher), 2 mM GlutaMAX (Thermo Fisher) and 1 mM sodium pyruvate (Thermo Fisher).

**Plasmid construction.** The plasmid pHIV-EGFP was a gift from B. Welm and Z. Werb (Addgene plasmid 21373) and pMD2.G and psPAX2 were gifts from D. Trono (Addgene plasmid 12259 and 12260). pLentiCRISPR v2 was a gift from F. Zhang (Addgene plasmid 52961). IL-13 (Uniprot ID P35225) residues 35–146 were cloned into the pHIV backbone containing the Igκ leader peptide, the platelet-derived growth factor receptor (PDGFR) stalk and transmembrane domain (Uniprot ID P09619 residues 449–497) with extracellular linkers listed in Supplementary Fig. 1b. The anti-CD19 scFv FMC63 was cloned into the IgG4 hinge-PDGFR display format in the pMD2 backbone. The anti-CD3 Fab UCHT1 was cloned into the PDGFR stalk-only format. Human CD80 (Uniprot ID P33681, residues 1–273) was cloned into the pMD2 backbone. pMHC single-chain trimers[47] were cloned into either the pMD2 backbone for individual infections or the pHIV backbone for library construction. To generate the pLeAPS backbone, the CMV core promoter was cloned into the pLenti backbone between the polypurine tract and the 3′ LTR, analogous to previous work[60]. For individual infections, SARS-CoV-2 RBD or HIV env CD4bs constructs were cloned into the pMD2 backbone on the PDGFR stalk-only display architecture. Prefusion-stabilized SARS-CoV-2 spike (2P) was cloned into the pMD2 backbone for individual infections. For library assembly, viral constructs described in Supplementary Table 4 were cloned into the pLenti backbone, C-terminally fused to GFP via a P2A motif. For pLeAPS barcodes, mCherry with an 8-nt degenerate sequence was cloned into the pLeAPS backbone downstream of the Ef1α core promoter. The pLeAPS backbone plasmid and pMD2-VSVGmut are available on Addgene.

**Transfection for lentiviral production.** Lentiviruses were prepared by transient transfection of HEK293T cells with linear 25 kDa polyethylenimine (PEI; Santa Cruz Biotechnology) at a 3:1 mass ratio of PEI to DNA. Briefly, DNA and PEI were diluted in Opti-MEM (Thermo Fisher) and mixed to form complexes. Complex formation was allowed to proceed for 15 min at room temperature before dropwise addition to cells. The medium was changed to complete DMEM + 25 mM HEPES after 3–6 h.

For individually targeted or VSVGwt viruses, plasmid mass ratios were 5.6:3:3:1 for transfer plasmid to psPAX2.1 to targeting plasmid (when used) to fusogen plasmid (either VSVGmut or VSVGwt). Targeting plasmids contain expression cassettes for virally displayed ligands in the pMD2 backbone. Total plasmid amounts are indicated in Supplementary Table 5. For LeAPS-based virus production, packaging cells were transfected with a 3:1 mass ratio of psPAX2.1:fusogen.

**Viral purification.** Unconcentrated viruses were filtered (0.45-µm polyethersulfone) and used directly. If needed, they were stored at 4 °C for up to 2 weeks or at −80 °C indefinitely. Concentrated viruses were filtered (0.45-µm polyethersulfone) and concentrated 200× by ultracentrifugation for 90 min at 100,000*g* at 4 °C. The supernatant was discarded and viral pellets were resuspended in Opti-MEM overnight at 4 °C.

**Single lentiviral infections.** Individual infections were carried out with the indicated amounts of virus and cells, in the presence of 8 µg ml$^{-1}$ of either polybrene or diethylaminoethyl-dextran (Sigma-Aldrich). After 24 h, an additional 1× volume of medium was added. Cells were analyzed by flow cytometry 48–72 h after infection for Jurkat or J76 infections and 24–48 h after infection for Ramos cell infections. If cells were only assessed for viral infection, cells were washed once in FACS buffer (PBS + 0.1% BSA and 1 mM EDTA) before analysis on an Accuri C6 or Cytoflex S flow cytometer. In experiments where an additional cell lineage or activation marker such as CD25 or CD69 were assessed, cells were washed once in FACS buffer, stained for 10 min in FACS buffer containing a marker-specific antibody, then washed twice with FACS buffer before analysis for infection (GFP or mCherry) via flow cytometry.

**Human primary T cell activation and transduction.** Peripheral blood mononuclear cells from healthy donors were purified from leukopaks purchased from Stem Cell Technologies using Ficoll-Paque PLUS (GE Healthcare) density gradient centrifugation with SepMate tubes (Stem Cell Technologies) as per manufacturer instructions. Primary CD8$^+$ T cells were isolated using EasySep Human CD8$^+$ T Cell Enrichment kits (Stem Cell Technologies) and cultured in RPMI-1640 (ATCC) supplemented with 10% FBS, 100 U ml$^{-1}$ pen/strep (Corning) and 30 IU ml$^{-1}$ recombinant human IL-2 (R&D Systems). Before transduction with VSVGwt viruses, T cells were activated using a 1:1 ratio of DynaBeads Human T-Activator CD3/CD28 (Thermo Fisher) for 24 h, after which 8 µg ml$^{-1}$ of polybrene (Santa Cruz Biotechnology) and concentrated lentivirus were added to culture at a multiplicity of infection of 1. For targeted viruses, the same protocol was used, but DynaBeads were omitted. For cell proliferation tracking, cells were stained with CellTrace dye (Thermo Fisher) according to the manufacturer's instructions on day 0, before viral infection.

**Antigen–receptor cell line generation.** Lentiviral TCR cassettes were formatted as TCRβ-P2A-TCRα and cloned into the pHIV backbone. TCR KO J76 cells were transduced as described above and sorted based on TCR expression to establish monoclonal cell lines. Ramos BCR cells were established according to the published protocol[70].

**Lentiviral infections in mixed cell populations.** Cells were labeled with CellTrace dyes (Thermo Fisher) according to the manufacturer's instructions, counted and mixed at the indicated ratios. After labeling, viral infections were carried out as indicated above. After 48 h, cells were washed in FACS buffer and analyzed via flow cytometry to examine infection in CellTrace$^+$ versus CellTrace$^-$ cells. After flow cytometry, the selectivity was calculated as follows:

$$\text{Selectivity ratio} = \frac{\%\ \text{target cells transduced}}{\%\ \text{off-target cells transduced}} \times (\text{frequency of target cells})$$

**Generation of LeAPS libraries.** To generate LeAPS production cell lines, 293T cells were transduced as outlined above while being seeded at 20% confluency on either six-well or 96-well plates (Thermo Fisher). For six-well plates, 500 µl each of unconcentrated, VSVGwt-pseudotyped LeAPS barcode and ligand expression viruses were used for infection. For 96-well plates, 50 µl of each virus was used.

For libraries, cells were transduced in duplicate. One of the duplicates was assessed by flow cytometry to determine the proportion of cells transduced with both ligand (GFP$^+$) and LeAPS barcode (mCherry$^+$). Cells from the remaining duplicate were then pooled, with each library member normalized to include an equivalent number of ligand-expressing (GFP$^+$) and LeAPS barcode-transduced (mCherry$^+$) cells for each library member. This pool of cells was then sorted based on GFP and mCherry expression to make library packaging cell pools. To produce lentiviruses, library packaging cells were transfected as described above, but using only psPAX2.1 and pMD2-VSVGmut. For library creation, both transfer plasmids and the pMD2 targeting plasmid were omitted, as they were replaced by the LeAPS packaging line. After 48 h, virus was then collected and concentrated as described above.

**Library screening of monoclonal TCR lines.** Concentrated pMHC library virus (10 µl) was used to transduce one million TCR-expressing J76 cells as outlined above. After infection was confirmed by flow cytometry as described above, genomic DNA was isolated using the PureLink Genomic DNA kit (Thermo Fisher). Barcode inserts were then amplified via 25 cycles of PCR and submitted for Amplicon-EZ analysis by Genewiz. Enrichment was calculated for each barcode as the fraction of total barcode-containing reads divided by the barcode frequency in the packaging cells.

**Library screening of monoclonal Ramos BCR lines.** Unconcentrated viral antigen library (500 µl) was used to transduce one million BCR-expressing Ramos cells as outlined above. After infection was confirmed by flow cytometry, genomic DNA was isolated using the PureLink Genomic DNA kit. Barcode inserts were then amplified via 25 cycles of PCR and submitted for Amplicon-EZ analysis by Genewiz or analyzed via a MiSeq v3 150×150-nt PE Nano kit (MIT BioMicro Center). Enrichment was calculated for each barcode as the fraction of total barcode-containing reads divided by the barcode frequency in the packaging cells.

**Tetramer enrichment of TCR libraries.** To enrich a polyclonal pool of cells from the CSS-930 TCR library for known specificities, 50 million library cells were co-stained with an anti-TRBC antibody (clone IP26) and a pool of HLA-A2 tetramers (National Institutes of Health (NIH) Tetramer Core) presenting the following peptides: NLVPMVATV (NLV), GILGFVFTL (GL9) and GLC, all at 1:800 dilution. Sorting for double-positive cells was performed on a Sony MA900 FACS.

**Library-versus-library TCR–pMHC screen.** Concentrated pMHC library virus (10 µl) was used to transduce two million tetramer-enriched cells. Cells were sorted for infection based on mCherry expression and then submitted for analysis using the 10X Genomics Chromium 5′ v2 V(D)J kit with a barcode construct-specific primer spiked in before droplet encapsulation (Supplementary Methods provide further details). TCR amplicons were prepared and sequenced according to the manufacturer's instructions. Following complementary DNA amplification,

mCherry barcodes were enriched via separate PCRs and sequenced on an Illumina MiSeq (150×150-nt paired end reads). CellRanger V(D)J was used to assign TCR clone identities for each cell. Cell barcodes were used to match TCRs with associated pMHC barcodes counts.

**Antibodies used in flow cytometry.** All antibodies were used at a 1:50 dilution from the stock concentration. Gating strategies shown in Extended Data Fig. 7.

**Software.** Graphs were generated using GraphPad Prism (v.8). Flow cytometry data were analyzed by FlowJo (10.8.0).

**Statistical analysis.** Statistical analyses (calculation of s.d.) were performed using the Prism 8 (GraphPad) software. Sample sizes were not predetermined using statistical methods.

**Reporting Summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this article.

## Data availability

The next-generation sequencing datasets generated during and/or analyzed during the current study are available in the National Center for Biotechnology Information Sequence Read Archive (accession no. PRJNA796957). The UniProt database was used to design constructs in this study (https://www.uniprot.org) for determining the peptide sequence of human IL-13 (accession no. P35225), human PDGFR (accession no. P09619) and human CD80 (accession no. P33681). All data generated or analyzed during this study are included in this published article, the Sequence Read Archive and its supplementary information files. Source data are provided with this paper.

## Code availability

A sample R workspace for extracting barcode counts based on CITEseq[76] is available in the Birnbaum laboratory code repository (https://github.com/birnbaumlab/Dobson-et-al-2022).

## References

88. Rosskopf, S. et al. A Jurkat 76 based triple parameter reporter system to evaluate TCR functions and adoptive T cell strategies. *Oncotarget* **9**, 17608–17619 (2018).

## Acknowledgements

We thank staff at the Koch Institute Robert A. Swanson Biotechnology Center for their technical support, especially staff in the Flow Cytometry Facility, MIT BioMicro Center and High-Throughput Sciences Facility. We thank C. Whittaker of the Barbara K. Ostrom (1978) Bioinformatics and Computing Facility for helpful discussion and aid in implementing single-cell sequencing data analysis pipelines. We thank G. Paradis and S. Levine for many helpful discussions and suggestions. We thank A. Winkler, K. Kravarik and M. Wadsworth for helpful discussions and suggestions. We thank the NIH Tetramer Core Facility (contract no. 75N93020D00005) for providing pMHC tetramer reagents used in this study.

This work was supported by the Koch Institute Frontier Research Program through the Michael (1957) and Inara Erdei Fund and the Casey and Family Foundation Cancer Research Fund, the Packard Foundation, the Damon Runyon Cancer Research Foundation, the Michelson Medical Research Foundation, Schmidt Futures, Pfizer, Inc. and the Department of Defense (W81XWH-18-1-0208) to M.E.B.; the NIH (DP2AI158126 to M.E.B. and R01AI137057, R01AI153098 to D.L.); a National Science Foundation Graduate Research Fellowship and a Siebel Scholarship to C.S.D.; a Canadian Institutes of Health Research Doctoral Foreign Study Award to S.G.; a graduate research fellowship from the Ludwig Center at MIT's Koch Institute to E.J.K.; and a Medical Scientist Training Program grant (T32 GM007753) from the National Institute of General Medical Sciences at Harvard Medical School to B.E.S. M.E.B and S.K.D. were funded by a Technology Impact Award from the Cancer Research Institute and are Pew-Stewart Scholars for Cancer Research. Core facilities in the Koch Institute are partially supported by Cancer Center Support (core) grant P30-CA14051 from the National Cancer Institute. This research is additionally supported by the National Research Foundation, Prime Minister's Office, Singapore under its Campus for Research Excellence and Technological Enterprise program, through Singapore MIT Alliance for Research and Technology: Critical Analytics for Manufacturing Personalized-Medicine Inter-Disciplinary Research Group.

## Author contributions

C.S.D. and M.E.B. conceived the RAPTR system. C.S.D., A.R., S.G., B.E.S., E.J.K., J.D., L.R. and V.O. conducted all experiments. C.S.D., A.R., S.G., B.E.S., E.J.K. and J.D. analyzed data. D.L., M.D., S.K.D. and M.E.B. supervised research. C.S.D. and M.E.B. wrote the manuscript. All authors edited the manuscript.

## Competing interests

The lentiviral targeting approach in this manuscript is the subject of US patent applications with M.E.B., S.G. and C.S.D. as co-inventors. M.E.B. is a founder, consultant and equity holder of Viralogic Therapeutics and Abata Therapeutics, is an equity holder in 3T Biosciences, has received consulting fees from Repertoire Immune Medicines and has received research funding from Pfizer that partially funded this work. S.K.D. received unrelated research funding from Novartis Pharmaceuticals, Eli Lilly and Company and Bristol-Myers Squibb and is a founder, science advisory board member and equity holder in Kojin. M.D. is a science advisory board member for Neoleukin, has research funding from Novartis and Eli Lilly, has received consulting fees from Tillotts Pharma, ORIC Pharmaceuticals, Partner Therapeutics, SQZ Biotech, AzurRx, Mallinckrodt Pharmaceuticals and Moderna. The remaining authors declare no competing interests.

## Additional information

**Extended data** is available for this paper at https://doi.org/10.1038/s41592-022-01436-z.

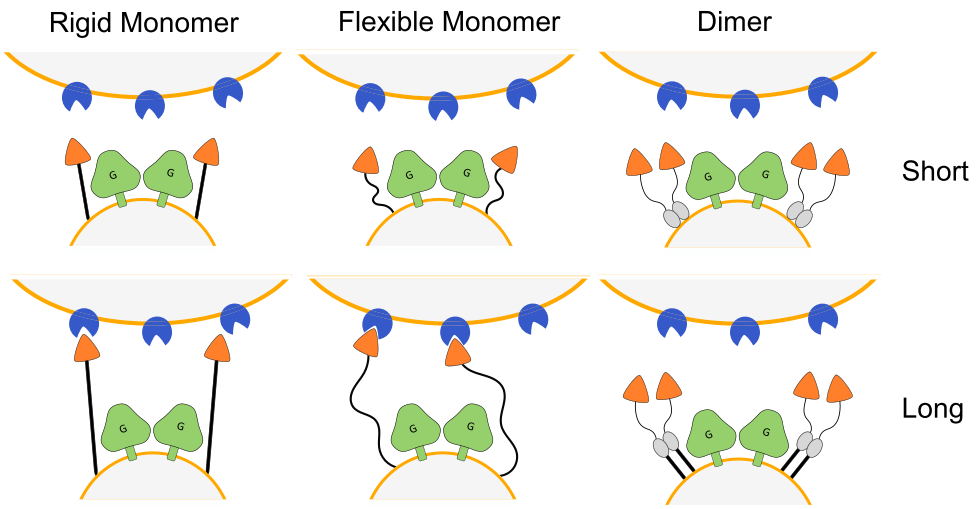
**Supplementary information** The online version contains supplementary material available at https://doi.org/10.1038/s41592-022-01436-z.

**Correspondence and requests for materials** should be addressed to Michael E. Birnbaum.

**Peer review information** *Nature Methods* thanks David Bending and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available. Madhura Mukhopadhyay was the primary editor on this article and managed its editorial process and peer review in collaboration with the rest of the editorial team.

**Reprints and permissions information** is available at www.nature.com/reprints.

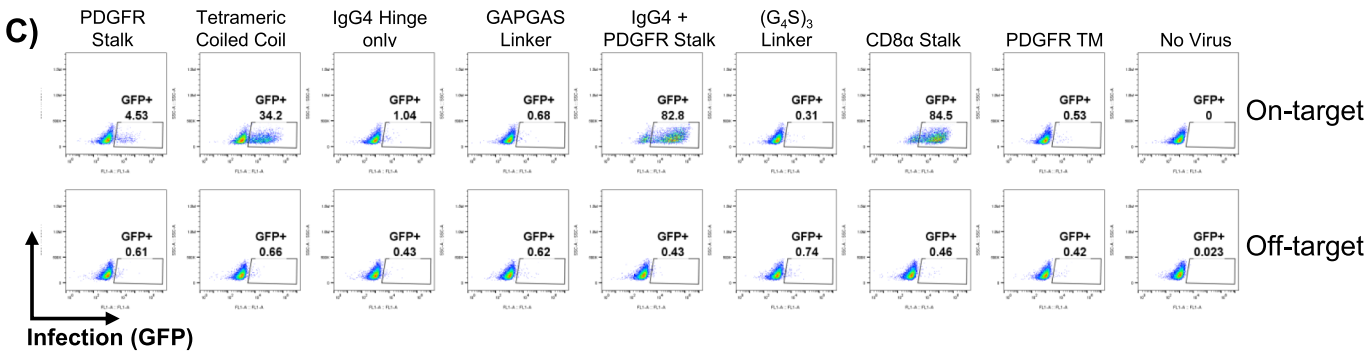
A)



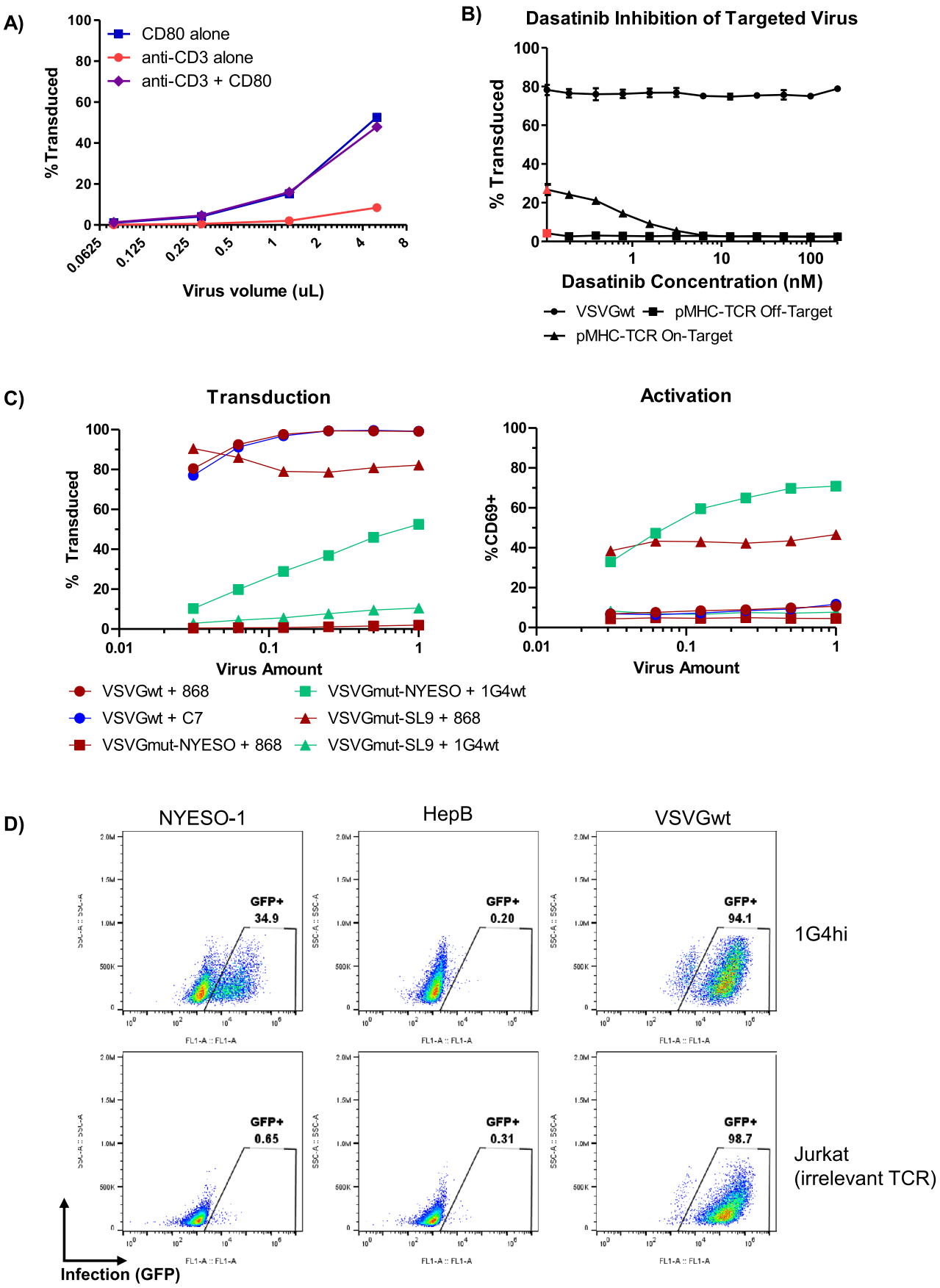

B)

| Architecture Type | Name | Unconcentrated Titer ($10^6$ TU/mL) |
|---|---|---|
| Short, rigid monomer | PDGFR transmembrane | 0.01 |
| Short, flexible monomer | GAPGAS linker | <0.01 |
| Short dimer | IgG4 hinge only | 0.01 |
| Extended, rigid monomer | PDGFR transmembrane and stalk | 0.046 |
| Extended, flexible monomer | $(G_4S)_3$ linker | <0.01 |
| Extended dimer | IgG4 hinge + PDGFR stalk | 0.83 |
| Extended dimer | CD8α stalk | 0.85 |
| Tetramer | Tetrameric coiled coil | 0.35 |

C)

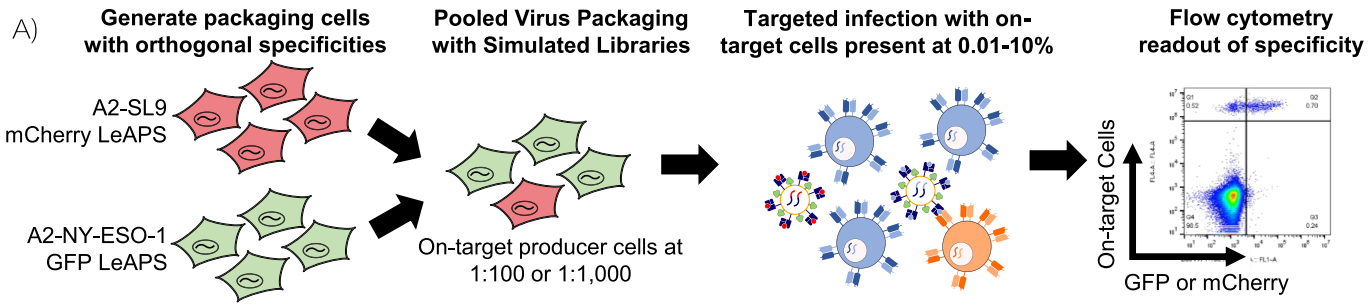



**Extended Data Fig. 1 | IL-13 Surface Architecture Development. a**, Schematic diagram of the different targeting ligand architectures used. **b**, Summary table with details of each individual construct and titer achieved. **c**, Flow cytometry plots showing infection of IL-13Rα1-expressing Jurkat cells (top, on-target) and parental Jurkat cells which don't express the receptor (bottom, off-target).

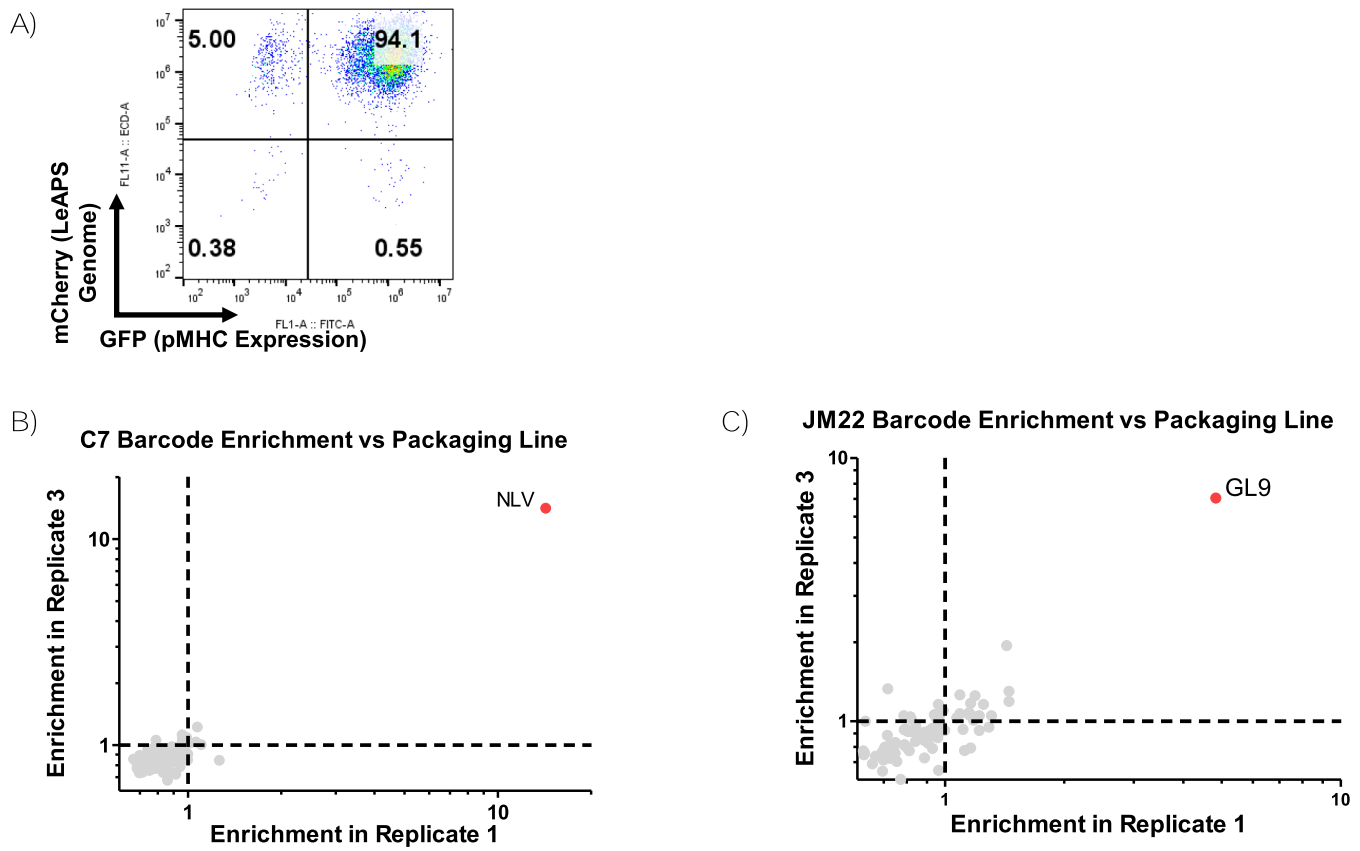



**Extended Data Fig. 2 | See next page for caption.**

**Extended Data Fig. 2 | Viral entry via the TCR complex and costimulatory receptors, and inhibition by dasatinib. a**, $1\times10^5$ Jurkat cells were infected with the indicated amount of VSVGmut-pseudotyped virus displaying either anti-CD3 Fab, CD80, or a combination of the two; data shown as mean + SD of three biological replicates **b**, $5\times10^4$ J76 cells expressing the 868 TCR were pre-incubated with the indicated concentrations of dasatinib for 2 hours to inhibit TCR signaling prior to virus addition (red dots indicate no dasatinib condition). Next, 5 μL of virus displaying either VSVGmut and HLA-A2-SL9 (on-target), VSVGmut and HLA-A2-NY-ESO-1 (off-target), or VSVGwt were added. Infection was measured 48 h post-infection. Results represent the mean and standard deviation of 3 biological replicates. **c**, The indicated amount of concentrated virus was added to each TCR-expressing J76 cell (100,000 cells/well in a 96-well plate). After 24 h, transduction was measured by fluorescent reporter (left) and activation was measured by CD69 upregulation (right) relative to unstimulated cells. **d**, 500,000 1G4hi-expressing J.RT3 cells or parental Jurkat cells were infected with 500 μL of unconcentrated VSVGmut-pseudotyped virus displaying HLA-A2 presenting either NYESO-1 or an irrelevant hepatitis B peptide (SLYAVSPSV), or VSVGwt-pseudotyped virus.

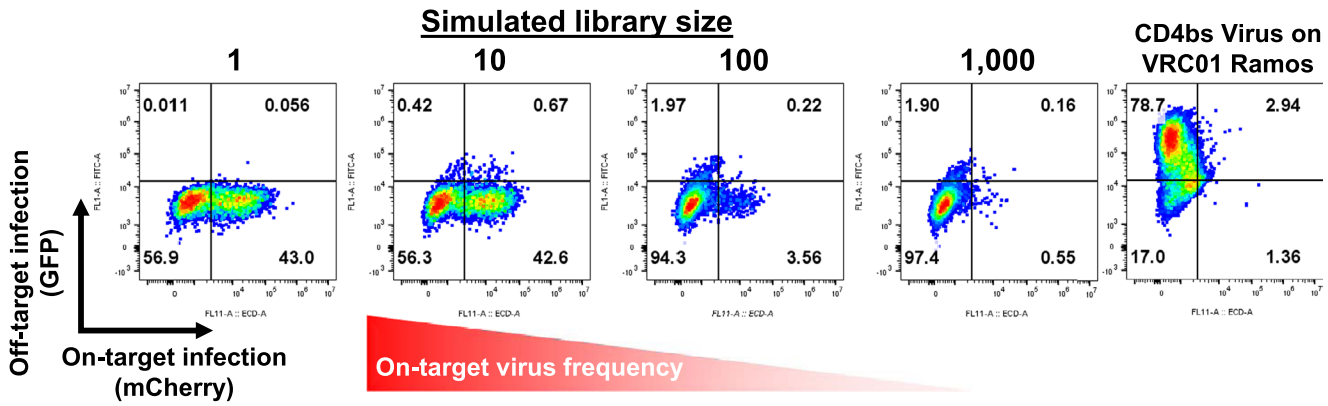



**Extended Data Fig. 3 | Schematic diagram of LeAPS proof of concept experiments.** Packaging cells transduced with (i) SL9 presented on HLA-A2 and LeAPS-mCherry and (ii) NY-ESO-1 presented on HLA-A2 and LeAPS-GFP were mixed at either 1:100 or 1:1000 for each to represent libraries containing 100 constructs and 1,000 constructs, respectively. These mixtures were then used as packaging lines, transfected with helper plasmids to produce viruses, and used to transduce mixtures of cell tracking dye-labeled J76 cells expressing either 868 TCR or 1G4wt TCR at defined ratios.




**Extended Data Fig. 4 | Quality control of LeAPS packaging line and reproducibility of infection.** A, LeAPS-barcode expression (mCherry) and pMHC expression (GFP) of the RAPTR 96-member packaging line after 9⁻passages post-sorting. B, Correlation between infection replicates of the C7 TCR cell line. C, Correlation between infection replicates of the JM22 TCR cell line.

A)




B)

**CR3022 Cell Dilution Selectivity**

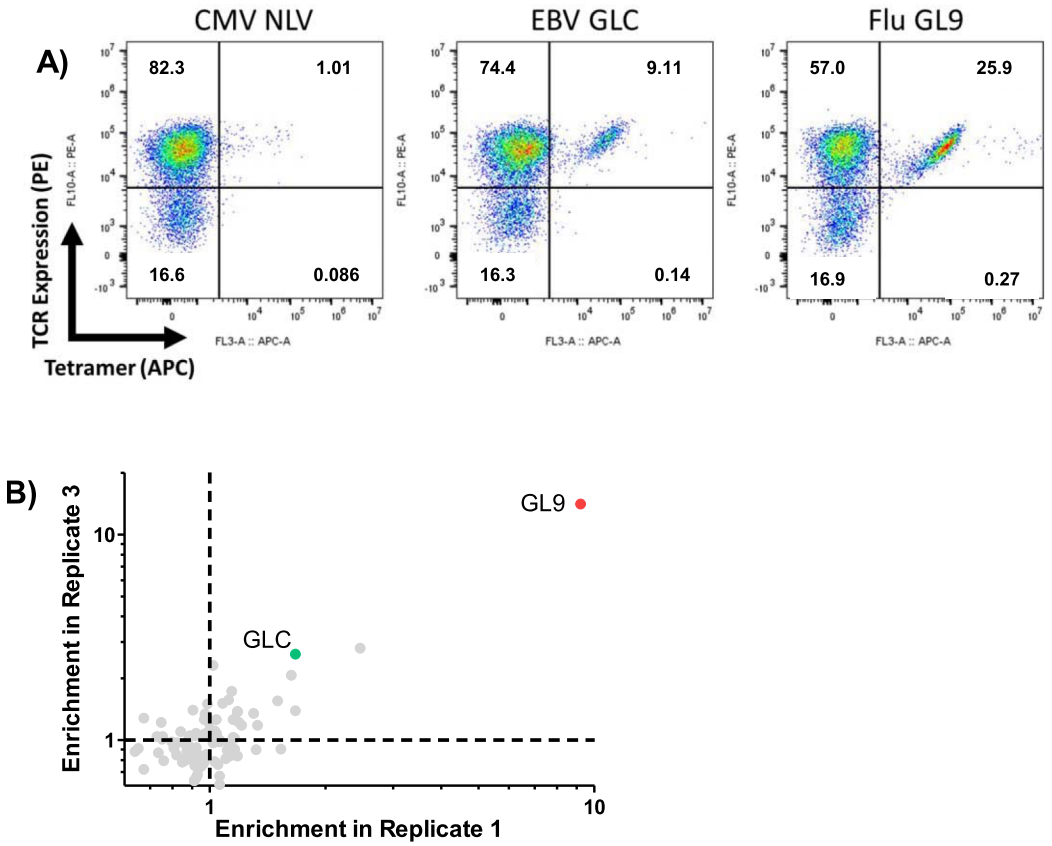
| Frequency | Selectivity |
|---|---|
| 0.1 | 1092 |
| 0.01 | 2917 |
| 0.001 | 1038 |
| 0.0001 | 224 |
| 0.00001 | 262 |
| 0.000001 | 16 |

C)

**Enrichment Ratios for SARS-CoV-2 Spike/RBD-displaying Viruses Mixed with HIV Env-displaying Viruses**

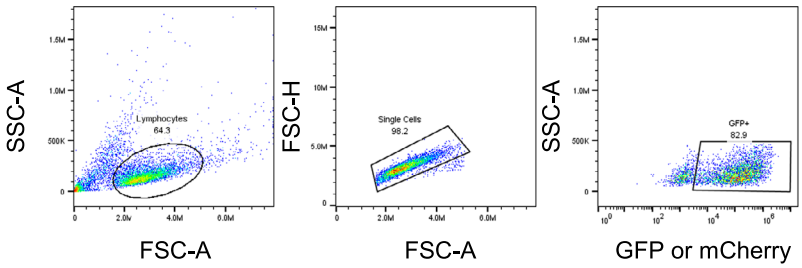
| Name | Target Virus Frequency | Enrichment | Max Enrichment Possible |
|---|---|---|---|
| RBD 1 | 1 | 1.00 | 1 |
| RBD 10 | 0.1 | 9.99 | 10 |
| RBD 100 | 0.01 | 98.38 | 100 |
| RBD 1000 | 0.001 | 738.58 | 1000 |
| Spike 1 | 1 | 1.00 | 1 |
| Spike 10 | 0.1 | 9.90 | 10 |
| Spike 100 | 0.01 | 65.74 | 100 |
| Spike 1000 | 0.001 | 272.03 | 1000 |

**Extended Data Fig. 5 | Signal to noise in virus mixtures containing SARS-CoV-2 full spike and HIV env. a**, Infection of CR3022 BCR-expressing Ramos cells with mixtures of SARS-CoV-2 prefusion-stabilized spike (on-target, expressing mCherry) and HIV env (off-target, expressing GFP) viruses; variant library size indicates the ratio of off-target to on-target virus present. Infection of VRC01-expressing Ramos cells with HIV env hybrid pseudotyped viruses (right) demonstrates that the viral particles remain functional. **b**, Quantification of selectivity ratios for the SARS-CoV-2 RBD infections of CR3022 Ramos presented in Fig. 5c. Selectivity ratio is calculated as the proportion of on-target cells infected divided by the proportion of off-target cells infected times the frequency of on-target cells (see Methods). **c**, Quantification of on-target enrichment for mixtures of HIV env and either SARS-CoV-2 RBD (presented in Fig. 5d) or full spike (presented in Supplementary Fig. 5a). Enrichment is calculated by fraction of cells expressing on-target signal divided by fraction of cells expressing off-target signal times on-target virus frequency.

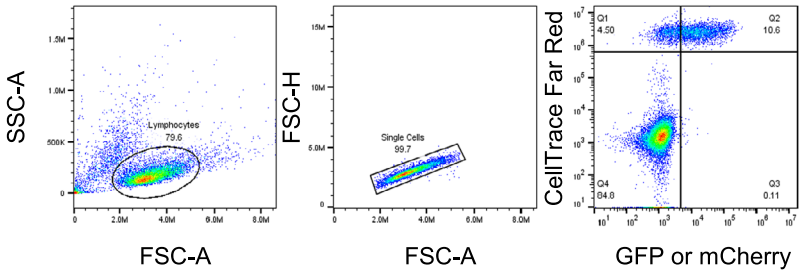



**Extended Data Fig. 6 | Quality control and reproducibility for the library vs library experiment. a)** Tetramer staining for each of the pooled specificities used for enrichment. **b)** Correlation between infection replicates in bulk sequencing data for the library vs library experiment.

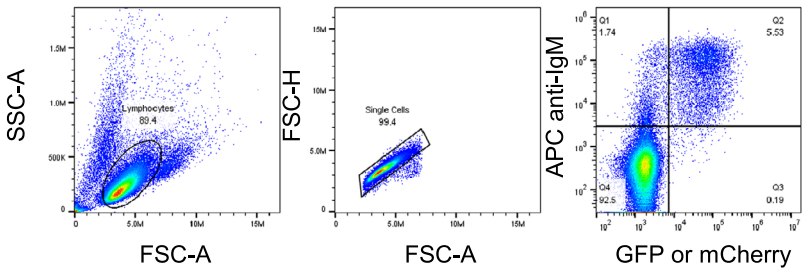
**A) Infection of Jurkat and J76 cell lines**



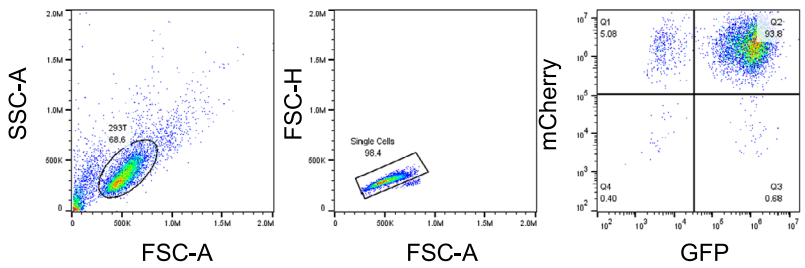

**B) Infection of mixed J76 or Jurkat cell lines labeled with tracking dye**



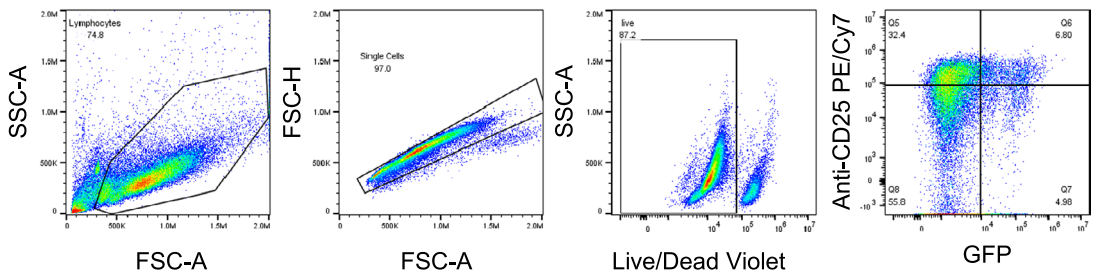

**C) Infection of mixed Ramos cell lines**




**D) 293T packaging line quality control**




**E) Primary T cell infection and activation**




**Extended Data Fig. 7 | Gating strategies used throughout manuscript.** Flow cytometry gating strategies for experiments described.

nature portfolio

Corresponding author(s): Birnbaum, Michael Edward

Last updated by author(s): Sep 16, 2021

# Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

## Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

| n/a | Confirmed |
|---|---|
| ☐ | ☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement |
| ☐ | ☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly |
| ☒ | ☐ The statistical test(s) used AND whether they are one- or two-sided<br>*Only common tests should be described solely by name; describe more complex techniques in the Methods section.* |
| ☒ | ☐ A description of all covariates tested |
| ☒ | ☐ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons |
| ☐ | ☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals) |
| ☒ | ☐ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted<br>*Give P values as exact values whenever suitable.* |
| ☒ | ☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings |
| ☒ | ☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes |
| ☒ | ☐ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated |

*Our web collection on statistics for biologists contains articles on many of the points above.*

## Software and code

Policy information about availability of computer code

| | |
|---|---|
| Data collection | MiSeq 3.1, CellRanger VDJ (3.0) |
| Data analysis | FlowJo (10.8.0), Graphpad Prism (v8). Custom code can be found at https://github.com/birnbaumlab |

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

## Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

The next-generation sequencing datasets generated during and/or analyzed during the current study are available in the NCBI sequence read archive (Accession Number PRJNA796957). All data generated or analyzed during this study are included in this published article, the sequence read archive, and its supplementary information files.

# Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

# Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

| | |
|---|---|
| Sample size | No sample-size calculation was performed. For flow cytometry and sequencing experiments, the minimum number of cells analyzed was determined as ten times the inverse of the frequency of the rarest population (e.g. 10,000 cells for a population expected at 0.1%). |
| Data exclusions | No data were excluded |
| Replication | All attempts at replication were successful with replicates as marked (typically 2-3x) |
| Randomization | This was not relevant to this study as no clinical or animal studies were performed |
| Blinding | This was not relevant to this study as no clinical or animal studies were performed |

# Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

## Materials & experimental systems

| n/a | Involved in the study |
|---|---|
| ☐ | ☒ Antibodies |
| ☐ | ☒ Eukaryotic cell lines |
| ☒ | ☐ Palaeontology and archaeology |
| ☒ | ☐ Animals and other organisms |
| ☒ | ☐ Human research participants |
| ☒ | ☐ Clinical data |
| ☒ | ☐ Dual use research of concern |

## Methods

| n/a | Involved in the study |
|---|---|
| ☒ | ☐ ChIP-seq |
| ☐ | ☒ Flow cytometry |
| ☒ | ☐ MRI-based neuroimaging |

## Antibodies

| | |
|---|---|
| Antibodies used | All antibodies were obtained from BioLegend unless indicated otherwise: anti-human CD69 (clone FN50, cat. 310910, lot B258744); anti-human TCRbeta (clone IP26, cat. 306708, lot B259839); anti-human EGFR (clone AY13, cat. 352911, lot B321041); anti-human HLA-A2 (clone BB7.2, cat. 343308, lot B277594); anti-human IgM (clone MHM-88, cat. 314510, lot B294323) |
| Validation | Antibodies were validated by commercial vendors using flow cytometry of antibody-stained cells and this information is available on their websites. |

## Eukaryotic cell lines

Policy information about cell lines

| | |
|---|---|
| Cell line source(s) | Wild type Jurkat cells [clone: E6-1, from ATCC TIB-152]<br>TCRbeta deficient Jurkat cells [clone: J.RT3-T3.5, from ATCC TIB-153]<br>HEK293T/17 cells (ATCC CRL-11268)<br>Ramos Cells (ATCC CRL-1596)<br>J76 Cells (a gift from Mark Davis) |
| Authentication | J76 cells were validated by staining for surface TCR beta chain expression. Cell lines have also been authenticated by morphology checking and through use in other experiments. |
| Mycoplasma contamination | Tested negative for mycoplasma contamination |

| | |
|---|---|
| Commonly misidentified lines (See ICLAC register) | No commonly misidentified cell lines were used. |

# Flow Cytometry

## Plots

Confirm that:

☒ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☒ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☒ All plots are contour plots with outliers or pseudocolor plots.

☒ A numerical value for number of cells or percentage (with statistics) is provided.

## Methodology

| | |
|---|---|
| Sample preparation | Tissue culture samples were resuspended to single cells for staining. |
| Instrument | Analysis: data were collected on either the Accuri C6 or the BD Cytoflex S. Sorting: Cells were sorted using the Sony MA900 |
| Software | Manual gating was used in Flowjo (10.8.0) for analysis |
| Cell population abundance | Sorted cells were directly analyzed by single-cell sequencing. Any cells maintained in culture were subsequently verified for >95% purity by flow cytometry analysis for the marker(s) used to sort originally. |
| Gating strategy | The gating strategies used are provided in Supplementary Figure 7 and all applicable figures. In brief, gating strategy was: FSC-A vs SSC-A to capture the cell population of interest --> FSC-A vs FSC-H to identify single cells --> Marker(s) of interest |

☒ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.