nature portfolio

Peer Review File

Identification of broadly tumor-reactive γδ TCRs from multiple myeloma

Corresponding Author: Dr Tak Mak

Any redactions in this file are there to maintain patient confidentiality, the confidentiality of unpublished data, or to remove third-party material.

This file contains all reviewer reports in order by version, followed by all author rebuttals in order by version.

Version 0:

Reviewer comments:

Referee #1

(Remarks to the Author)
Summary
In this manuscript by St. Paul et al., the authors explore if γδ T cells use their γδ TCR to recognize and kill cancer cells, and what defines tumor-reactive γδ T cells. The authors start by studying multiple myeloma (MM) material of (a limited set of) patients, obtained before and during treatment with the ADC belamaf in a clinical trial. Through a combination of scRNA-seq, CITE-seq and TCR-seq they find that the abundance of γδ T cells was associated with better outcomes. As this suggested clinical relevance of γδ T cells in this setting, the sequencing data were subsequently used to identify the paired γδ TCR sequences in the cohort and a large set was cloned into Jurkat cells, of which a selection showed tumor-reactivity against MM cell lines, as well as cross-reactivity against some cell lines of other cancers. By mapping the reactive TCRs on the initial single cell dataset, the authors then classified the γδ T cells as MM reactive, nonreactive, or borderline. Machine learning was then used to identify patterns in the scRNA/CITE-seq profiles that were associated with tumor reactivity. The resulting tool (coined “PreGame”) was benchmarked against tools developed for ab T cells and validated on unseen data including 14 additional MM samples and 8 samples of other (non-MM) cancer types. By studying feature importance in PreGame, as well as differential expression analyses of reactive vs nonreactive clones, the authors performed a phenotypic comparison (looking at RNA/protein expression and γ/δ chain usage) of MM-reactive vs nonreactive γδ T cells. Next, the authors revisited the trial data and studied the clonal dynamics of γδ T cells during belamaf treatment, showing the expansion of tumor-reactive clones was associated with favorable outcomes. Having screened many γδ TCRs for reactivity against a collection of cancer cell lines, the authors were then able to select several tumor-reactive clones and set out to study which antigens they recognized. To this end, they started with testing if tumor recognition was dependent on a few key proteins: B2M (through knockout), BTN3A1 (through knockout) and CD1D (through overexpression). This showed that the two Vγ9 TCRs both recognized a B2M-dependent target other than CD1D, and follow-up experiments with 11 additional tumor-reactive Vγ9 TCRs showed that roughly half of them also recognized B2M-dependent targets. Because of its broad tumor reactivity, they then selected “TCR65” (a Vδ3Vγ9 TCR) and performed HLA knockout experiments and co-cultures with APCs expressing single HLA alleles, to conclude that TCR65 recognized a HLA-C epitope. By mapping sequence differences among differentially recognized HLA-C allotypes, coupled with targeted mutation, the authors could identify the V103L codon as essential for recognition. Finally, the authors demonstrated that expression of KIR3DL1 (functioning together with CD8a as coreceptor) could overcome autoreactivity of TCR65 expressing cells, suggesting novel avenues to create cellular therapies with a “logic switch” to specifically recognize and kill cells that downregulate HLA-A and HLA-B, but at the same time (over)express HLA-C.

General remarks
This is an exciting and exceptionally data-rich manuscript which addresses multiple highly relevant questions in the field of cancer immunology, with clear translational implications. Especially for Vδ1 and Vδ3 cells, which have innate-like phenotypes, it has been a longstanding question whether these cells carry tumor-reactive γδ TCRs. With strong data, the authors proof that (especially in the context of MM) this is the case for at least a fraction of these cells, which I find a very intriguing, also given its potential relevance for “off-the-shelf” γδ T cell therapies. The identification of a “logic gate” that could be engineered by overexpressing a γδ TCR plus KIR3DL1 and CD8 to create a cellular product that specifically targets cells with HLA allele imbalances is also interesting, although I do think the biological interpretation of these data is somewhat

overstated (see below). In terms of biology, it would greatly increase the relevance of the findings if the authors would test if the γδ TCR is not only sufficient but also essential for tumor recognition by γδ T cells. It has long been postulated that the large repertoire of activating innate immune receptors expressed by γδ T cells play dominant roles in tumor recognition, but the findings of St. Paul et al raise the question what their relative importance really is? Regarding PreGame: this tool is interesting, but I do have important concerns regarding its accuracy and generalizability across cancer types and γδ T cell subsets, as well as its added value compared to simpler alternatives. Related to this, a concern of this manuscript is that some described phenotypic characteristics of tumor-reactive γδ T cells in MM contrast key observations made by others in other cancers, posing questions regarding the generalizability to other disease contexts. Also in other instances, there are findings which contrast/challenge well-established knowledge (e.g. Vδ2Vγ9 T cells which are described by the authors to act independently of BTN3A1); such unexpected claims need particularly strong data but are not always double checked with alternative approaches or being put into perspective. Furthermore, while multiple exciting findings are broadly relevant, it could also be noted that the relevance of the clinical cohorts is limited by the fact that these center around the drug belamaf whose regulatory approval has been retracted. A detailed point-by-point review below.

Major points
- The relevance of the study would be much greater if the authors could robustly test if the γδ TCR is not only sufficient but also essential for tumor recognition by γδ T cells. Especially since γδ T cells are known to express various activating innate immune receptors, which also represent some of the main features that PreGame uses to identify tumor-reactive cells (NKG2C, CD94, NCR3, FCGR3A/CD16, NKG7).
- Related to the above, it is not entirely clear which fraction of experimentally tested γδ TCRs were found to be tumor reactive? And are γδ T cells from all other clonotypes not tumor reactive, or are they tumor reactive via other (innate) receptors?
- It is also important to better understand the diversity of the tested γδ TCR repertoire. Looking at the cell line reactivity data, it seems that there are clear clusters of γδ TCRs recognizing exactly the same target cell lines suggesting that there may be very low diversity.
- The authors present PreGame as a tool to identify tumor-reactive γδ T cells across γδ T cell subsets and tumor types, but I do have concerns regarding its generalizability:
o Does it work within Vd1/2/3 separately? A quantitative analysis is needed. Especially as TRDV1 is one of the strongest features in the model, suggesting that part of its performance simply comes from separating Vd1 cells vs Vd2 cells (whose are known to show much higher tumor reactivity than Vd2 cells). This is an essential point.
o Does PreGame really work for gdT cells in other cancers? Especially in cancer types where gdT cells are known to play a significant role, such as MSI CRC (de Vries et al, 2023), pediatric neuroblastoma (Castenmillen et al, 2025), etc. The numbers of non-MM malignancies are very low and more data is needed to support this claim. Especially as some findings suggest that tumor-reactive gdT cells in MM have very different phenotypes than what has been observed in other cancers (e.g. high PD-1 expression clearly marks tumor-reactive gdT cells in MSI CRC, while the authors find that tumor-reactivity is largely confined within the PD-1 low population).
o Related to the above, how is it explained that previously described RNA expression markers of tumor-reactive gdT cells in solid tumors are not always overexpressed in PreGame-high cells in MM?
- More broadly, generalizability beyond MM is not always robustly supported by data. For such cases it is important to either provide supporting data or narrow the claims.
- To understand the utility of PreGame, it is essential to show quantitative comparisons to simpler approaches. How does the “hit rate” of PreGame compare flow sorting based on established protein-level markers (e.g. ki67, Vδ1, CD39, 4-1BB, PD-1)? Or a simple combination such as ki67+ Vδ1 cells?
- The validation of PreGame vs competitors is not entirely fair: 10-fold cross validation (PreGame) vs completely unseen (competitor algorithms). This should be addressed by performing the comparison on the unseen set. Of note, a fair comparison of PreGame with simpler approaches (as suggested above) should also be performed on data of patients that were completely unseen during training.
- “We observed a significant positive correlation between γδ TCR repertoire diversity and the fraction of tumor-reactive γδ TCRs in circulation, suggesting that diversity can serve as a proxy for measuring the proportion of tumor-reactive γδ TCRs (Extended Data Fig. 8g,h).” This seems counter-intuitive as tumor-reactivity would be expected to drive clonal expansion and reduced diversity. Could this be confounded by higher γδ T cell levels in patients with tumor-reactive γδ TCRs, which in turn results in the detection of more clonotypes? More broadly, one could argue that comparing measures of diversity and clonal expansion between responders and nonresponders cannot be reliably performed because γδ T cells were virtually absent in nonresponders.
- To test if the cloned γδ TCRs recognize phosphoantigens presented by BTN3A1, the authors knockout BTN3A1 and found no effect (extended data fig 9a). In cases of Vδ2Vγ9 TCRs this is surprising as it is well established that Vδ2Vγ9 TCRs recognize phosphoantigens in the context of butyrophilins. To confirm that phosphoantigens are indeed irrelevant for these TCRs, it would be important to add the phosphoantigen (e.g. HMBPP or IPP) to the culture, and repeat similar analyses (KO & phosphoantigen addition) in a second model system (ideally one that previously has been used to demonstrate BTN3A1-based activation of Vδ2Vγ9 T cells, to proof the claim that these T cells are fundamentally different).
- In general, it would be important to provide evidence that the knockouts (BTN3A1, B2M, HLA’s) and overexpression (CD1D) perturbations were successful.
- Fig 5: the authors seem to overinterpret these data, as they mainly show data suggestive of different γδ T cell dynamics in responders vs nonresponders, but do not provide a predictive biomarker.
- Regarding Fig 5e the authors mention strong correlation between tumor burden and clonal expansion of a reactive γδ T cell clone. However, it seems counterintuitive that the tumor burden is already 0 at timepoint C2D1 and the γδ T expansion only occurs two timepoints later (C12D1). How is this possible? Also, the conclusion that this patient achieved long term tumor control due to this γδ T cell clone seems overstated.
- The clinical cohort has a very low number of nonresponding MM pts, limiting the level of evidence, which should be

acknowledged in the claims.
- Belamaf was granted accelerated regulatory approval in 2020 but was withdrawn from the market in 2022 when the pivotal trial missed its primary endpoint. This limits the relevance of the clinical biomarker findings (Figure 1 and Figure 5).
- The observation that Vγ9 TCRs may recognize B2M-dependent targets is interesting. However, it is unclear why in figure 6A the authors classified TCR v10 as a B2M-dependent TCR, as the fold activation change with B2M knockout is ~0? And related: for this TCR the reactivity strongly increased with CD1D overexpression, but since CD1D is B2M-dependent it seems counterintuitive that B2M knockout has no effect? How could these seemingly contradicting observations be explained?
- I do believe that the authors overinterpret their data of the proposed "logic switch" of the γδ TCR + KIR3DL1. These data are intriguing as it shows that such a logic switch could be created by overexpressing these receptors in Jurkat cells, which has potential implications for the design of cellular therapies. However, there is no proof showing that this logic gate is actually being used by real γδ T cells for tumor recognition of "missing self" sensing. Indeed, γδ T cells express a whole repertoire of both activating and inhibiting HLA-specific receptors (e.g. inhibitory KIRs, activating KIRs, NKG2A, NKG2C, LILRBs, etc), all with different affinities for different HLA alleles. Hence, the interactions of γδ T cells with the HLA's of tumor cells are highly complex and the net effect of different HLA balances is unclear. To proof that this "logic gate" is a mechanism defining tumor-reactivity of γδ T cells, strong functional data with nonengineered/patient-derived γδ T cells are needed. Alternatively, the authors could narrow the claim and put the emphasis on the ability to synthesize logic gates for cellular therapies with these receptors, which is exciting on itself but is a narrower biological interpretation.

Minor points:
- More patient characteristics and details of the clinical cohorts are necessary.
- Figure 1f – the significance in not readable in the figure, similar as extended figure 2b-c
- In Fig 2e it is unclear to what extent the tumor reactivity is driven by CD1D overexpression by the lines.
- In Fig 3c-d it is confusing that different scales have been used for the PreGame score.
- When defining tumor reactivity, why did the authors take a cut-off of 13%? This cut-off should be clarified. If up until 13% is background, then an increase from 13% to 15% may not be major recognition by the TCR clone.
- In legends showing cytotoxicity assays, it would help to precise the readout.
- In general, the figure legends are not always complete and extensive enough to understand the figure – for example in figure legend 4 is never mentioned what type of data the authors are showing (from the text it seems to be CITE-seq but that is not shown in the figure/legend itself).
- Colors of figure 4f are unclear.

Reviewed by:
Joris van de Haar, MD, PhD
Department of Molecular Oncology and Immunology
The Netherlands Cancer Institute

(Remarks on code availability)

Referee #2

(Remarks to the Author)
In this manuscript, St Paul et al. analyze a Multiple Myeloma (MM) patient cohort from which they isolate bone marrow–infiltrating T cells to perform single-cell TCR and RNA sequencing, together with CITE sequencing. The authors experimentally test the reactivity of selected γδ TCRs by cloning them into Nur77-GFP Jurkat cells. The experimentally validated clones are then used to train a machine learning algorithm that predicts tumor reactivity of γδ T cells based on single-cell CITE-seq data. The trained model is subsequently applied to clonotypes that were not experimentally tested, with validation experiments demonstrating good predictive performance. Moreover, the authors use this algorithm to suggest that expansion of tumor-reactive γδ T cells serves as a biomarker of therapeutic response in MM patients treated with belantamab mafadotin. Finally, they identify a γδ TCR epitope in HLA-C and KIR3DL1 as an important inhibitory receptor that prevents recognition of healthy tissue.

Overall, this manuscript is innovative, methodologically strong, and presents original and clinically relevant findings, with significance extending beyond the γδ T-cell community to the broader cancer immunotherapy field. The work is elegant and well-executed, employing cutting-edge techniques and valuable clinical samples. However, several key conclusions are insufficiently supported or not clearly explained, and additional analyses, improved data interpretation, figure reorganization, and experimental validation are needed to strengthen the manuscript's impact before it can be considered for publication.

SPECIFIC ISSUES

Figure 1
Claim: γδ T cells mediate durable responses to ADC combination therapy
• No direct evidence of causality is provided. To support this claim, mechanistic evidence demonstrating mediation is required.
• The purpose of Fig. 1c is unclear. Extended Fig. 1d appears more informative in showing that γδ T cells ($GzmB^+$ and $GzmK^+$) are more frequent and associated with response. The authors should clarify how the frequency of $GzmB^+$ and $GzmK^+$ γδ T cells correlates with PFS, and include $GzmB^+$ $CD4^+$ T cells for comparison.
• Lines 90–92: "Comparison of γδ TCR repertoires between baseline and C2D1 revealed an increased proportion of

clonotypes that expanded on treatment compared to the more polyclonal αβ TCR response, suggesting recognition of tumor antigens."
This statement is unclear and potentially misleading. Are the authors suggesting that only γδ T cells recognize tumor antigens because αβ TCRs show higher diversity upon treatment? The reasoning should be clarified, as higher αβ T cell diversity does not imply absence of tumor recognition.
• Lines 95–97: The authors state, "an increase in γδ, but not αβ, TCR diversity at C2D1 was significantly associated with longer PFS." It should be clarified whether this refers to an increase from baseline or higher TCR diversity within the cohort.
• Additional analyses are requested:
o Compare TCR diversity (αβ and γδ) at C2D1 vs baseline in blood, using similar plots as Fig. 1f.
o Provide Kaplan–Meier curves for bone marrow, analogous to Fig. 1g/h.
o Examine whether higher γδ TCR diversity in bone marrow correlates with poorer prognosis, and if so, whether this relates to limited TCR reactivity or expansion capacity.

Figure 2
Claim: γδ T cells are tumor-specific and recognize broad tumor antigens via their TCRs.
• The data presented do not directly demonstrate tumor specificity; those results are in Extended Fig. 5e, which should be moved to the main figure.
• In Fig. 2c/d, the rationale for comparing $GzmB^+$ and $GzmK^+$ subsets is unclear. Are the authors proposing that reactivity is higher in $GzmB^+$ γδ T cells due to clonal expansion? If so, reduced diversity would be expected, which is not shown. This point is further weakened by Fig. 4a/b, which demonstrates that tumor-reactive clones exist in both subsets.
• Clarify whether all tumor-reactive γδ TCRs are private or if some are shared between patients.
• The cutoff for tumor reactivity in the Jurkat system is inconsistently defined:
o Line 1062: "We defined a tumor-reactive γδ TCR as causing a ≥ 13% increase in $\%GFP^+$ cells."
o Lines 768–769: > 15% for strongly reactive vs 0–13% for non-reactive.
o Line 505: "Only the subset of γδ TCRs eliciting a response (> 13% activation strength)".
These thresholds must be harmonized and justified.
• In Fig. 2g, TCRs 95 and 100 are labeled tumor-reactive despite < 13% GFP increase. Clarify or correct this inconsistency.
• In Fig. 2i, biological replicates should be presented instead of technical replicates. The CD1C negative control should be explicitly labeled in the graph, as in Fig. 2e.
• Include data showing non-reactive clones (TCR 98, 99, 102, 103) in the cytotoxicity assay.
• Explain why not all tumor-reactive clones were tested against healthy cells (in Extended Fig. 5e).
• Assess whether the presence or expansion of tumor-reactive γδ TCRs correlates with PFS or OS.

Figure 3
• The cutoff for tumor reactivity (15%) may be too permissive. Re-evaluate this threshold and test how increasing stringency affects the specificity of the PreGame algorithm, especially in melanoma and MM, where false negatives occur.

Figure 4
• If the $GzmK^+$ vs $GzmB^+$ cluster comparison is maintained, its biological relevance must be clarified. Clusters should be clearly labeled in Fig. 4a.
• Determine whether a surface marker signature (e.g., $CX3CR1^+$, $CD57^+$, $GPR56^+$, $TIGIT^+$) can enrich tumor-reactive γδ T cells. It would be highly valuable to the field if the authors could sort signature$^+$ vs signature$^-$ cells and experimentally test tumor reactivity.

Figure 5
Claim: Tumor-reactive γδ T cells predict favorable response to BPd
• No data currently demonstrates this conclusion. A ROC curve analysis should be performed to test whether the presence of expanded tumor-reactive clones predicts therapeutic response.
• Fig. 5c should be repeated using autologous tumor cells for more meaningful interpretation.
• The interpretation of Fig. 5e is overstated. Across timepoints (C2D1, C6D1, C18D1), tumor burden and γδ TCR frequency are not consistently inversely correlated.

Figure 6
• If TCR 65 expresses KIR3DL1 (Fig.6i), why is it reactive to healthy samples? Were all healthy donor cells (Extended Fig. 5e) $Bw4^-$? Are PBMCs and small airway epithelial cells (non-reactive targets) $Bw4^+$? Clarify this discrepancy.
• In Fig. 6l:
o Justify using CD69 as a readout when Nur77-GFP was used elsewhere as a reactivity surrogate.
o Explain the rationale for CD8 overexpression in this experiment.
o Provide quantification and statistical testing of the $Nur77\text{-}GFP^+$ percentages across the three experiments.
• According to the model in Fig. 6m, if AMO1 expresses Bw4, TCR 65 should be non-reactive, yet Fig. 2f shows reactivity. This inconsistency requires explanation.
• To fully demonstrate the roles of Bw4 and KIR3DL1, the following experiments are suggested:
o Knockout Bw4 in $Bw4^+$ cells to test for increased reactivity, then reintroduce Bw4 to confirm inhibition.
o Block KIR3DL1 with antibodies in Jurkat cells in parallel with overexpression.
o Test TCR 65 and KIR3DL1 relevance in γδ T cells: introduce TCR 65 into Vδ2 cells (normally non-reactive) with or without KIR3DL1 expression and measure cytotoxicity against $Bw4^+$, $Bw4^-$, and $Bw4^-$ + Bw4-rescued cells.

(Remarks on code availability)

Referee #4

(Remarks to the Author)
The manuscript titled ‘Identification of broadly tumor reactive γδ TCRs using machine learning’ introduces PreGame, which prioritizes tumor-reactive γδ T cells and validates candidate function. In multiple myeloma (MM) patients treated with Belamaf, the authors identify a broadly reactive γδ TCR that recognizes HLA-C and has a KIR3DL1-mediated "logic gate" that spares normal cells. This study emphasizes the underappreciated roles of non-Vγ9Vδ2 γδ T cells in anti-tumor immunity and suggests the translational potential of MM treatment.

This is an impressive study, especially the epitope identification part, and I fully support publication in nature. However, several concerns remain and should be addressed, as detailed below.
Major points A-F

A) Title
To avoid overstatement, please limit the disease scope in the title. The current wording implies generalizability beyond multiple myeloma.

B) Figure 1
The central claim—that γδ T cells mediate sustained responses to BPd via tumor-reactive γδTCRs—is not supported by the presented data.
1) In Fig. 1d, both GZMB$^+$ γδ and GZMK$^+$ γδ T-cell proportions decrease at C2D1 versus baseline. This pattern does not align with the proposed activation/expansion model. Please reconcile this with your conclusion, and clarify whether analyses are normalized to total lymphocytes, CD3$^+$ T cells, or γδ frequency. Provide paired per-patient trajectories and statistics.

2) Unequal sequencing depth can bias Shannon/normalized-Shannon estimates. Please provide a QC summary and a brief sensitivity analysis to exclude depth artifacts.
a) QC summary (per sample):
i) Total productive γδTCR counts.
ii) Unique γδTCR clonotypes (TCRγ chain, TCRδ chain, paired γδTCR chain).
iii) Cell counts and proportions contributing to TCRs.
b) Report correlations to demonstrate robustness to depth:
i) Shannon vs. Unique clonotypes (expect partial but not purely depth-driven association).
ii) Normalized Shannon vs. Total productive γδTCR counts. (expect minimal dependence if normalization is adequate).
c) State the inclusion and exclusion criteria for samples in Fig. 1e, f and Extended Figure 2.

3) Show direct clonotype dynamics rather than a non-significant diversity index. You claim antigen-driven clonal expansion. The primary evidence should therefore be clone-level tracking, not an indirect non-significant Shannon index. Please report:
a) Alluvial plots of γδ clonotypes (baseline → C2D1);
b) Clonality metrics (Simpson/Gini);
c) Report the definition of expanded, contracted γδ clonotypes (Fig. 1e).

4) The authors report (i) expanded γδ clones in BM mediating durable responses, (ii) increased BM–blood TCRδ overlap, and (iii) higher blood TCRδ diversity correlating with longer PFS. Taken together, these can appear contradictory. Please clarify

C) Figure 2
The results do not support the claim that γδ T cells are tumor-specific and recognize broadly expressed tumor-antigens via their TCR.
The healthy tissues and tumor cells lines are from different donors. There is still a possibility that donor variation, aka alloreactivity, may cause the recognition of gdTCR unless the inclusion of donor-matched controls—ideally tumor and peri-tumor/adjacent normal tissues from the same patient.
Still, the presented data in Fig. 2 does not support the ‘tumor-antigens’ recognized by the gdTCRs. Instead, these tumor-reactive gdTCRs recognize antigens expressed on tumor cells.

D) Figure 3 and Materials and Methods the PreGame model
The authors employed five categories of features: (1) expression values of a 19-gene signature, (2) surface protein expression of a 5-protein signature, (3) S phase score, (4) G2/M phase score, (5) cell cycle stage from the discovery cohort, which consisted of a 27-dimensional feature matrix containing 88 tumor-reactive T cells (positive class) and 361 non-reactive T cells (negative class) to train the ML algorithm-PreGame to identify tumor-reactive γδ TCRs.
The current claim is misleading. The model is useful, but trained on non-TCR features (transcript/protein/metadata) and therefore can only identify tumor-reactive γδ T cells from which corresponding γδTCRs can be recovered—not directly predict tumor-reactive γδTCRs.
This should be stated clearly throughout the ms.

E) Details ML algorithm
Regarding the PreGame development, there are several concerns regarding data leakage and robustness.
1) Please report the γδTCR clonotype overlap across donors and, if any, whether the same clonotype exhibits a consistent

phenotype.
2) Cell-level cross-validation mixes cells from the same clonotype/patient/batch across train/test. Donor-held-out and clonotype-disjoint splits are required to exclude data leakage.
3) Please train and evaluate the model within each predefined category, using donor-held-out splits. Report AUROC, AUPRC (with 95% CIs), and F1 score. This tests whether surface markers and cytotoxic programs robustly distinguish tumor-reactive vs bystander γδ T cells without a fancy model.
4) Cosider add Leave-one-category-out validation. Train on all categories except one.
5) clarify if all preprocessing, feature selection, and any oversampling were done only on the training split in each fold.
6) Please provide no-SMOTE, class-weighted baselines alongside any SMOTENC results.
Additionally, the benchmark compares PreGame to algorithms trained on αβ T cell neoantigen reactivity. Please clarify whether these existing models were re-trained on the γδ dataset using identical inputs, preprocessing, and donor-held-out/clonotype-disjoint splits.

F) Figure 6
Are the HLA-C alleles of the donor of TCR65 recognized by TCR65?

To support translational potential, it would of course be great demonstrate on-target cytotoxicity with sparing of normal tissues using γδTCR65-transduced primary human T cells. If possible within the scope of this ms, on-target MM killing and no reactivity to patient-matched normal cells here would substantially strengthen claims of specificity and safety for γδTCR65.

Minor:
Abstract: “we uncover a tumor-reactive γδ TCR epitope” , consider “we uncover a γδ TCR epitope involved in tumor reactivity”
Line 65-66: Please specify single-cell RNA and CITE sequencing data.
Line 67-68: The confirmation of tumor-reactive gdTCRs is based on in vitro functional assay instead of PreGame.
Line 67-69: The terms “adaptive-like” and “NK-like” may read as contradictory. It may be useful to (i) define “adaptive-like” operationally (evidence for TCR-dependent, clonotype-restricted tumor recognition; clonal expansion; memory markers), (ii) clarify that “NK-like” refers to a cytotoxic transcriptional program rather than NK lineage, and (iii) show representative genes/modules supporting the NK-like state.
Line 69-71: ‘clinical utility of this algorithm’? Or you just showed ‘utility of this algorithm on this samples’.
Line 84: Please write down the exact days post–baseline corresponding to Cycle 2 Day 1 (C2D1) to make it friendly for readers who are not MM experts.
Line 87: Please select a proper palette to visualize the cell populations in Fig1c and Extended Data Fig. 1b,c.
Line 88-90: cell type enrichment as shown in Fig. 1d is difficult to digest and appears to show high enrichment of gd T cells, which may not be the case, rather for CD4+ GZMB+ cells.
Line 90-93: rephrase the claim “suggesting recognition of tumor antigens.” The presented data do not support tumor-antigen recognition at this stage. Specifically: (i) GZMB$^+$ γδ and GZMK$^+$ γδ proportions decrease at C2D1 vs baseline (Fig. 1d); (ii) Shannon diversity and D50 of γδTCR show no material change; and (iii) elsewhere in the manuscript, the clonotypes of interest are shown to recognize HLA-C, which is not a tumor antigen and is framed as part of a “logic-gate” interaction. Taken together, these results indicate treatment-associated repertoire remodeling/trafficking rather than direct evidence of tumor-antigen recognition.
Line 111-115: ‘GZMB+ γδ T cells outnumbered GZMK+ γδ T cells’ are basically dominated by donor MM15. Even proportionally, it is hard to conclude that GZMB+ γδ T cells outnumbered GZMK+ γδ T cells. ‘suggestive of increased clonal expansion in some patients’ is not supported by the results (Fig. 2c, d, Extended Data Fig. 4f,g).
Line 115-116: Please report the γδ T cell number and unique paired γδ TCR clonotype number per sample.
Line 130-132: Would it be possible to add statistic results here? And could you please report the phenotype (GZMK or GZMB) of these tested gdTCR from the original scRNAseq data?
Line 140-141: ‘broadly expressed tumor antigens’ should be ‘broadly expressed antigens on tumor cell’.
Line 168-175: Why not co-culture with patient-derived tumor cells instead of cell line?
Line 175-180: It is contradictory that tumor-reactive TCRs are from clonotypes with relatively low clonal expansion. Are these tumor-reactive gd T cells ‘by-standing’?
Line 190-192: ‘Distinguish tumor-reactive from by-stander non-Vg9Vd2 γδ T cells’ would be more precise.
Line 214-216: There is actually a function in Seurat called addmodulescore to compare gene signatures.
Line 219-220: Benchmark PreGame against these re-trained public models. Only if PreGame still outperforms under leakage-robust conditions should you claim it’s superiority. Otherwise, the current comparison largely benchmarks feature selection, not the modeling approach itself, and comparative claims should be rephrased accordingly.
Line 260-261: This is not consistent with the claim in line 90-93.
Line 287-288: Are the samples independent in Extended Fig. 8h?
Line 349 ff: Where in the scheme of Fig 6m would be a place for CD8a?

discussion line 387 ff: Include key reference Rock et al from Chien JEM 1994, doi: 10.1084/jem.179.1.323.

(Remarks on code availability)
code not available

Following the link https://github.com/bigmaklab/TCRmodel did not yield any results, at least for me

Referee #5

(Remarks to the Author)
I co-reviewed this manuscript with one of the reviewers who provided the listed reports.

(Remarks on code availability)

Version 1:

Reviewer comments:

Referee #1

(Remarks to the Author)
The authors comprehensively addressed most of my points and I would like to congratulate the team with this wonderful manuscript.

There are, however, a few points insufficiently addressed that I believe deserve more attention before publication:

- Major point 1: "The relevance of the study would be much greater if the authors could robustly test if the γδ TCR is not only sufficient but also essential for tumor recognition by γδ T cells. Especially since γδ T cells are known to express various activating innate immune receptors, which also represent some of the main features that PreGame uses to identify tumor-reactive cells (NKG2C, CD94, NCR3, FCGR3A/CD16, NKG7)."

The authors now provide data showing that blocking the gdTCR reduces 4-1BB expression on expanded TR gdT cells exposed to target cells, and use this to demonstrate that the TCR is not only sufficient but also essential for tumor recognition. These data are of great interest, but currently seem insufficiently robust given the importance of this claim. Specifically, these experiments were currently only performed with 3 gdT cell donors and only one activation marker as readout. To convincingly support the claim, I believe it is essential to (1) increase the cohort size for this experiment to levels that allow performing robust statistics (paired Wilcoxon), and, critically, (2) complement the 4-1BB readout with a functional tumor cell killing readout and a more comprehensive set of activation markers.

- Minor point 19 "More patient characteristics and details of the clinical cohorts are necessary."

While supplementary table 1 provides extensive clinical information as patient-level data, I believe the manuscript would benefit from a cohort-level baseline table so that the reader understands what kind of patients were included.

- Minor point 23. "When defining tumor reactivity, why did the authors take a cut-off of 13%? This cut-off should be clarified. If up until 13% is background, then an increase from 13% to 15% may not be major recognition by the TCR clone."

The authors addressed this point by introducing a new category of 13-15% activation. However, this does not address the background of my question, which was to understand the rationale for using these exact cutoffs. These cutoffs seem arbitrary as data is missing to support that / show why these cutoffs robustly separate real activation from background activation. If these cutoffs are indeed arbitrary, one will wonder what the impact of setting different thresholds would be on the results in this manuscript. So, I believe the reader needs solid data to substantiate these cutoffs, or, if arbitrary, an analysis showing that the results are robust to choosing a range of other arbitrary cutoffs.

- Finally, a minor point that escaped my attention previously: Fig 6H currently shows the protein structure of HLA-C to pinpoint the location of the relevant residues, but (if I understand the figure correctly) this structure lacks B2M and hence the molecule has been shown in "open conformation". However, the data shows that tumor recognition by TCR65 is strongly dependent on B2M and hence TCR65 is expected to recognise HLA-C in "closed conformation" (that is, in complex with B2M). Hence, it seems more intuitive to show the HLA-C-B2M complex in Fig 6H.

(Remarks on code availability)

Referee #2

(Remarks to the Author)
We must congratulate the authors on their very comprehensive and solid point-by-point reply, and thorough revision of the manuscript. It has improved (from an already very good starting point), and it will certainly be an excellent contribution to the field once published. To ensure the best possible version of the final paper, we have some minor suggestions based on their insightful response and new data provided:

1. The original Extended Data Fig 1d, which they removed but is quite visual and thus helpful, should be added back in;

2. Revisions Fig. 2.5. is interesting and important, so it should be included in the paper for the benefit of the reader;

3. In the new Fig. 2f, the TCRs that have NOT been tested against healthy cells (as acknowledge by the authors in their reply) should be clearly noted as such in the figure (using some sort of shading/ stripes or adding NT in the respective squares).

4. The rationale for CD8 over expression in Fig. 6 provided in their reply should also be added to the paper for the benefit of the reader;

5. In Fig. 6m, the label "gd T cell" should be removed as it does not accurately reflect the experiment performed in Junket cells, instead stressing the ectopic expression of the gd TCR in that system;

6. Revisions Fig. 2.8. is rather important, so it should be included in the paper for the benefit of the reader (with some stats if possible).

Upon making these small changes, we fully support the publication of the paper in Nature. Congratulations on your outstanding work!

Bruno Silva-Santos & Sofia Mensurado

(Remarks on code availability)

Referee #4

(Remarks to the Author)
The authors have adequately addressed all the concerns raised by this and the three other reviewers.

- very minor point, as not included in the revised ms: page 53 of rebuttal, Revisions Figure 3.3. b) : It is not clear how the amount of clonotypes identified with paired TCR can exceed those identified based on either TRD or TRG only.

(Remarks on code availability)

Referee #5

(Remarks to the Author)
I co-reviewed this manuscript with one of the reviewers who provided the listed reports.

(Remarks on code availability)

Version 2:

Reviewer comments:

Referee #1

(Remarks to the Author)
The authors have convincingly addressed my final minor points.

To answer the authors' question regarding how to address the fact that an appropriate experimental HLA structure in complex with B2M is unavailable: I do believe that showing the structure with highlighted residues is of value, hence my suggestion would be to use AlphaFold 3 to model the HLA-B2M structure and cross reference this with the experimental HLA-only structure as a sanity check. When both are consistent (especially in terms of the location of the relevant residues), one could envision including the AF3 HLA-B2M structure in the main figures and the experimental HLA-only structure in the supplementals. When the two show substantially different results, it may be wise to refrain from showing any structure.

Congratulations with this landmark study!

Joris van de Haar
Netherlands Cancer Institute

(Remarks on code availability)

Open Access This Peer Review File is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.
In cases where reviewers are anonymous, credit should be given to 'Anonymous Referee' and the source.
The images or other third party material in this Peer Review File are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.
To view a copy of this license, visit https://creativecommons.org/licenses/by/4.0/

Responses to the Reviewers

We would like to thank all four Reviewers for taking the time to review our manuscript, for their enthusiastic appraisals of our work, and for providing thoughtful feedback that has greatly strengthened the revised manuscript and our conclusions. Please find below a point-by-point response to each Reviewer's comments. Response to Reviewer #1 begins on Page 1, response to Reviewer #2 starts on Page 31, and responses to Reviewer #3 and #4 begin on Page 49.

Reviewer #1 (Remarks to the Author):

Summary

In this manuscript by St. Paul et al., the authors explore if γδ T cells use their γδ TCR to recognize and kill cancer cells, and what defines tumor-reactive γδ T cells. The authors start by studying multiple myeloma (MM) material of (a limited set of) patients, obtained before and during treatment with the ADC belamaf in a clinical trial. Through a combination of scRNA-seq, CITE-seq and TCR-seq they find that the abundance of γδ T cells was associated with better outcomes. As this suggested clinical relevance of γδ T cells in this setting, the sequencing data were subsequently used to identify the paired γδ TCR sequences in the cohort and a large set was cloned into Jurkat cells, of which a selection showed tumor-reactivity against MM cell lines, as well as cross-reactivity against some cell lines of other cancers. By mapping the reactive TCRs on the initial single cell dataset, the authors then classified the γδ T cells as MM reactive, nonreactive, or borderline. Machine learning was then used to identify patterns in the scRNA/CITE-seq profiles that were associated with tumor reactivity. The resulting tool (coined "PreGame") was benchmarked against tools developed for ab T cells and validated on unseen data including 14 additional MM samples and 8 samples of other (non-MM) cancer types. By studying feature importance in PreGame, as well as differential expression analyses of reactive vs nonreactive clones, the authors performed a phenotypic comparison (looking at RNA/protein expression and γ/δ chain usage) of MM-reactive vs nonreactive γδ T cells. Next, the authors revisited the trial data and studied the clonal dynamics of γδ T cells during belamaf treatment, showing the expansion of tumor-reactive clones was associated with favorable outcomes. Having screened many γδ TCRs for reactivity against a collection of cancer cell lines, the authors were then able to select several tumor-reactive clones and set out to study which antigens they recognized. To this end, they started with testing if tumor recognition was dependent on a few key proteins: B2M (through knockout), BTN3A1 (through knockout) and CD1D (through overexpression). This showed that the two Vγ9 TCRs both recognized a B2M-dependent target other than CD1D, and follow-up experiments with 11 additional tumor-reactive Vγ9 TCRs showed that roughly half of them also recognized B2M-dependent targets. Because of its broad tumor reactivity, they then selected "TCR65" (a Vδ3Vγ9 TCR) and performed HLA knockout experiments and co-cultures with APCs expressing single HLA alleles, to conclude that TCR65 recognized a HLA-C epitope. By mapping sequence differences among differentially recognized HLA-C allotypes, coupled with targeted mutation, the authors could identify the V103L codon as essential for recognition. Finally, the authors demonstrated that expression of KIR3DL1 (functioning together with CD8a as coreceptor) could overcome autoreactivity of TCR65 expressing cells, suggesting novel avenues to create cellular therapies with a "logic switch" to specifically recognize and kill cells that downregulate HLA-A and HLA-B, but at the same time (over)express HLA-C.

General remarks

This is an exciting and exceptionally data-rich manuscript which addresses multiple highly relevant questions in the field of cancer immunology, with clear translational implications. Especially for Vδ1 and Vδ3 cells, which have innate-like phenotypes, it has been a longstanding question whether these cells carry tumor-reactive γδ TCRs. With strong data, the authors proof that (especially in the context of MM) this is the case for at least a fraction of these cells, which I find a very intriguing, also given its potential relevance for "off-the-shelf" γδ T cell therapies. The identification of a "logic gate" that could be engineered by overexpressing a γδ TCR plus KIR3DL1 and CD8 to create a cellular product that specifically targets cells with HLA allele imbalances is also interesting, although I do think the biological interpretation of these data is somewhat overstated (see below).

In terms of biology, it would greatly increase the relevance of the findings if the authors would test if the γδ TCR is not only sufficient but also essential for tumor recognition by γδ T cells. It has long been postulated that the large repertoire of activating innate immune receptors expressed by γδ T cells play dominant roles in tumor recognition, but the findings of St. Paul et al raise the question what their relative importance really is? Regarding PreGame: this tool is interesting, but I do have important concerns regarding its accuracy and generalizability across cancer types and γδ T cell subsets, as well as its added value compared to simpler alternatives. Related to this, a concern of this manuscript is that some described phenotypic characteristics of tumor-reactive γδ T cells in MM contrast key observations made by others in other cancers, posing questions regarding the generalizability to other disease contexts. Also in other instances, there are findings which contrast/challenge well-established knowledge (e.g. Vδ2Vγ9 T cells which are described by the authors to act independently of BTN3A1); such unexpected claims need particularly strong data but are not always double checked with alternative approaches or being put into perspective. Furthermore, while multiple exciting findings are broadly relevant, it could also be noted that the relevance of the clinical cohorts is limited by the fact that these center around the drug belamaf whose regulatory approval has been retracted. A detailed point-by-point review below.

Reply:

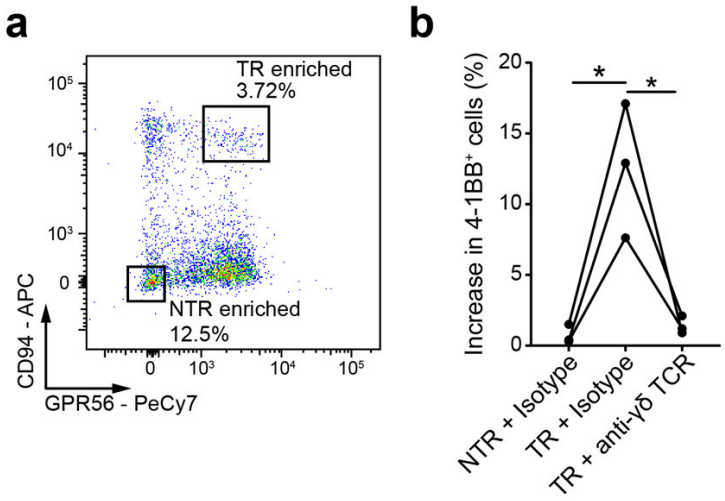
We thank the Reviewer for these enthusiastic comments about our manuscript and valuable suggestions to improve it.

Major points

1. The relevance of the study would be much greater if the authors could robustly test if the γδ TCR is not only sufficient but also essential for tumor recognition by γδ T cells. Especially since γδ T cells are known to express various activating innate immune receptors, which also represent some of the main features that PreGame uses to identify tumor-reactive cells (NKG2C, CD94, NCR3, FCGR3A/CD16, NKG7).

Reply:

The Reviewer raises an important point. We have now conducted additional experiments demonstrating that primary tumor-reactive γδ T cells expanded from the bone marrow (BM) of multiple myeloma (MM) patients are indeed recognizing tumor cells using their TCRs. Using the top two PreGame surface markers, GPR56 and CD94, we flow-sorted MM patient BM samples to obtain the double positive γδ T cell fraction (enriched for tumor-reactive γδ T cells) and the double negative γδ T cell fraction (enriched for NTR γδ T cells). We found that the population enriched for tumor-reactive γδ T cells up-regulated 4-1BB in response to co-culture with MM cells, and that this tumor-reactivity was abrogated upon addition of an anti-γδ TCR blocking antibody. These findings indicate that TCR signaling is essential for the activation of tumor-reactive γδ T cells. These data are presented below in **Revisions Fig. 1.1a,b**, which we have added to our revised manuscript as **Fig. 4f,g**. We thank the Reviewer for this thoughtful question, the answering of which we feel has greatly strengthened our claims. In addition to improving our manuscript, we hope this demonstration of using PreGame markers to flow-sort and enrich for tumor-reactive γδ T cells will be of great interest and benefit to the field.

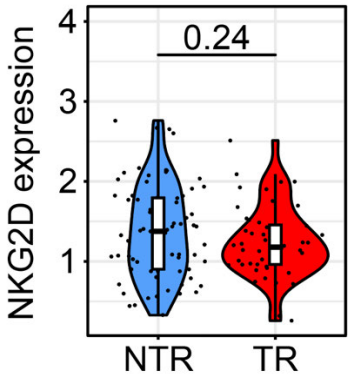


Revisions Figure 1.1. **(a)** Tumor-reactive (TR) enriched (GPR56$^+$CD94$^+$) and non tumor-reactive (NTR) enriched (GPR56$^-$CD94$^-$) primary γδ T cells were sorted from the BM of n = 3 MM patients. **(b)** Cells were expanded and incubated with MM cell lines in the presence of anti-γδ TCR blocking antibody or isotype control. Results show an increase in 4-1BB-expressing γδ T cells following culture with MM cells compared to unstimulated γδ T cells with each line representing a different patient. Significance was assessed using repeated-measures ANOVA, * $p < 0.05$.

2. Related to the above, it is not entirely clear which fraction of experimentally tested γδ TCRs were found to be tumor reactive? And are γδ T cells from all other clonotypes not tumor reactive, or are they tumor reactive via other (innate) receptors?

Reply:

We have now clarified in the text that the term “TR γδ T cells” refers to γδ T cells bearing tumor-reactive (TR) γδ TCRs. It is possible that non tumor-reactive (NTR) γδ T cells contribute to anti-tumor immunity via innate receptors such as NKG2D. Indeed, NTR γδ T cells show increased surface expression of NKG2D compared to TR γδ T cells (**Revisions Fig. 1.2**). However, PreGame is able to accurately identify only γδ T cells that express TR TCRs, demonstrating that these cells are transcriptionally different from NTR cells.



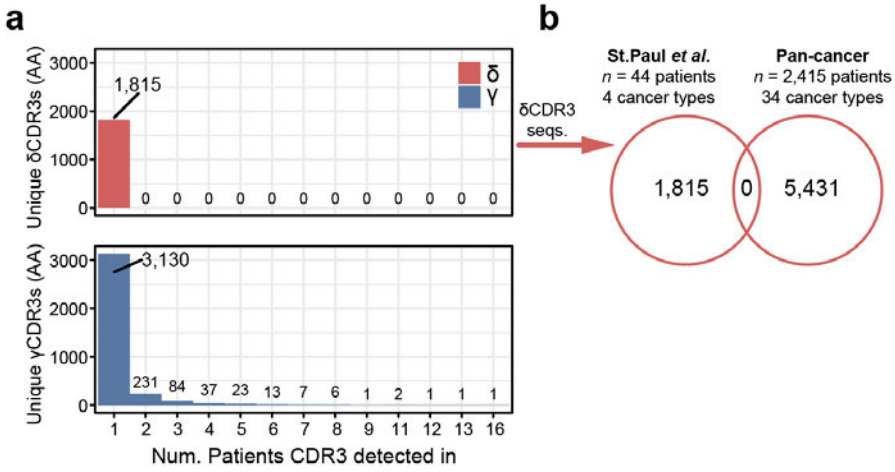
Revisions Figure 1.2. Average surface protein expression of NKG2D in TR and NTR γδ T cell clonotypes as determined by scCITE+VDJ-seq. Significance was assessed using a Wilcoxon rank-sum test.

3. It is also important to better understand the diversity of the tested γδ TCR repertoire. Looking at the cell line reactivity data, it seems that there are clear clusters of γδ TCRs recognizing exactly the same target cell lines suggesting that there may be very low diversity.

Reply:

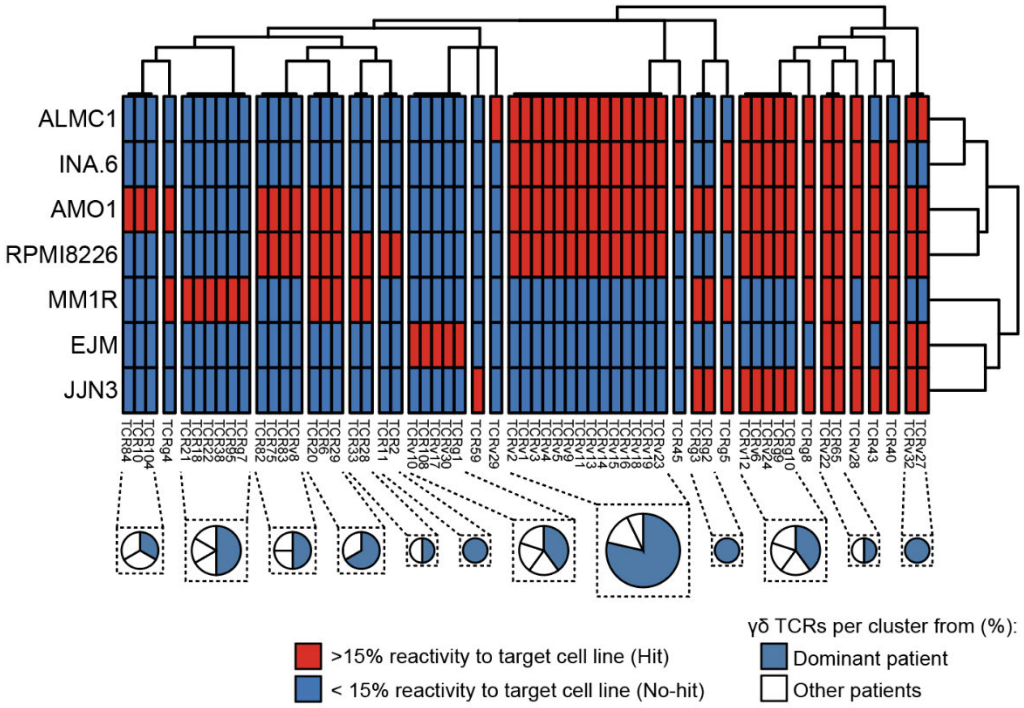
Reviewer #2 posed a question along similar lines: *"Clarify whether all tumor-reactive γδ TCRs are private or if some are shared between patients"*. We agree with the Reviewers that this is an important point that pertains to the overall diversity of the γδ TCR repertoire and requires further examination.

We did not observe any instances of γδ TCRs being shared between patients presented in this manuscript (*n* = 44). The complementarity-determining region 3 (CDR3) sequences of the δ TCR repertoire was 100% private whereas the γ CDR3 repertoire exhibited a higher degree of overlap between patients (88.5% private) (**Revisions Fig 1.3a**). We compared the 1,815 unique δ CDR3 sequences identified in our study with 5,431 unique δ CDR3 sequences from two recent investigations spanning > 2,000 patients (**PMID: 39368482, PMID: 39058036**) and again found no overlaps (**Revisions Fig. 1.3b**). These striking results suggest that the δ CDR3 repertoire is highly private. Taken together with other features of δ CDR3s, namely their long amino acid (AA) length and more variable length distribution (**Extended Data Fig. 7b,c**), the overall diversity of δ CDR3s is expected to be exceptionally high, a hypothesis in line with the existing literature (**PMID: 39368482**).



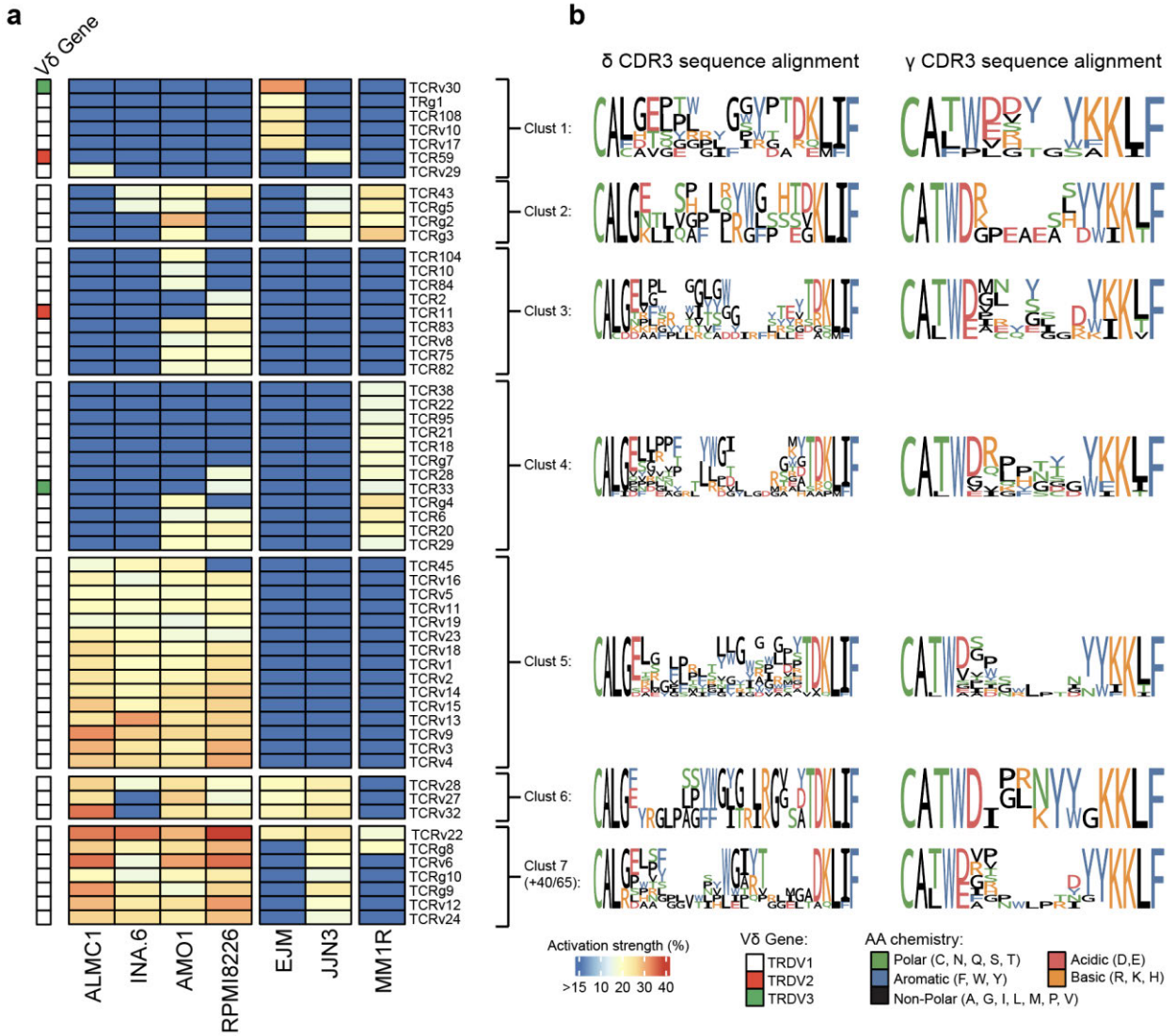
Revisions Fig.1.3. (a) Total unique δ CDR3 (top) and γ CDR3 (bottom) AA sequences across our MM, MCC, melanoma, and NSCLC patient sample single-cell sequencing cohorts (*n* = 44), and the number of patients in which the CDR3 sequence was detected. **(b)** Overlap of unique δ CDR3 AA sequences identified in this manuscript (as in **a**) with unique δ CDR3 AA sequences identified in two recent pan-cancer analyses (**PMID: 39368482, PMID: 39058036**).

To the Reviewer's point about the appearance of clusters of γδ TCRs recognizing the same target cell lines: we agree, but wondered whether the original presentation of our data might have emphasized the appearance of these clusters. We therefore generated a combined reactivity heatmap for every tumor-reactive γδ TCR (>15% activation strength) identified from our MM cohorts (i.e. the Discovery, Validation 1, Validation 2, and Algonquin cohorts; 59 γδ TCRs in total). We found that 22 clusters were required to perfectly sort these 59 γδ TCRs based on binarized cell line reactivity (i.e. hit or no-hit), with a median cluster membership of just 2 γδ TCRs (**Revisions Fig. 1.4, top**). The largest of these clusters was dominated by γδ TCRs from just one patient (11/14, 79%) (**Revisions Fig. 1.4, bottom**). This analysis highlights that the range of possible target cell line combinations is greater than it may have appeared visually and at least partially driven by clustering of γδ TCRs from the same patient, which is not unexpected.



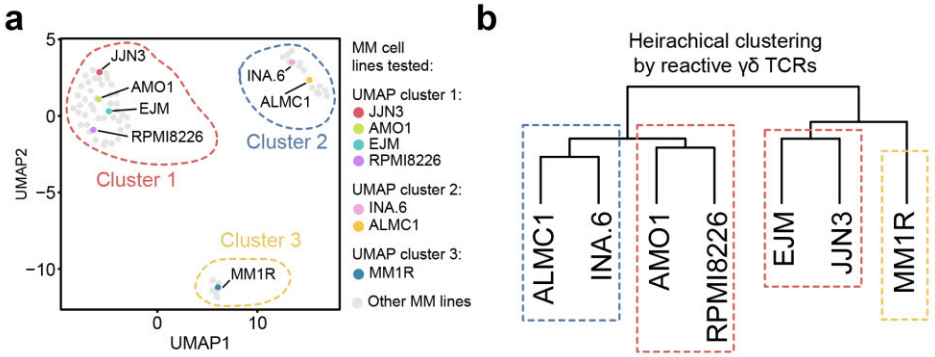
Revisions Fig.1.4. Top: Heatmap depicting hierarchical clustering of binarized γδ TCR reactivity from *in vitro* screens performed in our Discovery and Validation 1 and 2 cohorts (**Fig. 1d, Fig. 3c,d**). Reactivity towards a given MM cell line was binarized using a 15% activation strength cut-off to assign either reactivity (red) or non-reactivity (blue). Bottom: Pie charts depicting the proportion of γδ TCRs per cluster that were recovered from different patients, with the proportion of TCRs from the most represented patient being colored in blue in each pie.

If low γδ TCR diversity was giving rise to the observed target cell line clustering, we hypothesized that these TCRs would show commonalities in their CDR3 sequences. To glean insights into potential antigenic similarities among γδ TCR clusters, we first reduced the number of clusters to allow for more CDR3 sequences per cluster. Using the average silhouette width method, we determined an optimal resolution of 7 clusters (**Revisions Fig. 1.5a**) (*note: TCR40 and TCR65 were excluded from clustering analysis and added to cluster 7*). We performed multiple sequence alignment on a per cluster basis and, outside of the leader and trailing sequences, there was only one instance of a common AA residue in δ CDR3s, which occurred in the smallest cluster (Cluster 6, 3 γδ TCRs, G at residue 22) (**Revisions Fig. 1.5b**). Indeed, no discernable patterns arose from this analysis, not even in terms of AA chemistry at a single given position. These results suggest that even though there is the appearance of clustering by TCR reactivity, low sequence diversity does not appear to be at its root.



Revisions Fig.1.5. (a) Heatmap depicting hierarchical clustering of TR γδ TCRs identified from our *in vitro* screens from the Discovery and Validation 1 and 2 cohorts (**Fig. 1d, Fig. 3c,d**). Dendrogram is not shown to simplify visualization. Activation strength corresponds to the mean absolute percent increase in Nur77^{-}GFP^{+} Jurkat cells following co-culture with the indicated MM cell line compared to unstimulated Jurkat cells. **(b)** Logo plot depicting the frequency of common AA residues at each position following multiple sequence alignment of δ CDR3 (left) and γ CDR3 (right) sequences from the γδ TCRs in each reactivity cluster.

We next considered the hypothesis that the observed clustering might be attributable to cell line antigenic similarity rather than to low γδ TCR diversity. We originally selected our panel of seven MM cell lines on the basis of unbiased transcriptional clustering, which had revealed three distinct clusters for which we wanted to ensure representation (**Extended Data Fig. 5a**). Interestingly, the clustering of target cell lines by γδ TCR reactivity broadly recapitulated the cell lines' transcriptional clustering (**Revisions Fig. 1.6a,b**). Collectively, these results suggest that, rather than owing to low γδ TCR sequence diversity, the observed clustering of target cell lines may be due to these cell lines appearing more antigenically similar to a given γδ TCR. This notion is supported by the observation that TR γδ TCRs from the same patient tend to cluster. In addition, the high sequence diversity per cluster supports the idea that each individual γδ TCR recognizes a specific and different antigen.

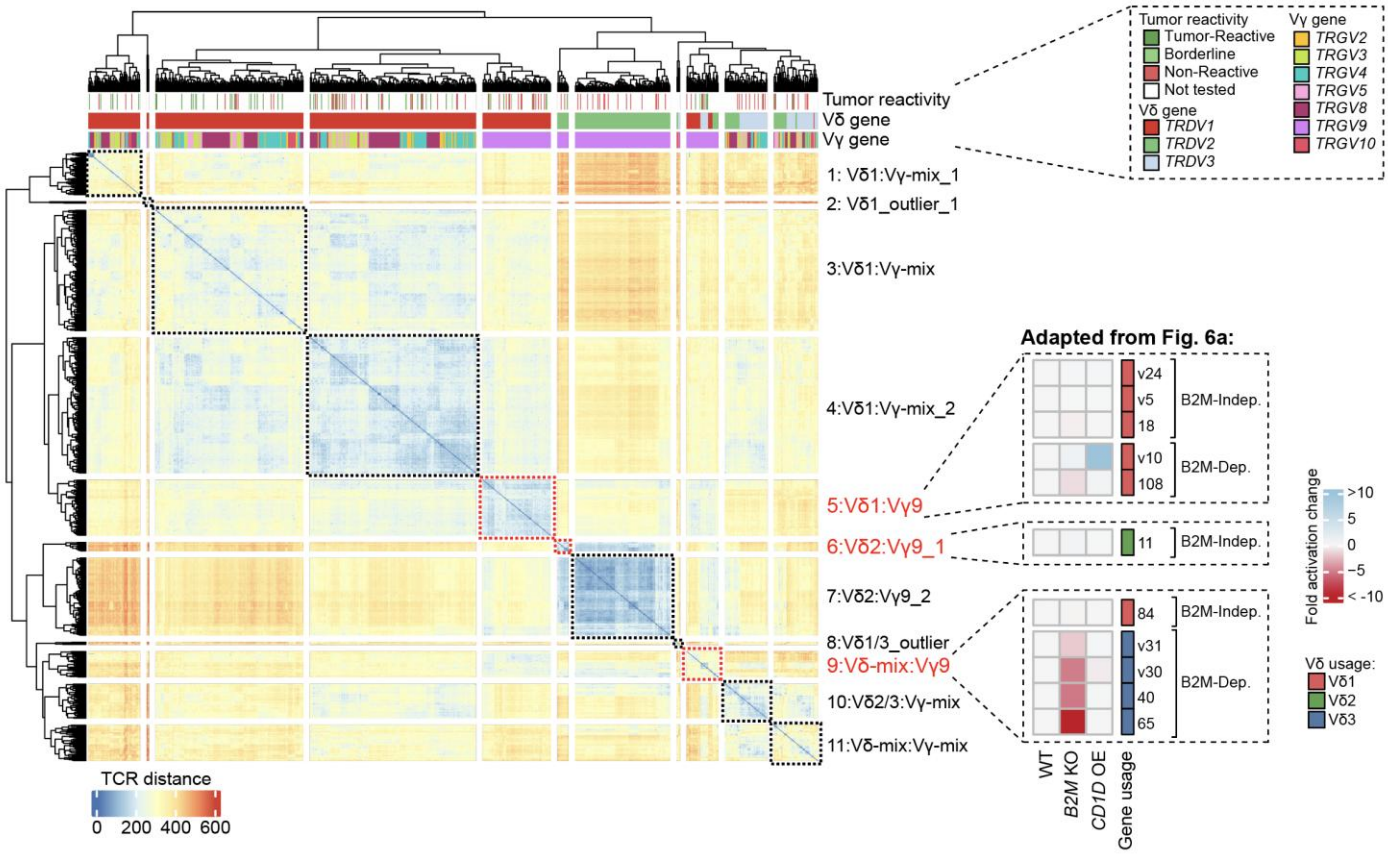


Revisions Fig.1.6. (a) UMAP embedding of bulk RNAseq data obtained from *n* = 66 MM cell lines revealing three transcriptionally distinct clusters (adapted from **Extended Data Fig. 5a**). **(b)** Dendrogram from the hierarchical clustering performed in **Revisions Fig.1.5a.** A dashed box overlay highlights each MM cell line's transcriptional cluster membership as determined in **a**.

Despite these results, we do not want to suggest that efforts to identify γδ TCR ligands on the basis of CDR3 sequence similarity would be in vain. While we have hopefully provided a better understanding of the diversity of the *tested* γδ TCR repertoire for this Reviewer, it is underpowered compared to our sequencing of full γδ TCR repertoires. We clustered our tested and untested γδ TCRs from our collection of *n* = 22 MM samples using TCRdist3 (**PMID: 34845983**), a sequence-based similarity approach for TCR analysis (**Revisions Figure 1.7**). In general, we observed that:

1. γδ TCRs predominantly clustered by the Vδ gene.
2. Vγ9 TCRs form distinct clusters more profoundly than other Vγ gene TCRs.
3. 'Jump off the page' clustering by tumor-reactivity is not observed (though cluster 3:Vδ1:Vγ-mix appears to be the most likely candidate).
4. The lone tumor-reactive Vδ2Vγ9 TCR identified in our screens clustered distinctly from the majority of Vδ2Vγ9 TCRs; however, this cluster also contained two NTR Vδ2Vγ9 TCRs.

We had hypothesized, based on our initial antigen-specificity screens, that B2M-dependent γδ TCRs would be confined to Vγ9 TCRs. Leveraging our γδ TCR specificity results (**Fig. 6a**), we observed that cluster 5:Vδ1:Vγ9 comprised three B2M-independent γδ TCRs, the sole C1D1-OE γδ TCR, and one B2M-dependent TCR. Cluster 9:Vδ-mix:Vγ9 comprised one B2M-independent and four B2M-dependent TCRs (**Revisions Figure 1.7**). These results, while preliminary, tend to support the idea that predicting γδ TCR specificity from TCR sequence information could be successful in the future.



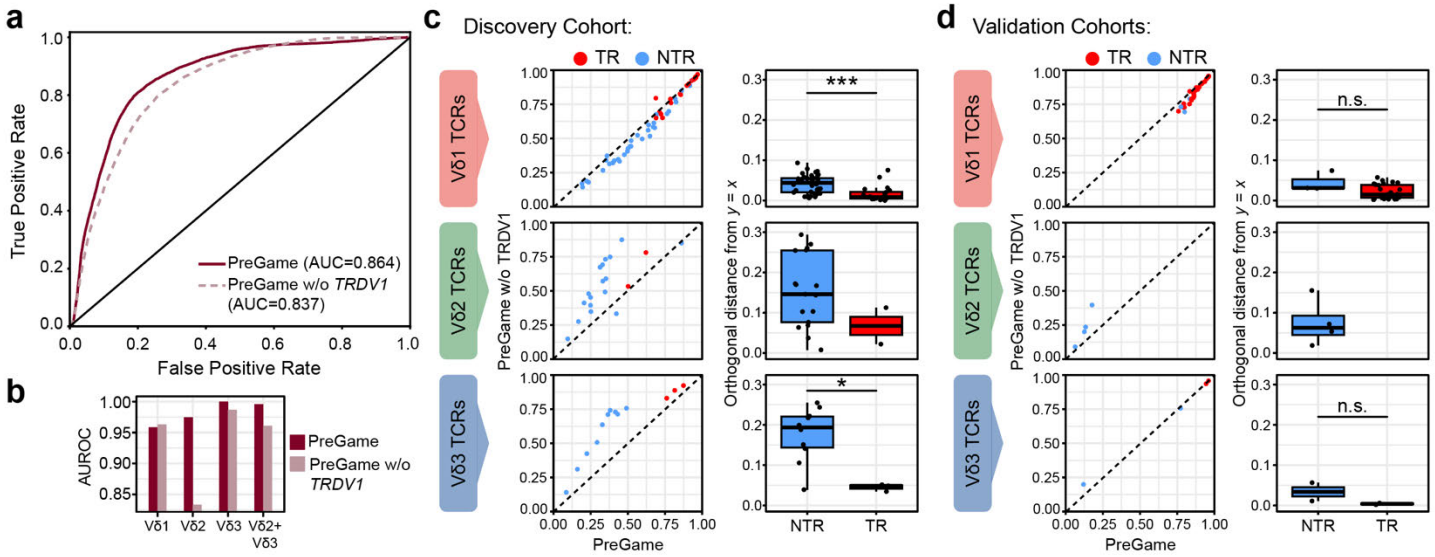
Revisions Fig.1.7. (a) Right: Heatmap depicting hierarchical clustering of paired γδ TCR distance, as computed by TCRdist3, for all paired γδ CDR3 sequences recovered in our *n* = 22 MM cohort. γδ CDR3 sequences that are closer in AA similarity formed 11 clusters largely by Vδ and then Vγ gene usage. Left: Heatmaps depicting B2M and CD1D dependencies for γδ TCRs belonging to the indicated TCRdist3 clusters (highlighted in red text), where available (adapted from **Fig. 6a**).

4. The authors present PreGame as a tool to identify -reactive γδ T cells across γδ T cell subsets and tumor types, but I do have concerns regarding its generalizability:
Does it work within Vd1/2/3 separately? A quantitative analysis is needed. Especially as TRDV1 is one of the strongest features in the model, suggesting that part of its performance simply comes from separating Vd1 cells vs Vd2 cells (whose are known to show much higher tumor reactivity than Vd2 cells). This is an tumor essential point.

Reply:

We agree with the Reviewer that this is an essential point, and one that we considered early on during our development of PreGame. Our *in vitro* tumor-reactivity screen of the Discovery cohort γδ TCRs revealed an enrichment for Vδ1 TCRs, which unsurprisingly translated into *TRDV1* being a significantly differentially expressed gene between TR and NTR γδ T cells in our unbiased analysis. While our objectives were to keep PreGame free from bias, we also recognized that *TRDV1* might introduce a particularly potent downstream bias. We therefore trained and tested two versions of PreGame, one with *TRDV1* and one without *TRDV1* (PreGame w/o TRDV1). Both models exhibited comparable area under the curve (AUC) during cross-validation (**Revisions Fig. 1.8a**), but the predominance of Vδ1 TCRs in the training set seemed to give an edge to the model with *TRDV1* included.

Surprisingly, when we computed area under the receiver operating characteristic (AUROC) values for the resubstitution prediction results of each individual Vδ chain, we found that PreGame w/o TRDV1 performed quite a bit worse on Vδ2 and Vδ3 TCRs (**Revisions Fig. 1.8b**). The driver underlying this poor performance was a slightly increased rate of false positives on Vδ2 and Vδ3 TCRs. This result has been visualized in **Revisions Fig. 1.8c, left**, and quantified by calculating each TCR's orthogonal distance from $y = x$ (dashed lines) (**Revisions Fig. 1.8c, right**). The Vδ2 and Vδ3 prediction results of PreGame w/o TRDV1 diverged from $y = x$ to a higher degree in NTR T cells compared to TR T cells (**Revisions Fig. 1.8c**). These data indicated to us that the inclusion of *TRDV1* as a feature in PreGame controlled spuriously high prediction values in the Vδ2 and Vδ3 compartments (perhaps due to differing weights of other features). The number of TR Vδ2 and Vδ3 TCRs was low in both the Discovery and Validation cohorts (**Revisions Fig. 1.8d**), so that we can only speculate whether this result would hold true in (*future*) larger training cohorts. Ultimately, the superior AUC from cross-validation coupled with the observed improvements on Vδ2 and Vδ3 TCRs led us to move forward with our model including *TRDV1* as a feature.



Revisions Fig.1.8. (a) Receiver operating characteristic (ROC) curve comparing false positive and true positive rates at various thresholds for PreGame (as in **Fig. 3b**), and PreGame re-trained without *TRDV1* (PreGame w/o TRDV1) as a feature. **(b)** Area under the ROC (AUROC) values for PreGame and PreGame w/o TRDV1 resubstitution prediction results when applied to the indicated subsets of Discovery cohort γδ TCRs. Higher AUROC values indicate better overall performance. **(c,d)** Left: Scatter plots of **(c)** resubstitution prediction results of Discovery cohort γδ TCRs, and **d)** prediction results of Validation 1 and 2 cohort γδ TCRs, by PreGame versus PreGame w/o TRDV1 for each indicated Vδ subset of *in vitro*-validated γδ TCRs. Points are colored by tumor-reactivity; dashed line indicates $y = x$. Right: Boxplots depicting orthogonal distance of each TR or non tumor-reactive (NTR) γδ TCR from $y = x$. Significance was assessed by Wilcoxon rank-sum tests, ***$p < 0.0005$, *$p < 0.05$, n.s. not significant.

These findings should be re-evaluated in future investigations after large independent datasets have been generated by us or others. We are excited about the prospect of our dataset being leveraged to train future prediction models, or to serve as an independent validation cohort for others.

5. Does PreGame really work for gdT cells in other cancers? Especially in cancer types where gdT cells are known to play a significant role, such as MSI CRC (de Vries et al, 2023), pediatric neuroblastoma (Castenmillen et al, 2025), etc. The numbers of non-MM malignancies are very low and more data is needed to

support this claim. Especially as some findings suggest that tumor-reactive gdT cells in MM have very different phenotypes than what has been observed in other cancers (e.g. high PD-1 expression clearly marks tumor-reactive gdT cells in MSI CRC, while the authors find that tumor-reactivity is largely confined within the PD-1 low population).

Reply:

We agree with the Reviewer that this important point warrants further investigation. However, given the lack of existing datasets combining single-cell CITE sequencing with single-cell γδ TCR sequencing that would be required to truly assess PreGame performance, we have toned down our claim in the revised text regarding PreGame's ability to work in other cancer types.

Initial manuscript: "Collectively, these results indicate that PreGame can accurately distinguish tumor-reactive from by-stander γδ T cells in multiple cancer types using a prediction threshold of 0.85 (Fig. 3g,h)."

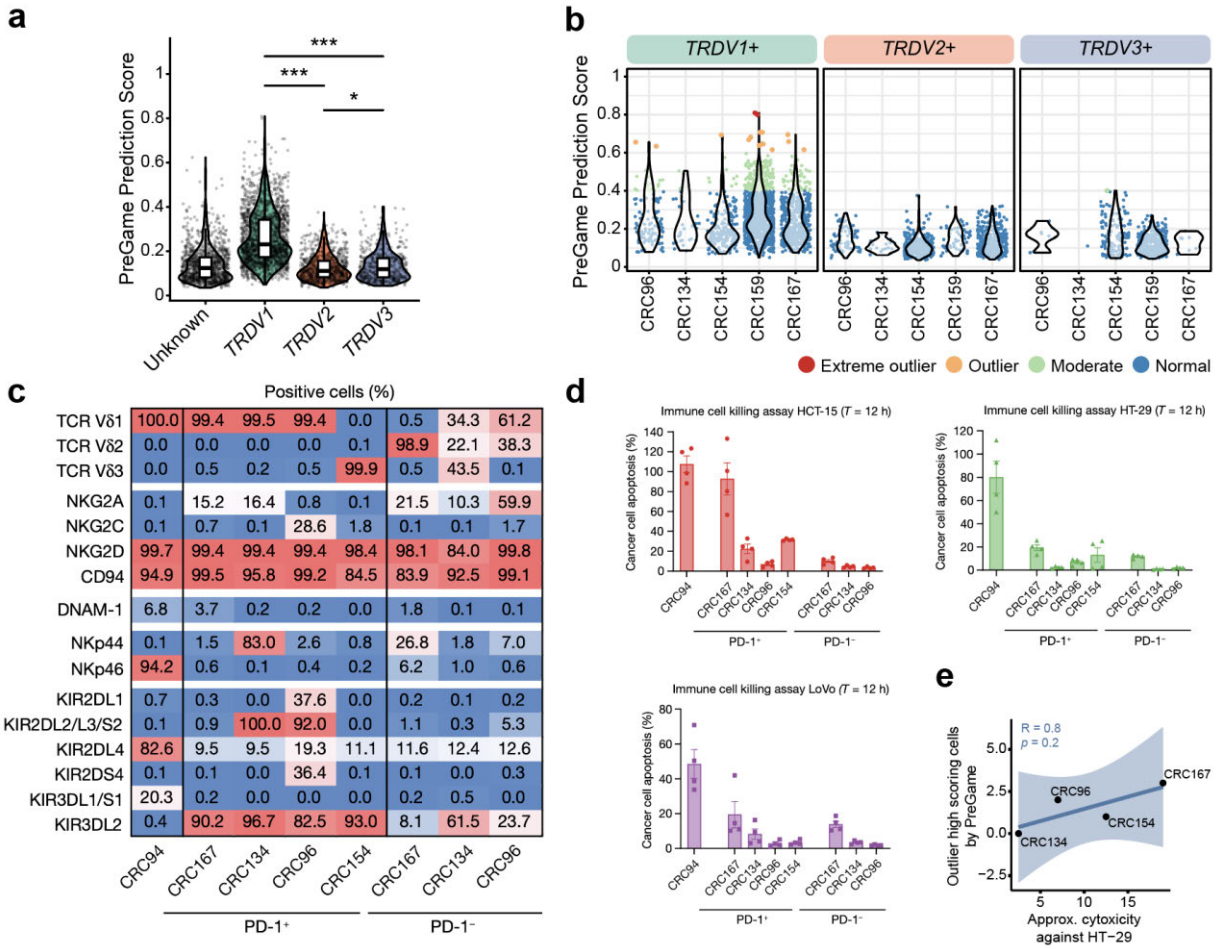
Revised manuscript: "This analysis of other cancer types indicated that, while a PreGame score threshold of ≥0.85 retained a perfect TPR, a less conservative cut-off or feature re-evaluation might be necessary to control false negatives outside of MM (Fig. 3h). However, our results support the notion that PreGame can **assist** in distinguishing TR from by-stander γδ T cells in MM and **in some** types of solid tumors"

Of relevance, we have performed additional analyses to explore whether PreGame works in cancer types where γδ T cells are known to play significant roles. We accessed and analyzed the single-cell RNA sequencing data generated in De Vries *et al*., 2023 (**PMID: 36631610**) and observed significantly higher PreGame prediction scores in *TRDV1*$^+$ cells compared to *TRDV2*$^+$ and *TRDV3*$^+$ cells, and significantly higher scores in *TRDV3*$^+$ cells compared to *TRDV2*$^+$ cells (**Revisions Fig. 1.9a**). These results are consistent with the findings presented in De Vries *et al.,* and with our findings that the Vδ1 and Vδ3 subsets represent the primary reservoir of TR γδ T cells. Overall, the observed prediction scores were generally lower than in our MM cohorts, but consistent with those observed in our melanoma cohort (**Revisions Fig. 1.9b, Fig. 3g**). While we did not observe any individual cells reaching the high confidence TR threshold of ≥0.85, we did observe extreme score outliers in CRC159, along with several statistical outliers in most samples, which were suggestive of tumor-reactivity (**Revisions Fig. 1.9b**). An important caveat is that this dataset from De Vries *et al.* did not contain surface antibody expression (scCITEseq) data, which could underly reduced prediction scores across the board. Still, the results presented in **Revisions Fig. 1.9b** highlight several key consistencies (see below) between the findings of De Vries *et al*. and those of PreGame, suggesting that PreGame can indeed 'work' in other cancer types:

1. Compared to other samples, Patient CRC154 exhibited the highest scores in the *TRDV3*$^+$ compartment, and the second fewest high-scoring cells in the *TRDV1*$^+$ compartment. This result is consistent with Vδ3$^+$ cells comprising 99.9% of expanded PD-1$^+$ cells from this patient (**Fig. 3a** from De Vries *et al*., 2023, adapted into **Revisions Fig. 1.9c** below), and with the demonstrated cytotoxicity of these cells (**Fig. 3e** from De Vries *et al*., 2023, adapted into **Revisions Fig. 1.9e** below).
2. Patient CRC154 was also the only case of >0% Vδ2$^+$ content in the expanded PD-1$^+$ fraction (**Fig. 3a** from De Vries *et al*., 2023, adapted into **Revisions Fig. 1.9c** below), and, consistent with this observation, CRC154 was found by PreGame to have the highest scoring *TRDV2*$^+$ cell (**Revisions Fig. 1.9b**).
3. Patients CRC96 and CRC167 both featured a handful of outlier high-scoring cells in the *TRDV1*$^+$ compartment, and their PD-1$^+$ fractions were consistently dominated by Vδ1$^+$ T cells following expansion (**Fig. 3a** from De Vries *et al*., 2023, adapted into **Revisions Fig. 1.9c** below). PD1$^+$ γδ T cells from Patient CRC167 performed better than those from Patient CRC96 in cytotoxicity assays

against all three cell lines presented in **Fig. 3e** (adapted into **Revisions Fig. 1.9e** below) from De Vries *et al.*, 2023, in line with Patient CRC167 being found by PreGame to have more high-scoring cells than Patient CRC96 (**Revisions Fig. 1.9b**).

4. The sample from Patient CRC159 was the only one found to have extreme outliers (**Revisions Fig. 1.9b**), the highest number of outliers, and the highest number of cells in the moderate score range. Unfortunately, this sample was not assessed in downstream functional experiments performed in De Vries *et al.*, 2023, which precludes us from assessing PreGame's performance in what might have been the most informative sample.



Revisions Fig.1.9. (a) PreGame prediction scores and distributions within the indicated γδ T cell subsets for single cells obtained from and annotated by De Vries *et al.*, 2023. A re-trained gene-only version of PreGame was used to account for the absence of surface protein (CITE) expression values in this dataset. Significance was assessed using Wilcoxon rank sum tests, ***$p < 0.0005$, *$p < 0.05$. **(b)** PreGame prediction scores as in **a** and further divided by donor patient within each γδ T cell subset. Points are colored by the following score ranges: Moderate, scores over 0.4 corresponding to approximately the top 20% of scores; Outlier, statistical outliers (scores over 75th percentile + 1.5 times the interquartile range); Extreme outliers, notably high statistical outliers with scores ≥ 0.8; Normal, the remaining scores. **(c)** The percentage of cells positive for the indicated markers among γδ T cells expanded from MMR-d colon cancers ($n = 5$), adapted from **Fig. 3a** of De Vries *et al.*, 2023. **(d)** Quantification of the killing of three CRC cell lines after co-culture with γδ T cells from each donor, adapted from **Fig. 3e** of De Vries *et al.,* 2023. **(e)** Scatter plot of approximate cytotoxicity of expanded PD-1⁺ γδ T cells against HT-29 as determined in De Vries *et al.*, 2023 (**d,** right, green bars) compared to the number of PreGame tumor-reactivity prediction score statistical outliers as determined in **b**. Cytotoxicity values on the x-axis are approximate and meant for illustrative purposes only. Blue line denotes Spearman correlation and 95% confidence interval; Rho value and significance as determined by Spearman rank correlation tests are presented within the plot.

These findings are consistent with PreGame 'working' in MSI CRC. Although the data are correlative in nature, PreGame would have nominated Vδ1$^+$ γδ T cells as the primary TR compartment and would have roughly predicted the observed cytotoxicity of expanded γδ T cells from each patient against HT-29 (**Revisions Fig. 1.9e**). Ultimately, the only way to prove whether PreGame works in new cancer types is to screen the γδ TCRs of high-scoring cells for tumor-reactivity, a task that requires single-cell γδ TCR sequencing data. Indeed, our single-cell cohorts are the first that we are aware of to combine (full panel) CITEseq with single-cell γδ TCR sequencing. However, with the growing excitement surrounding γδ T cells in the cancer immunotherapy and immunology fields (thanks to a handful of fantastic recent publications, including De Vries *et al.,* 2023), we anticipate that single-cell γδ TCR sequencing may soon become routine. PreGame stands ready to assist in such future studies and will itself benefit greatly from the increasing amounts of training/validation data emerging from studies of diverse cancer types.

6. Related to the above, how is it explained that previously described RNA expression markers of tumor-reactive gdT cells in solid tumors are not always overexpressed in PreGame-high cells in MM?

Reply:

We will address this point in combination with the latter part of the previous question: "*Especially as some findings suggest that tumor-reactive gdT cells in MM have very different phenotypes than what has been observed in other cancers (e.g. high PD-1 expression clearly marks tumor-reactive gdT cells in MSI CRC, while the authors find that tumor-reactivity is largely confined within the PD-1 low population).*"

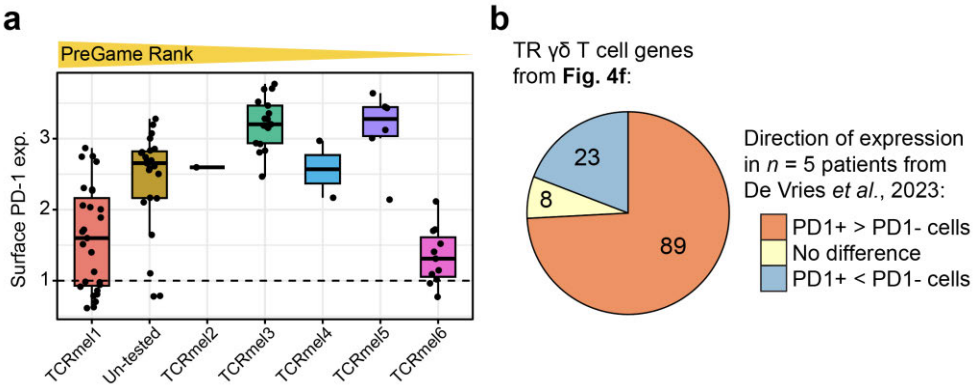
Our data are largely consistent with previous results. Single cell analysis performed by Harmon et al., 2023 (PMID: **37474835**) of γδ T cells found in TIL of endometrial and colon cancer found that the Vδ1 in the TIL are PD-1 low and express higher levels of TIGIT and Granzyme B similar to what we have found in our TR T cells in MM. Moreover, in our data despite finding lower levels of PD-1 in our expanded TR γδ T cells (**Fig. 4i**), we want to clarify that these cells still had surface PD-1 protein expression, thereby allowing for targeting by PD-1 blockade. We have now clarified this matter in our revised text as follows:

Initial manuscript: "Indeed, expanded tumor- reactive γδ T cells expressed significantly less PD-1 compared to low frequency tumor-reactive cells, with the opposite pattern occurring in non-reactive γδ T cells."

Revised manuscript: "Indeed, expanded TR γδ T cells expressed significantly less, **but still present**, PD-1 compared to low frequency TR cells."

Importantly, we observed this same phenomenon in our melanoma cohort (**Revisions Fig. 1.10a**). Because our cohorts are among the first to combine CITEseq with single-cell γδ TCR sequencing, it may well be that this phenomenon was present in previously published studies—but not observable. We suspect that our findings are a result of capturing (sequencing) these expanded clonotypes at a stage where PD-1 upregulation has not yet caught up with that of exhausted TR γδ T cells. The proliferative capacity of these latter cells might then be diminished so that they comprise a smaller and smaller fraction of sequenced cells over time. Indeed, the 'Non-expanded' cells might be a combination of true non-expanded cells, and 'previously expanded, now exhausted' cells.

Previous studies found KIR were upregulated in TR γδ T cells, which we also found to be the case (e.g. the TR TCR65 γδ clonotype was the only cell to express *KIR3DL1*). Nearly three-quarters of the genes we observed to be significantly upregulated in TR γδ T cells (in **Fig. 4e**) exhibited higher expression in PD-1$^+$ cells compared to PD-1$^-$ cells derived from De Vries *et al.*, 2023, consistent with an increased presence of TR γδ T cells within the PD-1$^+$ fraction (**Revisions Fig. 1.10b**).



Revisions Fig.1.10. (a) Surface protein expression of PD-1 (CITEseq) in six γδ T cell clonotypes whose TCRs were validated by PreGame as tumor-reactive (TR), and one additional high scoring clonotype (untested). Tumor-reactivity was assessed by *in vitro* validation against six melanoma cell lines (**Extended Data Fig. 6g,h**). **(b)** Expression of significant differentially expressed genes in TR γδ T cells (from **Fig. 4e**) compared between *PDCD1*$^+$ (PD1$^+$) and *PDCD1*$^-$ (PD1$^-$) cells from De Vries *et al.,* 2023. Pie chart depicts the proportion of TR genes whose average expression was higher in *PDCD1*$^+$ cells compared to *PDCD1*$^-$ cells (orange) and vice versa (blue).

In addition to these many similarities, there are some variations that could be due to differences in the population of γδ T cells studied. In De Vries *et al.,* sorted PD-1$^+$ γδ T cells were found to be largely dependent on NKG2D for tumor recognition, while blocking the γδ TCR was described as having inconsistent inhibitory effects; these data suggest a mixture of TCR-reactive and NKG2D-reactive cells were examined there. In our work, PreGame was developed to specifically identify cells that are dependent on γδ TCR signaling.

7. More broadly, generalizability beyond MM is not always robustly supported by data. For such cases it is important to either provide supporting data or narrow the claims.

Reply:

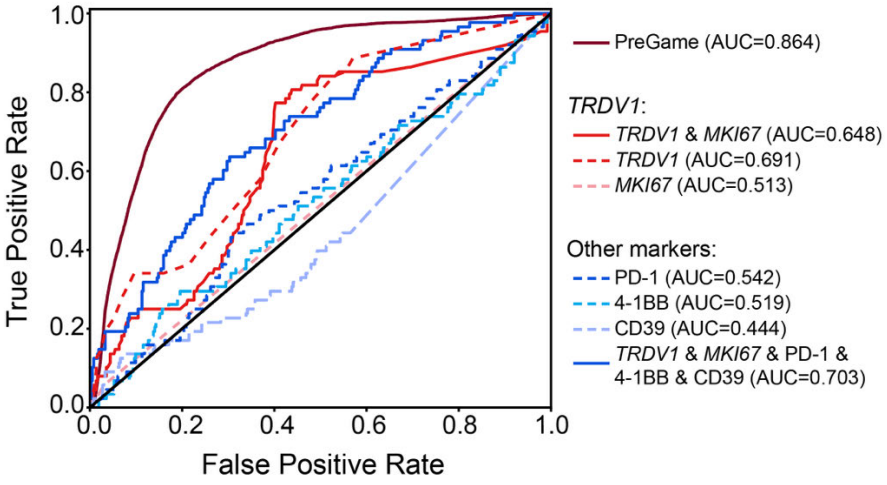
We have revised our text to narrow our claims as detailed above. We now state: “our results support the notion that PreGame can **assist** in distinguishing TR from by-stander γδ T cells in MM and in **some types** of solid tumors."

8. To understand the utility of PreGame, it is essential to show quantitative comparisons to simpler approaches. How does the “hit rate” of PreGame compare flow sorting based on established protein-level markers (e.g. ki67, Vδ1, CD39, 4-1BB, PD-1)? Or a simple combination such as ki67+ Vδ1 cells?

Reply:

We have now performed and presented additional receiver operating characteristic (ROC) analyses to address this important point. While the standard CITEseq antibody cocktail (used in our investigation) did not

contain Vδ1 or KI67, we examined PD-1, 4-1BB, and CD39 at the protein level, and *TRDV1* and *MKI67* at the gene expression level. Of these, *TRDV1* exhibited the highest AUC (AUC = 0.691), which was not improved when combined with *MKI67* (AUC = 0.513 alone; AUC = 0.648 in combination with *TRDV1*) (**Revisions Fig. 1.11**). The predictive performances of PD-1, 4-1BB, and CD39 alone were not impressive; however, in combination these three protein markers with *TRDV1* and *MKI67* achieved an AUC of 0.703 (**Revisions Fig. 1.11**). Still, none of these markers alone or in combination were able to outperform PreGame.



Revisions Fig.1.11. ROC curve comparing false positive and true positive rates at various thresholds for PreGame (as in **Fig. 3b**) and various other genes, proteins, and their combinations, as indicated. Single genes or proteins were modelled using normalized expression values, while combinations were modelled using random-forest classifiers as for PreGame. Higher AUC values indicate better overall performance.

9. The validation of PreGame vs competitors is not entirely fair: 10-fold cross validation (PreGame) vs completely unseen (competitor algorithms). This should be addressed by performing the comparison on the unseen set. Of note, a fair comparison of PreGame with simpler approaches (as suggested above) should also be performed on data of patients that were completely unseen during training.

Reply:

We agree with the Reviewer that the comparison is not perfect. To improve the fairness of the presented comparison, we have now updated **Fig. 3b** (copied below) to include the 95% confidence interval from cross-validation. We note that a fundamental goal of cross-validation is to estimate true (real-world) predictive performance. Even at the lower bound of the confidence interval PreGame outperforms existing algorithms.

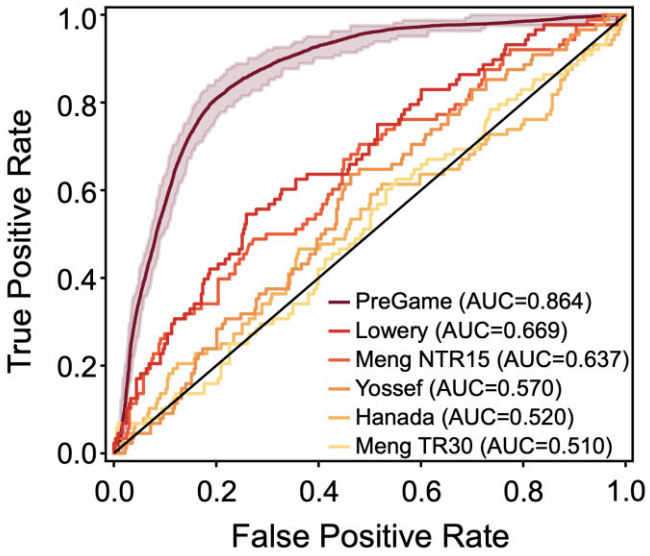
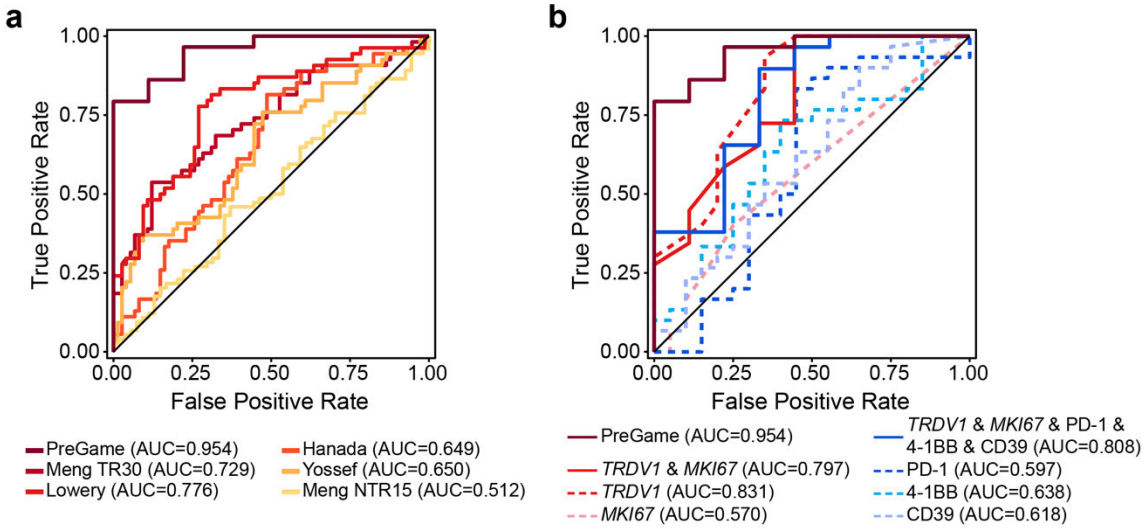


Fig. 3b. ROC curve comparing false positive and true positive rates at various thresholds for our model, PreGame, and several existing tumor-reactivity prediction models. The 95% confidence interval is shown for PreGame. A higher AUC value indicates overall better performance.

We also agree that it is important to compare each algorithm's performance on unseen data. Consistent with the performance observed in cross-validation, the performance of PreGame on completely unseen data was substantially better than other existing algorithms, with a robust AUC value of 0.954 compared to the next best algorithm at 0.776 (**Revisions Fig 1.12a**). Indeed, PreGame also outperforms the simpler approaches presented above on un-seen data (**Revisions Fig.1.12b**), consistent with PreGame's outperforming these approaches in cross-validation (**Revisions Fig.1.11**).



Revisions Fig.1.12. ROC curve comparing false positive and true positive rates at various thresholds for PreGame and **(a)** existing tumor-reactivity prediction models or **(b)** specific gene and protein markers and their combinations (as in **Revisions Fig.1.11**) on all un-seen validation data (excluding borderline TR γδ TCRs, which are found in **Extended Data Fig. 5e**).

The superior performance of PreGame in cross-validation and against unseen data is consistent with our conclusion that PreGame is superior to existing algorithms in predicting TR γδ T cells. However, it was also not

our intention to label these algorithms as 'competitors' since they were created for a different purpose, that is to predict TR αβ T cells, not TR γδ T cells. Accordingly, we have modified the tone of the relevant text as follows:

Initial manuscript: "To assess PreGame's performance, we benchmarked it against several recently published predictive algorithms trained on αβ TCRs."

Revised manuscript: "To assess **the utility of a bespoke γδ T cell prediction algorithm, we compared PreGame to existing TR αβ T cell prediction methods.**"

10. "We observed a significant positive correlation between γδ TCR repertoire diversity and the fraction of tumor-reactive γδ TCRs in circulation, suggesting that diversity can serve as a proxy for measuring the proportion of tumor-reactive γδ TCRs (Extended Data Fig. 8g,h)." This seems counter-intuitive as tumor-reactivity would be expected to drive clonal expansion and reduced diversity. Could this be confounded by higher γδ T cell levels in patients with tumor-reactive γδ TCRs, which in turn results in the detection of more clonotypes? More broadly, one could argue that comparing measures of diversity and clonal expansion between responders and nonresponders cannot be reliably performed because γδ T cells were virtually absent in nonresponders.

Reply:

This is an excellent point, and along similar lines to questions raised by the other Reviewers. We therefore performed substantial new analyses that we believe have significantly improved our revised manuscript. In this response we provide: (1) the related questions from the other Reviewers (for this Reviewer's benefit); (2) a summary of the new sequencing performed and samples added; (3) an in-depth description of our new analyses and results; and (4) a condensed summary of figure changes in the revised manuscript.

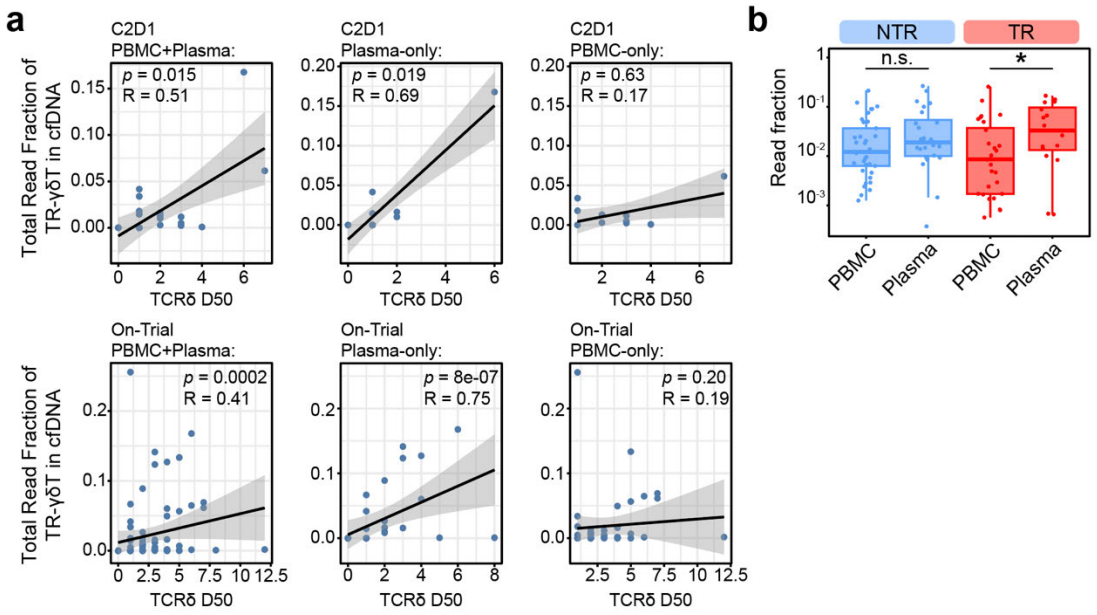
- Reviewer #2: *The authors state, "an increase in γδ, but not αβ, TCR diversity at C2D1 was significantly associated with longer PFS." It should be clarified whether this refers to an increase from baseline or higher TCR diversity within the cohort.*
- Reviewer #3 and #4: *The authors report (i) expanded γδ clones in BM mediating durable responses, (ii) increased BM–blood TCRδ overlap, and (iii) higher blood TCRδ diversity correlating with longer PFS. Taken together, these can appear contradictory. Please clarify*

To address these concerns, and advance our goal of reporting the most up-to-date results of the in-progress Algonquin clinical trial, we have now performed additional sequencing experiments and additional bioinformatic analyses while incorporating the most up-to-date clinical response data available. Specifically, we have now included capTCRseq on an additional 24 plasma cfDNA and 36 PBMC samples across 20 patients. Importantly, this additional capTCRseq data increases the number of patients for which we have matched PBMC and plasma data at C2D1 by seven patients. These additional samples afford increased power for comparisons of the plasma and PBMC TCR repertoires of Belamaf (BPd)-treated patients. Furthermore, we have now included additional scCITE+VDJseq data from the BM samples of an additional n = 2 patients (1 responder, 1 non-responder). Crucially, this additional sequencing brings our non-responder cohort up to n = 3 patients. The number of patients for which we can correlate BM and blood γδ TCR repertoires is now increased by two as well (to n = 13 patients total).

Indeed, we agree with the points raised by this Reviewer: patients with virtually absent γδ T cells in their BM should not have detectable levels of γδ TCRs in the circulation, consistent with a low-to-zero diversity readout.

On the other hand, patients with pronounced expansion of certain γδ TCR clonotypes in their BM should show more clonotypes in circulation (high diversity readout), or display a TCR repertoire dominated by that expanded clone (low diversity readout, but perhaps higher than baseline). In both cases, we would need to characterize the circulating γδ TCR repertoires of normal/healthy individuals and consider how patients' γδ TCR repertoires change between baseline and following treatment.

It was recently reported that CDR3 sequences of tumor-derived αβ TCR clonotypes were particularly enriched in cfDNA compared to PBMCs in patients receiving pembrolizumab (**PMID: 40875293**). In addition, higher levels of TCRs in the cfDNA of large B cell lymphoma patients receiving anti-CD19 CAR-T cells were found to correlate with improved response (**PMID: 36584673**). In our initial manuscript, we observed stronger correlations between γδ TCR diversity and the proportion of TR γδ TCRs in the circulation (initially **Extended Data Fig. 8g,h**, and **Revisions Fig.1.13a** below, now updated with additional sequencing data for the Reviewer's benefit) and a significantly higher read fraction of TR γδ TCRs in plasma cfDNA compared to PBMCs (initially **Extended Data Fig. 8e,** and **Revisions Fig.1.13b** below, again, updated with additional sequencing data). These findings all hint at a particularly informative role for plasma cfDNA as a proxy for TR TCR dynamics.

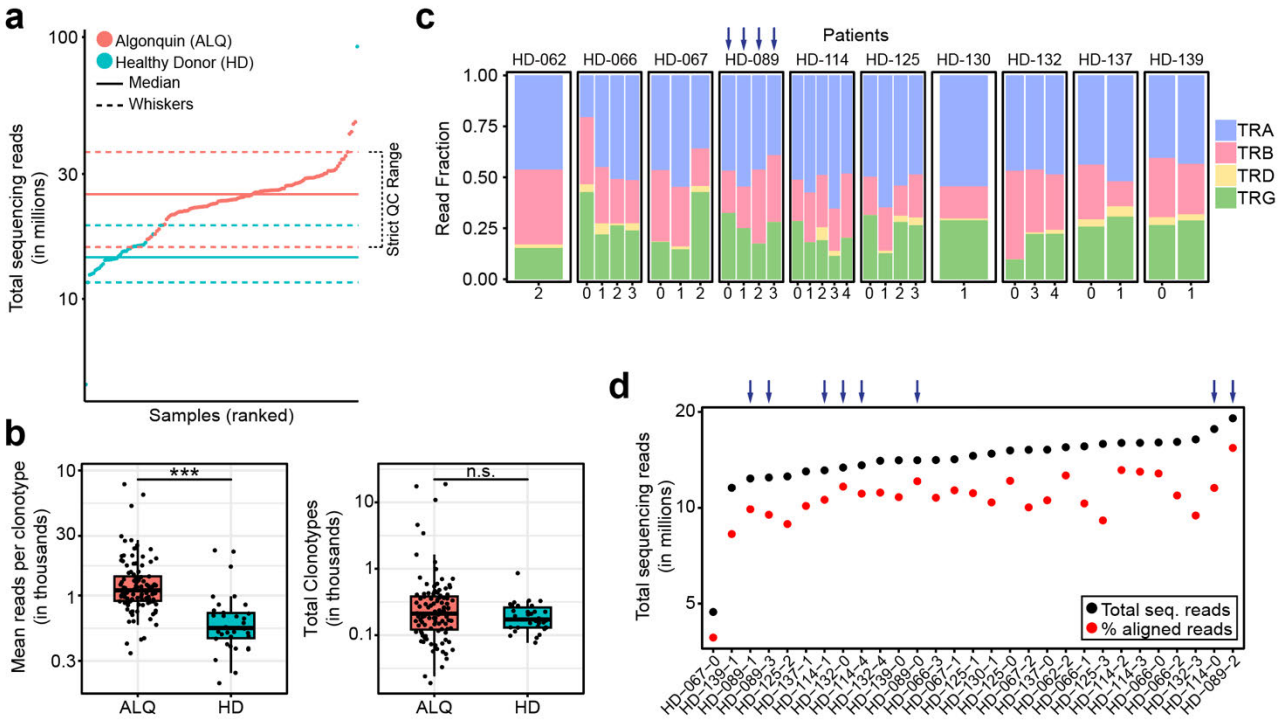


Revisions Fig.1.13. (a) Read fraction of TR γδ TCRs in PBMCs and plasma (left), plasma alone (middle), and PBMCs alone (right) at C2D1 (top) and all on-trial timepoints (bottom). Each point indicates a CapTCR-seq sample from a patient where scCITE+VDJseq was performed. Black lines denote Spearman correlation, and the 95% confidence interval is indicated in grey; Rho value and significance as determined by Spearman rank correlation tests are presented within the plot. **(b)** Read fraction of non tumor-reactive (NTR) and TR γδ TCRs in PBMCs and plasma cfDNA across on-trial. Each point indicates a CapTCR-seq sample from a patient where scCITE+VDJseq was performed. Significance was assessed using Wilcoxon rank-sum tests, *p = 0.049, n.s. not significant.

To delineate the relationship between BM expansion of TR γδ TCRs and cfDNA γδ TCR diversity read-outs, we examined previously generated, longitudinally-collected plasma cfDNA CapTCRseq data from a cohort of n = 10 healthy donors (HD) (**PMID: 40875293**). We re-processed the raw data using the methods described in our manuscript for the ALQ cohort, resulting in an HD cohort of 29 samples from these ten donors. It was of particular importance to assess and quality control (QC) any differences in sequencing depth, alignment, and total clonotypes between this HD cohort and our ALQ plasma cfDNA cohort to ensure that downstream comparisons

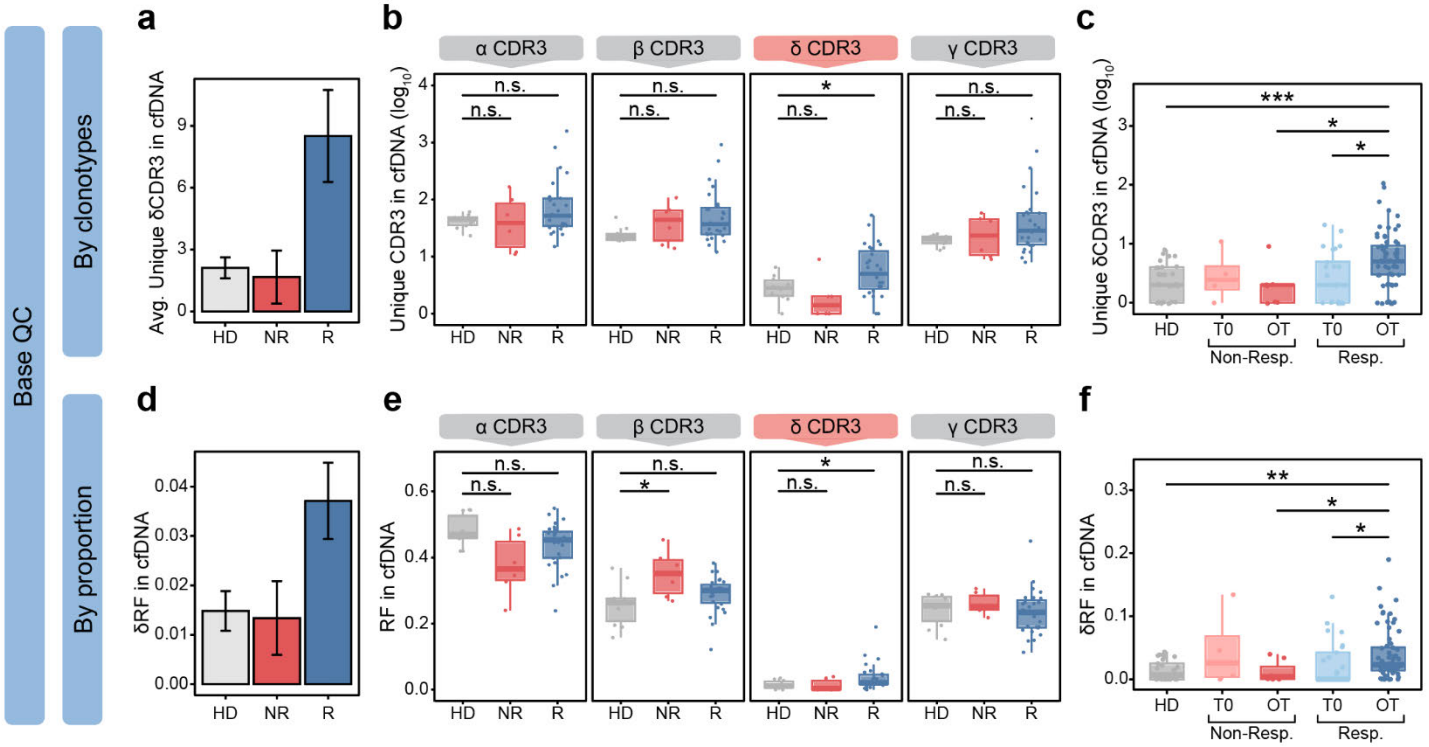
were technically robust. While the sequencing depth and average number of reads per clonotype tended to be higher in our ALQ cohort compared to the HD cohort, the resulting numbers of clonotypes were comparable (**Revisions Fig. 1.14a,b**). These results suggested that sequencing depth saturation could have been achieved in the HD cohort—the point at which increasing sequencing depth would not yield an increasing number of clonotypes. To be conservative, we opted to perform downstream analyses at two different QC thresholds, namely "Base" and "Strict" (**Revisions Fig. 1.14a, Strict QC range**). We show below that both QC methods led to very consistent results.

For the Base QC method, we analyzed 29 samples from all 10 donors and found that δ CDR3s comprised the lowest fraction of TCR repertoires (**Revisions Fig 1.14c**). One patient (HD-089) had a complete absence of δ CDR3 detection across all four longitudinal samples (**Revisions Fig 1.14c, arrows**). Eight samples (28%) had a complete absence of δ CDR3 detection; two of these, HD-089-T2 and HD-114-T0, had the first and second highest sequencing read depth, respectively (**Revisions Fig 1.14d, arrows**). These results suggested that the baseline cfDNA levels of γδ TCRs in the healthy population are very low, and that even at sequencing read depths comparable to that of the ALQ cohort, γδ TCRs can be virtually absent.



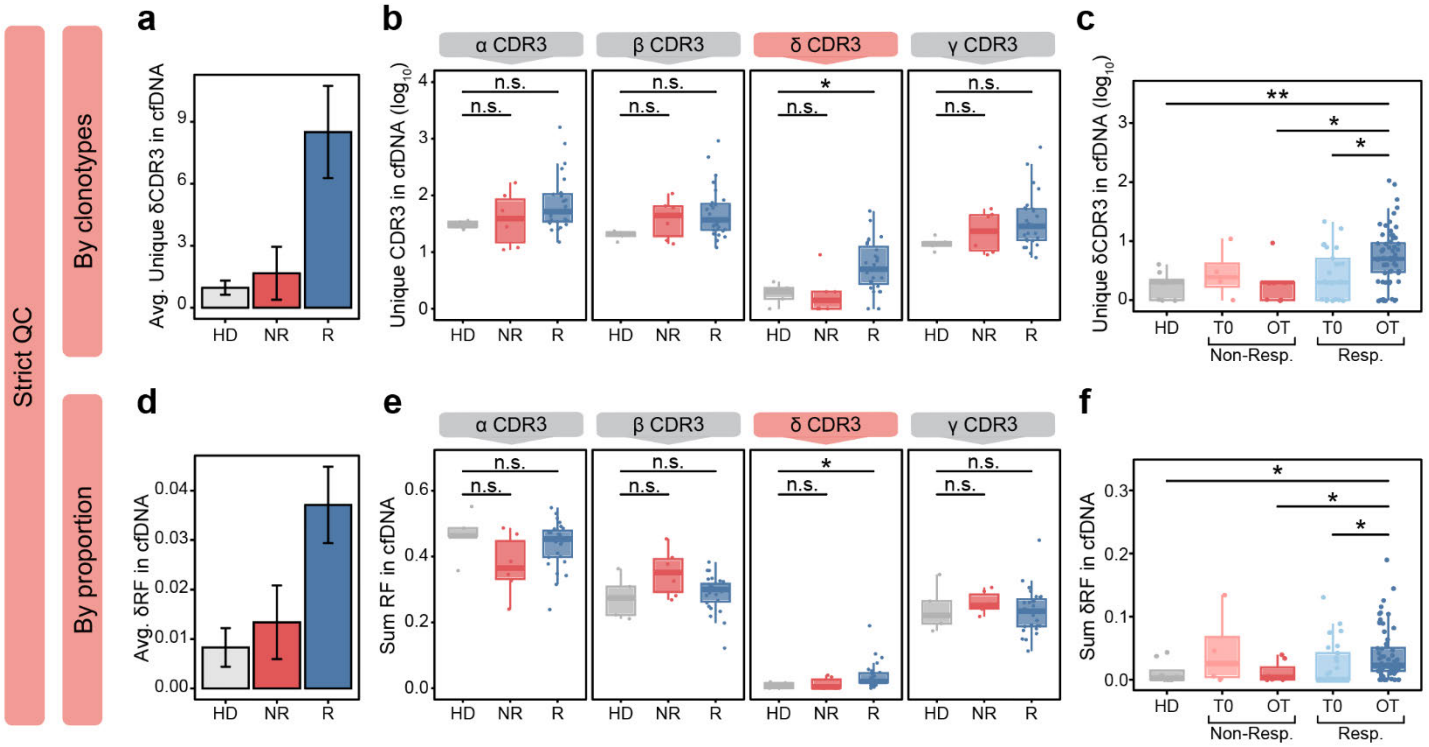
Revisions Fig.1.14. (a) Total sequencing reads obtained from CapTCR-seq performed on plasma cfDNA fractions. Points indicate samples and are colored by the sequencing cohort: either Algonquin (generated in this study) or healthy donors (HD) (previously generated in Soleimani *et al.*, 2025 PMID: **40875293**). The median number of sequencing reads per cohort is indicated with solid lines. The upper and lower whiskers per cohort, corresponding to the 75th and 25th percentile ± 1.5 times the IQR, respectively, are indicated with dashed lines. HD samples with total sequencing reads within the whiskers of the Algonquin cohort were used in the 'Strict QC' analysis described below. **(b)** Boxplots comparing the mean reads per clonotype (left) and total clonotypes (right) from CapTCR-seq samples between each cohort. Significance was assessed using Wilcoxon rank sum tests, ***p < 0.0005, n.s. not significant. **(c)** Proportions of the plasma cfDNA TCR repertoires occupied by each TCR chain in longitudinally collected samples from each of the n = 10 HDs. Arrows highlight the absence of detectable δ CDR3 sequences in four samples collected from HD-089. **(d)** Total sequencing reads (black points) and percentage of aligned reads (red) in CapTCR-seq samples from HDs. Arrows highlight samples devoid of δ CDR3 sequences, which are not strictly biased towards samples with lower sequencing depth.

To test the hypothesis that the presence of γδ TCRs in the cfDNA of BPd-treated MM patients is elevated compared to basal levels, we first compared the cfDNA TCR repertoires of HDs to that of responders (R; *n* = 27) and non-responders (NR; *n* = 6) from the ALQ trial. We found that Rs had nearly 6-fold higher numbers of unique δ CDR3 sequences (i.e. repertoire richness) in their cfDNA compared to both HDs and NRs (**Revisions Fig. 1.15a**). Strikingly, the δ TCR compartment was the only one to exhibit significant differences between Rs and HDs (**Revisions Fig. 1.15b**). More compelling still, this increase in cfDNA δ TCRs corresponded to treatment initiation and was only observed in Rs (**Revisions Fig. 1.15c**). In addition to this spike in δ TCR richness, we observed the same patterns when examining the proportional abundances (read fractions) of δ TCRs (**Revisions Fig. 1.15d-f**). These results are thus far consistent with the model proposed by this Reviewer; that is, *"higher γδ T cell levels in patients with TR γδ TCRs, in turn resulting in the detection of more clonotypes [compared to NRs where] γδ T cells were virtually absent."* By leveraging these HD samples, we can further elaborate on this model: that there is a surge of γδ TCRs in the cfDNA of Rs, rather than there being a decrease in NRs. This is an important distinction for the translation of these findings into a clinical prognostic test.



Revisions Fig.1.15. (a,d) Bar plots depicting the average **(a)** read fraction (RF) and **(d)** richness (number of unique CDR3 sequences) of δ CDR3 sequences in plasma cfDNA CapTCR-seq samples across healthy donors (HD) (*n* = 10), non-responders (NR) to BPd (*n* = 6), and responders (R) to BPd (*n* = 27). Longitudinally collected samples for each patient (excluding T0 and EOT samples for ALQ patients) were first summarized by calculating the mean of the total RF of Vδ1/3 CDR3s in **a**, and the mean richness in **b** across patient collections. Data are presented as the mean ± standard error. **(b,e)** Box plots comparing the average **(b)** RF and **(e)** richness, per TCR chain, from all longitudinally collected plasma cfDNA CapTCR-seq samples for each patient (excluding T0 and EOT samples for ALQ patients), as in **a**. Each point represents a patient. **(c,f)** Box plots comparing the plasma cfDNA **(e)** RF of Vδ1/3 CDR3s and **(f)** richness in all longitudinal samples from HDs, T0 (baseline) samples from NRs and Rs, and on-trial (OT) samples from NRs and Rs (i.e. all timepoints but excluding EOT samples), where CapTCR-seq data were available. Each point represents a blood collection. Significance was assessed by Kruskal-Wallis tests for each TCR chain followed by Wilcoxon rank-sum tests between groups, ***$p < 0.0005$, **$p < 0.005$, *$p < 0.05$, n.s. not significant.

For our second 'Strict QC' method, we set a more stringent QC threshold to ensure that our observations were not biased by the shallower sequencing read depth in HD samples. A 'Strict QC' threshold requiring that the total sequencing reads fell above the lower whisker of the ALQ cohort distribution was applied (**Revisions Fig. 1.16a, 'Strict QC range'**). This approach yielded a total of 8 samples from *n* = 5 healthy individuals. Despite using fewer samples, we observed strikingly consistent results (**Revisions Fig. 1.16a-f)**. Taken together, these data continue to support the model proposed by this Reviewer and suggest two more key take-aways: 1) any αβ TCR response elicited in response to BPd treatment was comparable in the cfDNA of Rs, NRs, and HDs alike; and 2) γδ TCR responses observed in the cfDNA appear to be attributable to BPd treatment initiation and a by-product of responsiveness to successful treatment.

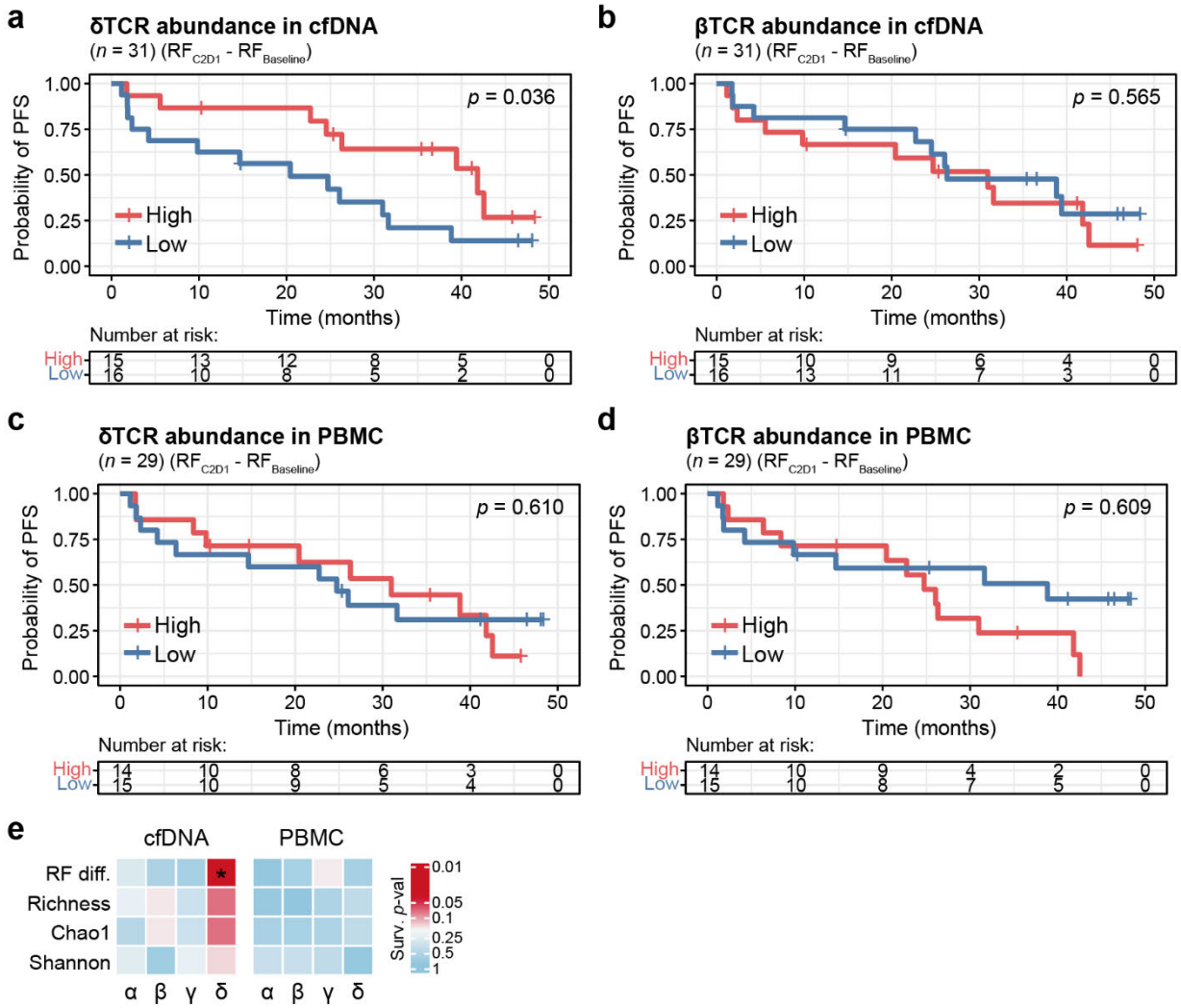


Revisions Fig.1.16. (a,d) Bar plots depicting the average **(a)** read fraction (RF) and **(d)** richness (number of unique CDR3 sequences) of δ CDR3 sequences in plasma cfDNA CapTCR-seq samples across healthy donors (HD) (*n* = 5), non-responders (NR) to BPd (*n* = 6), and responders (R) to BPd (*n* = 27). Longitudinally collected samples for each patient (excluding T0 and EOT samples for ALQ patients) were first summarized by calculating the mean of the total RF of Vδ1/3 CDR3s in **a**, and the mean richness in **b** across patient collections. Data are presented as the mean ± standard error. **(b,e)** Box plots comparing the average **(b)** RF and **(e)** richness, per TCR chain, from all longitudinally collected plasma cfDNA CapTCR-seq samples for each patient (excluding T0 and EOT samples for ALQ patients), as in **a**. Each point represents a patient. **(c,f)** Box plots comparing the plasma cfDNA **(e)** RF of Vδ1/3 CDR3s and **(f)** richness in all longitudinal samples from HDs, T0 (baseline) samples from NRs and Rs, and on-trial (OT) samples from NRs and Rs (i.e. all timepoints but excluding EOT samples), where CapTCR-seq data were available. Each point represents a blood collection. Significance was assessed by Kruskal-Wallis tests for each TCR chain followed by Wilcoxon rank-sum tests between groups, $^{**}p < 0.005$, $^{*}p < 0.05$, n.s. not significant.

These results collectively suggest that, as this Reviewer proposed, γδ TCRs are virtually absent in the cfDNA of both HDs and NRs to BPd. In Rs, there is a surge of γδ TCRs in the cfDNA both in terms of number of unique clonotypes (i.e. richness) and overall proportional abundance (read fraction), explaining our previous seemingly counter-intuitive observations of elevated diversity in circulation in parallel with expansion of γδ T cell clonotypes

in the BM. Noting that these increases in cfDNA richness and abundance went hand-in-hand, we wondered: 1) might the abundance of γδ TCRs in cfDNA at C2D1 recapitulate our initial PFS results but with a less counter-intuitive metric; 2) in relation to Reviewer #2's question, could it be that the *increase* in cfDNA γδ TCRs between baseline and C2D1 is more informative; and 3) is the plasma cfDNA compartment more informative as a biomarker than the PBMC compartment with respect to predicting PFS?

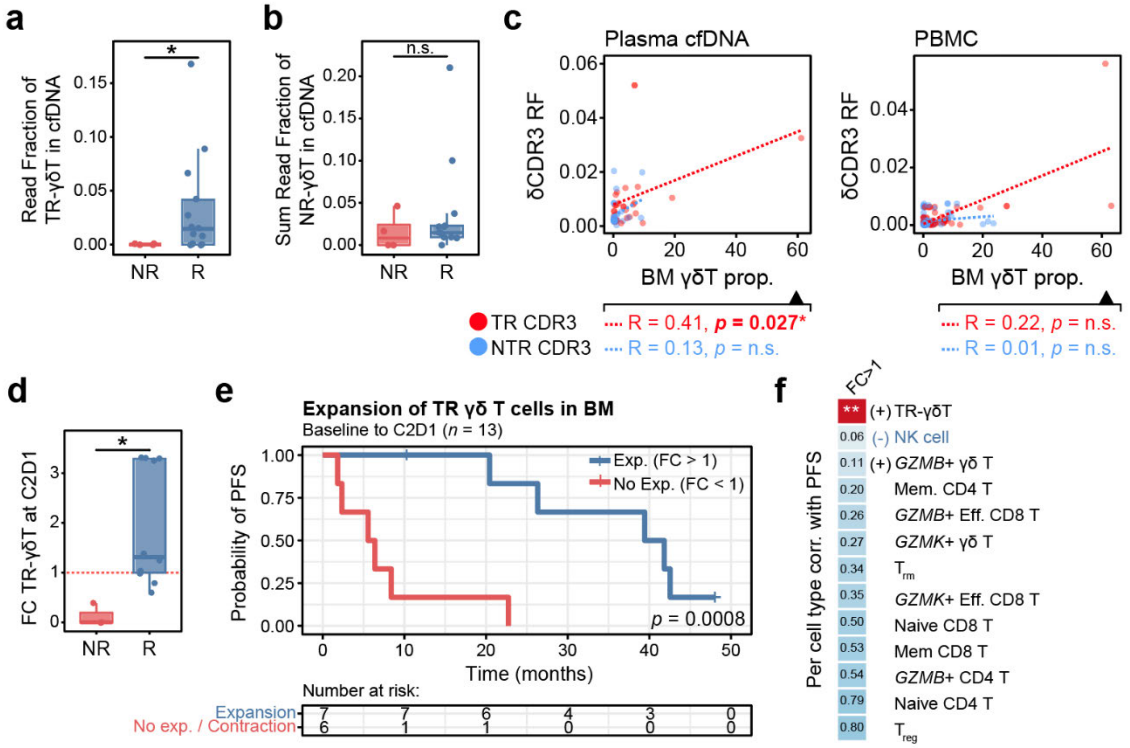
Crucially, our newly sequenced additions to the CapTCRseq cohort enabled us to perform these analyses in each compartment without losing too many patients to missing data. Consistent with our new findings above, we found that the feature correlating most strongly with favourable PFS was an increase in cfDNA δ TCR abundance between Baseline and C2D1 (**Revisions Fig. 1.17a**). The corresponding measurements were not significantly correlated with PFS for any of the other TCR chains (**Revisions Fig. 1.17b**); nor were any features of the PBMC repertoire (**Revisions Fig. 1.17c,d,e**). Richness of the cfDNA δ TCR repertoire at C2D1 ($p = 0.075$) narrowly missed significance at the $p < 0.05$ level, perhaps reflecting that increased diversity is secondary to an increased overall abundance of γδ TCRs in the cfDNA.



Revisions Fig.1.17. Kaplan-Meier survival curves showing PFS stratified by differences in **(a,c)** δTCR abundance or **(b,d)** βTCR abundance from baseline to cycle 2 day 1 (C2D1) for **(a,b)** *n* = 31 patients with available plasma CapTCR-seq data, and **(c,d)** *n* = 29 patients with available PBMC CapTCR-seq data. The abundance difference value was calculated as the total read fraction of CDR3 sequences at C2D1 minus the total read fraction of CDR3 sequences at baseline for the given TCR chain, where read fraction represents the proportional abundance of a given CDR3 sequence in the entire TCR repertoire. Patients were stratified into high and low abundance difference groups using the median value. Significance was assessed using Gehan-Breslow-Wilcoxon tests. **(e)** Heatmap depicting *p* values from Gehan-Breslow-Wilcoxon tests performed for the indicated metric and for the indicated TCR chain in plasma cfDNA samples (*n* = 31) and PBMC samples (*n* = 29) where CapTCR-seq data were available. **p* = 0.036, as depicted in **a**.

These new findings present a more intuitive model; increased abundance of γδ TCRs in circulation, which *may* coincide with elevated diversity, is a predictive biomarker of response to BPd treatment. However, the veracity of our proposed model still hinges on whether or not it is the expansion of TR γδ T cells in the BM driving this surge in γδ TCRs in the cfDNA. Consistent with our initial findings that Rs uniquely exhibited clonotypic expansion of TR γδ T cells (**Fig. 5a, Extended Data Fig. 8a**), we found that Rs also exhibited significantly higher proportions of TR, but not NTR, γδ TCRs in their cfDNA (**Revisions Fig. 1.18a,b**). We observed a significant correlation between the expansion level of TR γδ T cell clonotypes in the BM (i.e. frequency of cells with a given γδ TCR) and their respective CDR3s in the patient's cfDNA (**Revisions Fig. 1.18c**). Conversely, we did not observe that BM and cfDNA abundances were significantly correlated for NTR γδ TCR clonotypes, nor that significant correlations extended to the PBMC compartment (**Revisions Fig. 1.18c**). Whether these findings

simply reflect that TR γδ T cell clonotypes tend to expand following BPd treatment, or a more nuanced underlying biological mechanism, they add support to our claim that TR γδ T cells mediate responses to BPd across the broader clinical cohort (**Revisions Fig. 1.17a**). Indeed, with the addition of one more NR to our BM CITEseq cohort (now *n* = 3), we were able to quantify and observe that responders were significantly more likely to exhibit expansion of TR γδ T cells in their BM following BPd treatment (**Revisions Fig. 1.18d**). Strikingly, patients that exhibited an expansion of TR γδ T cells between baseline and C2D1 enjoyed a significantly and dramatically increased PFS, from a median PFS of just over 5 months to a median PFS of close to 40 months (**Revisions Fig. 1.18e**). TR γδ T cells were the only T cell subset that demonstrated a significant correlation with PFS (**Revisions Fig. 1.18f**). These results strongly suggest that TR γδ T cells play a crucial role in mediating long-term, durable responses to BPd.



Revisions Fig.1.18. Comparison of the total plasma cfDNA read fraction (RF) of δ CDR3 sequences from C2D1 or C6D1 that were detected in the patient-matched BM and labelled as **(a)** TR or **(b)** NR between clinical response groups. Significance was assessed using a Wilcoxon rank-sum test, **p* = 0.036, n.s. not significant. **(c)** Correlation between γδ TCR clonotype frequency in the BM and the RF of the corresponding δ CDR3 sequence in stage- and patient-matched plasma cfDNA (left) or PBMCs (right). Each point represents a unique δ CDR3 sequence detected in both compartments across *n* = 13 MM patients, where data were available. Points are colored by tumor-reactivity. Dashed lines show Spearman correlations for the TR and NTR subsets. Rho values and significance as determined by Spearman rank correlation tests are presented below each plot. **(d)** Fold-change (FC) of the proportion of TR γδ T cells, compared to total γδ T cells, between baseline and C2D1 for *n* = 3 NRs and *n* = 10 Rs. A fold-change > 1 indicates an increased proportion of TR γδ T cells at C2D1, which is indicated with a dashed red line. Significance was assessed using a Wilcoxon rank-sum test, **p* = 0.014. **(e)** Kaplan-Meier survival curve showing PFS stratified by FC in TR γδ T cells as determined in **d**. Patients were stratified by FC > 1 or FC ≤ 1, corresponding to expansion or no expansion of TR γδ T cells, respectively. This stratification was chosen for biological significance but also corresponds to median stratification. Significance was assessed using a log rank test. **(f)** Heatmap depicting *p* values from log rank tests performed as in **e** for each of the indicated T and NK cell subsets. Directionality is depicted as (+) representing correlation with longer PFS and (-) representing correlation with shorter PFS. Throughout, the tumor-reactivity of γδ T cells and clonotypes was assigned using PreGame (≥0.85) or *in vitro* screening results (**Extended Data Fig. 8c**), where applicable.

In summary, these findings are consistent with recent reports and support the idea that plasma cfDNA represents a proxy for the dynamics of TR TCR clonotypes (**PMID: 40875293**). We have now shown that, under normal healthy conditions, plasma cfDNA is relatively devoid of γδ TCRs. Following clinical responses to BPd, the plasma cfDNA becomes flooded with γδ TCRs that are predominantly TR. This surge can be exploited to predict PFS as early as 4 weeks following treatment initiation, but our findings do not preclude the possibility that the surge could be detected even sooner. Mechanistically, these new findings strongly suggest that the expansion of TR γδ T cell clonotypes drives long-term, durable responses to BPd treatment. Finally, these results demonstrate the usefulness of PreGame in a 'real-world' clinical setting.

We would like to thank all three Reviewers for their thoughtful questions on this counter-intuitive aspect of our original manuscript, and hope that we have clarified the matter to their satisfaction. In so doing, we believe we have significantly strengthened our revised manuscript.

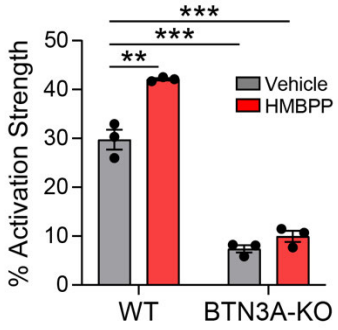
To summarize, please see below the list of changes to **Figure 1** and related Extended Data Figures:

Initial Manuscript	**Revised Manuscript**	**Remark**
Fig. 1e	Extended Data Fig. 3b	Updated with new patient samples.
Fig. 1f	Removed	No longer relevant and asked to remove/replace by Reviewer 3 and 4.
Fig. 1g	Fig. 1c	Now stratifying patients by abundance instead of diversity.
Fig. 1h	Extended Data Fig. 1c	Now stratifying patients by abundance instead of diversity.
Extended Data Fig. 2a	Extended Data Fig. 3a	Updated with new patient samples and added side-by-side γδ TCR repertoire alluvial plots per Reviewer 3 and 4.
Extended Data Fig. 2b	Removed	No longer relevant and asked to remove/replace by Reviewer 3 and 4.
Extended Data Fig. 2c	Removed	No longer relevant and asked to remove/replace by Reviewer 3 and 4.
Extended Data Fig. 2d	Extended Data Fig. 3c	Updated with new patient samples.
Extended Data Fig. 3a	Extended Data Fig. 1b	Updated with most recent patient response data.
Extended Data Fig. 3b-d	Replaced with Extended Data Fig. 1d-f.	Distinction between cfDNA and PBMC repertoires now more relevant.
-	Extended Data Fig. 2a-f	New
-	Extended Data Fig. 3d	New

11. To test if the cloned γδ TCRs recognize phosphoantigens presented by BTN3A1, the authors knockout BTN3A1 and found no effect (extended data fig 9a). In cases of Vδ2Vγ9 TCRs this is surprising as it is well established that Vδ2Vγ9 TCRs recognize phosphoantigens in the context of butyrophilins. To confirm that phosphoantigens are indeed irrelevant for these TCRs, it would be important to add the phosphoantigen (e.g. HMBPP or IPP) to the culture, and repeat similar analyses (KO & phosphoantigen addition) in a second model system (ideally one that previously has been used to demonstrate BTN3A1-based activation of Vδ2Vγ9 T cells, to proof the claim that these T cells are fundamentally different).

Reply:

The Reviewer makes an important point. Upon conducting the above experiments, we found that, although our BTN3 KO cells were flow-sorted to nearly 100% purity at the time of cell banking, over time in cell culture post-thaw, we noticed that a population of BTN3A1$^+$ tumor cells was growing out. These BTN3A1$^+$ cells could have escaped the original flow-sort if BTN3A1 is not constitutively expressed in RPMI-8226 cells. If true, some cells could have been BTN3A1-negative at the time of sorting despite not having a deletion of BTN3. Given that these emerging BTN3A1$^+$ cells may have potentially confounded our results, we re-sorted the cells to ensure extremely high purity **(Extended Data Fig. 9f)** and repeated the co-culture experiments using minimally passaged sorted cells to limit the potential for another emergence of a BTN3A1$^+$ population. Using these newly re-purified BTN3KO cells, we found that TCR11 was no longer able to respond to KO tumor cells. Moreover, we observed that the addition of HMBPP enhanced the response to TCR11 in WT tumor cells but not in BTN3A KO cells. Together, these findings indicate that the Vδ2Vγ9 TCR11 recognizes phosphoantigens similarly to the Vδ2Vγ9 T cells found in the blood. We thank the Reviewer for highlighting this inconsistency, which has led to a deeper characterization of TCR11. We have updated the text of the revised manuscript and added the panel below to **Extended Data Fig. 9b**.



Extended Data Fig. 9b. Percent activation strength for the Vδ2Vγ9 TCR11 cultured with WT or BTN3A KO RPMI-8226 cells treated with or without HMBPP (1 μM). Results shown are pooled from three independent experiments. Significance was assessed using ANOVA, $^{**}p < 0.01$, $^{***}p < 0.001$.

12. In general, it would be important to provide evidence that the knockouts (BTN3A1, B2M, HLA's) and overexpression (CD1D) perturbations were successful.

Reply:

We have now added these data as **Extended Data Fig. 9f,g.**

13. Fig 5: the authors seem to overinterpret these data, as they mainly show data suggestive of different γδ T cell dynamics in responders vs nonresponders, but do not provide a predictive biomarker.

Reply:

We have performed new analyses (and now with a larger clinical patient cohort) as requested by other Reviewers. We believe the results substantially increase support for our original claim: that expansion of TR γδ T cells is a driver of long-term BPd response. These data are presented in our revised manuscript as **Fig. 5b-d**, (copied below for the Reviewer's benefit). We have also toned-down language related to 'predicting' and 'mediating' responses throughout the revised manuscript.

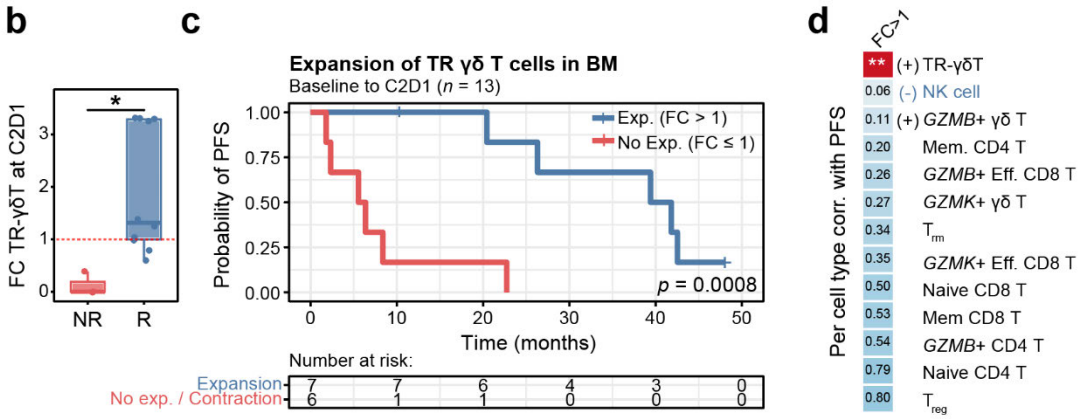


Figure 5. (b) Fold-change (FC) of the proportion of TR γδ T cells compared to total γδ T cells, between baseline and C2D1, for *n* = 3 NRs and *n* = 10 Rs. A fold-change > 1 indicates an increased proportion of TR γδ T cells at C2D1, which is indicated with a dashed red line. Significance was assessed using a Wilcoxon rank-sum test, **p* = 0.014. **(c)** Kaplan-Meier survival curve showing PFS stratified by FC > 1 or FC ≤ 1 (as determined in **b**) corresponding to expansion or no expansion of TR γδ T cells, respectively. This stratification was chosen for biological significance but also corresponds to median stratification. Significance was assessed using a log rank test. **(d)** Heatmap depicting *p* values from log rank tests performed as in **b** for the indicated T and NK cell subsets. Directionality is depicted as (+) representing correlation with longer PFS and (-) representing correlation with shorter PFS. Throughout, tumor-reactivity of γδ T cells and clonotypes was assigned using PreGame (≥0.85) or *in vitro* screening results (**Extended Data Fig. 8c**), where applicable.

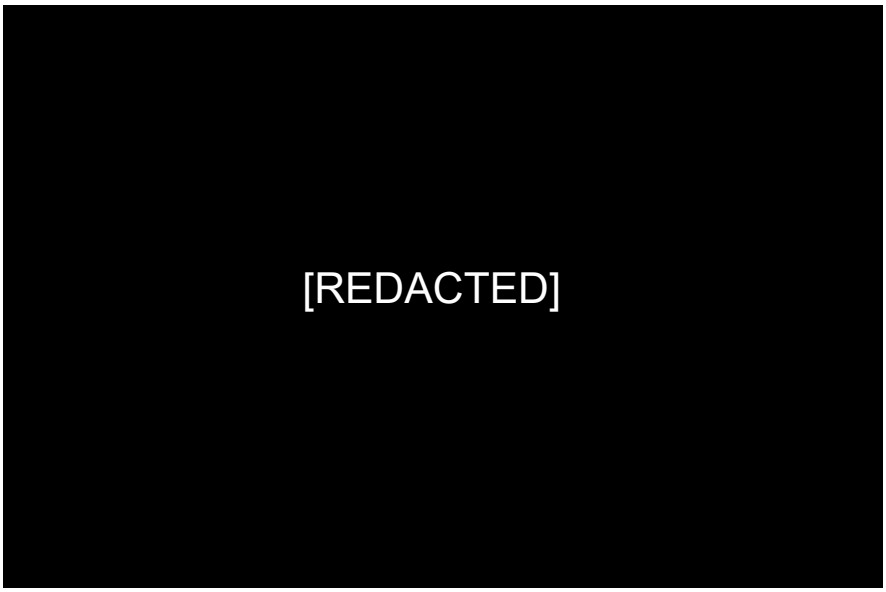
14. Regarding Fig 5e the authors mention strong correlation between tumor burden and clonal expansion of a reactive γδ T cell clone. However, it seems counterintuitive that the tumor burden is already 0 at timepoint C2D1 and the γδ T expansion only occurs two timepoints later (C12D1). How is this possible? Also, the conclusion that this patient achieved long term tumor control due to this γδ T cell clone seems overstated.

Reply:

We agree with the Reviewer's assessment and have conduction additional analysis to address these concerns. Given our new findings on the particular importance of the plasma cfDNA compartment (especially when compared to the PBMC compartment), we have split what was originally combined as 'Blood' into specific lines for plasma and PBMCs. Moreover, given this Reviewer's justified concern about our statement that long-term tumor control was due to a single TR γδ TCR clonotype (TCRg8), we have now replaced TCRg8 with the total proportion of TR γδ T clonotypes (**Revisions Fig. 1.19**).

These changes revealed the strikingly different γδ TCR dynamics between the PBMC and plasma cfDNA compartments. At baseline, the PBMC compartment in one patient (ALQ-17-010) had a high fraction of TR γδ TCRs, with nearly 15% of PBMCs being TCRg9 γδ T cells (**Revisions Fig. 1.19a,** light blue dashed line). At C2D1, 28 days following BPd treatment, this TCRg9 exhibited a massive decline, both in the circulation and in the biopsied BM (**Revisions Fig. 1.19a**). This steep decline of a very large TR γδ T cell clonotype resulted in this patient being assigned to the FC<1 group of **Fig. 5c**.

However, when we separated TCRg9 from the other TR γδ TCRs that we identified using PreGame (e.g. TCRg8, which was previously highlighted in **Fig. 5e**), we observed that there was in fact expansion of other TR γδ T cells in this patient's BM, indicative of response. This BM expansion was also reflected by a huge spike in the patient's plasma cfDNA at C2D1 (**Revisions Fig. 1.19b**). These increases in TR γδ TCRs coincided with a steep drop in tumor burden (plasma ctDNA) to below detectable levels. We hope our revised presentation of this data is now more convincing.



The detection of TR γδ TCRs in the PBMC compartment displays a significant lag behind that of the cfDNA (2-3 cycles), suggesting that plasma cfDNA is a more immediate reservoir for detection of BM TR γδ TCRs, consistent with the biomarker findings of **Fig. 1c**. Importantly, we know this patient had a PFS period that was: a) much longer than that of the next best patient in the FC<1 group (22.7 vs 8.3 months); and b) more reminiscent of the FC>1 group, in which PFS values were all over 20 months (**Fig. 5c**). We believe that these results do support our now revised conclusion that patient ALQ-17-010 benefited from the expansion of TR γδ T cells. Finally, it is noteworthy that, despite the drop-off in TCRg9 in this patient's plasma at C2D1, there was still a net increase in the proportion of TR γδ TCRs (12.2 to 16.8% read fraction), which saw this patient land in the High group of **Fig. 1c**. From the point of view of a prognostic test, it is therefore encouraging that the 'masking' effect of TCRg9 did not have as material of an effect on the plasma cfDNA compartment as it did on the PBMC and the BM compartments.

We again thank the Reviewer for this thoughtful question, which we hope has now been answered with more convincing results. Ultimately, due to space constraints we have decided to remove this from the manuscript. However if the Reviewer disagrees and would like us to include it in the manuscript we are happy to do so.

15. The clinical cohort has a very low number of nonresponding MM pts, limiting the level of evidence, which should be acknowledged in the claims.

Reply:

We acknowledge that there is a relatively low number of non-responding MM patients in this study, and agree that this limits the level of evidence, and point out that we are restricted in this aspect by the high ORR of BPd (**PMID: 38177852**). With the previously stated addition of 50 more CapTCRseq samples (across *n* = 20 patients) and additional scCITE+VDJseq samples (across *n* = 2 patients), we have now sequenced baseline and

C2D1 for *n* = 5 (71%) NR patients at the CapTCRseq level, and *n* = 3 (43%) at C2D1 at the single-cell level (**Extended Data Fig. 1a**). Having *n* = 3 NRs at C2D1 affords us increased power to test for statistically significant differences between Rs and NRs at C2D1, and we now provide statistical significance in **Figs. 1d, 5b,e**.

We have added wording to our revised Discussion to acknowledge this limitation: "Because the high response rate to BPd in the ALGONQUIN trial **led to a limited number of NR BM samples**, and because an alternative combination of Belamaf with bortezomib (BVd) has demonstrated similar response durability, **future studies should be aimed at validating our findings in larger MM cohorts that include patients treated with BVd and potentially other anti-MM therapies**."

16. Belamaf was granted accelerated regulatory approval in 2020 but was withdrawn from the market in 2022 when the pivotal trial missed its primary endpoint. This limits the relevance of the clinical biomarker findings (Figure 1 and Figure 5).

Reply:

Although BPd was withdrawn from the US market in 2022, it was subsequently re-approved for the US market by the FDA in 2025. Moreover, BPd has been approved and is available in many other jurisdictions, including: 1) The European Union (EMA): Approved in July 2025, 2) Canada: Approved by Health Canada in July 2025, 3) United Kingdom (MHRA): Authorized in April 2025, 4) Japan: Approved in May 2025, 5) Approved in Switzerland, Brazil, and most recently in the UAE. In Canada alone, over 100 MM patients have received BPd since its approval in July 2025. Given its worldwide regulatory approval and clinical utilization over the past year, we believe the identification of a clinical biomarker (Fig. 1c) is indeed relevant. Moreover, our downstream insights into the mechanisms underlying long-term, durable responses to BPd are highly relevant because they point towards a potential role for γδ T cells in mediating response to ADCs in general.

17. The observation that Vγ9 TCRs may recognize B2M-dependent targets is interesting. However, it is unclear why in figure 6A the authors classified TCR v10 as a B2M-dependent TCR, as the fold activation change with B2M knockout is ~0? And related: for this TCR the reactivity strongly increased with CD1D overexpression, but since CD1D is B2M-dependent it seems counterintuitive that B2M knockout has no effect? How could these seemingly contradicting observations be explained?

Reply:

We apologize for the confusion. We classified this TCRv10 as B2M-dependent because it recognized the RPMI-8226 MM cell lines that overexpressed CD1D, which is a B2M-dependent molecule. The B2M knockout was done using WT cells and so it was expected that no reactivity would be observed for either the WT or the B2M KO cells (the WT cells do not express CD1D). We have now clarified in our revised Materials and Methods that the KO was done in WT cells.

18. I do believe that the authors overinterpret their data of the proposed "logic switch" of the γδ TCR + KIR3DL1. These data are intriguing as it shows that such a logic switch could be created by overexpressing these receptors in Jurkat cells, which has potential implications for the design of cellular therapies. However, there is no proof showing that this logic gate is actually being used by real γδ T cells for tumor recognition of "missing self" sensing. Indeed, γδ T cells express a whole repertoire of both activating and inhibiting HLA-specific receptors (e.g. inhibitory KIRs, activating KIRs, NKG2A, NKG2C, LILRBs, etc), all with different affinities for different HLA alleles. Hence, the interactions of γδ T cells with the HLA's of tumor cells are highly complex and the net effect of different HLA balances is unclear. To proof that this "logic gate" is a mechanism

defining tumor-reactivity of γδ T cells, strong functional data with nonengineered/patient-derived γδ T cells are needed. Alternatively, the authors could narrow the claim and put the emphasis on the ability to synthesize logic gates for cellular therapies with these receptors, which is exciting on itself but is a narrower biological interpretation.

Reply:

We agree with the Reviewer and have now reframed the text to address these concerns. Instead of proposing a KIR-dependent logic gate as being the gatekeeper of autoreactivity, we instead propose that KIR-dependent signaling "may be one component of a regulatory mechanisms required to safely target certain 'self' antigens." We also note that this regulatory mechanism is likely to be highly complex, with KIRs being only one component. Lastly, we have included a sentence highlighting its potential clinical use in the context of cell therapies that states "these data also position KIR3DL1 as a logic-gate that could be leveraged to enhance the safety of certain types of cell therapies."

Minor points:

19. More patient characteristics and details of the clinical cohorts are necessary.

We have included all clinical information available to us in the current manuscript. If the Reviewer has specific additional patient characteristics in mind, we can consult with the treating clinicians to determine whether these details can be added.

20. Figure 1f – the significance in not readable in the figure, similar as extended figure 2b-c

These figures have been removed from the manuscript due to multiple Reviewers' comments. However, we have made the significance more readable in all newly added or revised figures.

21. In Fig 2e it is unclear to what extent the tumor reactivity is driven by CD1D overexpression by the lines.

Of the 22 TCRs that we tested to determine whether they react to WT and/or CD1D OE cells (**Fig. 6a, Extended Data Fig. 9a**), we found that 21 TCRs recognized WT cells and therefore were CD1D-independent. Based on this result, we believe that tumor-reactivity resulting from overexpression of CD1D accounts for only a small minority of TCRs that recognize a CD1D-dependent ligand.

22. In Fig 3c-d it is confusing that different scales have been used for the PreGame score.

We have now harmonized the scale bar so that the same scale is used in both **Fig. 3c** and **Fig. 3d**.

23. When defining tumor reactivity, why did the authors take a cut-off of 13%? This cut-off should be clarified. If up until 13% is background, then an increase from 13% to 15% may not be major recognition by the TCR clone.

We agree with the Reviewer and have updated the text to revise the cut-offs. We now indicate that >15% is classified as tumor-reactive, 13-15% is classified as borderline tumor-reactive, and <13% is considered non-reactive. We clarify that all figures involving bioinformatic analysis of tumor-reactive vs non-reactive T cells (i.e. **Fig. 3** and onward) are referring to those with >15% activation vs. <13% activation, respectively.

24. In legends showing cytotoxicity assays, it would help to precise the readout.

We have updated the legends to indicate that tumor cell lines expressing luciferase were used, and that cytotoxicity was determined by measuring differences in luminescence between tumor cells cultured with or without T cells.

25. In general, the figure legends are not always complete and extensive enough to understand the figure – for example in figure legend 4 is never mentioned what type of data the authors are showing (from the text it seems to be CITE-seq but that is not shown in the figure/legend itself).

We apologize for our lack of detail and now clarify in the **Figure 4** legend that it is based on scCITE+VDJ-seq data. We have also further clarified legends throughout the manuscript regarding data usage.

26. Colors of figure 4f are unclear.

We have now adjusted the colour scheme for this figure panel.

Reviewer #2 (Remarks to the Author):

In this manuscript, St Paul et al. analyze a Multiple Myeloma (MM) patient cohort from which they isolate bone marrow–infiltrating T cells to perform single-cell TCR and RNA sequencing, together with CITE sequencing. The authors experimentally test the reactivity of selected γδ TCRs by cloning them into Nur77-GFP Jurkat cells. The experimentally validated clones are then used to train a machine learning algorithm that predicts tumor reactivity of γδ T cells based on single-cell CITE-seq data. The trained model is subsequently applied to clonotypes that were not experimentally tested, with validation experiments demonstrating good predictive performance. Moreover, the authors use this algorithm to suggest that expansion of tumor-reactive γδ T cells serves as a biomarker of therapeutic response in MM patients treated with belantamab mafadotin. Finally, they identify a γδ TCR epitope in HLA-C and KIR3DL1 as an important inhibitory receptor that prevents recognition of healthy tissue.

Overall, this manuscript is innovative, methodologically strong, and presents original and clinically relevant findings, with significance extending beyond the γδ T-cell community to the broader cancer immunotherapy field. The work is elegant and well-executed, employing cutting-edge techniques and valuable clinical samples. However, several key conclusions are insufficiently supported or not clearly explained, and additional analyses, improved data interpretation, figure reorganization, and experimental validation are needed to strengthen the manuscript's impact before it can be considered for publication.

Reply:

We thank the Reviewer for these kind words and enthusiasm regarding the methodological strength, elegance, and broad impact of our manuscript.

SPECIFIC ISSUES

Figure 1
Claim: γδ T cells mediate durable responses to ADC combination therapy

• No direct evidence of causality is provided. To support this claim, mechanistic evidence demonstrating mediation is required.

Reply:

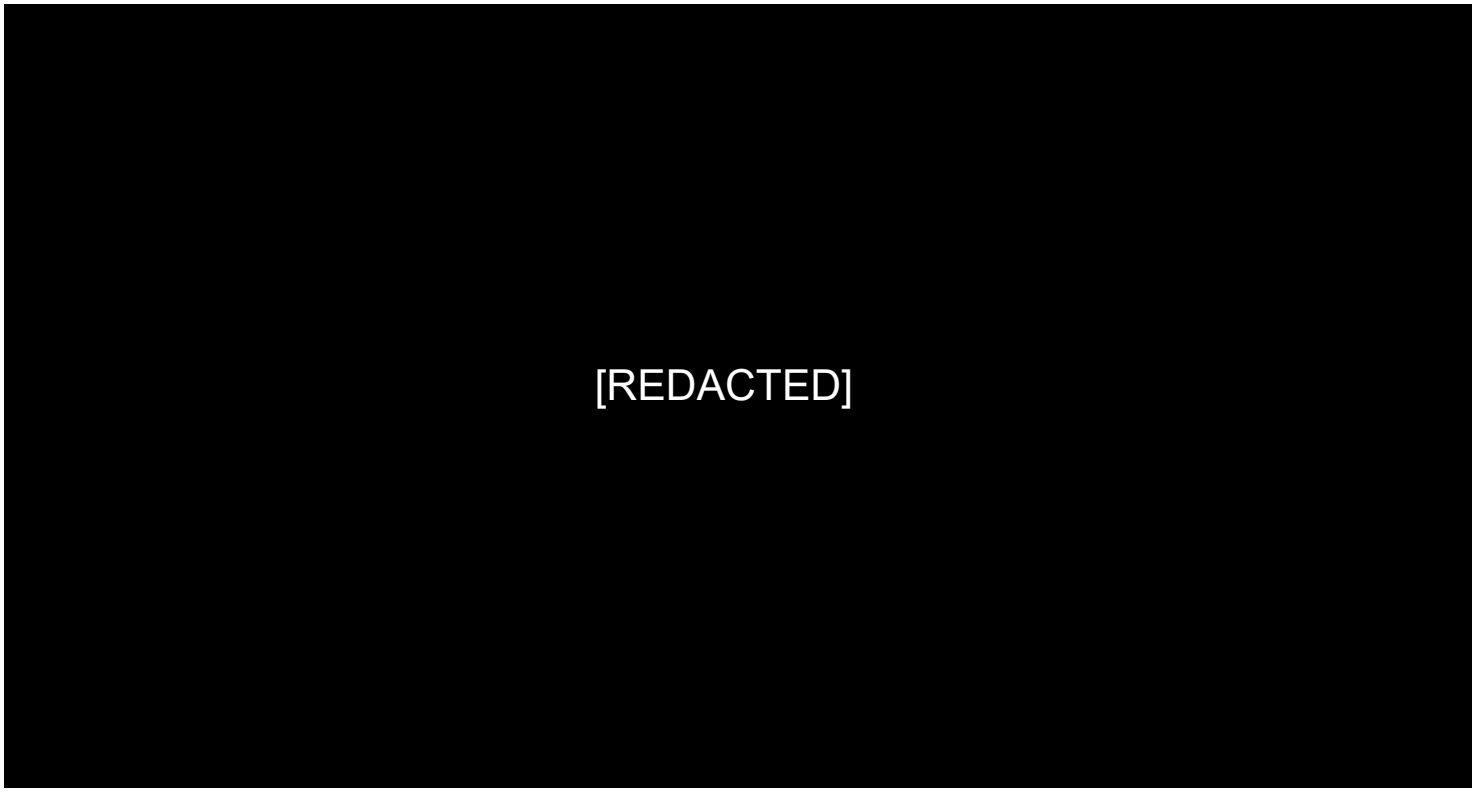
We agree with the Reviewer that in **Fig. 1** we have not provided mechanistic evidence to conclude that γδ T cells mediate responses to Belamaf + Pd (BPd). We have now adjusted the text accordingly to indicate that the presence of γδ T cells correlates with responses to BPd. Notably, we have now performed additional experiments and computational analyses that we believe strengthen our claim that γδ T cells (specifically tumor-reactive (TR) γδ T cells) are involved in mediating responses to BPd; these data are presented further along in the revised manuscript. We now show in that TR γδ T cells can be flow-sorted using PreGame protein markers (**Fig. 4g,h**), and that expansion of TR γδ T cells in patients' BM after only one treatment cycle strongly predicts long-term PFS (**Fig. 5c**).

The purpose of Fig. 1c is unclear. Extended Fig. 1d appears more informative in showing that γδ T cells ($GzmB^+$ and $GzmK^+$) are more frequent and associated with response. The authors should clarify how the frequency of $GzmB^+$ and $GzmK^+$ γδ T cells correlates with PFS, and include $GzmB^+$ $CD4^+$ T cells for comparison.

Reply:

We thank the Reviewer for this constructive feedback, which is similar to that of Reviewer #3 and #4. We have now performed additional sequencing and computational analyses that have resulted in a substantially rearranged (and hopefully more compelling) **Figure 1** in our revised manuscript. Specific changes are:

- We have now moved **Fig. 1c** out of the main figures and into **Extended Data Fig. 2h,i**, where we provide additional detail regarding the distinction between *GZMB*$^+$ and *GZMK*$^+$ γδ T cells, as requested below.
- At the request of Reviewer #3 and #4, we now include a quantitative analysis (in place of the original **Fig. 1d** panel) as **Fig. 1d** and **Extended Data Fig. 2k** in our revised manuscript. We hope this change will address this Reviewers concern as well, and have copied this new **Fig. 1d** below for the Reviewer's benefit.
- We have removed the original **Extended Data Fig. 1d**, since the new **Fig. 1d** and **Extended Data Fig. 2k** now provide a quantitative analysis of the same data. However, if the Reviewer feels this figure was helpful, we are happy to add it back in.



To the Reviewer's point about cellular correlates of PFS from our BM scCITE+VDJseq cohort, we have now performed additional survival analyses to investigate this important aspect. Notably, we have now included additional scCITE+VDJseq data from *n* = 2 patients (1 responder and 1 non-responder to BPd) in the revised manuscript. As suggested, we examined whether BM proportions of *GZMB*$^+$ γδ T cells, *GZMK*$^+$ γδ T cells, and

GZMB⁺ CD4⁺ T cells correlated with PFS in *n* = 13 patients at baseline, C2D1, and the change between timepoints. We did not find significant ($p < 0.05$) correlations, but, and as suggested by this Reviewer below, we did observe a particularly striking difference in PFS between patients with vs. without expansion of TR γδ T cells (shown in **Fig. 5b-d** of the revised manuscript and copied below for the Reviewer's benefit). We extended the requested analysis to test for PFS correlations with each T and NK cell subset, and found that TR γδ T cells provided the only significant correlation (**Fig. 5d,** and below).

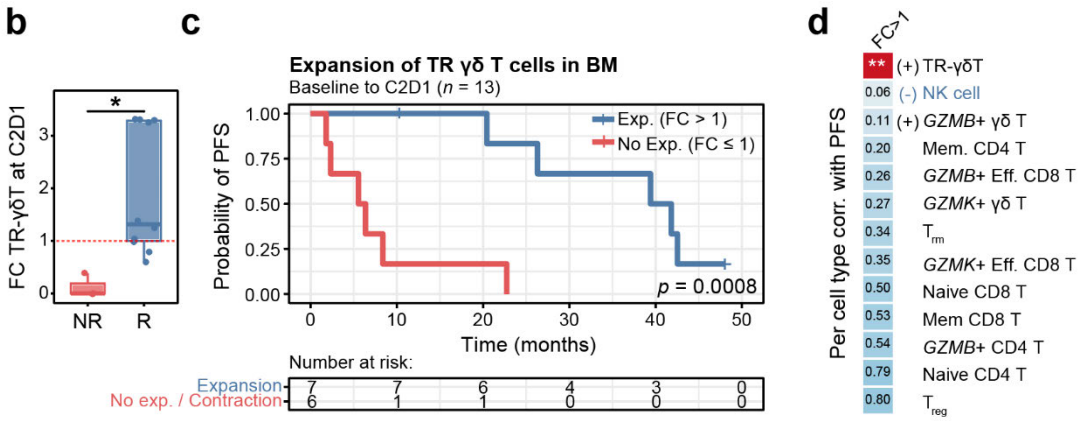


Figure 5. (b) Fold-change (FC) of the proportion of TR γδ T cells compared to total γδ T cells, between baseline and C2D1, for *n* = 3 NRs and *n* = 10 Rs. A fold-change > 1 indicates an increased proportion of TR γδ T cells at C2D1, which is indicated with a dashed red line. Significance was assessed using a Wilcoxon rank-sum test, **p* = 0.014. **(c)** Kaplan-Meier survival curve showing PFS stratified by FC > 1 or FC ≤ 1 (as determined in **b**) corresponding to expansion or no expansion of TR γδ T cells, respectively. This stratification was chosen for biological significance but also corresponds to median stratification. Significance was assessed using a log rank test. **(d)** Heatmap depicting *p* values from log rank tests performed as in **b** for the indicated T and NK cell subsets. Directionality is depicted as (+) representing correlation with longer PFS and (-) representing correlation with shorter PFS. Throughout, tumor-reactivity of γδ T cells and clonotypes was assigned using PreGame (≥0.85) or *in vitro* screening results (**Extended Data Fig. 8c**), where applicable.

Lines 90–92: "Comparison of γδ TCR repertoires between baseline and C2D1 revealed an increased proportion of clonotypes that expanded on treatment compared to the more polyclonal αβ TCR response, suggesting recognition of tumor antigens."
This statement is unclear and potentially misleading. Are the authors suggesting that only γδ T cells recognize tumor antigens because αβ TCRs show higher diversity upon treatment? The reasoning should be clarified, as higher αβ T cell diversity does not imply absence of tumor recognition.

Reply:

We have now revised this sentence to read "*Comparison of γδ TCR repertoires between baseline and C2D1 revealed a higher proportion of expanding clonotypes, which contrasted with a comparatively polyclonal αβ TCR response, suggesting treatment associated repertoire remodelling*". We have also moved **Fig. 1e** to **Extended Data Fig. 3b** in the revised manuscript, reflecting its non-quantitative nature and de-emphasizing its conclusions.

Lines 95–97: The authors state, "an increase in γδ, but not αβ, TCR diversity at C2D1 was significantly associated with longer PFS." It should be clarified whether this refers to an increase from baseline or higher TCR diversity within the cohort.

Reply:

We thank the Reviewer for highlighting this confusing aspect of our manuscript. However, all four Reviewers expressed concerns and proposed ideas for additional analyses on this topic of diversity (copied below for this Reviewer's benefit). These lines of questioning have now led to very informative new findings, and we have now substantially revised the data presented in **Figure 1** which we hope will assuage all Reviewers' concerns. Pertinently, the figure and text corresponding to Lines 95-97 in the revised manuscript (**Fig. 1c**) have been updated to show γδ TCR abundance, rather than diversity, and we have now clarified that this data represents an increase from baseline in both **Fig. 1c** and the text:

"We found that an increased abundance of the δ TCR chain in plasma cfDNA at C2D1 **compared to baseline**, but not the β chain, was significantly correlated with longer progression-free survival (PFS) (Fig. 1c, Extended Data Fig. 1b,c)." (Lines 92-95)

Concerns from other Reviewer's related to this concern:

- Reviewer #1: *We observed a significant positive correlation between γδ TCR repertoire diversity and the fraction of tumor-reactive γδ TCRs in circulation, suggesting that diversity can serve as a proxy for measuring the proportion of tumor-reactive γδ TCRs (Extended Data Fig. 8g,h)." This seems counter-intuitive as tumor-reactivity would be expected to drive clonal expansion and reduced diversity. Could this be confounded by higher γδ T cell levels in patients with tumor-reactive γδ TCRs, which in turn results in the detection of more clonotypes? More broadly, one could argue that comparing measures of diversity and clonal expansion between responders and nonresponders cannot be reliably performed because γδ T cells were virtually absent in nonresponders.*
- Reviewer #3 and #4: *The authors report (i) expanded γδ clones in BM mediating durable responses, (ii) increased BM–blood TCRδ overlap, and (iii) higher blood TCRδ diversity correlating with longer PFS. Taken together, these can appear contradictory. Please clarify*

For a full, in-depth response surrounding changes made to our TCR diversity analyses in our revised manuscript please see our response to Major Concern #10 from Reviewer #1 which begins on **Page 16** of this document.

Additional analyses are requested:

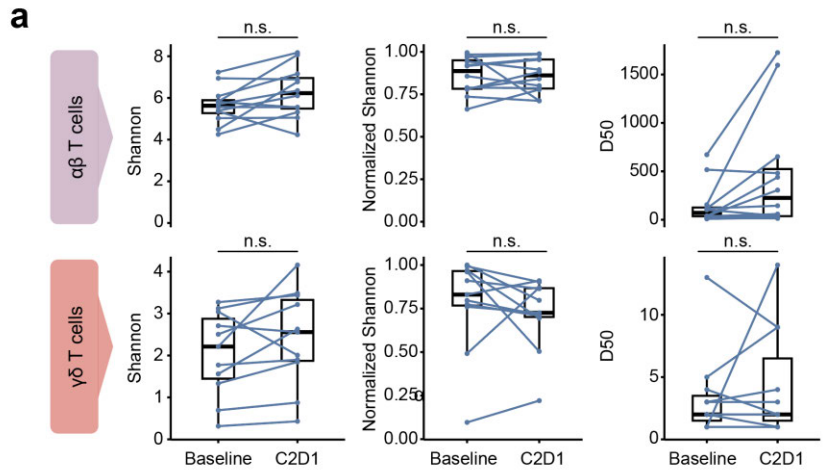
Reply:

Based on feedback from all four Reviewers, we have now largely removed analyses of TCR diversity in our revised manuscript (as alluded to above). However, we are providing the below requested analysis for the Reviewer's benefit.

o Compare TCR diversity (αβ and γδ) at C2D1 vs baseline in blood, using similar plots as Fig. 1f.

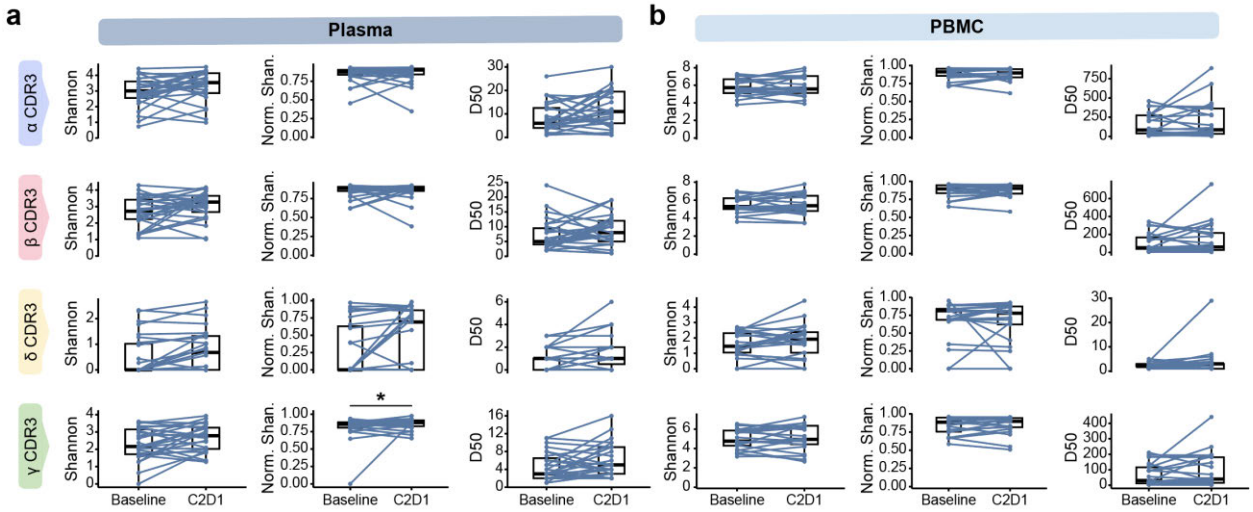
Reply:

First, we updated the analyses originally presented in **Fig. 1f** to include data from newly sequenced samples and found that all relationships were not significant (**Revisions Fig. 2.1**). For this reason and others mentioned, we have removed **Fig. 1f** from the revised manuscript.



Revisions Fig. 2.1. (a) Shannon diversity, normalized Shannon diversity, and D50 indices of the αβ (top) and γδ (bottom) BM TCR repertoires in each MM patient sample at baseline and C2D1 (n = 11). Points connected by a line indicate measurements from the same MM patient. Significance was assessed using Wilcoxon signed-rank tests, n.s. not significant.

Similarly, for both plasma cfDNA and PBMC, we did not find many significant relationships, save for the normalized Shannon index in plasma for γCDR3 (p = 0.034). While not significant at $p < 0.05$, the related δCDR3 difference had a p-value of 0.051 (**Revisions Fig. 2.2a,b**).

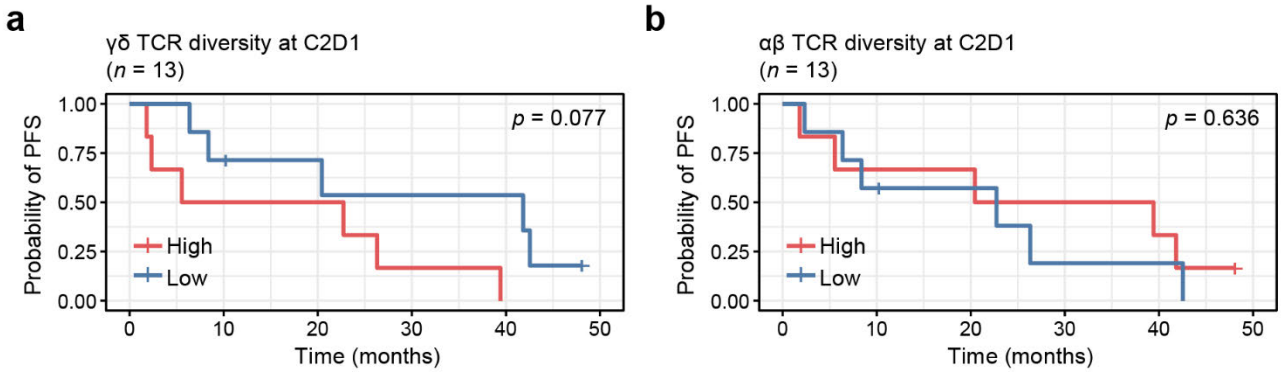


Revisions Fig. 2.2. Shannon diversity, normalized Shannon diversity, and D50 indices by TCR chain for TCR repertoires in the **(a)** plasma cfDNA (n = 27) and **(b)** PBMC (n = 21) compartments of each MM patient at either baseline or C2D1 (when data were available). Points connected by a line indicate measurements from the same MM patient. Significance was assessed using Wilcoxon signed-rank tests, *p = 0.034, all other comparisons were not significant.

o Provide Kaplan–Meier curves for bone marrow, analogous to Fig. 1g/h.

Reply:

We generated the requested plots and found no significant differences in PFS in the n = 13 BM samples for which scCITEseq data were available (**Revisions Fig. 2.3a,b**).



Revisions Fig. 2.3. Kaplan-Meier survival curves showing PFS stratified by **(a)** γδ TCR diversity or **(b)** αβ TCR diversity for *n* = 13 patients with scCITE+VDJseq data obtained at C2D1. Shannon diversity was calculated and normalized by the total number of unique clonotypes obtained for each TCR type. Patients were stratified into high and low diversity groups using the median diversity value. Significance was assessed using log rank tests.

o Examine whether higher γδ TCR diversity in bone marrow correlates with poorer prognosis, and if so, whether this relates to limited TCR reactivity or expansion capacity.

Reply:

We did not observe any statistically significant correlations between TCR diversity metrics and PFS, although there did appear to be a trend in the direction proposed by the Reviewer for the γδ TCR normalized Shannon diversity parameter (**Revisions Fig. 2.3a**). These results are consistent with our new findings demonstrating that expansion of TR γδ T cells is significantly correlated with long-term PFS following BPd treatment (**Fig. 5c,d**).

Figure 2
Claim: γδ T cells are tumor-specific and recognize broad tumor antigens via their TCRs.

• The data presented do not directly demonstrate tumor specificity; those results are in Extended Fig. 5e, which should be moved to the main figure.

Reply:

We have now moved **Extended Data Fig. 5e** into the main figures as **Fig. 2f**. We have also adjusted the text to read tumor-reactive, rather than tumor-specific, to address the concerns of a different Reviewer.

• In Fig. 2c/d, the rationale for comparing GzmB$^+$ and GzmK$^+$ subsets is unclear. Are the authors proposing that reactivity is higher in GzmB$^+$ γδ T cells due to clonal expansion? If so, reduced diversity would be expected, which is not shown. This point is further weakened by Fig. 4a/b, which demonstrates that tumor-reactive clones exist in both subsets.

Reply:

We apologize for our lack of clarity. Our rationale for highlighting these distinct subsets of γδ T cells was based on their distinct (and unbiased) transcriptional clustering, which mirrored that of their αβ T cell counterparts; namely, that the *GZMB*$^+$ subsets represent T cells with a cytotoxic phenotype characterized by expression of *GZMB*, *PRF*, *GNLY*, *NKG7*, and more. While this phenotype is well described in CD8$^+$ αβ T cells as an “Effector” phenotype, it has been less characterized in γδ T cells and as such we opted for the more conservative approach of labelling them as *GZMB*$^+$ γδ T cells. *This rationale also applied to GZMK$^+$ γδ T cells, which mirrored GZMK$^+$ Effector CD8$^+$ T cells but featured a less pronounced cytotoxic signature and tissue-*

resident phenotype. Thus, our intention was to demarcate these two distinct clusters while also alluding to the observation that γδ T cells transcriptionally mirror the clustering of $CD8^+$ T cells. We have now sought to clarify this issue in the text related to **Figure 1**, and have adjusted the presentation of the data in the new **Extended Data Fig. 2h,i**.

As stated above, we have largely moved away from analyses of diversity in our revised manuscript due to certain aspects of our data being perceived as counter-intuitive and confusing by other Reviewers. As such, we have removed the panel originally presented in **Fig. 2d** and the related Extended Data figures from our revised manuscript. Our initial intension with this figure was to present the trend of lower diversity in the *GZMB*$^+$ compartment of γδ T cells (which mirrored, though non-significantly, that of the $CD8^+$ αβ T cells), as a precursor to our observations that TR γδ T cells were more frequent in the *GZMB*$^+$ compartment, tended to be more expanded, and concordantly exhibited lower diversity (**Fig. 4a-c, Extended Data Fig. 7a**). Based on feedback from all Reviewers, we have removed **Fig. 2c** and **Extended Data Figs. 2b,c,4f,g** from the revised manuscript, retained **Fig. 4c**, and retained but moved **Fig. 4d** to **Extended Data Fig. 7a**, results which we feel are still noteworthy and stand on their own. Finally, we have modified the presentation of the data in **Fig. 2b** (initially **Fig. 2c**) to make the comparison clearer.

• Clarify whether all tumor-reactive γδ TCRs are private or if some are shared between patients.

Reply:

We have performed additional analysis of this subject at the request of all four Reviewers. For TR γδ TCRs, the δ complementarity-determining region 3 (CDR3) sequences were found to be 100% unique, having no overlap between patients. While the γ CDR3 sequence was predominantly non-overlapping between patients, one common γ CDR3 sequence was found in three separate TR γδ TCRs (from 3 separate patients): TCR43, TCRv29, and TCRmel5. Interestingly, while the γ CDR3 sequence was identical at the amino acid (AA) level, these three sequences were all distinct at the nucleotide (NT) level and comprised three distinct variable gene usages (**Revisions Fig 2.4**). Since the δ CDR3s of these three TCRs were completely unique, the paired TCR sequences of all of our tumor-reactive γδ TCRs were found to be unique (i.e. private).

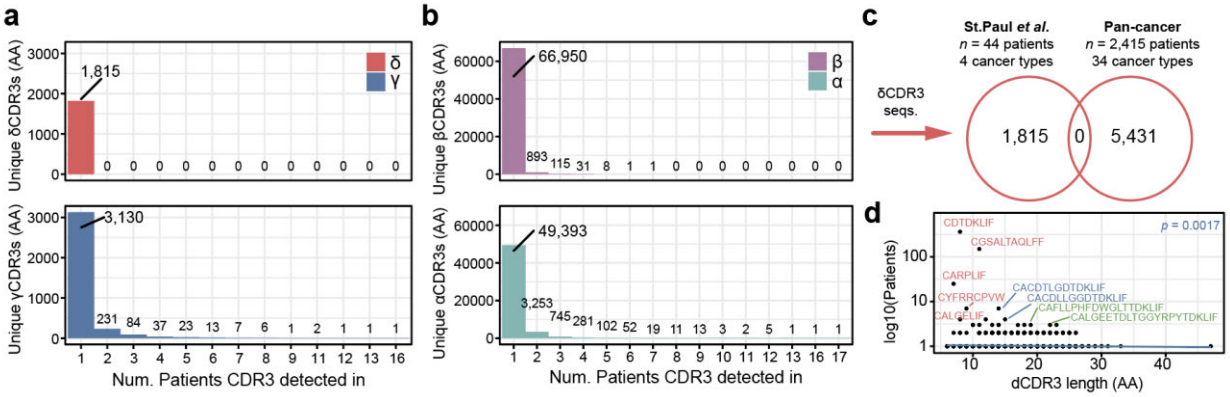
γ CDR3 AA:	C A T W D R Y Y K K L F	
TCR43:	TGTGCCACCTGGGA**CA**G**G**TATTATAAGAAACTCTTT	TRGV3
TCRv29:	TGTGCCACCTGGGA**CA**G**A**TATTATAAGAAACTCTTT	TRGV5
TCRmel5:	TGTGCCACCTGGGA**TC**G**G**TATTATAAGAAACTCTTT	TRGV8

Revisions Fig. 2.4. Alignment of nucleotide γ CDR3 sequences of TCR43, TCRv29, and TCRmel5 compared to their amino acid sequences, which were identical. Bolded nucleotides show variability between TCRs, with blue representing the most common nucleotide and red representing nucleotides unique to one of the TCRs.

More generally, we have not yet found a single instance of a shared δ CDR3 sequence between the *n* = 44 patients sequenced in this manuscript (across myeloma, melanoma, MCC, NSCLC, and including tumor-reactive γδ TCRs or otherwise), suggesting a highly private repertoire (**Revisions Fig. 2.5a**). In comparison, γ CDR3 sequences exhibited a higher degree of overlap between patients (11.5% overlapping), suggesting a less private repertoire. Still, there were no instances of a shared γδ TCR between patients. While the total number of

sequenced αβ TCRs vastly outnumbered γδ TCRs, it is noteworthy that CDR3 sequence overlap between patients was observed for αβ TCRs (**Revisions Fig. 2.5b;** 1.54% and 8.33% overlap for β CDR3s and α CDR3s, respectively). Since δ CDR3 sequences are longer than those of α, β, or γ chains, and feature a wider distribution of lengths (**Extended Data Fig. 7c**), it stands to reason that overlap between patients would be far less frequent and the possible diversity significantly higher. TR γδ TCRs, which contain the significantly longer δ CDR3 sequences (**Extended Data Fig. 7b**), would be mathematically expected to exhibit the highest level of sequence diversity amongst all TCR chains. Whether this hypothesis holds true biologically, and what implications this could have for the scope of target antigens/ligands, will be very intriguing questions for future investigations.

For an added layer of confidence, we sought to increase the number of δ CDR3 sequences subjected to overlap analysis. We compared the 1,815 unique δ CDR3 sequences identified in our study against 5,431 unique δ CDR3 sequences from two recent γδ TCR investigations spanning > 2,000 patients (**PMID: 39368482, PMID: 39058036**) but found no overlaps (**Revisions Fig. 2.5c**). Shared δ CDR3 sequences between patients are present in these external datasets, although the most highly abundant overlapping sequences are significantly shorter and lack the leader sequences typical of productive δ CDR3 chains (**Revisions Fig. 2.5d, red text**), suggesting that they might be artifactual alignments. Still, there were about 127 δ CDR3 sequences (2.33%) that possessed the expected lengths and leader sequences and were shared between patients: Vδ2 was the most frequent (**Revisions Fig. 2.5d, blue text**), but instances of Vδ1 and Vδ3 occurred as well (**Revisions Fig. 2.5d, green text**). Collectively, these results suggest that the δ repertoire is highly private, which in turn gives rise to highly private γδ TCRs.



Revisions Fig. 2.5. (a) Total unique δ CDR3 (top) and γ CDR3 (bottom) AA sequences across our cohorts of MM, MCC, melanoma, and NSCLC patient samples subjected to single-cell sequencing (*n* = 44), and the number of patients in whom such CDR3 sequences were detected. **(b)** Total unique β CDR3 (top) and α CDR3 (bottom) AA sequences across our cohorts of MM, MCC, melanoma, and NSCLC patient samples subjected to single-cell sequencing (*n* = 44), and the number of patients in whom such CDR3 sequences were detected. **(c)** Overlap between the unique δ CDR3 AA sequences identified in this manuscript (as in **a**) and the unique δ CDR3 AA sequences identified in two recent pan-cancer analyses (**PMID: 39368482, PMID: 39058036**). **(d)** Scatter plot comparing δ CDR3 AA length to the number of patients in which the sequence was detected in the external dataset of 2,415 patients (from **c**). Highly abundant sequences are labelled with their respective CDR3 AA sequences. Labels in red indicate likely artifacts; labels in blue indicate Vδ2 sequences; and labels in green indicate Vδ1 or Vδ3 sequences. The Spearman correlation is shown with a blue line, and significance as determined by a Spearman rank correlation test is presented within the plot.

• The cutoff for tumor reactivity in the Jurkat system is inconsistently defined:
o Line 1062: “We defined a tumor-reactive γδ TCR as causing a ≥ 13% increase in %GFP$^+$ cells.”
o Lines 768–769: > 15% for strongly reactive vs 0–13% for non-reactive.
o Line 505: “Only the subset of γδ TCRs eliciting a response (> 13% activation strength)”.

These thresholds must be harmonized and justified.

Reply:

We apologize for this confusion and have updated the text and figure legends to be internally consistent. We now state that a γδ TCR causing a >15% increase in GFP$^+$ cells was classified as tumor-reactive; a γδ TCR causing a 13-15% increase was classified as borderline tumor-reactive; and a γδ TCR causing a 0-13% increase was classified as non-reactive. Borderline tumor-reactive γδ TCRs were excluded from the algorithm development and analyses done in **Fig. 3 and 4** to prevent false positive/negatives.

• In Fig. 2g, TCRs 95 and 100 are labeled tumor-reactive despite < 13% GFP increase. Clarify or correct this inconsistency.

Reply:

We apologize for this confusion. In the original **Fig. 2g**, the labels indicating tumor-reactive vs non-reactive were the labels assigned from our *in vitro* screening assay (originally found in **Fig. 2f,** now **Fig. 2d**). In the original **Fig. 2g**, our intention was to show that these γδ TCRs previously flagged as tumor-reactive exhibited significantly higher reactivity to patient-matched tumor cells compared to γδ TCRs flagged as non-reactive. For improved clarity, we have now updated the main text, adjusted the figure legend, and modified the figure, which now includes statistical analysis. These new data are presented in **Fig. 2e** of our revised manuscript and copied below for the Reviewer's benefit.

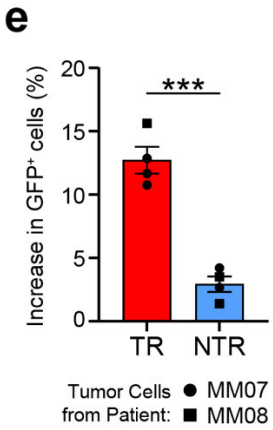


Figure 2e. Jurkat clones each expressing a γδ TCR from either donor MM07 or donor MM08 were cultured with the respective donor's matched primary MM cells (n = 4 TR; n = 4 borderline TR; and n = 4 non-reactive TCRs, as determined in **Fig. 2d)**. Data points are the mean absolute % increase in GFP$^+$ cells. Significance was assessed by Student t-test, ***p < 0.001.

• In Fig. 2i, biological replicates should be presented instead of technical replicates. The CD1C negative control should be explicitly labeled in the graph, as in Fig. 2e.

Reply:

We have repeated this experiment with n = 3 biological replicates as indicated in **Fig. 2h** copied below. We have also updated the figure to label the CD1C-negative control TCR as the "control TCR".

• Include data showing non-reactive clones (TCR 98, 99, 102, 103) in the cytotoxicity assay.

Reply:

We have repeated the CTL assays originally presented in **Fig. 2i** (now **Fig. 2h** in the revised manuscript and copied below for the Reviewer's benefit) to include TCRs 98, 99, 102, and 103, in addition to conducting *n* = 3 independent experiments with T cells derived from a different healthy donor each time. Overall, these new results demonstrate that non tumor-reactive γδ TCRs exhibit low cytotoxicity similar to that of the negative-control TCR, while TR γδ TCRs exhibit significantly higher cytotoxicity compared to the negative-control TCR. We thank the Reviewer for these thoughtful two questions, the answering of which we believe has significantly improved our revised manuscript.

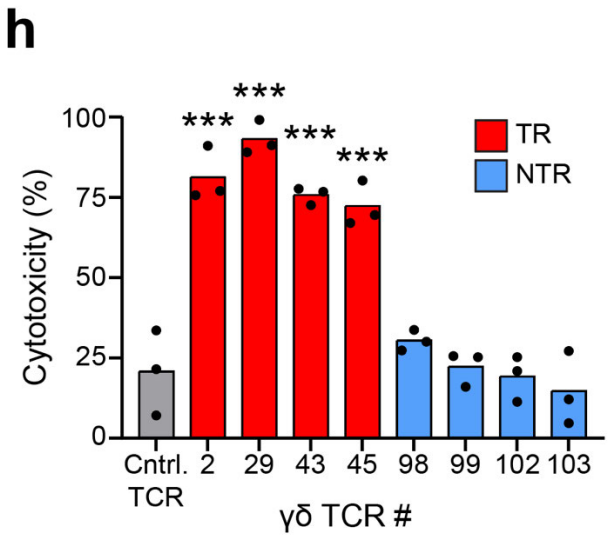


Figure 2h. Percent cytotoxicity in the RPMI-8226-luciferase cell line after overnight culture with primary T cells transduced with the indicated γδ TCR. Percent cytotoxicity was calculated by measuring the difference in luminescence between tumor cells cultured with vs. without T cells. Results shown are the mean % cytotoxicity from *n* = 3 independent donors. Significance was assessed using ANOVA, ***$p < 0.001$ compared to the CD1C-restricted control TCR.

• Explain why not all tumor-reactive clones were tested against healthy cells (in Extended Fig. 5e).

Reply:

In the original **Extended Data Fig. 5e** (now **Fig. 2f** in the revised manuscript), we tested only the TCRs that had a >15% activation response to tumor cells. In line with feedback from this Reviewer and others, we have revised the text to indicate that >15% activation is the cut-off for tumor reactivity, with the TCRs in the 13-15% range labelled as borderline tumor-reactive. We have now clarified in the text that only TCRs with demonstrated tumor-reactivity (>15%) were tested against healthy cells.

• Assess whether the presence or expansion of tumor-reactive γδ TCRs correlates with PFS or OS.

Reply:

We agree with the Reviewer that the suggested additional analysis would be important to perform. However, survival analysis of the cohort of patients presented in **Fig. 2** would be fraught with confounding factors due to the mix of newly diagnosed vs relapse/recurrence patients and the variety of different treatments and prior lines of therapy that these patients received (**Supplementary Table 1**). In addition, we were unable to obtain reliable survival data for many of these patients. Instead, and as suggested by this Reviewer below, we have performed survival analysis for the presence and expansion of TR γδ T cells in the Algonquin clinical trial cohort.

These new results, included as **Fig. 5b-d** (and copied again below) demonstrate that the expansion of TR γδ T cells conferred a remarkable advantage in PFS after only one cycle of BPd treatment.

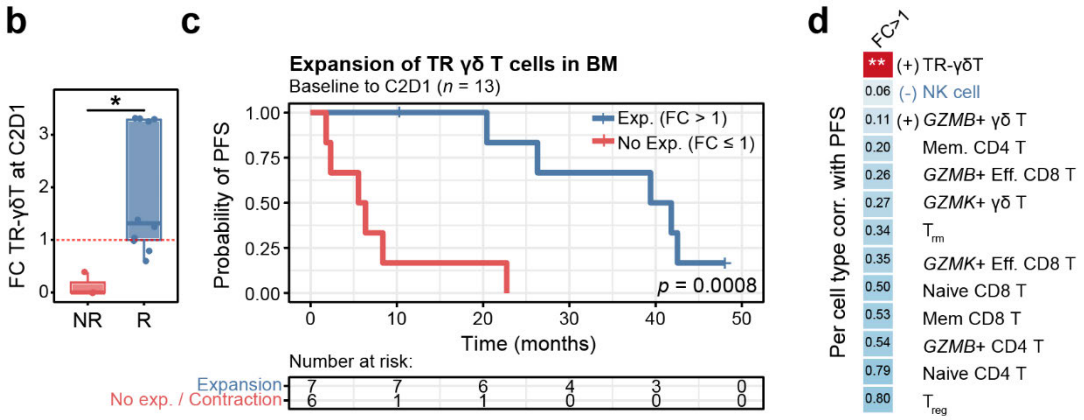


Figure 5. (b) Fold-change (FC) of the proportion of TR γδ T cells compared to total γδ T cells, between baseline and C2D1, for *n* = 3 NRs and *n* = 10 Rs. A fold-change > 1 indicates an increased proportion of TR γδ T cells at C2D1, which is indicated with a dashed red line. Significance was assessed using a Wilcoxon rank-sum test, **p* = 0.014. **(c)** Kaplan-Meier survival curve showing PFS stratified by FC > 1 or FC ≤ 1 (as determined in **b**) corresponding to expansion or no expansion of TR γδ T cells, respectively. This stratification was chosen for biological significance but also corresponds to median stratification. Significance was assessed using a log rank test. **(d)** Heatmap depicting *p* values from log rank tests performed as in **b** for the indicated T and NK cell subsets. Directionality is depicted as (+) representing correlation with longer PFS and (-) representing correlation with shorter PFS. Throughout, tumor-reactivity of γδ T cells and clonotypes was assigned using PreGame (≥0.85) or *in vitro* screening results (**Extended Data Fig. 8c**), where applicable.

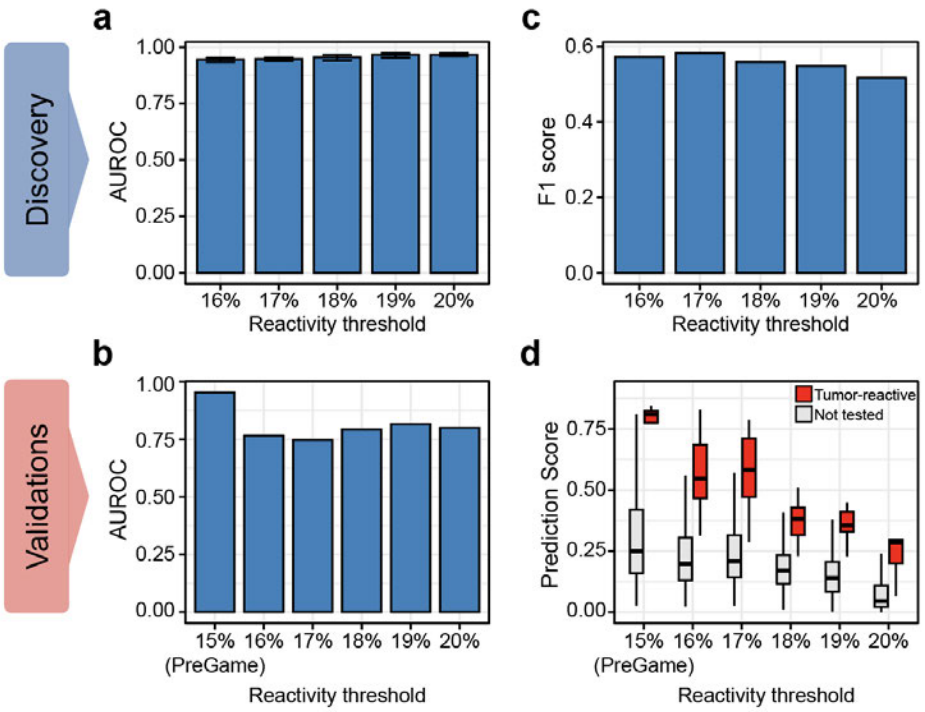
Figure 3
• The cutoff for tumor reactivity (15%) may be too permissive. Re-evaluate this threshold and test how increasing stringency affects the specificity of the PreGame algorithm, especially in melanoma and MM, where false negatives occur.

Reply:

We thank the reviewer for this important suggestion. We have now re-evaluated PreGame at several increasingly stringent tumor-reactivity thresholds, namely 16%, 17%, 18%, 19%, and 20%. Notably, this increasing stringency reduces the number of tumor-reactive TCRs (positive class) available from 24 to 18, 17, 12, 10, and 8, respectively, in an already class-imbalanced training cohort. We observed increasingly high cross-validation area under the receiver operating characteristic curve (AUROC) values with increasing reactivity stringency (**Revisions Fig. 2.6a**). However, increasing AUROC values (especially approaching 1), in combination with narrowing confidence intervals and reduced positive class sizes, are characteristic signs of overfitting. Indeed, we observed notable performance drops (~0.2 drop in AUROC) on the Validation 1 and Validation 2 cohorts (unseen validation data) when comparing these increasingly stringent models to PreGame (**Revisions Fig. 2.6b**). Consistent with this finding, the F1 scores decreased with increasingly stringent reactivities (**Revisions Fig. 2.6c**). The combination of these two observations strongly suggested that increasing overfitting occurred with increasing reactivity stringency. Therefore, models trained with increasingly stringent reactivity thresholds exhibit increasingly poorer generalizability to independent (unseen) validation datasets.

We then applied these models to the melanoma validation dataset. Because this validation dataset has only positive classes (all predicted tumor-reactive TCRs were validated as such), we could not compute and

compare AUROC values. We instead examined the distribution of prediction scores. PreGame consistently outperformed the higher stringency models, assigning higher and more discriminant predictions to true tumor-reactive TCRs with a far narrower distribution of scores (**Revisions Fig. 2.6d**). Indeed, the false negative rate observed in the melanoma cohort was increased when these higher stringency models were used.



Revisions Fig. 2.6. (a) Area Under the Receiver Operating Characteristic Curve (AUROC) values for models of PreGame trained at increasingly stringent activation strengths to define tumor-reactivity. Mean AUROC values from cross-validation are presented with the 95% confidence intervals as error bars. **(b)** AUROC values depicting prospective predictive performance on unseen data from Validation cohorts 1 and 2 for models as in **a**. Lower AUROC values indicate worse overall performance. **(c)** Mean F1 score derived from cross-validation of the models as in **a**. **(d)** Boxplot comparing the distributions of each model's prediction scores on validated TR and untested γδ TCRs from our melanoma sequencing cohort (**Extended Data Fig. 6g,h**).

Collectively, these results are consistent with overfitting and a lack of generalizability due to the increasingly sparse positive cases that remain when the reactivity threshold is increased. Thus, the reactivity threshold of 15% (i.e. PreGame) represents the optimal stringency for classifying tumor-reactive γδ T cells, at least in the contexts examined to date.

Figure 4

• If the GzmK$^+$ vs GzmB$^+$ cluster comparison is maintained, its biological relevance must be clarified. Clusters should be clearly labeled in Fig. 4a.

Reply:

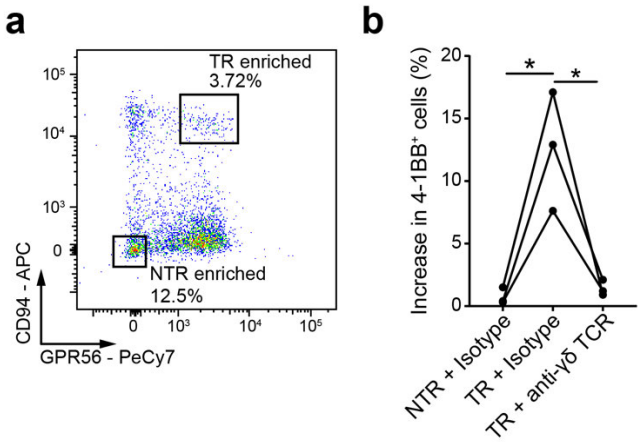
We have now added labels to the *GZMB*$^+$ and *GZMK*$^+$ clusters in **Fig. 4a**. As stated above, we have sought to clarify in our revised text the reasoning behind the distinction of *GZMB*$^+$ and *GZMK*$^+$ clusters of γδ T cells. Beyond what is stated above, the biological relevance of this distinction is highlighted in our own findings through the specific enrichment of *GZMB*$^+$ γδ T cells in BPd responders compared to non-responders (**Fig. 1d**), and the increased representation of TR γδ T cells in the *GZMB*$^+$ cluster (**Fig. 4a,b**). Both findings imply that

GZMB$^+$ γδ T cells have a heightened proliferative and/or clonal expansion capacity that may be tied to their biological role and high expression of cytotoxic marker genes (as for cytotoxic CD8$^+$ effector T cells).

• Determine whether a surface marker signature (e.g., CX3CR1$^+$, CD57$^+$, GPR56$^+$, TIGIT$^+$) can enrich tumor-reactive γδ T cells. It would be highly valuable to the field if the authors could sort signature$^+$ vs signature$^-$ cells and experimentally test tumor reactivity.

Reply:

We have now conducted additional experiments demonstrating that primary TR γδ T cells expanded from the bone marrow (BM) of multiple myeloma (MM) patients are indeed recognizing tumor cells using their TCRs. Using the top two PreGame surface markers, GPR56 and CD94, we flow-sorted MM patient BM samples to obtain the double positive γδ T cell fraction (enriched for TR γδ T cells) and the double negative γδ T cell fraction (enriched for NTR γδ T cells). We found that the population enriched for TR γδ T cells up-regulated 4-1BB in response to co-culture with MM cells, and that this tumor-reactivity was abrogated upon addition of an anti-γδ TCR blocking antibody. These findings indicate that TCR signaling is essential for the activation of TR γδ T cells. These data are presented below in **Revisions Fig. 2.7a,b**, which we have added to our revised manuscript as **Fig. 4f,g**. We thank the Reviewer for this idea and hope that our newly added method for sorting TR γδ T cells will be valuable for the field.



Revisions Figure 2.7. **(a)** Tumor-reactive (TR) enriched GPR56$^+$ CD94$^+$ and non tumor-reactive (NTR) enriched GPR56$^-$ CD94$^-$ primary γδ T cells were sorted from the BM of n = 3 MM patients. **(b)** Cells were expanded and incubated with MM cell lines in the presence of anti-γδ TCR blocking antibody or isotype control. 4-1BB-expressing γδ T cells were increased following culture with MM cells compared to unstimulated γδ T cells. Significance was assessed using repeated-measures ANOVA, *$p < 0.05$.

Figure 5

Claim: Tumor-reactive γδ T cells predict favorable response to BPd

• No data currently demonstrates this conclusion. A ROC curve analysis should be performed to test whether the presence of expanded tumor-reactive clones predicts therapeutic response.

Reply:

We thank the Reviewer for pointing out this crucial gap in our initial manuscript. We have now analyzed the correlation between PFS and expansion of TR γδ T cells at C2D1 following BPd treatment and have uncovered an exquisitely strong correlation. These new results are included in our revised manuscript as **Fig. 5b-d** and are copied below for the Reviewer's benefit. Patients who exhibited an expansion of TR γδ T cells at

C2D1 compared to baseline (fold change > 1) responded exceptionally well to treatment. Patients who did not mount an adequate TR γδ T cell response (fold change ≤ 1) had a comparatively dismal median PFS of just over 5 months, compared to a median PFS of nearly 40 months in patients where TR γδ T cells expanded. Importantly, even when patients exhibited a clinical response (PR or better, i.e. Responder (R) in **Fig. 5b**), the duration of response was significantly shorter if they did not show expansion of TR γδ T cells. These results greatly bolster our claim that TR γδ T cells underlie the long-term, durable responses observed in BPd-treated MM patients. We thank the Reviewer for this suggestion which we believe has greatly improved the strength of our revised manuscript.

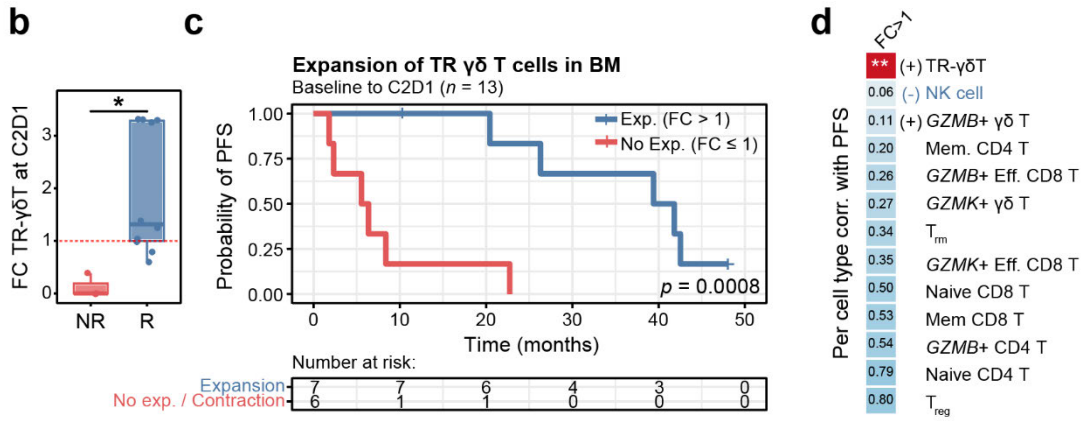


Figure 5. (b) Fold-change (FC) of the proportion of TR γδ T cells compared to total γδ T cells, between baseline and C2D1, for *n* = 3 NRs and *n* = 10 Rs. A fold-change > 1 indicates an increased proportion of TR γδ T cells at C2D1, which is indicated with a dashed red line. Significance was assessed using a Wilcoxon rank-sum test, **p* = 0.014. **(c)** Kaplan-Meier survival curve showing PFS stratified by FC in TR γδ T cells as determined in **b**. Patients were stratified by FC > 1 or FC ≤ 1, corresponding to expansion of, or no expansion of, TR γδ T cells, respectively. This stratification was chosen for biological significance but also corresponds to median stratification. Significance was assessed using a log rank test. **(d)** Heatmap depicting *p* values from log rank tests performed as in **b** for the indicated T and NK cell subsets. Directionality is depicted as (+) representing correlation with longer PFS and (-) representing correlation with shorter PFS. Throughout, tumor-reactivity of γδ T cells and clonotypes was assigned using PreGame (≥0.85) or *in vitro* screening results (**Extended Data Fig. 8c**), where applicable.

• Fig. 5c should be repeated using autologous tumor cells for more meaningful interpretation.

Reply:

Unfortunately, due to limitations in sample availability, we do not have patient-matched tumor cells to test for reactivity within the Algonquin clinical trial cohort. However, our findings in **Fig. 2e** demonstrate consistent reactivity between our curated MM cell line panel and patient-matched tumor cells, which we would suggest is a good indication that the tumor-reactivity determined in **Fig. 5c** (now **Extended Data Fig. 8c** in the revised manuscript) can be extrapolated to reactivity towards autologous tumor cells.

• The interpretation of Fig. 5e is overstated. Across timepoints (C2D1, C6D1, C18D1), tumor burden and γδ TCR frequency are not consistently inversely correlated.

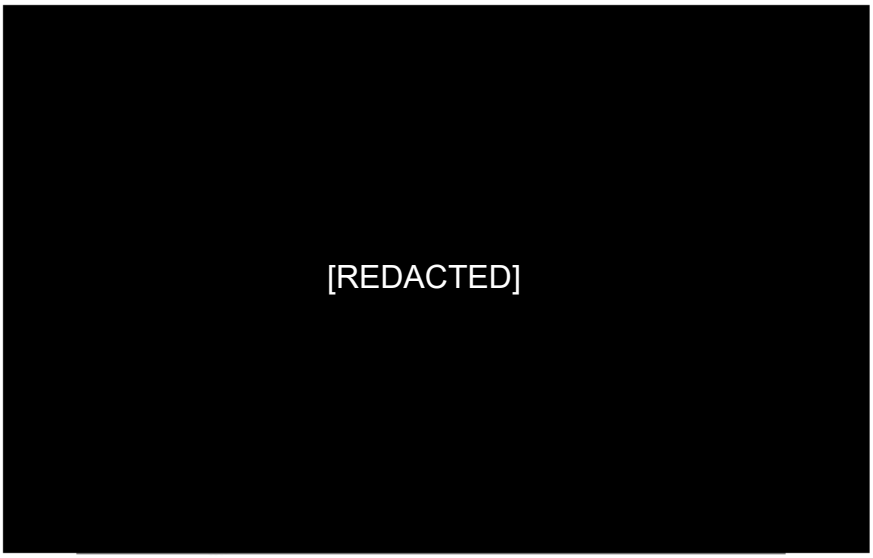
Reply:

This was a concern also raised by Reviewer #1, and we agree that additional analyses are required to support our claim. Copied below is our response to Reviewer #1 for convenience.

Given our new findings on the particular importance of the plasma cfDNA compartment (especially when compared to the PBMC compartment), we have split what was originally combined as 'Blood' into specific lines for plasma and PBMCs. Moreover, given this Reviewer's justified concern about our statement that long-term tumor control was due to a single TR γδ TCR clonotype (TCRg8), we have now replaced TCRg8 with the total proportion of TR γδ T clonotypes (**Revisions Fig. 1.19**).

These changes revealed the strikingly different γδ TCR dynamics between the PBMC and plasma cfDNA compartments. At baseline, the PBMC compartment in one patient (ALQ-17-010) had a high fraction of TR γδ TCRs, with nearly 15% of PBMCs being TCRg9 γδ T cells (**Revisions Fig. 1.19a,** light blue dashed line). At C2D1, 28 days following BPd treatment, this TCRg9 exhibited a massive decline, both in the circulation and in the biopsied BM (**Revisions Fig. 1.19a**). This steep decline of a very large TR γδ T cell clonotype resulted in this patient being assigned to the FC<1 group of **Fig. 5c**.

However, when we separated TCRg9 from the other TR γδ TCRs that we identified using PreGame (e.g. TCRg8, which was previously highlighted in **Fig. 5e**), we observed that there was in fact expansion of other TR γδ T cells in this patient's BM, indicative of response. This BM expansion was also reflected by a huge spike in the patient's plasma cfDNA at C2D1 (**Revisions Fig. 1.19b**). These increases in TR γδ TCRs coincided with a steep drop in tumor burden (plasma ctDNA) to below detectable levels. We hope our revised presentation of this data is now more convincing.



The detection of TR γδ TCRs in the PBMC compartment displays a significant lag behind that of the cfDNA (2-3 cycles), suggesting that plasma cfDNA is a more immediate reservoir for detection of BM TR γδ TCRs, consistent with the biomarker findings of **Fig. 1c**. Importantly, we know this patient had a PFS period that was: a) much longer than that of the next best patient in the FC<1 group (22.7 vs 8.3 months); and b) more reminiscent of the FC>1 group, in which PFS values were all over 20 months (**Fig. 5c**). We believe that these results do support our now revised conclusion that patient ALQ-17-010 benefited from the expansion of TR γδ T

cells. Finally, it is noteworthy that, despite the drop-off in TCRg9 in this patient's plasma at C2D1, there was still a net increase in the proportion of TR γδ TCRs (12.2 to 16.8% read fraction), which saw this patient land in the High group of **Fig. 1c**. From the point of view of a prognostic test, it is therefore encouraging that the 'masking' effect of TCRg9 did not have as material of an effect on the plasma cfDNA compartment as it did on the PBMC and the BM compartments.

We again thank the Reviewer for this thoughtful question, which we hope has now been answered with more convincing results. Ultimately, due to space constraints we have decided to remove this from the manuscript. However if the Reviewer disagrees and would like us to include it in the manuscript we are happy to do so.

Figure 6

• If TCR 65 expresses KIR3DL1 (Fig.6i), why is it reactive to healthy samples? Were all healthy donor cells (Extended Fig. 5e) $Bw4^{-}$? Are PBMCs and small airway epithelial cells (non-reactive targets) $Bw4^{+}$? Clarify this discrepancy.

Reply:

We apologize for the confusion. The Jurkat cells used in **Extended Data Fig. 5e** (now **Fig. 2f** in the revised manuscript) did not express KIR3DL1 and so the presence or absence of Bw4 on target cells had no effect on their reactivity. Only in **Fig. 6l** were Jurkat cells expressing KIR3DL1 used in co-cultures with primary renal epithelial cells, and therefore, in this context, the presence of Bw4 alleles influenced their reactivity.

• In Fig. 6l:

o Justify using CD69 as a readout when Nur77-GFP was used elsewhere as a reactivity surrogate.

Reply:

CD69 was used for a readout in this experiment because CD69 is substantially downstream of TCR engagement, whereas Nur77 activity occurs very early on in the TCR signaling cascade. As a result, CD69 levels are more reflective of the functional response of the Jurkat cells. We believe this function is more biologically significant in the context of limiting autoreactivity than are very early TCR signaling events. We have now clarified this in the text.

• Explain the rationale for CD8 overexpression in this experiment.

Reply:

CD8 was overexpressed in our Jurkat cells based on a report by Geng *et al.,* PNAS 2019 (**PMID: 31420518**), which indicated that CD8aa is required as a co-receptor allowing KIR3DL1 to function in Jurkats. Moreover, we had observed that γδ T cells possessing B2M-dependent TCRs were $CD8^{+}$ (**Fig. 6b**). Therefore, CD8 overexpression in the Jurkats better reflected what had occurred naturally in the MM patient in which TCR65 was found.

• Provide quantification and statistical testing of the Nur77-GFP^{+} percentages across the three experiments.

Reply:

We have updated **Fig. 6g** to include a panel showing statistical testing of the data pooled from *n* = 3 independent replicates. We have copied the updated panel below for the Reviewer's benefit.

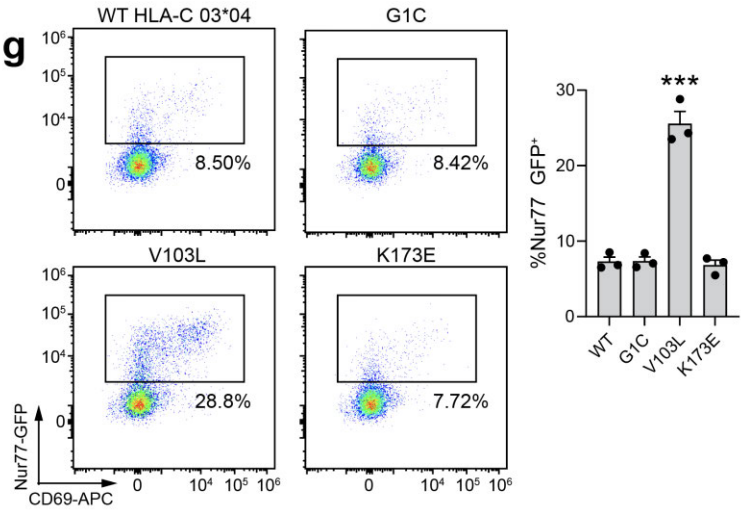


Figure 6g. Nur77-GFP expression in γδ TCR65 Jurkat cells after overnight co-culture with RPMI-8226 cells engineered to express WT HLA-C 03*04 or the G1C, V103L, and K173E HLA-C 03*04 mutants. Results shown are (left) flow cytometry plots from a representative single experiment, and (right) pooled data from 3 independent experiments. Significance was assessed using ANOVA, ***$p < 0.001$ compared to WT.

• According to the model in Fig. 6m, if AMO1 expresses Bw4, TCR 65 should be non-reactive, yet Fig. 2f shows reactivity. This inconsistency requires explanation.

Reply:

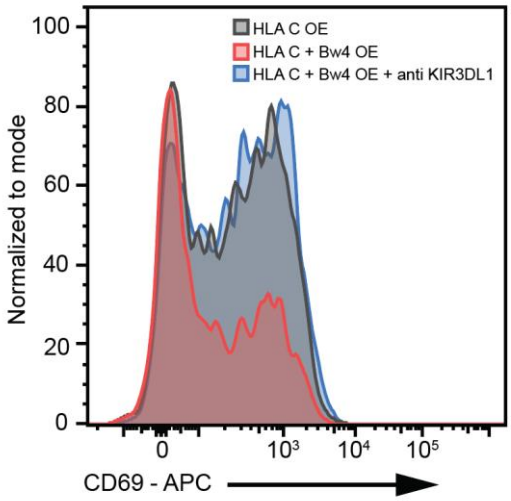
We apologize for the confusion. In the original **Fig. 2f** (now **Fig. 2d** in the revised manuscript), the Jurkat cells did not express KIR3DL1 and so responded to the $BW4^+$ and $BW4^-$ cell lines equally well. Only when we engineered the Jurkats to overexpress KIR3DL1 did we find that the TCR65 Jurkat cells no longer responded to AMO1.

• To fully demonstrate the roles of Bw4 and KIR3DL1, the following experiments are suggested:
o Knockout Bw4 in $Bw4^+$ cells to test for increased reactivity, then reintroduce Bw4 to confirm inhibition.
o Block KIR3DL1 with antibodies in Jurkat cells in parallel with overexpression.
o Test TCR 65 and KIR3DL1 relevance in γδ T cells: introduce TCR 65 into Vδ2 cells (normally non-reactive) with or without KIR3DL1 expression and measure cytotoxicity against $Bw4^+$, $Bw4^-$, and $Bw4^-$ + Bw4-rescued cells.

Reply:

Given the feedback from the other Reviewers and the reshaping of our paper, we have de-emphasized the biological significance of KIR3DL1 in the revised manuscript. Nevertheless, we have conducted the suggested experiments and observed that overexpression of HLA-C + Bw4 alleles can inhibit responsiveness compared to cells overexpressing HLA-C only **(Revisions Fig. 2.8)**. We also show that the inclusion of an anti-KIR3DL1 blocking antibody abrogated the inhibitory effects of this Bw4 allele. We have made multiple attempts to introduce TCR65 + KIR3DL1 + CD8 into primary Vδ2 cells isolated from the blood of healthy donors, but have been unable to isolate a population of γδ T cells that was infected by all three lentiviral constructs and expressed

all three required genes. In addition to the challenges of infecting primary T cells with 3 separate viruses, given the ubiquitous nature of HLA-C and its expression on the T cells themselves, it could be that TCR65 is toxic in primary T cells when overexpressed without the entire suite of regulatory mechanisms. Accordingly, given that we have de-emphasized the KIR aspect in the revised manuscript, we feel that the experiments aimed at evaluating the KIR logic gate in primary cells for potential use in cell therapy are more suited for a follow-up investigation. We have now revised the text to highlight the potential clinical utility of the KIR logic gate in the context of TCR-T and CAR-T therapy.



Revisions Figure 2.8. HLA-C*01:02 with or without HLA-B*52:01 (contains Bw4) were overexpressed (OE) in RPMI HLA ABC Total KO cells and incubated with TCR65 + CD8 + KIR3DL1+ Jurkat cells overnight, with or without the addition of an anti-KIR3DL1 blocking antibody (10 µg/mL clone DX9). The results showed that these Jurkat cells reacted to HLA-C OE cells, but that this reactivity was inhibited in HLA-C + Bw4 OE cells in a manner that was reversed upon addition of an anti-KIR3DL1 blocking antibody.

Referee #4 (Remarks to the Author):

The manuscript titled ‘Identification of broadly tumor reactive γδ TCRs using machine learning’ introduces PreGame, which prioritizes tumor-reactive γδ T cells and validates candidate function. In multiple myeloma (MM) patients treated with Belamaf, the authors identify a broadly reactive γδ TCR that recognizes HLA-C and has a KIR3DL1-mediated "logic gate" that spares normal cells. This study emphasizes the underappreciated roles of non-Vγ9Vδ2 γδ T cells in anti-tumor immunity and suggests the translational potential of MM treatment.

This is an impressive study, especially the epitope identification part, and I fully support publication in nature. However, several concerns remain and should be addressed, as detailed below.

Reply:

We thank the Reviewer for this positive appraisal of our manuscript and support for its publication in *Nature*.

Major points A-F

A) Title
To avoid overstatement, please limit the disease scope in the title. The current wording implies generalizability beyond multiple myeloma.

Reply:

We have changed the title to read “Identification of broadly tumor-reactive γδ TCRs **from multiple myeloma**”.

B) Figure 1
The central claim—that γδ T cells mediate sustained responses to BPd via tumor-reactive γδTCRs—is not supported by the presented data.

Reply:

We have now adjusted the text to indicate that γδ T cells are correlated with responses to ADC combination therapy, rather than mediating it. While still correlative, and not presented until later in **Figure 5**, we have now uncovered a very striking association between expansion of tumor-reactive (TR) γδ T cells and progression free survival (PFS), providing further substantiation of our hypothesis that TR γδ T cells play an important role in mediating sustained responses to BPd. These findings are presented in revised **Fig. 5b-d** and copied below for the Reviewer’s benefit.

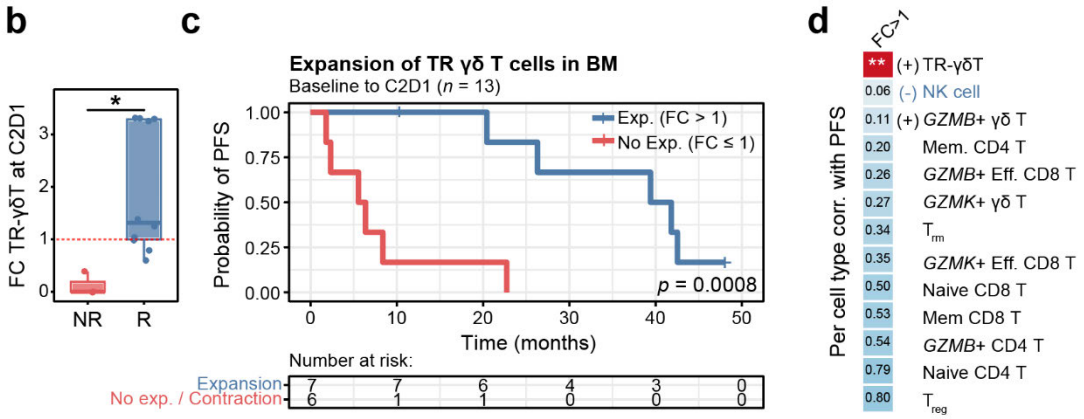


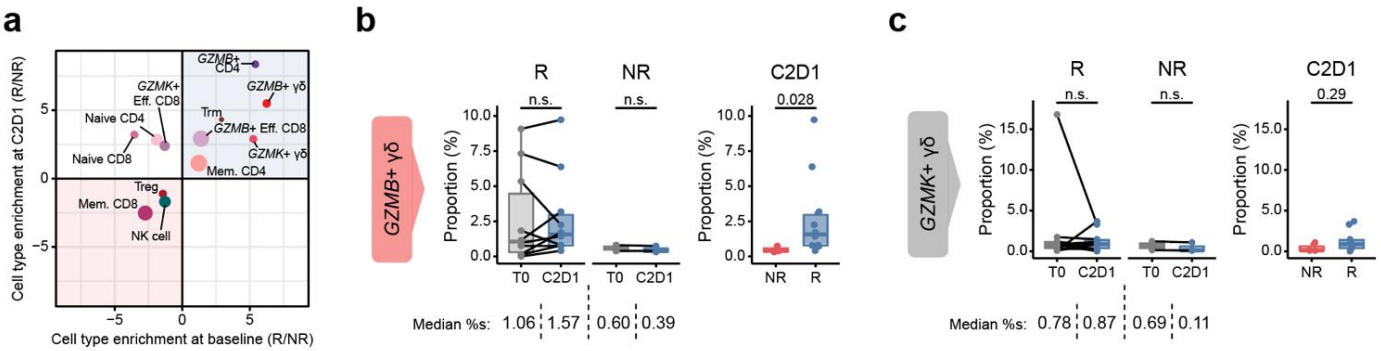
Figure 5. (b) Fold-change (FC) of the proportion of TR γδ T cells compared to total γδ T cells, between baseline and C2D1, for *n* = 3 NRs and *n* = 10 Rs. A fold-change > 1 indicates an increased proportion of TR γδ T cells at C2D1, which is indicated with a dashed red line. Significance was assessed using a Wilcoxon rank-sum test, **p* = 0.014. **(c)** Kaplan-Meier survival curve showing PFS stratified by FC in TR γδ T cells as determined in **b**. Patients were stratified by FC > 1 or FC ≤ 1, corresponding to expansion of, or no expansion of, TR γδ T cells, respectively. This stratification was chosen for biological significance but also corresponds to median stratification. Significance was assessed using a log rank test. **(d)** Heatmap depicting *p* values from log rank tests performed as in **b** for the indicated T and NK cell subsets. Directionality is depicted as (+) representing correlation with longer PFS and (-) representing correlation with shorter PFS. Throughout, tumor-reactivity of γδ T cells and clonotypes was assigned using PreGame (≥0.85) or *in vitro* screening results (**Extended Data Fig. 8c**), where applicable.

1) In Fig. 1d, both GZMB$^+$ γδ and GZMK$^+$ γδ T-cell proportions decrease at C2D1 versus baseline. This pattern does not align with the proposed activation/expansion model. Please reconcile this with your conclusion, and clarify whether analyses are normalized to total lymphocytes, CD3$^+$ T cells, or γδ frequency. Provide paired per-patient trajectories and statistics.

Reply:

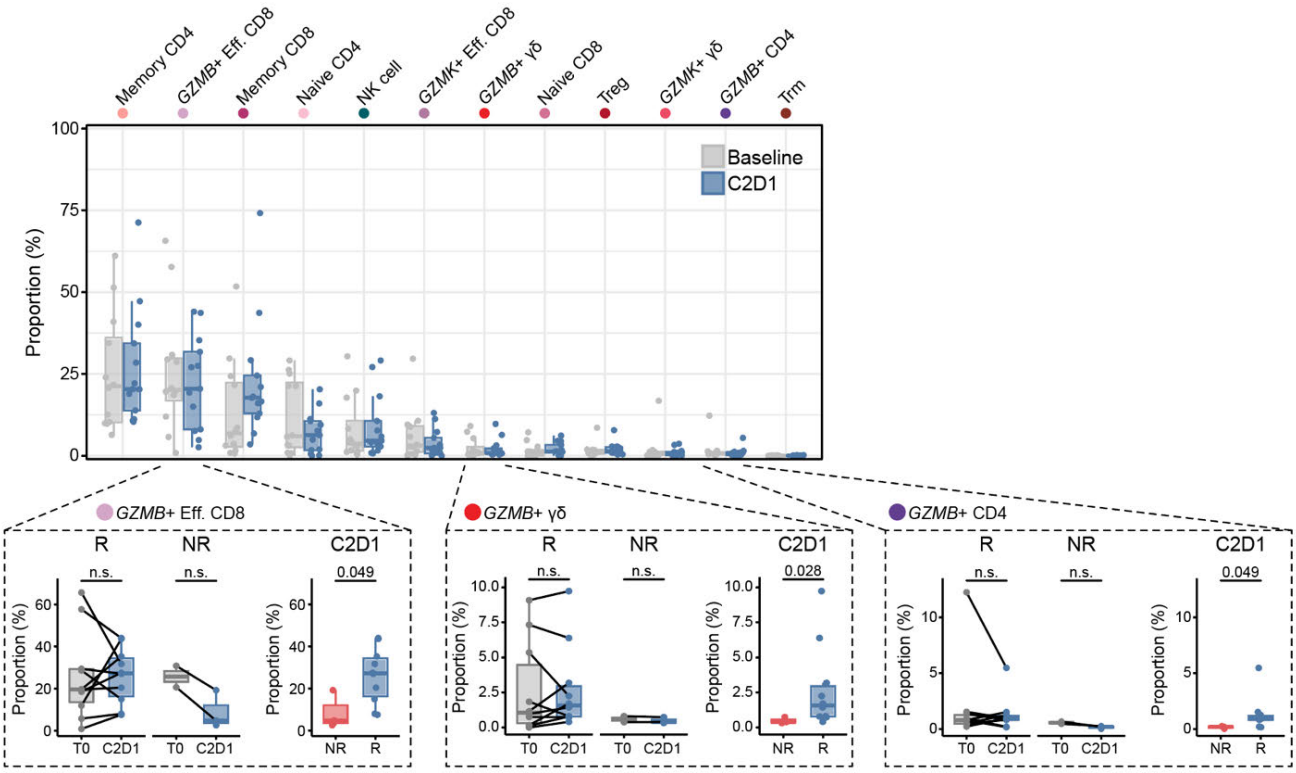
We thank the Reviewer for this comment and for suggesting how we can improve clarity in our figures and manuscript. We have added *n* = 2 additional patients to our scCITE+VDJseq clinical cohort, one new responder (R) with both baseline and C2D1 samples, and one new non-responder (NR) where only the C2D1 BM collection was viable. These additions bring our sample size for NRs up to *n* = 3 at C2D1, allowing us more power to present quantitative analysis. We have therefore updated the result originally presented in **Fig. 1d** (and provided it below as **Revisions Figure 3.1a** for the Reviewer's benefit) but have replaced it in our revised manuscript.

In line with the Reviewer's suggestion, we now present the data as paired per-patient trajectories with statistics, and clarify that proportions were normalized to total lymphocytes (**Revisions Figure 3.1b,c**). The results show that both *GZMB*$^+$ and *GZMK*$^+$ γδ T cell proportions increased between baseline and C2D1 in Rs, and decreased in NRs, albeit non-significantly for both (**Revisions Figure 3.1b,c**). We found that Rs displayed significantly elevated proportions of *GZMB*$^+$ γδ T cells compared to NRs at C2D1 (*p* = 0.028, **Revisions Figure 3.1b, right**), whereas *GZMK*$^+$ γδ T cell proportions were not significantly increased (*p* = 0.29, **Revisions Figure 3.1c, right**). These results therefore fit with the broad conclusions of our manuscript, where we might have expected to observe expansion of *GZMB*+ γδ T cells in particular, as we show that the cytotoxic *GZMB*+ phenotype is the predominant phenotype of TR γδ T cells (**Fig. 4b**).



Revisions Figure 3.1. (a) The frequency of each cell type in BPd Rs (*n* = 10) divided by the frequency of the cell type in NRs (*n* = 3) in baseline samples (x-axis) and C2D1 samples (y-axis). Points are colored by cell type and sized by their relative frequency across the cohort. Points in the upper right quadrant represent cell types that were enriched in Rs at both baseline and C2D1, while points in the bottom left quadrant represent cell types that were enriched in NRs at both baseline and C2D1. **(b,c)** Comparison of **(b)** *GZMB*$^+$ and **(c)** *GZMK*$^+$ γδ T cell cluster proportions at baseline and C2D1 split by clinical response (left), and comparison of proportions at C2D1 between clinical response groups (right), for *n* = 10 Rs and *n* = 3 NRs. Significance was assessed using Wilcoxon rank-sum tests, n.s. not significant.

We performed this analysis across all lymphocyte populations, but we did not find any individual cell type to be significantly increased between baseline and C2D1, even when splitting for response (**Revisions Fig. 3.2, top,** all n.s.). However, we did observe significant differences for three cell types between Rs and NRs at C2D1: *GZMB*$^+$ CD8$^+$ T cells, *GZMB*$^+$ γδ T cells, and *GZMB*$^+$ CD4$^+$ T cells (**Revisions Fig. 3.2, bottom**). *GZMB*$^+$ γδ T cells represented the most significantly enriched cell type in Rs compared to NRs post-treatment, with all three NRs exhibiting low proportions.



Revisions Figure 3.2. Top: Comparison of transcriptionally defined T and NK cell cluster proportions in the BM at baseline and C2D1 for *n* = 13 patients. No significant differences were identified. Bottom: Comparison of *GZMB*⁺ T cell cluster proportions at baseline and C2D1 split by clinical response, and comparison of proportions at C2D1 between clinical response groups, for *n* = 10 Rs and *n* = 3 NRs. Significance was assessed using a Wilcoxon rank-sum tests, n.s. not significant.

We thank the Reviewer for highlighting this area for improvement. We have included a modified version of **Revisions Fig. 3.1 and 3.2** as main **Fig. 1d** and **Extended Data Fig. 2k** in our revised manuscript.

2) Unequal sequencing depth can bias Shannon/normalized-Shannon estimates. Please provide a QC summary and a brief sensitivity analysis to exclude depth artifacts.

Reply:

We agree that this is an important consideration and have responded to each individual bullet point below. However, given the several points of concern and advice for improvements raised by all four Reviewers, we have largely shifted away from and removed aspects of diversity from our revised manuscript. Nevertheless, we want to ensure that each and every point raised by all Reviewers has been addressed. We are therefore providing the following analyses for this Reviewer's benefit only.

a) QC summary (per sample):

i) Total productive γδTCR counts.

Reply:

Total production γδ TCR counts were calculated as the total number of reads in the γδ sequencing library times the fraction of reads which were mapped to TRG or TRD genes from the cellranger alignment outputs. Since we used two different γδ TCR primer sets and performed three separate alignments for optimal γδ TCR

clonotype recovery (see Methods), we report the total production γδ TCR counts for each alignment for each sample in **Revisions Fig. 3.3a**.

ii) Unique γδTCR clonotypes (TCRγ chain, TCRδ chain, paired γδTCR chain).

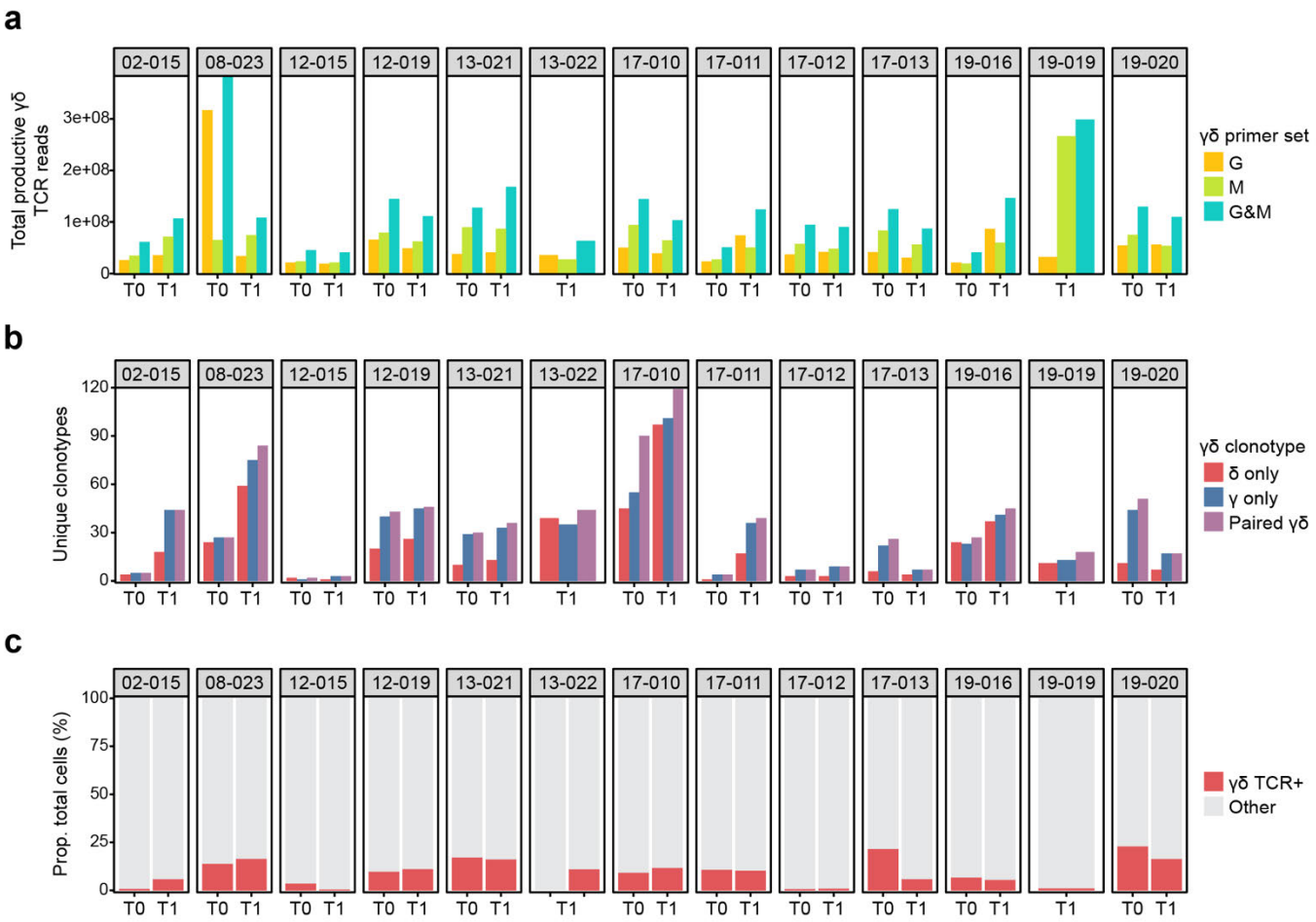
Reply:

The number of unique clonotypes for γ, δ, and paired γδ TCR chains per sample are depicted in **Revisions Fig. 3.3b**.

iii) Cell counts and proportions contributing to TCRs.

Reply:

The proportion of total cells contributing γδ TCRs per sample are depicted in **Revisions Fig. 3.3c**.



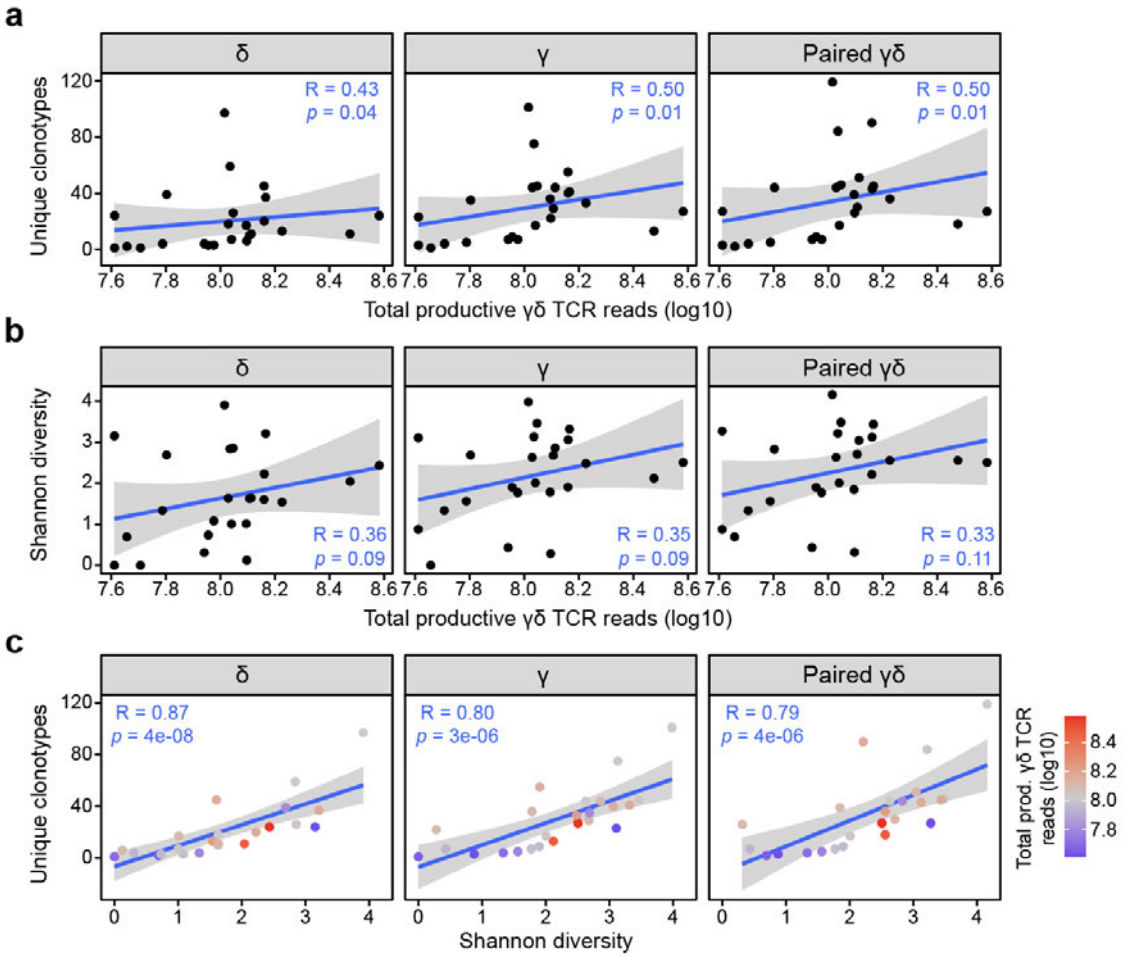
Revisions Figure 3.3. (a) Total productive γδ TCR reads per library for each sample where single-cell γδ TCR sequencing was performed. G: Gherardin *et al*. 2021 primer set, M: Mimitou *et al*. 2019 primer set. **(b)** Number of unique clonotypes post-QC in each sample when considering δ, γ, or paired γδ CDR3 sequences. **(c)** Proportion of total cells with a sequenced γδ TCR post-QC in each sample.T0: baseline, T1: C2D1.

b) Reporting of correlations to demonstrate robustness to depth:

i) Shannon vs. Unique clonotypes (expect partial but not purely depth-driven association).

Reply:

We performed the requested correlations which are depicted in **Revisions Fig. 3.4** below. We performed the correlations at the level of γ, δ, and paired γδ TCR clonotypes and observed consistent conclusions across chains. We observed significant correlations between sequencing read depth and number of unique γδ TCR clonotypes (not unexpected) (**Revisions Fig. 3.4a**). We did not observe significant correlations between Shannon diversity and number of unique γδ TCR clonotypes (not unexpected) (**Revisions Fig. 3.4b**). We observed significant correlations between Shannon diversity and the number of unique γδ TCR clonotypes (**Revisions Fig. 3.4c**). However, this correlation was not purely read depth-driven as evidenced by the mix of high and low read depth samples and the results of **Revisions Fig. 3.4b** (as expected).

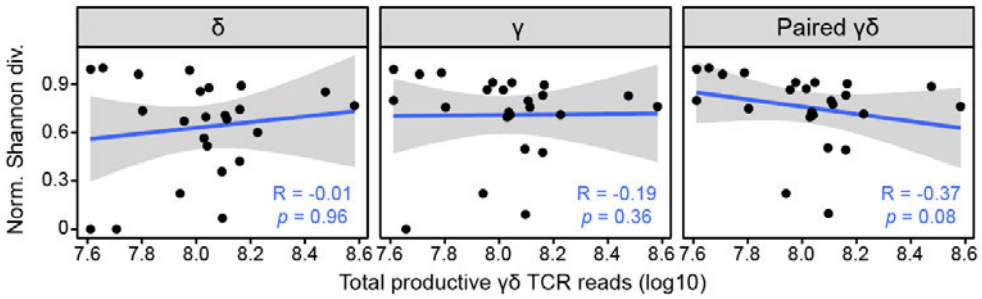


Revisions Figure 3.4. (a) Total productive γδ TCR reads from the G&M library vs the number of unique γ, δ, or paired γδ clonotypes per sample. **(b)** Total productive γδ TCR reads from the G&M library vs Shannon diversity per sample for the indicated group of clonotypes. **(c)** Shannon diversity vs number of unique clonotypes per sample for the indicated group of clonotypes. Points are colored by total productive γδ TCR reads from the G&M library. Blue lines denote Spearman correlation, and the 95% confidence interval is indicated in grey; Rho value and significance as determined by Spearman rank correlation tests are presented within the plot.

ii) Normalized Shannon vs. Total productive γδ TCR counts. (expect minimal dependence if normalization is adequate).

Reply:

We performed the requested correlations and did not observe significant correlations nor consistent conclusions across chains (**Revisions Fig. 3.5**). These results suggested minimal dependence between sequencing read depth and normalized Shannon diversity, consistent with adequate normalization.



Revisions Figure 3.5. Total productive γδ TCR reads from the G&M library vs normalized Shannon diversity per sample for the indicated group of clonotypes. Normalized Shannon diversity was calculated as Shannon diversity divided by the total number of unique clonotypes. Blue lines denote Spearman correlation, and the 95% confidence interval is indicated in grey; Rho value and significance as determined by Spearman rank correlation tests are presented within the plot.

Overall, this sensitivity analysis, especially the moderate but significant correlation between sequencing read depth and resulting number of unique γδ TCR clonotypes (**Revisions Fig. 3.4a**), suggest that sequencing depth saturation was not achieved by our single-cell γδ TCR sequencing libraries. This is not unexpected given that single-cell γδ TCR sequencing is not officially supported by 10X. While our combination of two distinct primer sets and custom alignment pipeline improved γδ TCR clonotype recovery, it is likely that our results reflect an underestimation of the full γδ TCR repertoire in these patients. Since the presence of a sequenced γδ TCR was a large factor in our single-cell annotations, these too may reflect an underestimation of the true proportion of γδ T cells in these patients. Notably, such underestimations would not change our presented conclusions (which may even be strengthened by an underestimation in this regard). If the Reviewer feels any aspects of the presented sensitivity analysis should be included in our manuscript we are happy to do so.

c) State the inclusion and exclusion criteria for samples in Fig. 1e, f and Extended Figure 2.

Reply:

All baseline and C2D1 samples from all patients were included in analyses throughout the manuscript. The inclusion criteria have now been further clarified in the revised Methods.

3) Show direct clonotype dynamics rather than a non-significant diversity index. You claim antigen-driven clonal expansion. The primary evidence should therefore be clone-level tracking, not an indirect non-significant Shannon index. Please report:

a) Alluvial plots of γδ clonotypes (baseline → C2D1);

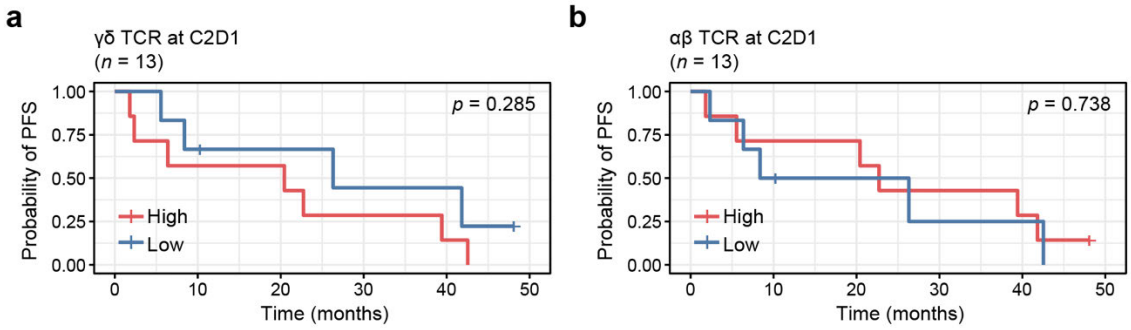
Reply:

We now present alluvial plots for both αβ and γδ BM TCR repertoires in **Extended Data Fig. 3a**. Alluvial plots for γδ clonotypes are also presented in **Fig. 5a** and **Extended Data Fig. 8b**.

b) Clonality metrics (Simpson/Gini);

Reply:

We did not observe any significant differences in the Gini Simpson metric, nor a significant correlation with PFS (**Revisions Fig. 3.6a,b**). We did note a trend in the direction consistent with our new findings demonstrating that expansion of tumor-reactive γδ T cells was significantly correlated with long-term PFS following BPd treatment (**Fig. 5c,d**).



Revisions Fig. 3.6. Kaplan-Meier survival curves showing PFS stratified by **(a)** γδ and **(b)** αβ Gini Simpson index for *n* = 13 patients with scCITE+VDJseq data obtained at C2D1. Patients were stratified into high and low diversity groups using the median diversity value. Significance was assessed using log rank tests.

c) Report the definition of expanded, contracted γδ clonotypes (Fig. 1e).

Reply:

Expanded clonotypes were defined as those present both at baseline and C2D1 but found at a higher proportion at C2D1. Contracted clonotypes were defined as those present at both baseline and C2D1 but found at a higher proportion at baseline. New clonotypes were defined as those detected only at C2D1. Lost clonotypes were defined as those detected only at baseline. These definitions are reported in the figure legend for **Extended Data Fig. 3b** in the revised manuscript.

4) The authors report (i) expanded γδ clones in BM mediating durable responses, (ii) increased BM–blood TCRδ overlap, and (iii) higher blood TCRδ diversity correlating with longer PFS. Taken together, these can appear contradictory. Please clarify

Reply:

We thank the Reviewer for pointing out this apparently contradictory aspect of our manuscript. Indeed, all four Reviewers expressed concerns and proposed ideas for additional analyses (copied below for the Reviewer's benefit). These lines of questioning have now led to very informative new findings, and we have now substantially revised the data presented in **Figure 1** which we hope will assuage all Reviewers' concerns.

Concerns from other Reviewer's related to this concern:

- Reviewer #1: *We observed a significant positive correlation between γδ TCR repertoire diversity and the fraction of tumor-reactive γδ TCRs in circulation, suggesting that diversity can serve as a proxy for measuring the proportion of tumor-reactive γδ TCRs (Extended Data Fig. 8g,h)." This seems counter-intuitive as tumor-reactivity would be expected to drive clonal expansion and reduced diversity. Could this be confounded by higher γδ T cell levels in patients with tumor-reactive γδ TCRs,*

which in turn results in the detection of more clonotypes? More broadly, one could argue that comparing measures of diversity and clonal expansion between responders and nonresponders cannot be reliably performed because γδ T cells were virtually absent in nonresponders.

- Reviewer #2: *Lines 95–97: The authors state, "an increase in γδ, but not αβ, TCR diversity at C2D1 was significantly associated with longer PFS." It should be clarified whether this refers to an increase from baseline or higher TCR diversity within the cohort.*

For a full, in-depth response surrounding changes made to our TCR diversity analyses in our revised manuscript please see our response to Major Concern #10 from Reviewer #1 which begins on **Page 16** of this document. Briefly, we leveraged recently published CapTCR-seq data from longitudinally collected plasma cfDNA samples from a cohort of *n* = 10 healthy donors (HD) (**PMID: 40875293**). Comparison of plasma cfDNA TCR repertoires from our Algonquin clinical trial samples to HDs revealed a treatment-induced surge of γδ TCRs in the cfDNA of responders, but not non-responders where detection of γδ TCRs in cfDNA remained as sparse as HDs. We found that an increase in abundance of γδ TCRs in specifically the plasma cfDNA was correlated with improved PFS. We tied this back to clonotypic expansion of γδ T cells in the BM of responders by demonstrating a correlation between clonotypic frequency in the BM and plasma cfDNA for TR γδ TCRs. These new findings present a more intuitive model: expansion in the BM precedes increased abundance of γδ TCRs in circulation—which *may* coincide with elevated diversity since basal levels are so lowly detectable—and represents a predictive biomarker of response to BPd treatment.

C) Figure 2
The results do not support the claim that γδ T cells are tumor-specific and recognize broadly expressed tumor-antigens via their TCR.

Reply:

We have updated the text to refer to these γδ TCRs as TR rather than tumor-specific. However, we have shown that many of these TCRs do not react to healthy cells, which supports their specificity for tumor cells (**Fig. 2f**). Nevertheless, we agree with all Reviewers that we cannot definitively say that these cells are tumor-specific until we know the antigen recognized, and so we now refer to them as TR throughout the revised manuscript. We have also updated our text to read: "antigens expressed broadly on tumor cells" rather than "tumor antigens". At the request of Reviewer #2, we have moved the data originally presented in **Extended Data Fig. 5e** to the main figures as **Fig. 2f** to improve clarity regarding which TR γδ TCRs do not react to healthy cells.

Furthermore, we have now conducted additional experiments demonstrating that primary TR γδ T cells expanded from the bone marrow (BM) of multiple myeloma (MM) patients are indeed recognizing tumor cells using their TCRs. Using the top two PreGame surface markers, GPR56 and CD94, we flow-sorted MM patient BM samples to obtain the double positive γδ T cell fraction (enriched for TR γδ T cells) and the double negative γδ T cell fraction (enriched for non tumor-reactive γδ T cells). We found that the population enriched for TR γδ T cells up-regulated 4-1BB in response to co-culture with MM cells, and that this tumor-reactivity was abrogated upon addition of an anti-γδ TCR blocking antibody. These findings indicate that TCR signaling is essential for the activation of TR γδ T cells. These data are presented below in our revised manuscript as **Fig. 4f,g** and copied below for the Reviewers benefit. These new experiments provide further evidence that TR γδ T cells recognize tumor-antigens via their TCR.

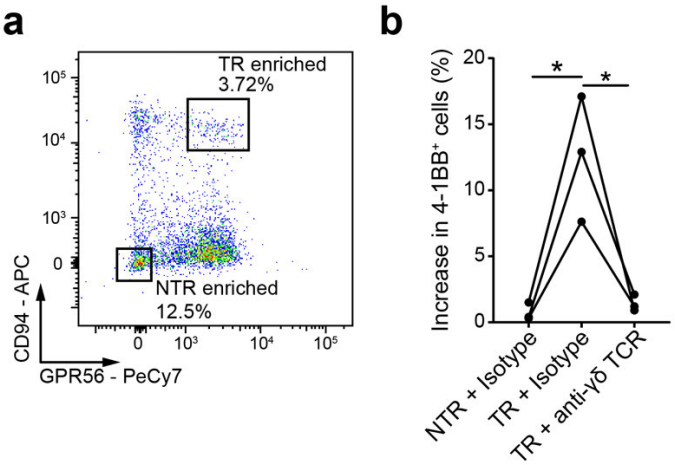


Figure 4. **(f)** Tumor-reactive (TR) enriched GPR56$^+$ CD94$^+$ and non tumor-reactive (NTR) enriched GPR56$^-$ CD94$^-$ primary γδ T cells were sorted from the BM of *n* = 3 MM patients. **(g)** Cells were expanded and incubated with MM cell lines in the presence of anti-γδ TCR blocking antibody or isotype control. 4-1BB-expressing γδ T cells were increased following culture with MM cells compared to unstimulated γδ T cells. Significance was assessed using repeated-measures ANOVA, $*p < 0.05$.

The healthy tissues and tumor cells lines are from different donors. There is still a possibility that donor variation, aka alloreactivity, may cause the recognition of gdTCR unless the inclusion of donor-matched controls—ideally tumor and peri-tumor/adjacent normal tissues from the same patient.

Reply:

Due to limitations in what patient samples could be collected as approved by our research protocol and ethics board, we were unable to assess TCR responses to peri-tumor/adjacent normal tissues in a patient-matched manner. However, we have shown in **Fig. 2e** that tumor-reactive TCRs were able to respond to patient-matched primary tumor cells, suggesting that the responses observed in cell lines are not due to allogeneic responses. This conclusion is supported by our findings presented in **Extended Data Fig. 9a,** namely that the vast majority of γδ T cells recognize HLA-independent antigens and are not susceptible to mounting allogeneic responses due to mismatched HLAs.

Still, the presented data in Fig. 2 does not support the 'tumor-antigens' recognized by the gdTCRs. Instead, these tumor-reactive gdTCRs recognize antigens expressed on tumor cells.

Reply:

We have updated the text to indicate that the TCRs are recognizing antigens expressed on tumor cells.

D) Figure 3 and Materials and Methods the PreGame model
The authors employed five categories of features: (1) expression values of a 19-gene signature, (2) surface protein expression of a 5-protein signature, (3) S phase score, (4) G2/M phase score, (5) cell cycle stage from the discovery cohort, which consisted of a 27-dimensional feature matrix containing 88 tumor-reactive T cells (positive class) and 361 non-reactive T cells (negative class) to train the ML algorithm-PreGame to identify tumor-reactive γδ TCRs.
The current claim is misleading. The model is useful, but trained on non-TCR features (transcript / protein / metadata) and therefore can only identify tumor-reactive γδ T cells from which corresponding γδTCRs can be recovered—not directly predict tumor-reactive γδTCRs.
This should be stated clearly throughout the ms.

Reply:

We thank the Reviewer for these kind words regarding the usefulness of our PreGame algorithm. We agree with the Reviewer about the importance of clarity regarding its nature. It is true that PreGame is only able to identify tumor-reactive γδ T cells rather than tumor-reactive γδ TCRs directly, and we have updated the language throughout the text to ensure that this distinction is clear. We note that this distinction is sometimes overlooked in the literature, and that an algorithm able to accurately predict tumor-reactive TCRs (αβ or γδ) does not currently exist.

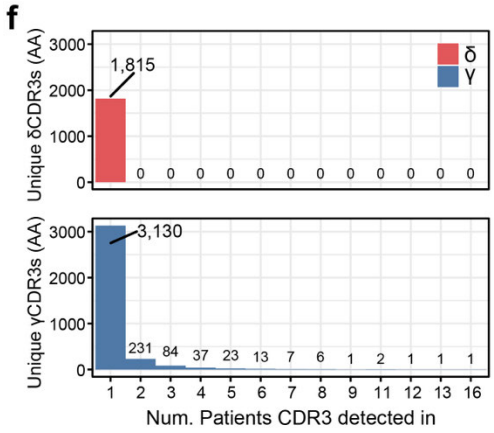
E) Details ML algorithm

Regarding the PreGame development, there are several concerns regarding data leakage and robustness.

1) Please report the γδTCR clonotype overlap across donors and, if any, whether the same clonotype exhibits a consistent phenotype.

Reply:

We have performed additional analysis of this subject at the request of all four Reviewers. Generally speaking, we did not observe any instances of a shared δ CDR3 sequence between any of the *n* = 44 patients sequenced in this manuscript (across MM, melanoma, MCC, NSCLC) (**Extended Data Fig. 4f,** and copied below for the Reviewer's benefit). While the repertoires of γ CDR3 sequences exhibited a higher degree of overlap between patients, when considering γ and δ CDR3 pairing as a prerequisite for a functional TCR, γδ TCRs were patient-unique (**Extended Data Fig. 4f**). Having said that, with specific relevance to this Reviewer's question, since we multiplexed patient tumor samples in our scCITE+VDJseq protocol, there were *rare* instances where a paired γδ TCR clonotype *appeared* to be part of the repertoire of both multiplexed samples. This was typically the case for larger clonotypes with an obvious bias towards one sample and in cells with lower demultiplexing confidence scores. This, in combination with the fact that we didn't observe any γδ TCR overlap between samples that were not multiplexed together, led us assign each γδ TCR clonotype to the most likely donor (rather than excluding γδ TCRs as we wanted to maximize this number). In this way, there was no overlap across donors, but with this in mind we opted for a more conservative 'batch-held-out' split to address the next point.

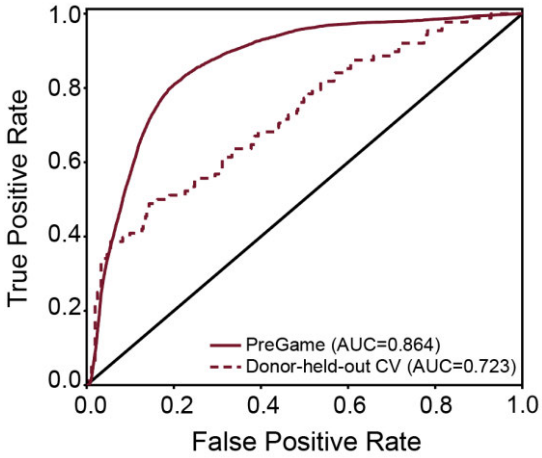


Extended Data Fig. 4. (f) Total unique δ CDR3 (top) and γ CDR3 (bottom) aa sequences across our MM, MCC, melanoma, and NSCLC patient sample single-cell sequencing cohorts (*n* = 44) and the number of patients that the CDR3 sequence was detected in.

2) Cell-level cross-validation mixes cells from the same clonotype/patient/batch across train/test. Donor-held-out and clonotype-disjoint splits are required to exclude data leakage.

Reply:

We re-ran all model evaluations using donor-held-out and clonotype-disjoint splitting, such that no clonotype present in the test set appeared in the training set, and donors were held out as the primary unit of generalization. We defined groups based on donor, but merged donors into the same held-out group if they were multiplexed together during single-cell sequencing (as outlined in previous reply). Model performance using these donor-held-out, clonotype-disjoint splits is depicted in **Revisions Fig. 3.7**. We note that, under these conditions, the training groups become particularly small and exacerbate the imbalanced nature of the training data. Therefore, the observed drop in AUC may not truly reflect data leakage, especially when taken together with PreGame's prospective performance on Validation Cohorts 1 and 2; the latter best demonstrate the model's generalizability.

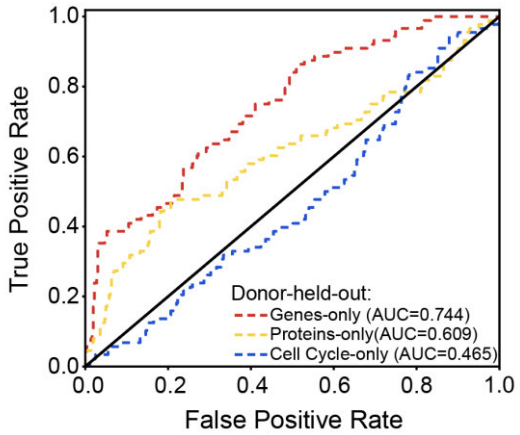


Revisions Fig.3.7. ROC curve comparing false positive and true positive rates at various thresholds for PreGame (as in **Fig. 3b**) and model training with donor-held-out clonotype-disjoint cross-validation. Overall AUC: 0.7229 (95% CI: 0.6588–0.7826); Overall AUPR: 0.4407 (95% CI: 0.3407–0.55400; Overall F1: 0.4722 (threshold=0.5). Confidence intervals were computed from 2000 bootstrap resamples.

3) Please train and evaluate the model within each predefined category, using donor-held-out splits. Report AUROC, AUPRC (with 95% CIs), and F1 score. This tests whether surface markers and cytotoxic programs robustly distinguish tumor-reactive vs bystander γδ T cells without a fancy model.

Reply:

We then evaluated the model within each predefined category (cell cycle signal, gene signatures, protein signatures) using donor-held-out clonotype-disjoint splits. The model performance is depicted in **Revisions Fig. 3.8**. The gene-only and protein-only models performed best under these donor-held-out conditions, but we note the same limitations as stated above.

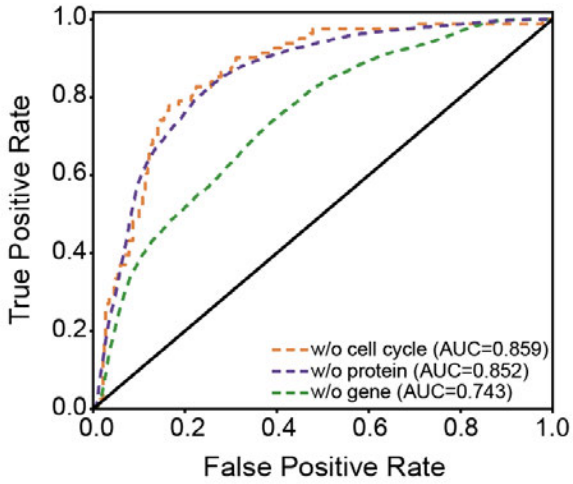


Revisions Fig.3.8. ROC curve comparing models trained using donor-held-out clonotype-disjoint cross-validation within each feature category. For genes-only, Overall AUC: 0.7441 (95% CI: 0.6866–0.7966); Overall AUPR: 0.4652 (95% CI: 0.3674–0.5805); Overall F1: 0.4755 (threshold=0.5). For proteins-only, Overall AUC: 0.6094 (95% CI: 0.5315–0.6784); Overall AUPR: 0.3443 (95% CI: 0.2580–0.4439); Overall F1: 0.3490 (threshold=0.5). For cell-cycle-only, Overall AUC: 0.4653 (95% CI: 0.4016–0.5290); Overall AUPR: 0.1817 (95% CI: 0.1457–0.2438); Overall F1: 0.2091 (threshold=0.5). Confidence intervals were computed from 2000 bootstrap resamples.

4) Cosider add Leave-one-category-out validation. Train on all categories except one.

Reply:

We next evaluated the model with leave-one-category-out validation; as depicted in **Revisions Fig. 3.9**. The model that left out cell cycle-related features performed best, although this performance was still marginally lower than when all categories were included (i.e. PreGame).



Revisions Fig.3.9. ROC curve comparing models using leave-one-category-out training. For “without cell cycle”, AUC: 0.8592 (95% CI: 0.8592-0.8592); AUPR: 0.4986 (95% CI: 0.4986-0.4986). For “without proteins”, AUC: 0.8534 (95% CI: 0.8396-0.8667); AUPR: 0.5090 (95% CI: 0.4786-0.5371). For “without genes”, AUC: 0.7431 (95% CI: 0.7159-0.7691); AUPR: 0.3669 (95% CI: 0.3386-0.3956). Confidence intervals were computed from 2000 bootstrap resamples.

5) clarify if all preprocessing, feature selection, and any oversampling were done only on the training split in each fold.

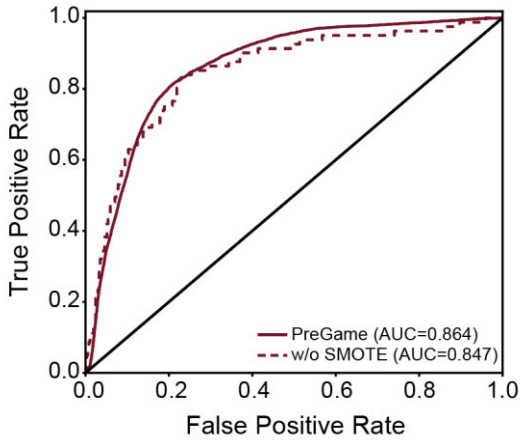
Reply:

Oversampling was done only on the training split in each fold. Gene and protein signatures were obtained from single-cell analysis, whole dataset. We have now clarified this matter in the revised Methods.

6) Please provide no-SMOTE, class-weighted baselines alongside any SMOTENC results.

Reply:

We evaluated the model with and without SMOTE and its performance is depicted in **Revisions Fig. 3.10**. We observed that SMOTE gave a marginal performance boost, justifying its usage with PreGame.



Revisions Fig.3.10. ROC curve comparing PreGame and a model trained without SMOTE. AUC: 0.8467 (95% CI: 0.8467-0.8467); AUPR: 0.5489 (95% CI: 0.5489-0.5489) ; F1: 0.5621.

7. Additionally, the benchmark compares PreGame to algorithms trained on αβ T cell neoantigen reactivity. Please clarify whether these existing models were re-trained on the γδ dataset using identical inputs, preprocessing, and donor-held-out/clonotype-disjoint splits.

Reply:

The existing algorithms used gene signatures to predict tumor-reactive T cells, rather than overt machine-learning. Therefore, re-training, cross-validation, and assessing donor-held-out/clonotype-disjoint splits are not applicable. The existing algorithms were run on identical inputs (although without train/test splits, whole dataset) with identical pre-processing steps.

We thank the Reviewer for this thoughtful line of questioning. We believe that these results, presented in combination with PreGame's robust prospective performance using unseen validation cohorts from MM and other cancer types, demonstrate that our current iteration of PreGame is sufficiently superior to other models and highly generalizable.

F) Figure 6

1. Are the HLA-C alleles of the donor of TCR65 recognized by TCR65?

Reply:

The HLA-C alleles expressed by the TCR65 donor are HLA-C*01:02 and HLA-C*07:01 (**Fig. 6j**). We tested these two alleles through overexpression assays and found them to be recognized by TCR65 (**Fig. 6e**). We have updated the text to clarify this important point.

2. To support translational potential, it would of course be great demonstrate on-target cytotoxicity with sparing of normal tissues using γδTCR65-transduced primary human T cells. If possible within the scope of this ms, on-target MM killing and no reactivity to patient-matched normal cells here would substantially strengthen claims of specificity and safety for γδTCR65.

Reply:

The Reviewer makes an important suggestion. Unfortunately, we do not have patient-matched tumor or normal cells for the patient who donated TCR65. We have now toned down our conclusions regarding this logic gate at the request of Reviewer #1. Our primary intent on highlighting this ligand and logic-gate was to provide insight into the biology of tumor-reactive γδ T cells, the potential ligands recognized, and how these cells might be regulated within the TME. However, we do believe that our proposed KIR logic gate could have translational potential for other antigens in the context of CAR-T (perhaps in a future 'if-then' or 'if-better' type of CAR-T) or TCR-T therapy. We have revised our text to mention this exciting prospect.

Minor:

1. Abstract: "we uncover a tumor-reactive γδ TCR epitope" , consider "we uncover a γδ TCR epitope involved in tumor reactivity"

We have now made this correction.

2. Line 65-66: Please specify single-cell RNA and CITE sequencing data.

We have specified these data to be obtained through single-cell CITE sequencing.

3. Line 67-68: The confirmation of tumor-reactive gdTCRs is based on in vitro functional assay instead of PreGame.

We have changed this sentence to read "We confirmed that **γδ T cells flagged by PreGame** bear TR TCRs and exhibit TCR-dependent activation, in support of their "adaptive-like" function".

4. Line 67-69: The terms "adaptive-like" and "NK-like" may read as contradictory. It may be useful to (i) define "adaptive-like" operationally (evidence for TCR-dependent, clonotype-restricted tumor recognition; clonal expansion; memory markers), (ii) clarify that "NK-like" refers to a cytotoxic transcriptional program rather than NK lineage, and (iii) show representative genes/modules supporting the NK-like state.

We agree with the Reviewer and have clarified this sentence to read: "We show that γδ T cells flagged by PreGame bear TR TCRs and exhibit TCR-dependent activation, in support of their "adaptive-like" function. Simultaneously, we establish that TR γδ T cells are transcriptionally NK-like **based on their cytotoxic transcriptional programs and surface markers**, and that they do not upregulate most exhaustion markers." The representative genes/modules that support the NK-like state can be found in **Fig. 4e**.

5. Line 69-71: 'clinical utility of this algorithm'? Or you just showed 'utility of this algorithm on this samples'.

With our substantial new evidence provided in revised **Figure 5** (thanks to all the Reviewers' excellent suggestions), we believe we have now demonstrated the clinical utility of our algorithm enough to keep these lines as-is. For example: while it may not be *feasible* in the clinic, patients' BM samples *could* be routinely sequenced by scCITEseq and assessed by PreGame. This scenario would be clinically relevant in so far as

informing patients about their prognosis (whether to expect 5 months or 40 months of PFS), and informing physicians about whether or not their patients are responding to BPd, or if a change in treatment regimen would be better for their patients. However, if the Reviewer feels that our new results still do not demonstrate clinical utility, we will make this change.

6. Line 84: Please write down the exact days post–baseline corresponding to Cycle 2 Day 1 (C2D1) to make it friendly for readers who are not MM experts.

We have indicated in the revised text that C2D1 is 28 days post-treatment initiation.

7. Line 87: Please select a proper palette to visualize the cell populations in Fig1c and Extended Data Fig. 1b,c.

We have now changed the color palette used in the figures relating to this scCITE+VDJseq data.

8. Line 88-90: cell type enrichment as shown in Fig. 1d is difficult to digest and appears to show high enrichment of gd T cells, which may not be the case, rather for CD4+ GZMB+ cells.

We have now replaced **Fig. 1d** with a clearer representation of proportional cell type differences, also fulfilling the request of Reviewer #2.

9. Line 90-93: rephrase the claim “suggesting recognition of tumor antigens.” The presented data do not support tumor-antigen recognition at this stage. Specifically: (i) GZMB$^+$ γδ and GZMK$^+$ γδ proportions decrease at C2D1 vs baseline (Fig. 1d); (ii) Shannon diversity and D50 of γδTCR show no material change; and (iii) elsewhere in the manuscript, the clonotypes of interest are shown to recognize HLA-C, which is not a tumor antigen and is framed as part of a “logic-gate” interaction. Taken together, these results indicate treatment-associated repertoire remodeling/trafficking rather than direct evidence of tumor-antigen recognition.

We have rephrased this claim to read “treatment-associated repertoire remodeling”.

10. Line 111-115: ‘GZMB+ γδ T cells outnumbered GZMK+ γδ T cells’ are basically dominated by donor MM15. Even proportionally, it is hard to conclude that GZMB+ γδ T cells outnumbered GZMK+ γδ T cells. ‘suggestive of increased clonal expansion in some patients’ is not supported by the results (Fig. 2c, d, Extended Data Fig. 4f,g).

We have modified the presentation of this data and it can be found in **Fig. 2b**. We have rephrased this claim to read “In general, GZMB+ γδ T cells outnumbered GZMK+ γδ T cells, **suggesting increased recruitment or expansion** in some patients.”

11. Line 115-116: Please report the γδ T cell number and unique paired γδ TCR clonotype number per sample.

We now include this information in **Supplementary Table 4**.

12. Line 130-132: Would it be possible to add statistic results here? And could you please report the phenotype (GZMK or GZMB) of these tested gdTCR from the original scRNAseq data?

We have added statistics to this figure, which corresponds to **Fig. 2e** in the revised manuscript. The phenotypes of the γδ T cells from which these TCRs were derived are indicated in the table below.

Patient	TCR	Reactivity	Phenotype
MM07	95	TR	Unknown
MM07	98	NTR	*GZMK*+
MM07	99	NTR	*GZMK*+
MM07	100	TR	*GZMB*+
MM07	101	TR	*GZMK*+
MM08	102	NTR	*GZMB*+
MM08	103	NTR	*GZMK*+
MM08	104	TR	*GZMK*+

13. Line 140-141: 'broadly expressed tumor antigens' should be 'broadly expressed antigens on tumor cell'.

We have made this correction as suggested.

14. Line 168-175: Why not co-culture with patient-derived tumor cells instead of cell line?

Unfortunately, due to limited sample availability, we did not have additional vials of cryopreserved BM cells to perform this experiment with matched patient-derived tumor cells.

15. Line 175-180: It is contradictory that tumor-reactive TCRs are from clonotypes with relatively low clonal expansion. Are these tumor-reactive gd T cells 'by-standing'?

This is an interesting observation and perhaps something unique to the biology of γδ T cells that we would like to further explore in a follow-up manuscript. We think that this lack of clonal expansion by tumor-reactive γδ T cells might be attributed to their high expression of CD57, which is a known marker of terminal differentiation in αβ T cells (**ref?).** It may be that many tumor-reactive γδ T cells are in fact terminally differentiated and so are unable to expand without some sort of intervention such as that imposed by checkpoint blockade. This scenario may tie in with our observation of lower PD-1 expression in highly expanded γδ T cell clonotypes compared to poorly expanded clonotypes (which we also observed in our melanoma samples, see **Revisions Fig. 1.10a**). Perhaps PD-1 expression initially lags behind that of CD57, but then catches up and limits further γδ T cell expansion.

Alternatively, γδ T cells may not follow the same 'rules' as αβ T cells. This situation may be illustrated by the TCR65 clonotype, a highly potent γδ TCR linked to relatively low clonal expansion. Knowing that the epitope recognized by this TCR is in a ubiquitously expressed self-protein may offer a glimpse into the reason behind the sometimes-discordant tumor-reactivity and clonal expansion capacity of γδ TCRs. We found that TCR65 had a very specific target: malignant cells that lacked Bw4 HLA proteins but retained HLA-C. These cells represented only a minor proportion (~14%) of the malignant cells in this patient. It may be that, in order to limit potential toxicity, the degree of γδ T cell expansion is more strictly controlled than that of αβ T cells so as to be proportional to the number of viable target cells present; ie. "With great power comes great responsibility". We believe this is an interesting phenomenon to introduce in our current manuscript, which we hope will spur interest from the field and lead to new discoveries and future publications. Importantly, PreGame will now be available to serve as step 1 in answering this critical question.

16. Line 190-192: ‘Distinguish tumor-reactive from by-stander non-Vg9Vd2 γδ T cells’ would be more precise.

Respectfully, we disagree with the Reviewer on this point since it would suggest that PreGame is unable to accurately label Vγ9Vδ2 T cells. Indeed, Reviewer #1 posed a similar question as to whether PreGame was able to accurately label Vδ2 T cells as TR vs NTR, given the inclusion of *TRDV1* in PreGame’s training matrix. We have found that the inclusion of *TRDV1* actually helps to control false positives for Vδ2 and Vδ3 TCRs; these data are now shown in **Revisions Fig. 1.8.**

We would also argue that, within the scope of our current manuscript (that is, developing a method to identify tumor-reactive γδ T cells), we have been very fair to Vδ2 TCRs. Our Discovery cohort, which was based on an unbiased validation of every single γδ TCR identified in our first eight patients, contained only 18% Vδ2 TCRs, a minority even before screening for tumor-reactivity. This finding is consistent with the literature, which points towards Vδ1 and Vδ3, and not Vδ2 T cells, as being tumor-reactive in patients (**PMID: 36631610**). This minority of Vδ2 TCRs corresponded to an even smaller fraction of total cells (9%), indicative of lesser expansion. Furthermore, tumor-reactivity screening revealed only two TR Vδ2 TCRs, and only one of these was a Vγ9Vδ2 TCR. Therefore, less than 1% of tumor-infiltrating γδ TCRs were TR Vγ9Vδ2s in an investigation that currently represents the largest unbiased inquiry into the TR of γδ T cells. While it was not our goal to specifically characterize the reactivity landscape of Vγ9Vδ2 TCRs, PreGame has still been >90% accurate at labelling Vδ2 TCRs, comparable with its overall performance. We therefore do not believe the added distinction proposed by this Reviewer is warranted.

17. Line 214-216: There is actually a function in Seurat called addmodulescore to compare gene signatures.

We have now removed this sentence from our revised text due to additions and changes suggested by other Reviewers.

18. Line 219-220: Benchmark PreGame against these re-trained public models. Only if PreGame still outperforms under leakage-robust conditions should you claim it’s superiority. Otherwise, the current comparison largely benchmarks feature selection, not the modeling approach itself, and comparative claims should be rephrased accordingly.

As stated above, re-training the public models used for comparison against PreGame would not be applicable as they are not based on machine-learning but rather on gene signatures. We note that, even under the harsh grouped (due to multiplexing, as detailed above) donor-held-out clonotype-disjoint splits, the performance of PreGame was still superior to the next best public algorithm (**Revisions Fig. 3.7**). However, we have now rephrased the relevant text throughout the manuscript as it was not our intention to disparage the modelling approaches used by others. Instead, we sought to highlight the need for parallel algorithms specifically aimed at classifying γδ T cells. In that sense, our tailored feature selection was externally superior, and our modelling approach internally superior (as demonstrated above).

19. Line 260-261: This is not consistent with the claim in line 90-93.

This sentence has been removed as part of our edits to the revised manuscript aimed at re-focusing on TCR abundance rather than diversity, a request of all Reviewers.

20. Line 287-288: Are the samples independent in Extended Fig. 8h?

This figure has been removed as part of our edits to the revised manuscript aimed at re-focusing on TCR abundance rather than diversity, a request of all Reviewers.

21. Line 349 ff: Where in the scheme of Fig 6m would be a place for CD8a?

We now include CD8a in revised **Figure 6m**.

discussion line 387 ff: Include key reference Rock et al from Chien JEM 1994, doi: 10.1084/jem.179.1.323.

We have added this reference.

Referee #4 (Remarks on code availability):

code not available

Reply:

We apologize for our oversight of keeping our uploaded code private from the Reviewers. Reviewer's should now be able to request access to our uploaded code here: https://doi.org/10.5281/zenodo.20028241 .

Responses to the Reviewers

Reviewer #1 (Remarks to the Author):

The authors comprehensively addressed most of my points and I would like to congratulate the team with this wonderful manuscript.

Reply:

We thank the Reviewer for their kind words and further suggestions for improvement, which we hope we have now addressed thoroughly.

There are, however, a few points insufficiently addressed that I believe deserve more attention before publication:

- Major point 1: "The relevance of the study would be much greater if the authors could robustly test if the γδ TCR is not only sufficient but also essential for tumor recognition by γδ T cells. Especially since γδ T cells are known to express various activating innate immune receptors, which also represent some of the main features that PreGame uses to identify tumor-reactive cells (NKG2C, CD94, NCR3, FCGR3A/CD16, NKG7)."

The authors now provide data showing that blocking the gdTCR reduces 4-1BB expression on expanded TR gdT cells exposed to target cells, and use this to demonstrate that the TCR is not only sufficient but also essential for tumor recognition. These data are of great interest, but currently seem insufficiently robust given the importance of this claim. Specifically, these experiments were currently only performed with 3 gdT cell donors and only one activation marker as readout. To convincingly support the claim, I believe it is essential to (1) increase the cohort size for this experiment to levels that allow performing robust statistics (paired Wilcoxon), and, critically, (2) complement the 4-1BB readout with a functional tumor cell killing readout and a more comprehensive set of activation markers.

Reply:

We thank the Reviewer for their thoughtful and constructive comments, and for recognizing the importance of these data. We fully agree that, given the significance of the claim, strengthening the robustness of the analysis is essential.

In response, we have expanded the cohort size from $n = 3$ to $n = 5$ independent patients, thereby improving the statistical power of the analysis. This has allowed us to perform a more rigorous paired statistical evaluation. Specifically, we applied a repeated-measures ANOVA with Tukey's post-hoc test, which is well suited for comparisons across multiple conditions for a given patient's cells. With this expanded dataset, we now observe a highly significant overall effect ($p = 0.0002$), with between-group comparisons reaching $p = 0.0005$ and $p = 0.0004$. We believe these

additions substantially strengthen the robustness and reliability of our conclusions. These new results with an expanded cohort of patient samples can be found as **Fig. 4g** in our revised manuscript, which is copied below for the Reviewer's benefit.

We also appreciate the Reviewer's suggestion to include additional activation markers. Unfortunately, these markers were not included in the original flow cytometry panel. Moreover, as we currently do not have access to additional patient material, performing new experiments with extended panels would require substantial time to procure and process fresh samples. Nevertheless, we would like to emphasize that 4-1BB is a highly relevant and informative marker in this context. It is widely recognized as a direct and sensitive indicator of recent TCR engagement in human T cells, with rapid and transient expression kinetics that make it particularly suitable for detecting antigen-specific activation (PMID **17371945**). Importantly, in contrast to other activation markers such as CD69 or CD25 (which can be upregulated by cytokines such as IL-2 and IL-15 even in the absence of TCR signaling) 4-1BB expression has been shown to depend specifically on TCR stimulation and is not induced by these cytokines alone (PMID **41567240**). Thus, we believe that 4-1BB provides a meaningful and specific readout of TCR-dependent activation in our system.

However, to further strengthen our conclusions, we have now also included a complementary tumor cell killing assay. In these experiments, TR γδ T cells kill CD138$^+$ tumor cells, and importantly, this cytotoxic activity is significantly reduced in the presence of the same anti-γδ TCR-blocking antibody (**Fig. 4h** and copied below for the Reviewer's benefit). These results provide functional evidence that tumor recognition and killing are dependent on the γδ TCR. Taken together, the combined evidence from the expanded cohort, the statistically strengthened 4-1BB analysis, and the newly added functional cytotoxicity assay strongly supports our conclusion that the γδ TCR is not only sufficient but also essential for tumor recognition by TR γδ T cells.

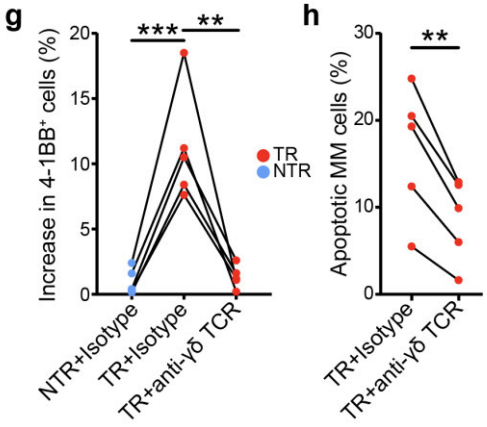


Fig.4 (g) Percent increase of 4-1BB positivity in TR (GPR56$^+$CD94$^+$) or NTR (GPR56$^-$CD94$^-$) γδ T cells following co-culture with MM cells compared to unstimulated γδ T cells, in the presence of an anti-γδ TCR blocking antibody or isotype control. Significance was assessed using repeated-measures ANOVA with Tukey's post-hoc test, $^{**}p < 0.005$, $^{***}p < 0.0005$. **(h)** Killing of CD138+ MM cells following co-culture with TR (GPR56$^+$CD94$^+$) γδ T cells in the presence of an anti-γδ TCR blocking antibody or isotype control. Significance was assessed using a two-sided paired t-test, $^{**}p < 0.005$.

- Minor point 19 "More patient characteristics and details of the clinical cohorts are necessary." While supplementary table 1 provides extensive clinical information as patient-level data, I believe the manuscript would benefit from a cohort-level baseline table so that the reader understands what kind of patients were included.

Reply:

We thank the Reviewer for clarifying, and we have now added a baseline characteristic table for patients in our clinical cohort as **Extended Data Fig. 1b**, which we have copied below for the Reviewer's benefit.

Baseline characteristic	All Patients (n = 37)
Age, median (range), years	69 (38-85)
Sex, n (%)	
Male	23 (62.2)
Female	14 (37.8)
Time since initial Dx, median (range), years	3.8 (1.1, 20.6)
ECOG status, n (%)	
0	9 (24.3)
1	26 (70.3)
2	0 (0.0)
Missing	2 (5.4)
Derived ISS stage at screening, n (%)	
I	6 (16.2)
II	13 (35.1)
III	10 (27.0)
Missing	8 (21.6)
Baseline Cytogenetics, n (%)	
High Risk	4 (10.8)
Standard risk	15 (40.5)
Missing	18 (48.6)
Prior therapies, median (range)	2 (2, 6)
Prior therapies, n (%)	
Autologous stem cell transplantation	17 (45.9)
Lenalidomide	37 (100.0)
Proteosome Inhibitor	37 (100.0)
Daratumamab	33 (89.2)
Triple class exposure	33 (89.2)
Refractory to, n (%)	
Lenalidomide	37 (100.0)
Proteosome inhibitor	33 (89.2)
Daratumamab	33 (89.2)
Triple class refractory	29 (78.4)

Extended Data Fig. 1 (b). Patient characteristics of the clinical cohort.

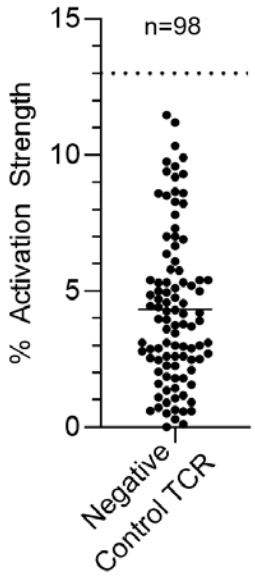
- Minor point 23. "When defining tumor reactivity, why did the authors take a cut-off of 13%? This cut-off should be clarified. If up until 13% is background, then an increase from 13% to 15% may not be major recognition by the TCR clone."

The authors addressed this point by introducing a new category of 13-15% activation. However, this does not address the background of my question, which was to understand the rationale for using these exact cutoffs. These cutoffs seem arbitrary as data is missing to support that / show why these cutoffs robustly separate real activation from background activation. If these cutoffs are indeed arbitrary, one will wonder what the impact of setting different thresholds would be on the results in this manuscript. So, I believe the reader needs solid data to substantiate these

cutoffs, or, if arbitrary, an analysis showing that the results are robust to choosing a range of other arbitrary cutoffs.

Reply:

We apologize for not fully addressing this point. Based on the activation strength distribution of the negative control TCR in the initial Discovery cohort experiments (n = 98, mean = 4.329, SD = 2.827, see **Revisions Fig. 1** below), we chose a threshold corresponding to approximately 3 standard deviations above the mean (rounded up from 12.81%). Experimentally, we didn't observe a negative control value over 13%, and statistically, the probability of observing a value greater than 13% from the negative control TCR would be approximately 0.11% (p = 0.0011). Therefore, observing an activation strength over 13% in an experimental TCR would lead us to reject the null hypothesis (at a significance level of $p < 0.005$), and favor it being from a distinct distribution of tumor-reactive TCRs. Our confidence in rejecting the null hypothesis improves to a significance level of $p < 7.8 \times 10^{-5}$ when using a 15% threshold, which was the threshold used in downstream analyses. We have now updated the Materials and Methods to more clearly describe how we selected the 13% activation threshold, and we are happy to include **Revisions Fig. 1** in the manuscript if the Reviewer believes it would be helpful for readers.



Revisions Figure 1. Percentage activation strength in Jurkat cells expressing the CD1C-restricted negative control TCR for all co-cultures with the MM cell line panel for all experiments involving the Discovery cohort samples (n = 98). Lower line indicates the mean, while the upper dotted line indicates the 13% boundary initially set to identify hits.

- Finally, a minor point that escaped my attention previously: Fig 6H currently shows the protein structure of HLA-C to pinpoint the location of the relevant residues, but (if I understand the figure correctly) this structure lacks B2M and hence the molecule has been shown in "open conformation". However, the data shows that tumor recognition by TCR65 is strongly dependent on B2M and hence TCR65 is expected to recognise HLA-C in "closed conformation" (that is, in complex with B2M). Hence, it seems more intuitive to show the HLA-C-B2M complex in Fig 6H.

Reply:

We thank the reviewer for this insightful comment. We agree that, given the observed dependence of TCR65 recognition on B2M, a structural representation of HLA-C in its closed conformation would be more appropriate. However, to our knowledge, no experimentally resolved structure of an HLA-C*03:04–B2M complex suitable for accurately mapping the relevant residues is currently available in the Protein Data Bank. We therefore chose to remove this figure panel from the revised manuscript to avoid presenting a potentially misleading structural context. If the reviewer feels that inclusion of the HLA-C-only structure would still add value, we would be happy to reinstate the figure.

Referee #2 (Remarks to the Author):

We must congratulate the authors on their very comprehensive and solid point-by-point reply, and thorough revision of the manuscript. It has improved (from an already very good starting point), and it will certainly be an excellent contribution to the field once published. To ensure the best possible version of the final paper, we have some minor suggestions based on their insightful response and new data provided:

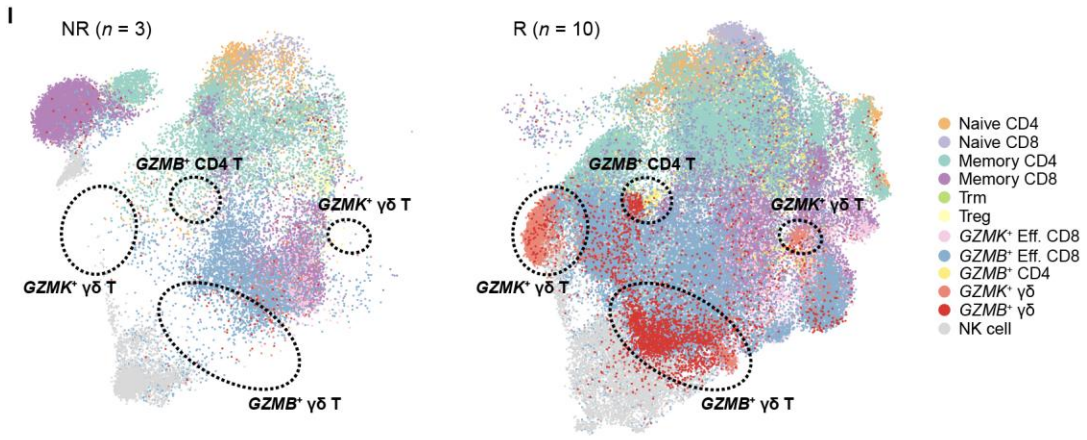
Reply:

We thank the Reviewers for their exceptionally kind words and suggestions for further improvement of our manuscript.

1. The original Extended Data Fig 1d, which they removed but is quite visual and thus helpful, should be added back in;

Reply:

We have now updated and re-included the original **Extended Data Fig. 1d** as **Extended Data Fig. 2l** in the revised manuscript, which we have copied below for the Reviewer's benefit.



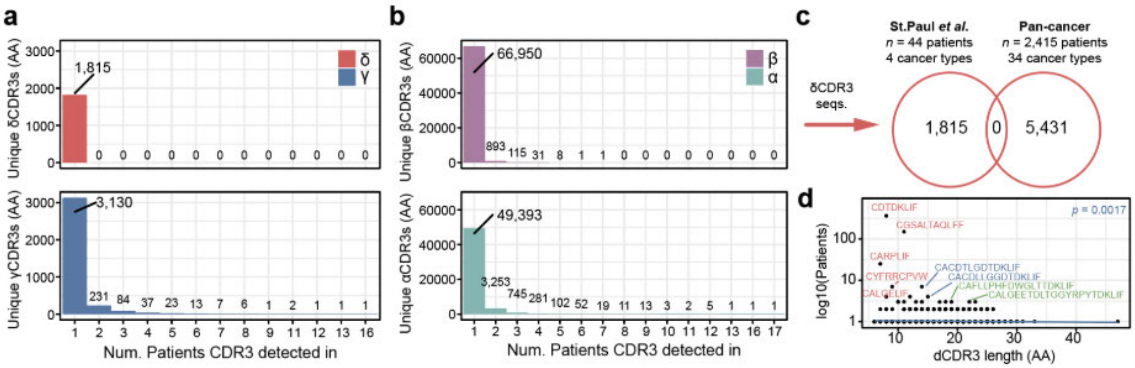
Extended Data Fig. 2. (l) UMAP embedding as in **Extended Data Fig. 2i,** showing cells obtained from NRs (left, *n* = 3) and Rs (right, *n* = 10) separately. Several regions (dashed lines) were devoid of cells in NRs which encapsulated γδ T cells and *GZMB*$^+$ CD4 T cells present in Rs.

2. Revisions Fig. 2.5. is interesting and important, so it should be included in the paper for the benefit of the reader;

Reply:

We thank the Reviewers for their interest in this finding and apologize that in our previous response letter we did not explicitly point out that **Revisions Fig. 2.5a** and **c** were added to the

revised manuscript already, as **Extended Data Fig. 7f,g**. We felt these were the most relevant panels to include in our manuscript, which, given length concerns, were the only panels added. However, if the Reviewer's still feel that panels **b** and **d** of **Revisions Fig. 2.5** should be included as well, we are happy to make that change. We have copied **Revisions Fig. 2.5** below for the Reviewer's benefit.



Revisions Fig. 2.5. (a) Total unique δ CDR3 (top) and γ CDR3 (bottom) AA sequences across our cohorts of MM, MCC, melanoma, and NSCLC patient samples subjected to single-cell sequencing (*n* = 44), and the number of patients in whom such CDR3 sequences were detected. **(b)** Total unique β CDR3 (top) and α CDR3 (bottom) AA sequences across our cohorts of MM, MCC, melanoma, and NSCLC patient samples subjected to single-cell sequencing (*n* = 44), and the number of patients in whom such CDR3 sequences were detected. **(c)** Overlap between the unique δ CDR3 AA sequences identified in this manuscript (as in **a**) and the unique δ CDR3 AA sequences identified in two recent pan-cancer analyses (**PMID: 39368482, PMID: 39058036**). **(d)** Scatter plot comparing δ CDR3 AA length to the number of patients in which the sequence was detected in the external dataset of 2,415 patients (from **c**). Highly abundant sequences are labelled with their respective CDR3 AA sequences. Labels in red indicate likely artifacts; labels in blue indicate Vδ2 sequences; and labels in green indicate Vδ1 or Vδ3 sequences. The Spearman correlation is shown with a blue line, and significance as determined by a Spearman rank correlation test is presented within the plot.

3. In the new Fig. 2f, the TCRs that have NOT been tested against healthy cells (as acknowledge by the authors in their reply) should be clearly noted as such in the figure (using some sort of shading/ stripes or adding NT in the respective squares).

Reply:

We apologize for any confusion from our previous reply and want to clarify that we did indeed test all 24 TCRs presented in **Fig. 2f** (now **Fig. 2d** in our revised manuscript) against the full panel of healthy cells.

4. The rationale for CD8 over expression in Fig. 6 provided in their reply should also be added to the paper for the benefit of the reader;

Reply:

We have now added this rationale to the main text of our revised manuscript as follows:

"Therefore, we overexpressed KIR3DL1 along with CD8a—a required co-receptor for KIR3DL1 and enriched in B2M-dependent TCRs (**Extended Data Fig. 9c**)—in γδ TCR65+ Jurkat cells and co-cultured them with Bw4- ALMC1 and Bw4+ AMO1 MM cells, as well as with Bw4+ primary renal epithelial cells (**Extended Data Fig. 9h,i**)."

5. In Fig. 6m, the label "gd T cell" should be removed as it does not accurately reflect the experiment performed in Jurkat cells, instead stressing the ectopic expression of the gd TCR in that system;

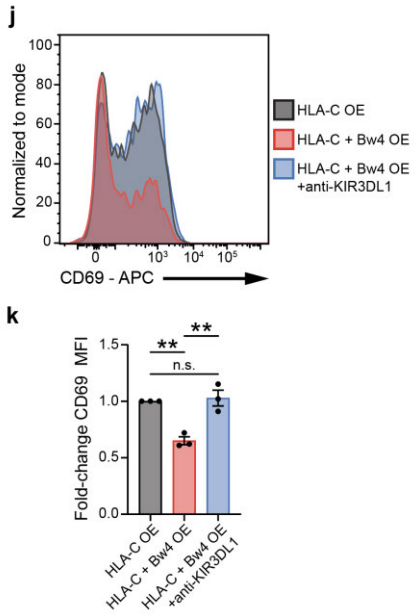
Reply:

We have now changed this label, in **Fig. 6i** of the revised manuscript, to read "TCR65^{+} T cell".

6. Revisions Fig. 2.8. is rather important, so it should be included in the paper for the benefit of the reader (with some stats if possible).

Reply:

We thank the Reviewers for their appraisal of this finding. We have now added **Revisions Fig. 2.8** into our revised manuscript, as well as a statistical quantification of this result, as **Extended Data Fig. 9j and k**, respectively. We have copied these new panels below for the Reviewer's benefit.



Extended Data Fig. 9. (j) Representative flow cytometric determination of CD69 expression in TCR65^{+}CD8^{+}KIR3DL1^{+} Jurkat cells after overnight co-culture with HLA-C*01:02^{+} RPMI-8226 MM cells with or without overexpressed HLA-B*52:01 (contains Bw4) in the presence or absence of an anti-KIR3DL1 blocking antibody. **(k)** Quantification of **j** comparing CD69 expression between groups. Values presented are fold-change compared to the HLA-C-OE group in n = 3 independent experiments, bars represent the mean ± s.e.m. Significance was assessed using ANOVA with Tukey's post-hoc test, ** $p < 0.01$, n.s. not significant.

Upon making these small changes, we fully support the publication of the paper in Nature. Congratulations on your outstanding work!

We thank the Reviewers once again for their time, helpful suggestions, and their endorsement of our manuscript for publication in Nature.

Bruno Silva-Santos & Sofia Mensurado

Referee #4 (Remarks to the Author):

The authors have adequately addressed all the concerns raised by this and the three other reviewers.

Reply:

We thank the Reviewer for taking their time to review our response, manuscript and the helpful input they have given us.

- very minor point, as not included in the revised ms: page 53 of rebuttal, Revisions Figure 3.3. b) : It is not clear how the amount of clonotypes identified with paired TCR can exceed those identified based on either TRD or TRG only.

Reply:

This arises from four definitions in the analysis; 1) defining clonotypes by their CDR3 aa sequence (as opposed to giving clonotypes names such as ‘clonotypeA’, ‘clonotypeB’, etc.), 2) excluding any single chain γ or δ clonotypes where an αβ TCR was sequenced, 3) treating dual TCRs (e.g. TRD:TRG:TRG) as a unique paired clonotype, and 4) plotting the number of unique clonotypes on the y-axis. There are essentially more unique combinations of γCDR3 and δCDR3 pairs (or combinations of 3 in the case of dual TCRs), than there are unique γCDR3 or δCDR3 sequences, in those instances. We hope this clarifies your question.

Referee #5 (Remarks to the Author):

I co-reviewed this manuscript with one of the reviewers who provided the listed reports.

Responses to the Reviewers:

Referee #1 (Remarks to the Author):

The authors have convincingly addressed my final minor points.

To answer the authors' question regarding how to address the fact that an appropriate experimental HLA structure in complex with B2M is unavailable: I do believe that showing the structure with highlighted residues is of value, hence my suggestion would be to use AlphaFold 3 to model the HLA-B2M structure and cross reference this with the experimental HLA-only structure as a sanity check. When both are consistent (especially in terms of the location of the relevant residues), one could envision including the AF3 HLA-B2M structure in the main figures and the experimental HLA-only structure in the supplementals. When the two show substantially different results, it may be wise to refrain from showing any structure.

Reply:

We thank the reviewer for this constructive suggestion. We agree that a structural depiction with the relevant residues highlighted adds value, and we followed the proposed workflow of generating an HLA-C*03:04–B2M complex with AlphaFold 3 to cross-reference against the experimental HLA-C*03:04-only structure.

However, the ipTM scores for the predicted complex fell within the intermediate ("gray zone") range, where the confidence in the relative positioning of the two proteins is neither clearly reliable nor clearly incorrect. Because the interface confidence was ambiguous, we do not believe the predicted arrangement of the units can be presented as accurate, and including it risks conveying a false sense of certainty about features that the model does not support.

For this reason, and in keeping with the Reviewer's note that "it may be wise to refrain from showing any structure" when the results are not dependable, we have opted not to include the AlphaFold 3 HLA-C*03:04–B2M model. We felt that presenting a prediction whose subunit arrangement could not be validated would not add reliable scientific insight and could be potentially misleading. We hope the reviewer agrees that this cautious approach is the most appropriate.

Congratulations with this landmark study!

Reply:

We thank the Reviewer for their time, valuable inputs, and kind words regarding our manuscript.

Joris van de Haar

Netherlands Cancer Institute