nature portfolio

Peer Review File

Liposomal mitoxantrone plus tislelizumab in patients with relapsed or refractory extranodal natural killer/T-cell lymphoma: a phase 1b/2 trial

Corresponding Author: Professor Qingqing Cai

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

Version 0:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
Overall, this is a well-designed and appropriately conducted early-phase clinical study, with generally sound statistical methods and adequate presentation of safety and efficacy data. While protocol amendments are acceptable in early-phase trials, they should be transparently documented and consistently reflected across the protocol, trial registration, and manuscript reporting. Clearer reporting of protocol changes, particularly regarding the primary efficacy endpoint, would strengthen internal consistency and interpretation.

Major concern:
It appears that the primary efficacy endpoint was modified in the ClinicalTrials.gov registration from ORR to CR on March 11, 2025. As changes to a primary outcome may materially influence the interpretation of trial results, please provide the protocol version history with corresponding dates and clarify whether this amendment was implemented before patient enrollment was completed and before data analysis began.

In the current protocol, the sample size calculation is based on CR as the primary efficacy endpoint. Please clarify what the originally prespecified primary efficacy endpoint was before this amendment, and provide the corresponding initial statistical assumptions and sample size justification when ORR was defined as the primary endpoint. A clearer description of the rationale, timing, and implications of this endpoint modification would improve transparency.

Abstract:
“DLTs and RP2D were the primary safety endpoints.”: The description of RP2D as a primary safety endpoint may be slightly imprecise. Consider rephrasing to indicate that DLTs were assessed to determine the RP2D.

Abstract:
For the primary efficacy endpoint, reporting confidence intervals would improve the interpretability and completeness of the results.

Results:
According to the protocol, if no MTD was identified at the highest prespecified dose level (20 mg/m²), further dose escalation could be considered following discussion between the investigator and sponsor. Please clarify whether such a discussion took place and provide the rationale for selecting 20 mg/m² as the RP2D without further dose escalation.

Results:
The description of induction completion and subsequent maintenance therapy in the text does not appear to be fully consistent with the patient disposition shown in Figure 1. In particular, the number of patients reported as completing induction therapy (23) does not match the numbers presented in the flow diagram, and the categorization of patients proceeding to or discontinuing maintenance therapy requires clarification. Please revise the text to ensure numerical accuracy and logical consistency with the study flow diagram.

Results:

Including more detailed information on dose modifications in the Results section—such as the frequency, type, and reasons for dose reductions, delays, or treatment interruptions of Lipo-MIT and tislelizumab—would enhance the description of treatment exposure and tolerability.

Results:
Given the exploratory and post hoc nature of the subgroup analyses, the use of inferential language such as “significantly longer” or “had longer” should be avoided. The authors are encouraged to adopt more cautious, descriptive wording (e.g., “associated with” or “tended to”) and to state that these findings are hypothesis-generating clearly.

Statistical analysis:
As this is a single-arm study, the statistical methods section should primarily emphasize estimation of the primary and secondary efficacy and safety endpoints, including effect measures and corresponding confidence intervals. Any between-subgroup comparisons should be clearly identified as post hoc and exploratory analyses. Explicitly stating that subgroup comparisons are intended for signal generation only and should be interpreted with caution would improve alignment between the statistical approach and the study design.

Table 2:
As the sample size calculation for the primary efficacy endpoint was based on a one-sided hypothesis test, it would be helpful to report the corresponding one-sided 95% confidence interval and one-sided P value for the complete response rate (e.g., in Table 2). This would help ensure internal consistency between the study design assumptions and the statistical reporting.

Reviewer #2

(Remarks to the Author)
The authors designed a new regimen of Lipo-MIT and tislelizumab and conducted a phase 1b/2 study in patients with rr ENKTL.

Major comments:
1. The primary endpoint of this study was best CRR. Regarding historical control, the authors set it at 28% based on Ref #28, but Ref #18 reported it as 33%. Please provide more details on the basis for this setting. Ref #28 is an abstract of a meeting, of which the rate was reported as 28.6% (6/21).

2. Since this is a phase 1b/2 study, please provide more details regarding safety. In Table 3, please show the frequency of each grade. Infections and other adverse events appear to occur more frequently during the dose-expansion phase. In this study, DLT evaluation was only conducted in the first cycle, but as this is a new treatment incorporating an ICI, longer-term toxicity evaluation including irAE and others is desirable.

3. Although many post hoc analyses of prognostic factors have been conducted, they are unlikely to be very useful due to insufficient observation periods. They should be omitted and only the results regarding survival should be described.

4. As the authors state in the introduction, MIT is a drug with cardiotoxicity concerns and is also known as an MDR-related drug. Do the authors have any data showing that combining it with tislelizumab (or other ICIs) is effective for this disease?

Reviewer #3

(Remarks to the Author)
The manuscript by Zou et al is a multicenter phase 1b/2 trial evaluating liposomal mitoxantrone in combination with tislelizumab in relapsed or refractory extranodal NK/T-cell lymphoma (ENKTL), a setting with no established standard of care and where clinical trials remain the preferred option. The study is supported by a biological rationale, with prior preclinical and clinical evidence for both agents. The study population appears representative of the epidemiology of ENKTL in Asia, with response rates comparable to other combinations with PD-1/PD-L1 inh and durable responses in a subset of patients. Furthermore, exploratory correlative analyses suggest potential immune-related predictors of response, although methodological limitations preclude firm conclusions. Toxicities are consistent with the known safety profiles of the individual agents, clinically manageable, and without treatment-related mortality. Overall, this work represents a meaningful contribution to the limited literature in relapsed/refractory ENKTL and supports this combination as a promising therapeutic option while highlighting important unanswered questions regarding optimal combinations, resistance mechanisms, patient selection, and generalizability across diverse populations. Here are some comments that I hope can strengthen the authors’ work:

Major comments:
- Introduction, line 95: The reference supporting the higher prevalence of ENKTL in Latin America should be revised and/or supplemented with an epidemiologic study rather than a review article. Malpica et al. (Lancet Haematology 2025, PMID: 40056928) recently reported primary epidemiologic data on peripheral T-cell lymphomas in Latin America with ENKTL representing the third most common subtype in this region of the world (second most common in Central America). This reference would more appropriately support this statement than a narrative review.
- Results, line 133: Clarification is needed regarding disease staging. Did patients undergo restaging at the time of trial enrollment, or does the statement refer to stage at initial diagnosis (e.g., “25 patients had stage III–IV disease at enrollment?

or at diagnosis")? In routine clinical practice, patients are not typically restaged at relapse unless this represents a new ENKTCL diagnosis after a remote history (very late relapse). Please clarify how staging was defined at enrollment.
- Results, line 191: In the subgroup analysis reporting longer overall survival in patients with stage I-II disease, it is unclear whether this refers to patients with localized disease at initial diagnosis or those experiencing locoregional relapse. This distinction is important, particularly given the earlier statement regarding stage III-IV disease at enrollment, which raises the question of whether restaging was performed at relapse or during initial staging of the disease.
- Results, line 134: The definition of "refractory disease" is not clearly described. While patients who progressed on therapy would be considered refractory, it is unclear how patients who relapsed shortly after completion of first-line therapy were classified. If these patients were considered refractory, the timeframe used should be specified. This clarification is particularly important given the reported differences in progression-free survival between relapsed and refractory disease patients (lines 185–187), and explicit criteria for both categories should be provided.
- Results, multiplex IHC analyses, lines 218–226: After reviewing Appendix pages 16 and 17, median values and interquartile ranges for CD8+, PD-1+CD8+, PD-L1+CD56+, PD-1+CD56+, and PD-L1+ cells were not clearly presented. Although the figures illustrate differences between responders and non-responders, additional detail would strengthen interpretation. Specifically, the authors should clarify how immune infiltration was quantified, how "high" versus "low" infiltration was defined, whether specific percentage thresholds were applied, how proportion and density were calculated, and which cells were used as denominators. These details are not clearly described in the Methods section or figure legends. Greater clarity would help assess whether these findings could have potential clinical relevance for predicting early progression using baseline IHC. I agree with the authors that future studies incorporating spatial transcriptomics and spatial proteomics would be highly informative.
- Safety section / Figure 1 (trial profile): Please clarify in the manuscript text why only 23 of the 40 enrolled patients completed therapy. In the Safety section, the authors state that 14 patients experienced treatment interruptions but were able to restart therapy. However, the trial profile indicates that two patients experienced disease progression during the phase 1b dose-escalation period (yet appear to be included in CR/PR best response assessments), and nine patients progressed during phase 2, with seven discontinuing due to investigator decision, withdrawal of consent, or loss to follow-up. Given the relatively small sample size of this phase 1b/2 study, additional clarification regarding these patients' safety, tolerability, and reasons for discontinuation is warranted, either in the Results or Discussion.
- Given the multiple associations reported, if possible, consideration should be given to performing multivariable analyses to better contextualize these findings.
- Are there on-treatment or post-progression biopsy samples available? If so and feasible, we would suggest incorporating a brief comment on future plans and the value of multi-omic approaches to further elucidate mechanisms of resistance and inform future combination strategies.
- Introduction/Discussion: Consider including a recent meta-analysis evaluating outcomes with anti-PD-1/PD-L1 monotherapy, which reported pooled overall response rates of approximately 50%/43% and 12-month progression-free survival of ~30%, compared with higher response rates and improved 12-month progression-free survival observed with combination approaches such as the regimen studied here (Yang et al., PMID: 40033281).

Minor comments:
- Introduction, line 108: would recommend rephrasing the statement "given the 'mild' toxicities of PD-1/PD-L1 antibodies" with "relatively safe" or "manageable" toxicities as immune checkpoint inhibitors can be associated with severe and potentially life-threatening immune-related adverse events.
- Introduction, line 123: Please correct "Haematology" to "Hematology," which reflects American English usage and the official name of the American Society of Hematology.
- Results, line 138: Recommend revising this sentence to clearly state that the dose-escalation phase involved combination therapy with a fixed dose of tislelizumab.
- Discussion, lines 228–229: There appears to be a missing word that affects sentence clarity. For example, the phrase may require insertion of "given" between "when" and "in" ("…the RP2D of lipo-MIT when given in…").

Reviewer #4

(Remarks to the Author)
Comments to the authors:

SUMMARY

Zou and colleagues report on a phase 1B/2 study of a novel formulation of a classicaly cytotoxic anthracycline – liposomal mitoxantrone – combined with tislelizumab. Relapsed/refractory ENKTL represent a group with unmet needs for more therapeutic options. The inclusion criteria is representative of difficult to treat patients and includes a breadth of exploratory analyses. Lipo-MIT + tislelizumab seem to offer comparable efficacy to other similar combinations in this setting. Toxicity seems to be a concern with high rates of febrile neutropenia, which notably, was not considered a dose-limiting toxicity and represents a limitation of this study. Below are my major and minor comments.

MAJOR COMMENTS:

Methods

Lines 338 and 339 are a bit unclear. Based on the following line, it seems like induction was for a definitive 6 cycles,

followed by maintenance until progression or toxicity. However lines 339 reads that induction was for 6 cycles or until progression or toxicity. Recommend clarification.
Results

Recommend including details on what previous therapies patients received either in the main text or appendix.
Only 60% of patients finished induction. Please report the reasons and numbers of patients that dropped out during induction.
Table 2 and figure 3 – it is not clear if the 95% confidence intervals reported represent the absolute number of participants or the percentage of participants. Please clarify this in the table and figure by adding units to the confidence intervals
Overall 18% patients developed febrile neutropenia. What was the frequency of primary and secondary granulocyte colony stimulating factor use? Recommend reporting.
Discussion

Older pts had markedly lower efficacy than younger. Is this due to older patients requiring more dose reductions or interruptions? This was also seen with ECOG of 2 in the post hoc analyses. Recommend reporting the rates of dose interruptions/dose reductions for older and younger patients, and for those with ECOG 0/1 and 2, and commenting on this in the discussion.
FN not resolving within 7 days was not considered a dose limiting toxicity (as well as some other grade 3+ AEs). However FN is a critical adverse event that requires urgent high-level care and as such is treated as a DLT in many other dose escalation studies. Had this been considered a DLT, there may have been more patients enrolled in the initial dose level because 33% developed FN in the dose escalation cohort. Additionally, 7 (~18%) needed a dose reduction or interruption in dose expansion phase for neutropenia or FN. I recommend commenting on this study design decision in the discussion section and possibly in the limitations section.
Some univariate methods to identify factors associated with higher and lower PFS have extravagant HR and 95% CI estimates, likely due to the small sample sizes compared. Recommend including this in the limitations and highlighting these associations, while interesting, are exploratory, even when using penalized cox regression as there may be underlying selection biases among subgroups (e.g. patients not fit enough for previous HSCT are likely to have more dose interruptions with lipo-MIT and as such have reduced efficacy).

MINOR COMMENTS:

Methods

Line 368 – enrollment is misspelt (spelt as “enrolment”)
Line 392 – enroll is misspelt (spelt as “enrol”)

Reviewer #5

(Remarks to the Author)
I co-reviewed this manuscript with one of the reviewers who provided the listed reports. This is part of the Nature Communications initiative to facilitate training in peer review and to provide appropriate recognition for Early Career Researchers who co-review manuscripts.

Reviewer #6

(Remarks to the Author)
I co-reviewed this manuscript with one of the reviewers who provided the listed reports. This is part of the Nature Communications initiative to facilitate training in peer review and to provide appropriate recognition for Early Career Researchers who co-review manuscripts.

Version 1:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
I thank the authors for their careful and constructive responses to my previous comments. Overall, the revised manuscript has addressed most of my statistical concerns. I have one remaining statistical clarification and one minor presentation suggestion.

1. Given the importance of the amendment to the primary efficacy endpoint, I recommend that the following clarification or similar wording be explicitly included in the Methods or Statistical analysis section, rather than only being provided in the supplementary protocol amendment table:

The primary efficacy endpoint was clarified from ORR to best CR rate in a protocol amendment dated February 10, 2025, after completion of enrollment but before database lock and any efficacy analyses; the sample size assumptions and planned statistical test were based on CR rate and remained unchanged throughout the study.

2. When reporting a one-sided 95% confidence interval in the main text, I suggest reporting it as a one-sided 95% lower confidence bound rather than as a confidence interval with an upper bound of 100%. For example, the current wording: The CR rate was 53% (21/40; one-sided 95% CI 38%–100%; p<0.001) could be revised to: The CR rate was 53% (21/40; one-sided 95% lower confidence bound, 38%; one-sided p<0.001).

Reviewer #2

(Remarks to the Author)
no further suggestions

Reviewer #3

(Remarks to the Author)
The authors have responded to my comments. I have no further comments.

Reviewer #5

(Remarks to the Author)
I co-reviewed this manuscript with one of the reviewers who provided the listed reports. This is part of the Nature Communications initiative to facilitate training in peer review and to provide appropriate recognition for Early Career Researchers who co-review manuscripts.

Point-by-point responses

Reviewer #1 (Remarks to the Author):

Overall, this is a well-designed and appropriately conducted early-phase clinical study, with generally sound statistical methods and adequate presentation of safety and efficacy data. While protocol amendments are acceptable in early-phase trials, they should be transparently documented and consistently reflected across the protocol, trial registration, and manuscript reporting. Clearer reporting of protocol changes, particularly regarding the primary efficacy endpoint, would strengthen internal consistency and interpretation.

Major concern

1) ***It appears that the primary efficacy endpoint was modified in the ClinicalTrials.gov registration from ORR to CR on March 11, 2025. As changes to a primary outcome may materially influence the interpretation of trial results, please provide the protocol version history with corresponding dates and clarify whether this amendment was implemented before patient enrollment was completed and before data analysis began.***

In the current protocol, the sample size calculation is based on CR as the primary efficacy endpoint. Please clarify what the originally prespecified primary efficacy endpoint was before this amendment, and provide the corresponding initial statistical assumptions and sample size justification when ORR was defined as the primary endpoint. A clearer description of the rationale, timing, and implications of this

endpoint modification would improve transparency.

Response: We thank the reviewer for this important comment. In the original version of the ClinicalTrials.gov registration (June 15, 2022), the primary efficacy endpoint was listed ORR, but the sample size calculation was based on CR rate. We noticed this error and amended the protocol on February 10, 2025, to clearly specify CR rate as the primary efficacy endpoint. This amendment was made after completion of patient enrollment but prior to database lock and before any efficacy analyses. After IRB approval of the change in trial protocol, we amended the ClinicalTrials.gov registration on March 11, 2025. The planned sample size and statistical assumptions remained unchanged throughout the study.

Details of the changes to trial protocol are now submitted in the Protocol (page 62-63). A statement that directs the readers to the change history has been added to the main text (Line 475-476).

Line 475-476: The trial protocol and a table of amendments are available in the supplementary materials.

Protocol (page 62-63):

Protocol amendments

Versions	Date	Original content	Amendment(s)	Reason(s)
V 1.0	2022-3-21	—	—	—
V 1.1	2023-10-12	extranodal NK/T-cell lymphoma, nasal type	extranodal NK/T-cell lymphoma	The terminology has been revised in accordance with "The 5th edition of the WHO Classification of Haematolymphoid Tumours".
		SCHEDULE OF ASSESSMENTS FACT-H&N Version 4	Deleted: "FACT-H&N Version 4"	The FACT-H&N was primarily developed for patients with head and neck squamous cell carcinoma and was not fully applicable to patients with systemic NK/T-cell lymphoma.
V 1.2	2025-2-10	Primary endpoints •Dose-expansion phase To evaluate the efficacy of tislelizumab combined with liposomal mitoxantrone in relapsed/refractory ENKTL, which is the objective response rate.	Primary endpoints •Dose-expansion phase To evaluate the efficacy of tislelizumab combined with liposomal mitoxantrone in relapsed/refractory ENKTL, which is the complete response rate.	The sample size calculation was based solely on the complete response rate, which served as the predefined efficacy benchmark of the trial. To maintain consistency between the statistical framework and the endpoint structure, the complete response rate was formally defined as the primary efficacy endpoint
		Anticipated date for first patient enrolment: May 2022; Anticipated completion of enrolment: January 2024; Anticipated end date of follow-up: January 2025.	Anticipated date for first patient enrolment: May 2022; Anticipated completion of enrolment: November 2024; Anticipated end date of follow-up: November 2025.	Updated based on the study enrolment progress and anticipated follow-up schedule.

2) Abstract:

"DLTs and RP2D were the primary safety endpoints.": The description of RP2D as a primary safety endpoint may be slightly imprecise. Consider rephrasing to indicate that DLTs were assessed to determine the RP2D.

Response: We thank the reviewer's suggestion, and have revised the statement to indicate the primary endpoint of the dose-escalation phase was DLTs, and the RP2D was derived based on DLTs (Line 76-78 and 410-412).

Line 76-78: The primary endpoint of the dose-escalation phase was dose-limiting toxicities (DLTs), which in turn were assessed to determine the RP2D.

Line 410-412: The primary endpoint of the dose-escalation phase was DLTs, which in turn were assessed to determine the RP2D of Lipo-MIT when given in combination with tislelizumab.

3) Abstract:

For the primary efficacy endpoint, reporting confidence intervals would improve the interpretability and completeness of the results.

Response: We appreciate the reviewer's comment, and have added 95% confidence interval to the primary efficacy endpoint of CR rate (Line 80-82).

Line 80-82: The CR rate was 53% (21/40; one-sided 95% CI 38%-100%), and the objective response rate was 73% (29/40; 95% CI 56%-85%).

4) Results:

According to the protocol, if no MTD was identified at the highest prespecified dose level (20 mg/m^2), further dose escalation could be considered following discussion between the investigator and sponsor. Please clarify whether such a discussion took place and provide the rationale for selecting 20 mg/m^2 as the RP2D without further dose escalation.

Response: We appreciate the reviewer's comments, and have expanded the manuscript (Line 140-142) to indicate that 20 mg/m^2 was selected as the RP2D following a discussion between the investigators and sponsor based on cumulative safety data and prior clinical evidence.

Briefly, a phase 1b trial of liposomal mitoxantrone in NHL showed that the incidence of ≥ 3 neutropenia was 57% (4/7) at doses of 22 or 24 mg/m^2.[1] Furthermore,

liposomal mitoxantrone monotherapy at 20 mg/m² has been shown to be efficacious in a prior phase 2 trial in patients with relapsed/refractory T-cell lymphoma, including ENKTL (reported in 2020),[2] and subsequently approved for relapsed/refractory T-cell lymphoma in China on January 7, 2022.[3] Considering that 20mg/m^2 is the recommended monotherapy dose of liposomal mitoxantrone and taking into account the potential risk of myelosuppression, this dose level was selected as the RP2D when used in combination with tislelizumab following a discussion between the investigators and sponsor.

Line 140-142: The recommended phase 2 dose (RP2D) of Lipo-MIT was determined as 20 mg/m^2 when given in combination with tislelizumab, following a discussion based on the cumulative safety data and prior clinical evidence.

Reference

[1] Li Simon, Gao YH, Ma GY, *et al.* Phase 1b safety and antitumor activity of PLM60 (pegylated liposomal mitoxantrone) in NHL. J Clin Oncol. 2018 36: 15_suppl, e19503-e19503.

[2] Gao Y, Huang HQ, Wang XX, *et al.* Safety and efficacy of mitoxantrone hydrochloride liposome in patients with relapsed or refractory peripheral T-cell lymphoma and extranodal NK/T-cell lymphoma: a prospective, single-arm, open-label, multi-center, phase II clinical trial. Blood. 2020 136 (Supple 1):36-37.

[3] https://www.nmpa.gov.cn/zwfw/sdxx/sdxxyp/yppjfb/20220111154458197.html

5) Results:

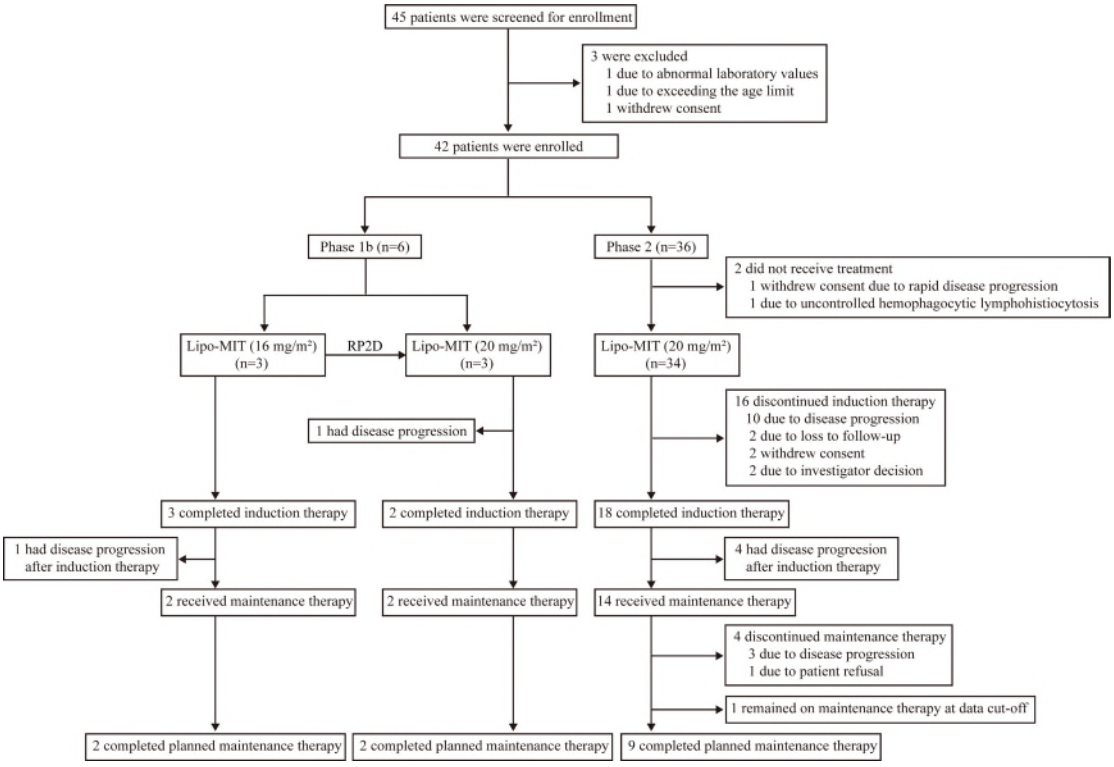
The description of induction completion and subsequent maintenance therapy in the text does not appear to be fully consistent with the patient disposition shown in Figure 1. In particular, the number of patients reported as completing induction therapy (23) does not match the numbers presented in the flow diagram, and the categorization of patients proceeding to or discontinuing maintenance therapy requires clarification. Please revise the text to ensure numerical accuracy and logical consistency with the study flow diagram.

Response: We appreciate the reviewer's careful comments, and have revised both the Results section (Line 146-154) and Figure 1 to present a clear, stepwise description of patient disposition during induction and maintenance therapy. Briefly, the 23 patients who completed induction therapy included 5 patients in the phase 1b and 18 patients in the phase 2. Four of 5 patients who completed induction therapy in phase 1b and 14 of 18 patients who completed induction therapy in phase 2 (a total of 18) proceeded to tislelizumab maintenance therapy.

Line 146-154: Twenty-three (58%) patients completed all 6 planned induction cycles, whereas 17 discontinued induction therapy due to disease progression (n=11), loss to follow-up (n=2), withdrawal of consent (n=2), or investigator decision (n=2). Among the 23 patients who completed induction therapy, 18 proceeded to tislelizumab maintenance therapy and 5 did not undergo maintenance because of disease progression at the end of induction. Of the 18 patients who received maintenance therapy, 13 completed the planned maintenance course, 1 remained on treatment at data cut-off,

and 4 discontinued maintenance therapy due to disease progression (n=3) or patient refusal (n=1) (Figure 1).

Figure 1:



6) Results:

Including more detailed information on dose modifications in the Results section—such as the frequency, type, and reasons for dose reductions, delays, or treatment interruptions of Lipo-MIT and tislelizumab—would enhance the description of treatment exposure and tolerability.

Response: We appreciate the reviewer's comments and have added detailed description of the frequency, type, and reasons for dose reduction/delay/interruption of Lipo-MIT

and tislelizumab in the Results section (Line 207-213) and Supplementary Table S7 and S8.

Line 207-213: Treatment interruptions were reported in 14 patients due to neutropenia or febrile neutropenia (n=6 [43%] of 14 patients), leukopenia (n=2 [14%]), COVID-19 infection (n=2 [14%]), thrombocytopenia (n=1 [7%]), abnormal liver function (n=1 [7%]), diarrhea (n=1 [7%]), and hypothyroidism (n=1 [7%]; appendix p 16). All 14 patients resumed planned treatment after recovery from these adverse events. Four patients required Lipo-MIT dose reductions due to neutropenia (n=2 [50%] of 4 patients), febrile neutropenia (n=1 [25%]), and thrombocytopenia (n=1 [25%]; appendix p 17).

Table S7. Reasons for treatment interruptions (n=14)

Treatment-related adverse event	Number (%)
Neutropenia or febrile neutropenia	6 (43%)
Leukopenia	2 (14%)
COVID-19 infection	2 (14%)
Thrombocytopenia	1 (7%)
ALT, AST and bilirubin elevated	1 (7%)
Diarrhea	1 (7%)
Hypothyroidism	1 (7%)

Data are shown as number (%).

Abbreviations: ALT, alanine aminotransferase; AST, aspartate aminotransferase.

Table S8. Reasons for Lipo-MIT dose reduction (n=4)

Treatment-related adverse event	Number (%)
Neutropenia	2 (50%)
Febrile neutropenia	1 (25%)
Thrombocytopenia	1 (25%)

Data are shown as number (%).

7) Results:

Given the exploratory and post hoc nature of the subgroup analyses, the use of inferential language such as "significantly longer" or "had longer" should be avoided. The authors are encouraged to adopt more cautious, descriptive wording (e.g., "associated with" or "tended to") and to state that these findings are hypothesis-generating clearly.

Response: We appreciate the reviewer's comment, and have replaced inferential language with description wordings when describing/discussing the results of subgroup analysis of the primary endpoint of CR (Line 165-167 and 169-171). Subgroup analysis of the survival outcomes has been deleted per comments by another reviewer.

Line 165-167: In the post-hoc subgroup analysis, a lower CR rate was observed in patients aged > 60 years (0% [0/7] vs 64% [21/33] for ≤ 60 years; p=0.003).

Line 169-171: Patients with prior exposure to anti-PD-1 antibodies tended to have a lower CR rate than those without (38% [8/21] vs 68% [13/19]; p=0.055; appendix p 20). These findings are hypothesis-generating and should be interpreted with caution.

8) Statistical analysis:

As this is a single-arm study, the statistical methods section should primarily emphasize estimation of the primary and secondary efficacy and safety endpoints, including effect measures and corresponding confidence intervals. Any between-subgroup comparisons should be clearly identified as post hoc and exploratory

analyses. Explicitly stating that subgroup comparisons are intended for signal generation only and should be interpreted with caution would improve alignment between the statistical approach and the study design.

Response: We appreciate the reviewer's comment, and have revised the statistical methods section to emphasize the estimation of the primary and secondary efficacy endpoints. Subgroup analyses are now explicitly labelled as exploratory and post-hoc and therefore should be interpreted with caution (Line 439-450).

Line 439-450: Categorical variables are summarized as frequencies and percentages, and continuous variables as medians with IQRs. The CR rate was estimated with one-sided 95% CI using the Clopper-Pearson exact method. A one-sided exact binomial test was used to compare the observed CR rate with prespecified historical control rate of 0.28. Secondary efficacy endpoints were reported with two-sided 95% CIs. PFS, OS, DFS, and DOR were estimated using the Kaplan-Meier method. Univariable Cox regression analysis was used to estimate HRs with corresponding 95% CIs for survival outcomes. Subgroup analyses were post-hoc and for signal generation only. Categorical variables were compared using chi-square test or Fisher's exact test as appropriate. Continuous variables were compared using Mann-Whitney U test. Two-sided tests were used, and nominal p values < 0.05 were considered statistically significant. Statistical analyses were performed using R version 4.5.0 and GraphPad Prism 8.

9) Table 2:

As the sample size calculation for the primary efficacy endpoint was based on a one-sided hypothesis test, it would be helpful to report the corresponding one-sided 95% confidence interval and one-sided P value for the complete response rate (e.g., in Table 2). This would help ensure internal consistency between the study design assumptions and the statistical reporting.

Response: We appreciate the reviewer's comment, and have revised the manuscript accordingly. The primary efficacy endpoint of CR rate is now reported with the corresponding one-sided 95% confidence interval and one-sided p value (Line 80-82, 156-158, 164-165, and 440-443), and Table 2.

Line 80-82: The CR rate was 53% (21/40; one-sided 95% CI 38%-100%), and the objective response rate was 73% (29/40; 95% CI 56%-85%).

Line 156-158: In the FAS (n=40), the best CR rate was 53% (21/40; one-sided 95% confidence interval [CI] 38%-100%; p<0.001) and the best ORR was 73% (29/40; 95% CI 56%-85%) (table 2).

Line 165-166: The CR rate was 55% (21/38; one-sided 95% CI 41%-100%) and the ORR was 76% (29/38; 95% CI 60%-89%) in the PPS (appendix p 10).

Line 440-443: The CR rate was estimated with one-sided 95% CI using the Clopper-Pearson exact method. A one-sided exact binomial test was used to compare the observed CR rate with prespecified historical control rate of 0.28. Secondary efficacy endpoints were reported with two-sided 95% CIs.

Table 2. Anti-tumor activity in the full analysis set (n=40)

	Dose-escalation phase (n=6)	Dose-expansion phase (n=34)	Total (n=40)
Best response*			
Complete response#	5 (83%; 42%-100%)	16 (47%; 32%-100%)	21 (53%; 38%-100%)
Partial response	1 (17%; 0-64%)	7 (21%; 9%-38%)	8 (20%; 9%-36%)
Stable disease	0	2 (6%; 1%-20%)	2 (5%; 1%-17%)
Progressive disease	0	7 (21%; 9%-38%)	7 (18%; 7%-33%)
Not evaluable	0	2 (6%; 1%-20%)	2 (5%; 1%-17%)
Objective response rate	6 (100%; 54%-100%)	23 (68%; 49%-83%)	29 (73%; 56%-85%)
Disease control rate	6 (100%; 54%-100%)	25 (74%; 56%-87%)	31 (78%; 62%-89%)

Data are shown as number (%, confidence interval). *Best response was defined as the best response achieved at any time during treatment according to the Lugano 2014 criteria. #Confidence intervals for the primary efficacy endpoint (complete response rate) are one-sided 95%, corresponding to the prespecified one-sided α of 0.05. Confidence intervals for secondary efficacy endpoints are two-sided 95%. Percentages may not add to 100 due to rounding.

Reviewer #2 (Remarks to the Author):

The authors designed a new regimen of Lipo-MIT and tislelizumab and conducted a phase 1b/2 study in patients with rr ENKTL.

Major comments:

1) The primary endpoint of this study was best CRR. Regarding historical control, the authors set it at 28% based on Ref #28, but Ref #18 reported it as 33%. Please provide more details on the basis for this setting. Ref #28 is an abstract of a meeting, of which the rate was reported as 28.6% (6/21).

Response: We appreciate the reviewer's inquiry on this issue. Briefly, at the time of trial initiation (March 2022), the only available data on liposomal mitoxantrone monotherapy in relapsed or refractory ENKTL were from the 2020 ASH abstract (Ref #28 in the original submission; Ref #41 in the revised manuscript). Accordingly, we used it as the historical control for sample size calculation. The 33% CR rate in Ref #18 (now Ref #19 in the revised manuscript) was published in 2025.

2) Since this is a phase 1b/2 study, please provide more details regarding safety. In Table 3, please show the frequency of each grade. Infections and other adverse events appear to occur more frequently during the dose-expansion phase. In this study, DLT evaluation was only conducted in the first cycle, but as this is a new treatment incorporating an ICI, longer-term toxicity evaluation including irAE and others is

desirable.

Response: We appreciate the reviewer's comments. As the reviewer pointed out, there were some differences in the incidence of AEs between the phase 1b and phase 2. Such differences may be partly attributable to the relatively small sample size in the phase 1b (n=6). To provide more detailed safety information, we have presented all adverse events in safety analysis set, and in phase 1b and 2, with the frequency of each grade reported, in the revised manuscript (Table 3 and Supplementary Tables S4 and S5).

In addition, we agree with the reviewer that toxicity evaluation should be conducted beyond the first treatment cycle, and added these data (AEs during the entire treatment period), with an emphasis on irAEs, to the revised manuscript (Line 204-206 and Supplementary Table S6).

Line 204-206: Immune-related AEs included hypothyroidism (14 [35%]), rash (6 [15%]), pruritus (4 [10%]), and hyperthyroidism (1 [3%]), all of which were grade 1-2 (appendix p 15).

Table 3 and Table S4, 5, and 6:

Table 3. Adverse events in the safety analysis set (n=40)

	Grade 1	Grade 2	Grade 3	Grade 4	Any grade
Hematological toxicities					
Leukopenia	1 (3%)	8 (20%)	19 (48%)	2 (5%)	30 (75%)
Neutropenia	1 (3%)	11 (28%)	11 (28%)	5 (13%)	28 (70%)
Anemia	17 (43%)	8 (20%)	3 (8%)	0	28 (70%)
Lymphopenia	4 (10%)	4 (10%)	12 (30%)	2 (5%)	22 (55%)
Thrombocytopenia	4 (10%)	4 (10%)	0	2 (5%)	10 (25%)
Febrile neutropenia	0	0	7 (18%)	0	7 (18%)
Non-hematological toxicities					
Hyperglycemia	14 (35%)	0	1 (3%)	0	15 (38%)
Pigmentation	14 (35%)	0	0	0	14 (35%)
Hypothyroidism	9 (23%)	5 (13%)	0	0	14 (35%)
Albumin decreased	13 (33%)	1 (3%)	0	0	14 (35%)
Fever	9 (23%)	3 (8%)	0	0	12 (30%)
AST increased	6 (15%)	3 (8%)	1 (3%)	0	10 (25%)
Infection#	1 (3%)	5 (13%)	3 (8%)	0	9 (23%)
ALT increased	6 (15%)	0	2 (5%)	0	8 (20%)
Anorexia	4 (10%)	2 (5%)	1 (3%)	0	7 (18%)
Hyponatremia	6 (15%)	0	1 (3%)	0	7 (18%)
Hypokalemia	6 (15%)	0	0	0	6 (15%)
Hypocalcemia	6 (15%)	0	0	0	6 (15%)
Rash	4 (10%)	2 (5%)	0	0	6 (15%)
Fatigue	4 (10%)	1 (3%)	0	0	5 (13%)
Nausea or vomiting	2 (5%)	1 (3%)	1 (3%)	0	4 (10%)
Diarrhea	3 (8%)	0	1 (3%)	0	4 (10%)
Infusion reaction	0	4 (10%)	0	0	4 (10%)
Alkaline phosphatase increased	4 (10%)	0	0	0	4 (10%)
Pruritus	3 (8%)	1 (3%)	0	0	4 (10%)
Blood bilirubin increased	2 (5%)	0	1 (3%)	0	3 (8%)
Headache	1 (3%)	1 (3%)	0	0	2 (5%)
Blood creatinine increased	2 (5%)	0	0	0	2 (5%)
Palpitations	2 (5%)	0	0	0	2 (5%)
Atrial fibrillation	0	1 (3%)	0	0	1 (3%)
QT interval prolonged	1 (3%)	0	0	0	1 (3%)
Oral ulcer	1 (3%)	0	0	0	1 (3%)
Hyperthyroidism	1 (3%)	0	0	0	1 (3%)
Eczema	0	1 (3%)	0	0	1 (3%)
Herpes zoster	0	1 (3%)	0	0	1 (3%)

Data are shown as number (%). #Infections included lung infection (n=4), COVID-19 infection (n=2), urinary tract infection (n=2), and intestinal infection (n=1).

Abbreviations: ALT, alanine aminotransferase; AST, aspartate aminotransferase.

Table S4. Adverse events in phase 1b (n=6)

	Grade 1	Grade 2	Grade 3	Grade 4
Hematological toxicities				
Leukopenia	0	1 (17%)	2 (33%)	2 (33%)
Neutropenia	0	2 (33%)	1 (17%)	2 (33%)
Anemia	4 (67%)	1 (17%)	0	0
Lymphopenia	1 (17%)	2 (33%)	0	1 (17%)
Thrombocytopenia	0	1 (17%)	0	1 (17%)
Febrile neutropenia	0	0	2 (33%)	0
Non-hematological toxicities				
Hyperglycemia	4 (67%)	0	0	0
Pigmentation	4 (67%)	0	0	0
Hypothyroidism	0	1 (17%)	0	0
Albumin decreased	2 (33%)	0	0	0
Fever	1 (17%)	0	0	0
AST increased	0	1 (17%)	0	0
Infection[#]	0	2 (33%)	0	0
ALT increased	1 (17%)	0	1 (17%)	0
Anorexia	0	1 (17%)	0	0
Hyponatremia	2 (33%)	0	0	0
Hypokalemia	3 (50%)	0	0	0
Hypocalcemia	1 (17%)	0	0	0
Rash	2 (33%)	0	0	0
Nausea or vomiting	0	1 (17%)	0	0
Diarrhea	0	0	1 (17%)	0
Alkaline phosphatase increased	1 (17%)	0	0	0
Pruritus	2 (33%)	0	0	0
Headache	0	1 (17%)	0	0

Data are shown as number (%). [#]Infections included COVID-19 infection (n=1) and intestinal infection (n=1).

Abbreviations: ALT, alanine aminotransferase; AST, aspartate aminotransferase.

Table S5. Adverse events in phase 2 (n=34)

	Grade 1	Grade 2	Grade 3	Grade 4
Hematological toxicities				
Leukopenia	1 (3%)	7 (21%)	17 (50%)	0
Neutropenia	1 (3%)	9 (26%)	10 (29%)	3 (9%)
Anemia	13 (38%)	7 (21%)	3 (9%)	0
Lymphopenia	3 (9%)	2 (6%)	12 (35%)	1 (3%)
Thrombocytopenia	4 (12%)	3 (9%)	0	1 (3%)
Febrile neutropenia	0	0	5 (15%)	0
Non-hematological toxicities				
Hyperglycemia	10 (29%)	0	1 (3%)	0
Pigmentation	10 (29%)	0	0	0
Hypothyroidism	9 (26%)	4 (12%)	0	0
Albumin decreased	11 (32%)	1 (3%)	0	0
Fever	8 (24%)	3 (9%)	0	0
AST increased	6 (18%)	2 (6%)	1 (3%)	0
Infection#	1 (3%)	3 (9%)	3 (9%)	0
ALT increased	5 (15%)	0	1 (3%)	0
Anorexia	4 (12%)	1 (3%)	1 (3%)	0
Hyponatremia	4 (12%)	0	1 (3%)	0
Hypokalemia	3 (9%)	0	0	0
Hypocalcemia	5 (15%)	0	0	0
Rash	2 (6%)	2 (6%)	0	0
Fatigue	4 (12%)	1 (3%)	0	0
Nausea or vomiting	2 (6%)	0	1 (3%)	0
Diarrhea	3 (9%)	0	0	0
Infusion reaction	0	4 (12%)	0	0
Alkaline phosphatase increased	3 (9%)	0	0	0
Pruritus	1 (3%)	1 (3%)	0	0
Blood bilirubin increased	2 (6%)	0	1 (3%)	0
Headache	1 (3%)	0	0	0
Blood creatinine increased	2 (6%)	0	0	0
Palpitations	2 (6%)	0	0	0
Atrial fibrillation	0	1 (3%)	0	0
QT interval prolonged	1 (3%)	0	0	0
Oral ulcer	1 (3%)	0	0	0
Hyperthyroidism	1 (3%)	0	0	0
Eczema	0	1 (3%)	0	0
Herpes zoster	0	1 (3%)	0	0

Data are shown as number (%). #Infections included lung infection (n=4), COVID-19 infection (n=1), and urinary tract infection (n=2).

Abbreviations: ALT, alanine aminotransferase; AST, aspartate aminotransferase.

Table S6. Immune-related adverse events in the safety analysis set (n=40)

	Grade 1	Grade 2	Grade 3	Grade 4	Any grade
Hypothyroidism	9 (23%)	5 (13%)	0	0	14 (35%)
Rash	4 (10%)	2 (5%)	0	0	6 (15%)
Pruritus	3 (8%)	1 (3%)	0	0	4 (10%)
Hyperthyroidism	1 (3%)	0	0	0	1 (3%)

Data are shown as number (%).

3) Although many post hoc analyses of prognostic factors have been conducted, they are unlikely to be very useful due to insufficient observation periods. They should be omitted and only the results regarding survival should be described.

Response: We appreciate the reviewer's comment, and have omitted exploratory post-hoc survival analyses of prognostic factors from the revised manuscript.

4) As the authors state in the introduction, MIT is a drug with cardiotoxicity concerns and is also known as an MDR-related drug. Do the authors have any data showing that combining it with tislelizumab (or other ICIs) is effective for this disease?

Response: We appreciate the reviewer's comment. At the time of trial initiation, no data were available to suggest that combination of mitoxantrone with immune checkpoint inhibitors is effective for relapsed or refractory ENKTL. But a subsequent small retrospective study reported an objective response rate of 67% with tislelizumab in combination with MIT-based regimens. This has been added to Discussion section (Line 252-256).

Line 252-256: In addition, a small retrospective study of 6 patients with R/R ENKTL reported a 67% ORR with tislelizumab in combination with MIT-based regimens, including MIT plus pegaspargase and dexamethasone (n=3), lipo-MIT (n=1), lipo-MIT plus chidamide (n=1), and lipo-MIT plus brentuximab vedotin (n=1), suggesting potential anti-tumor activity of this combination.[29]

Reviewer #3 (Remarks to the Author):

The manuscript by Zou et al is a multicenter phase 1b/2 trial evaluating liposomal mitoxantrone in combination with tislelizumab in relapsed or refractory extranodal NK/T-cell lymphoma (ENKTL), a setting with no established standard of care and where clinical trials remain the preferred option. The study is supported by a biological rationale, with prior preclinical and clinical evidence for both agents. The study population appears representative of the epidemiology of ENKTL in Asia, with response rates comparable to other combinations with PD-1/PD-L1 inhibitors and durable responses in a subset of patients. Furthermore, exploratory correlative analyses suggest potential immune-related predictors of response, although methodological limitations preclude firm conclusions. Toxicities are consistent with the known safety profiles of the individual agents, clinically manageable, and without treatment-related mortality. Overall, this work represents a meaningful contribution to the limited literature in relapsed/refractory ENKTL and supports this combination as a promising therapeutic option while highlighting important unanswered questions regarding optimal combinations, resistance mechanisms, patient selection, and generalizability across diverse populations. Here are some comments that I hope can strengthen the authors' work:

Major comments:

1) - Introduction, line 95: The reference supporting the higher prevalence of ENKTL in Latin America should be revised and/or supplemented with an epidemiologic study

rather than a review article. Malpica et al. (Lancet Haematology 2025, PMID: 40056928) recently reported primary epidemiologic data on peripheral T-cell lymphomas in Latin America with ENKTL representing the third most common subtype in this region of the world (second most common in Central America). This reference would more appropriately support this statement than a narrative review.

Response: We appreciate the reviewer's comment, and have added the study by Malpica et al. to the Introduction section as a reference to support the claim (Line 93 and Ref #5 in the revised manuscript).

Line 93: ENKTL is more common in the East Asian and Latin American population.[4,5]

Ref #5: Malpica, L. et al. Epidemiology, clinical features, and outcomes of peripheral T-cell lymphoma in Latin America: an international, retrospective, cohort study. Lancet Haematol 12, e258-e268 (2025).

2) - Results, line 133: Clarification is needed regarding disease staging. Did patients undergo restaging at the time of trial enrollment, or does the statement refer to stage at initial diagnosis (e.g., "25 patients had stage III–IV disease at enrollment? or at diagnosis")? In routine clinical practice, patients are not typically restaged at relapse unless this represents a new ENKTCL diagnosis after a remote history (very late relapse). Please clarify how staging was defined at enrollment.

Response: We appreciate the reviewer's comment. The manuscript has been revised to state that all patients underwent restaging at the time of trial enrollment using the Ann

Arbor staging system to assess disease status and disease burden (Line 402-404 and Table 1 footnote).

Line 402-404: Disease stage was determined using the Ann Arbor staging system based on imaging at trial enrollment.

Table 1 footnote: #Disease stage was determined according to the Ann Arbor staging system at trial enrollment.

3) - Results, line 191: In the subgroup analysis reporting longer overall survival in patients with stage I-II disease, it is unclear whether this refers to patients with localized disease at initial diagnosis or those experiencing locoregional relapse. This distinction is important, particularly given the earlier statement regarding stage III-IV disease at enrollment, which raises the question of whether restaging was performed at relapse or during initial staging of the disease.

Response: We appreciate the reviewer's comment, and have revised the manuscript to clearly state that all patients underwent restaging at the time of trial enrollment to assess disease status and disease burden (Line 402-404 and Table 1 footnote). The stage I-II classification reflects localized involvement at study entry rather than at initial diagnosis.

Line 402-404: Disease stage was determined using the Ann Arbor staging system based on imaging at trial enrollment.

Table 1 footnote: #Disease stage was determined according to the Ann Arbor staging

system at trial enrollment.

4) - Results, line 134: The definition of "refractory disease" is not clearly described. While patients who progressed on therapy would be considered refractory, it is unclear how patients who relapsed shortly after completion of first-line therapy were classified. If these patients were considered refractory, the time-frame used should be specified. This clarification is particularly important given the reported differences in progression-free survival between relapsed and refractory disease patients (lines 185–187), and explicit criteria for both categories should be provided.

Response: We appreciate the reviewer's comment, and have expanded the Methods section (Line 358-362) to clarify the definitions of refractory and relapsed diseases. Briefly, refractory disease was defined as failure to achieve a complete response to asparaginase-based chemotherapy or disease progression within 6 months after the last asparaginase-containing regimen; relapsed disease was defined as disease recurrence after achieving a complete response to initial therapy and occurring beyond 6 months of the last regimen.

Line 358-362: Refractory disease was defined as failure to achieve a CR to asparaginase-based chemotherapy or disease progression within 6 months after the last asparaginase-containing regimen. Relapsed disease was defined as disease recurrence after achieving a CR to initial therapy and occurring beyond 6 months of the last regimen.

5) - Results, multiplex IHC analyses, lines 218–226: After reviewing Appendix pages 16 and 17, median values and interquartile ranges for CD8+, PD-1+CD8+, PD-L1+CD56+, PD-1+CD56+, and PD-L1+ cells were not clearly presented. Although the figures illustrate differences between responders and non-responders, additional detail would strengthen interpretation. Specifically, the authors should clarify how immune infiltration was quantified, how "high" versus "low" infiltration was defined, whether specific percentage thresholds were applied, how proportion and density were calculated, and which cells were used as denominators. These details are not clearly described in the Methods section or figure legends. Greater clarity would help assess whether these findings could have potential clinical relevance for predicting early progression using baseline IHC. I agree with the authors that future studies incorporating spatial transcriptomics and spatial proteomics would be highly informative.

Response: We appreciate the reviewer's comment, and have revised the manuscript to:

1). report median values and interquartile ranges for $CD8^+$, $PD\text{-}1^+CD8^+$, $PD\text{-}L1^+CD56^+$, $PD\text{-}1^+CD56^+$, and $PD\text{-}L1^+$ cells in the overall cohort as well as for subgroups stratified by treatment response and progression status (Line 217-234, and Supplementary Table S9);

2). add details regarding quantification of immune infiltration (Supplementary Methods). Briefly, proportion was defined as the percentage of marker-positive cells

relative to total nucleated cells within the analyzed area, and density was defined as the number of marker-positive cells per mm^2. No specific percentage thresholds were applied, and analyses were conducted using continuous values rather than categorical "high" versus "low" infiltration.

Line 217-234: Multiplex immunohistochemistry was performed on baseline tumor specimens from 25 patients. Among these patients, the median proportion of $CD8^+$ T cells was 10.0% (IQR 5.9%-15.6%), with a median density of 821 (IQR 246-1267) cells/mm^2. The median proportion of PD-1^+CD8^+ T cells was 0.5% (IQR 0.1%-1.2%). The median proportions of PD-$L1^+CD56^+$ cells and PD-1^+CD56^+ cells were 1.4% (IQR 0.4%-2.2%) and 0.1% (IQR 0.1%-0.5%), respectively. The overall PD-$L1^+$ cell proportion was 9.0% (IQR 4.2%-16.8%). Detailed median values and IQRs for all markers, including both proportion and density, in these 25 patients and stratified by response and progression status, are shown in appendix p 18.

Patients who achieved CR had a median $CD8^+$ T cell proportion of 12.8% (IQR 9.9%-19.4%) and a median density of 1072 (IQR 816-1409) cells/mm^2, whereas those who did not achieve CR had a median proportion of 7.3% (IQR 2.2%-10.6%) and a median density of 454 (IQR 227-821) cells/mm^2 (appendix p 18). The level of PD-1^+CD8^+ T cell infiltration did not differ between patients with and without CR (appendix p 21). Additionally, patients without disease progression had a median $CD8^+$ T cell proportion of 16.3% (IQR 13.1%-19.6%) and a median density of 1320 (IQR 986-1571) cells/mm^2; among those with disease progression, the corresponding values were 8.7% (IQR 3.3%-11.6%) and 619 (235-960) cells/mm^2 (appendix p 22).

Supplementary Methods (page 4-5): Whole-slide imaging was conducted using the KF-FL-005 scanner (KFBIO, Ningbo, China). Image analysis was performed using Panoscore software (KFBIO, Ningbo, China) with automated cell segmentation and marker quantification. Analyses were performed across the entire scanned area (whole slide). Cell proportion was defined as the percentage of marker-positive cells relative to total nucleated cells within the analyzed area. Immune cell density was defined as the number of marker-positive cells per mm^2.

Table S9:

Table S9. Baseline immune cell infiltration according to response and disease progression

	Total (n=25)	Response		Disease progression	
		CR (n=12)	Non-CR (n=13)	PD (n=19)	Non-PD (n=6)
$CD8^+$ T cell proportion (%)	10.0 (5.9–15.6)	12.8 (9.9–19.4)	7.3 (2.2–10.6)	8.7 (3.3–11.6)	16.3 (13.1–19.6)
$CD8^+$ T cell density (cells /mm^2)	821 (246–1267)	1072 (816–1409)	454 (227–821)	619 (235–960)	1320 (986–1571)
PD-1^+CD8^+ T cell proportion (%)	0.5 (0.1–1.2)	1.1 (0.4–1.2)	0.2 (0.1–1.2)	0.2 (0.1–1.2)	1.2 (0.7–1.3)
PD-1^+CD8^+ T cell density (cells /mm^2)	26 (8–77)	57 (24–92)	10 (5–26)	13 (7–71)	62 (33–123)
PD-$L1^+CD56^+$ cell proportion (%)	1.4 (0.4–2.2)	2.1 (1.2–2.3)	0.4 (0.2–1.4)	1.2 (0.3–2.1)	1.8 (0.7–2.4)
PD-$L1^+CD56^+$ cell density (cells /mm^2)	77 (22–140)	115 (56–199)	29 (18–121)	57 (19–140)	83 (55–191)
PD-1^+CD56^+ cell proportion (%)	0.1 (0.1–0.5)	0.4 (0.1–1.3)	0.1 (0.1–0.1)	0.1 (0.1–0.4)	0.3 (0.1–0.6)
PD-1^+CD56^+ cell density (cells /mm^2)	11 (2–38)	39 (9–87)	7 (2–11)	9 (4–29)	25 (4–42)
PD-$L1^+$ cell proportion (%)	9.0 (4.2–16.8)	13.9 (7.8–31.1)	8.1 (2.4–14.5)	10.2 (5.8–21.4)	8.1 (3.8–14.9)
PD-$L1^+$ cell density (cells /mm^2)	757 (199–1534)	991 (471–2121)	709 (177–1245)	757 (281–1646)	598 (198–1032)

Data are shown as median (interquartile range). Cell proportion was defined as the percentage of marker-positive cells relative to total nucleated cells. Cell density was defined as the number of marker-positive cells per mm^2. Statistical comparisons between groups are shown in Supplementary Figure S2 and 3.
Abbreviations: CR, complete response; Non-CR, non-complete response; Non-PD, non-progressive disease; PD, progressive disease.

6) - Safety section / Figure 1 (trial profile): Please clarify in the manuscript text why only 23 of the 40 enrolled patients completed therapy. In the Safety section, the authors state that 14 patients experienced treatment interruptions but were able to restart therapy. However, the trial profile indicates that two patients experienced disease progression during the phase 1b dose-escalation period (yet appear to be included in CR/PR best response assessments), and nine patients progressed during

phase 2, with seven discontinuing due to investigator decision, withdrawal of consent, or loss to follow-up. Given the relatively small sample size of this phase 1b/2 study, additional clarification regarding these patients' safety, tolerability, and reasons for discontinuation is warranted, either in the Results or Discussion.

Response: We appreciate the reviewer's comment, and have revised the manuscript to clarify patient disposition, treatment completion, and reasons for discontinuation (L149-157). Briefly, 17 out of the 40 patients discontinued induction therapy for the following reasons: disease progression (n=11), loss to follow-up (n=2), withdrawal of consent (n=2), and investigator decision due to an increase of < 50% in the sum of the product of the diameters (SPD) despite stable disease (n=2). These details have been added to the Results section (L146-154) and Figure 1.

Disease progression occurred during phase 1b in two patients: one during the induction therapy and was included among the 11 patients who discontinued induction due to disease progression; the other after completing all six induction cycles but prior to maintenance therapy and was therefore not considered a discontinuation of induction therapy. All 14 patients who experienced treatment interruptions resumed planned treatment after recovery from adverse events. These interruptions were temporary and did not lead to permanent treatment discontinuation.

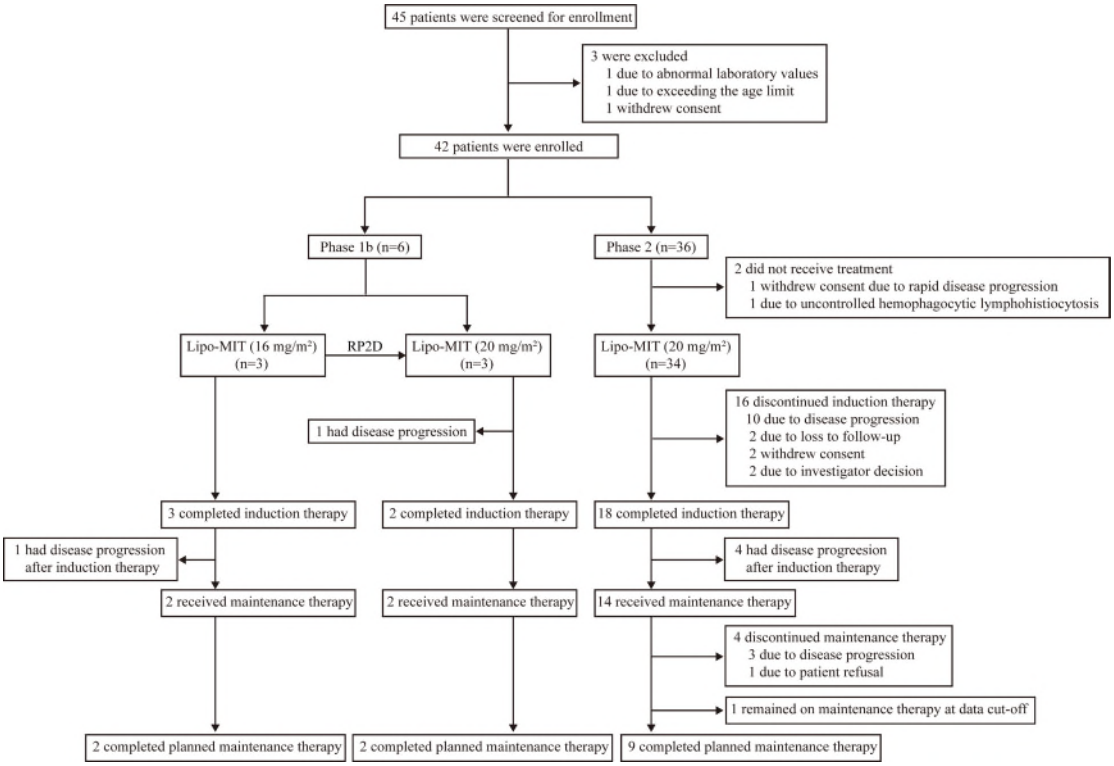
For efficacy assessment, the best response was evaluated using the Lugano 2014 criteria during the entire treatment period. The definition has been added to Methods section (Line 423-424) and Table 2 footnote.

Line 146-154: Twenty-three (58%) patients completed all 6 planned induction cycles, whereas 17 discontinued induction therapy due to disease progression (n=11), loss to follow-up (n=2), withdrawal of consent (n=2), or investigator decision (n=2). Among the 23 patients who completed induction therapy, 18 proceeded to tislelizumab maintenance therapy and 5 did not undergo maintenance because of disease progression at the end of induction. Of the 18 patients who received maintenance therapy, 13 completed the planned maintenance course, 1 remained on treatment at data cut-off, and 4 discontinued maintenance therapy due to disease progression (n=3) or patient refusal (n=1) (Figure 1).

Line 423-424: Best response was defined as the best response achieved at any time during treatment according to the Lugano 2014 criteria.

Table 2 footnote: Best response was defined as the best response achieved at any time during treatment according to the Lugano 2014 criteria.

Figure 1:



7) - Given the multiple associations reported, if possible, consideration should be given to performing multivariable analyses to better contextualize these findings.

Response: We appreciate the reviewer's suggestion, but deleted subgroup survival analyses from the manuscript due to small sample size and limited follow-up duration, per comment by another reviewer.

8) - Are there on-treatment or post-progression biopsy samples available? If so and feasible, we would suggest incorporating a brief comment on future plans and the value of multi-omic approaches to further elucidate mechanisms of resistance and inform future combination strategies.

Response: We appreciate the reviewer's suggestion. Post-progression biopsy samples

were indeed available in a subset of the patients in this trial (n=4). A brief statement has been added to the Discussion section to outline future plans of integrated multi-omic profiling on these samples (Line 335-339 and 345-350).

Line 335-339: Post-progression biopsy samples are available in 4 patients, which could be used for future integrated multi-omic analyses, including genomic, transcriptomic, proteomic, single-cell, and spatial approaches, to further elucidate mechanisms of resistance and inform rational combination strategies.

Line 345-350: Third, the findings of multiplex immunohistochemistry were preliminary and were limited by small sample size, lack of paired pre- and post-treatment analyses, and restricted cellular resolution. Validation using longitudinal sampling and multi-omic profiling is warranted to better elucidate immune mechanisms underlying treatment response and inform combination strategies.

9) - Introduction/Discussion: Consider including a recent meta-analysis evaluating outcomes with anti-PD-1/PD-L1 monotherapy, which reported pooled overall response rates of approximately 50%/43% and 12-month progression-free survival of ~30%, compared with higher response rates and improved 12-month progression-free survival observed with combination approaches such as the regimen studied here (Yang et al., PMID: 40033281)

Response: We appreciate the reviewer's suggestion, and have added the meta-analysis by Yang et al. (PMID: 40033281) to the Discussion section (Line 249-252) and Ref #

28.

Line 249-252: A meta-analysis reported a pooled ORR of 47% and a 12-month PFS rate of 30% with anti-PD-1/PD-L1 monotherapy in this setting.[28] In the current trial, the ORR and 12-month PFS rate were 73% and 50%, respectively.

Ref # 28: Yang, J., Xue, X., Ma, Y., Wang, X. & Xu, C. Efficacy and safety of PD-1/PD-L1 inhibitors for natural killer/T-cell lymphoma: a single-arm meta-analysis. *BMC Cancer* **25**, 385 (2025).

Minor comments:

10) - Introduction, line 108: would recommend rephrasing the statement "given the 'mild' toxicities of PD-1/PD-L1 antibodies" with "relatively safe" or "manageable" toxicities as immune checkpoint inhibitors can be associated with severe and potentially life-threatening immune-related adverse events.

Response: We appreciate the reviewer's suggestion, and have revised the statement accordingly (Line 106-108).

Line 106-108: Given the relatively manageable toxicity profile of anti-PD-1/PD-L1 antibodies, the combination strategies are worth further exploration.

11) - Introduction, line 123: Please correct "Haematology" to "Hematology," which reflects American English usage and the official name of the American Society of Hematology.

Response: We appreciate the reviewer's suggestion, and have revised the statement accordingly (Line 120-121).

Line 120-121: Preliminary findings have been presented at the 2023 American Society of Hematology Annual Meeting.[22]

12) - Results, line 138: Recommend revising this sentence to clearly state that the dose-escalation phase involved combination therapy with a fixed dose of tislelizumab.

Response: We appreciate the reviewer's suggestion, and have revised the statement to state that the dose-escalation phase involved combination therapy with a fixed dose of tislelizumab (Line 137-139).

Line 137-139: In the dose-escalation phase, 6 patients received escalating doses of Lipo-MIT (16 or 20 mg/m^2) in combination with a fixed dose of tislelizumab (200 mg), and no dose-limiting toxicities (DLTs) were observed.

13) - Discussion, lines 228–229: There appears to be a missing word that affects sentence clarity. For example, the phrase may require insertion of "given" between "when" and "in" ("…the RP2D of lipo-MIT when given in…").

Response: We apologize for the mistake, and have added "given" between "when" and "in combination with tislelizumab ...." (Line 236-238).

Line 236-238: In this single-arm, phase 1b/2 trial, we determined the RP2D of Lipo-MIT when given in combination with tislelizumab in patients with R/R ENKTL in the dose-escalation phase,

Reviewer #4 (Remarks to the Author):

Comments to the authors:

SUMMARY

Zou and colleagues report on a phase 1B/2 study of a novel formulation of a classicaly cytotoxic anthracycline – liposomal mitoxantrone – combined with tislelizumab. Relapsed/refractory ENKTL represent a group with unmet needs for more therapeutic options. The inclusion criteria is representative of difficult to treat patients and includes a breadth of exploratory analyses. Lipo-MIT + tislelizumab seem to offer comparable efficacy to other similar combinations in this setting. Toxicity seems to be a concern with high rates of febrile neutropenia, which notably, was not considered a dose-limiting toxicity and represents a limitation of this study. Below are my major and minor comments.

MAJOR COMMENTS

Methods

1) Lines 338 and 339 are a bit unclear. Based on the following line, it seems like induction was for a definitive 6 cycles, followed by maintenance until progression or toxicity. However lines 339 reads that induction was for 6 cycles or until progression or toxicity. Recommend clarification.

Response: We appreciate the reviewer's comment, and have revised the statement to clarify that the planned induction therapy consisted of six 28-day cycles of Lipo-MIT

plus tislelizumab (Line 377-378).

Line 377-378: All patients received induction therapy with intravenous Lipo-MIT plus tislelizumab on day 1 in 28-day cycles for 6 planned cycles.

Results

2) Recommend including details on what previous therapies patients received either in the main text or appendix.

Response: We appreciate the reviewer's comment, and have added details of prior therapies as Supplementary Table S1 to the revised manuscript, with an accompanying summary in the Results section (Line 131-132).

Line 131-132: Most patients had received P-GemOx-based regimens (appendix p 8).

Table S1:

Table S1. Prior treatment regimens before study enrollment (n=40)

Prior treatment regimens	Number
P-GemOx-based regimens	**36**
P-GemOx plus anti-PD-1	14
P-GemOx	7
P-GemOx in combination with radiotherapy	7
P-GemOx plus ASCT	4
P-GemOx plus anti-PD-1 in combination with radiotherapy	1
P-GemOx plus chidamide	1
P-GemOx plus etoposide	1
P-GemOx plus dexamethasone	1
Asparaginase-based regimens	**13**
Asparaginase plus anti-PD-1 and chidamide	2
Asparaginase plus anti-PD-1 in combination with radiotherapy	1
Asparaginase plus anti-PD-1	1
GEPD in combination with radiotherapy	1
GEPD plus anti-PD-1 in combination with radiotherapy	1
Asparaginase, anti-PD-1, dexamethasone and linperlisib	1
Asparaginase plus EPOCH in combination with ASCT	1
Asparaginase, anti-PD-1, cyclophosphamide, vincristine, and etoposide	1
Asparaginase, anti-PD-1, vincristine, and prednisone	1
Asparaginase, cyclophosphamide, vinorelbine, etoposide, and prednisone	1
Asparaginase, cyclophosphamide, vincristine, etoposide, and prednisone	1
Asparaginase, methotrexate, dexamethasone, and etoposide	1
Other regimens	**7**
Anti-PD-1 plus chidamide	3
Anti-PD-1 plus chidamide in combination with radiotherapy	1
Cyclophosphamide, doxorubicin, vincristine, and prednisone	1
Golidocitinib	1
Doxorubicin, etoposide, and methylprednisolone	1

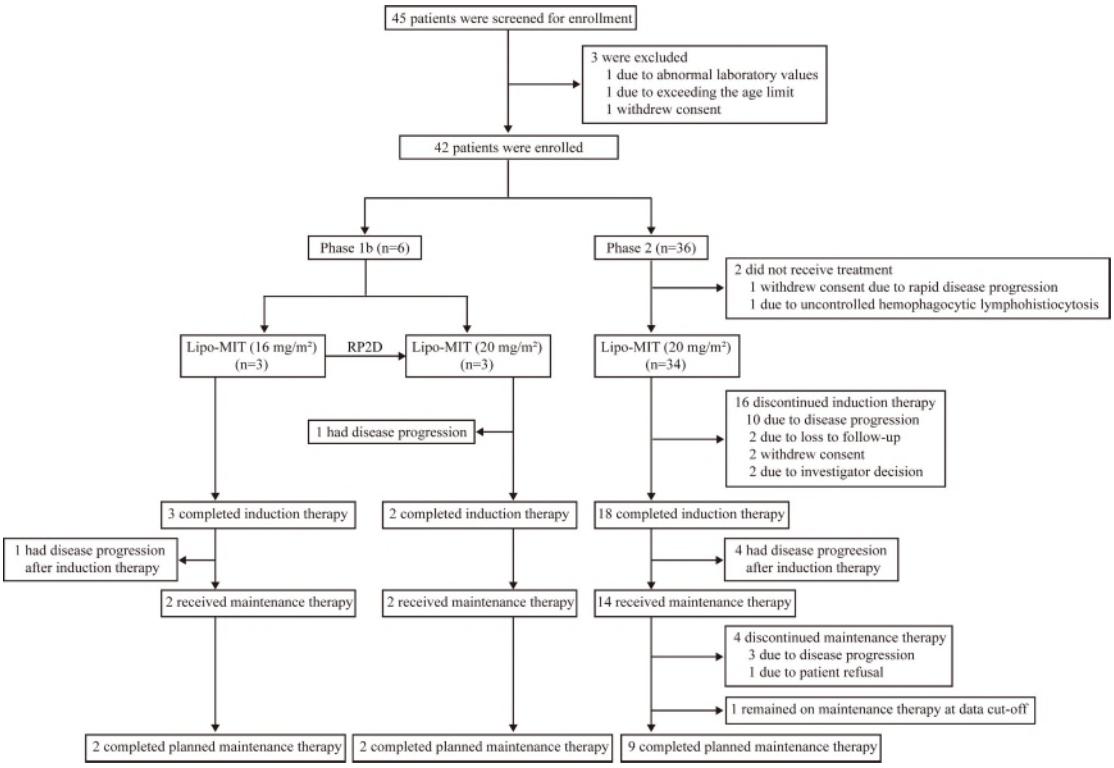
Data are shown as number.
ASCT: autologous stem cell transplantation; CHOP: cyclophosphamide, doxorubicin, vincristine, and prednisone; EPOCH: etoposide, prednisone, vincristine, cyclophosphamide, and doxorubicin; GEPD: asparaginase, gemcitabine, etoposide, and dexamethasone; P-GemOx: pegaspargase, gemcitabine, and oxaliplatin.

3) Only 60% of patients finished induction. Please report the reasons and numbers of patients that dropped out during induction.

Response: We appreciate the reviewer's comment, and added reasons and number of patients for treatment discontinuation during induction therapy to the Results section (Line 146-148), as well as in Figure 1.

Line 146-148: Twenty-three (58%) patients completed all 6 planned induction cycles, whereas 17 discontinued induction therapy due to disease progression (n=11), loss to follow-up (n=2), withdrawal of consent (n=2), or investigator decision (n=2).

Figure 1:

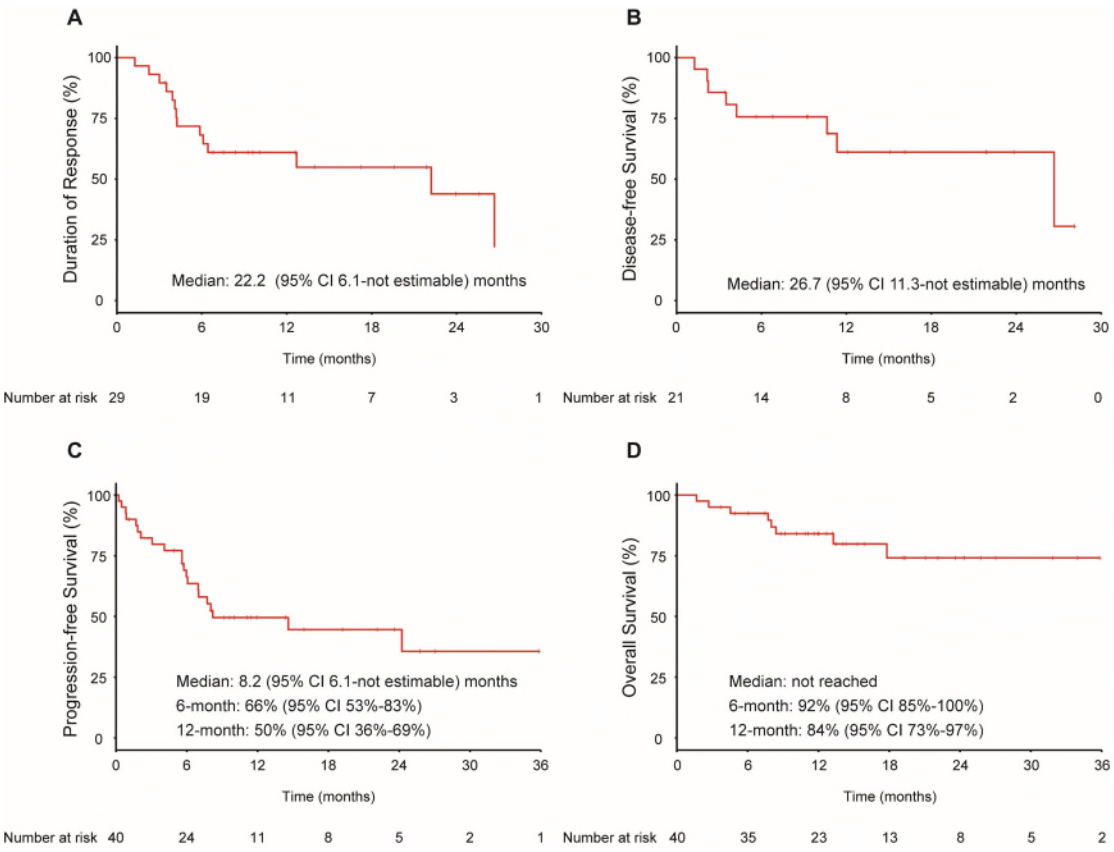


4) Table 2 and figure 3 – it is not clear if the 95% confidence intervals reported represent the absolute number of participants or the percentage of participants. Please clarify this in the table and figure by adding units to the confidence intervals

Response: We appreciate the reviewer's comment, and have added units to the 95% CI (percentage of participants in Table 2; time in months in Figure 3).

Table 2. Anti-tumor activity in the full analysis set (n=40)

	Dose-escalation phase (n=6)	Dose-expansion phase (n=34)	Total (n=40)
Best response*			
Complete response#	5 (83%; 42%-100%)	16 (47%; 32%-100%)	21 (53%; 38%-100%)
Partial response	1 (17%; 0-64%)	7 (21%; 9%-38%)	8 (20%; 9%-36%)
Stable disease	0	2 (6%; 1%-20%)	2 (5%; 1%-17%)
Progressive disease	0	7 (21%; 9%-38%)	7 (18%; 7%-33%)
Not evaluable	0	2 (6%; 1%-20%)	2 (5%; 1%-17%)
Objective response rate	6 (100%; 54%-100%)	23 (68%; 49%-83%)	29 (73%; 56%-85%)
Disease control rate	6 (100%; 54%-100%)	25 (74%; 56%-87%)	31 (78%; 62%-89%)



5) Overall 18% patients developed febrile neutropenia. What was the frequency of primary and secondary granulocyte colony stimulating factor use? Recommend reporting.

Response: We appreciate the reviewer's suggestion, and have expanded the Results section (Line 197-201) to describe the frequency of primary and secondary G-CSF use. Briefly, no patients received primary G-CSF use; secondary G-CSF use occurred in 6 patients for febrile neutropenia.

Line 197-201: No patients received primary granulocyte colony-stimulating factor (G-CSF) prophylaxis. Secondary G-CSF was administered in 6 patients following febrile neutropenia; the remaining 1 patient who developed febrile neutropenia did not receive G-CSF due to withdrawal of consent before further treatment.

Discussion

6) Older pts had markedly lower efficacy than younger. Is this due to older patients requiring more dose reductions or interruptions? This was also seen with ECOG of 2 in the post hoc analyses. Recommend reporting the rates of dose interruptions/dose reductions for older and younger patients, and for those with ECOG 0/1 and 2, and commenting on this in the discussion.

Response: We appreciate the reviewer's comments, and have revised the Discussion section (Line 281-289) to speculate on this issue. Briefly, the rate of dose reduction or treatment interruption due to adverse events was numerically lower in older patients

(1/7 [14%] in patients > 60 years vs 14/33 [42%] in patients ≤ 60 years; p=0.224). Among the 7 older patients, 4 discontinued treatment after one cycle (2 due to progression, 1 withdrew consent, and 1 lost to follow-up), 2 discontinued treatment after two to three cycles; the remaining 1 completed 6 induction cycles but experienced progression immediately thereafter. Given the high proportion of elderly patients who discontinued induction therapy, the association between age and treatment efficacy should be interpreted with caution. The rate of dose reduction or treatment interruption did not differ between the patients with ECOG PS 0 vs 1.

Line 281-289: In subgroup analyses, patients aged > 60 years had a lower CR rate than those aged ≤ 60 years. However, dose modification (dose reduction or treatment interruption) was not more frequent in older patients (1/7 [14%] vs 14/33 [42%]; p=0.224). Instead, most older patients discontinued treatment early, with only one patient completing all planned induction cycles (appendix p 18). Given the high proportion of elderly patients who discontinued induction therapy, the association between age and treatment efficacy should be interpreted with caution. In addition, Eastern Cooperative Oncology Group performance status (ECOG PS) was not associated with dose modification (0: 6/15 [40%] vs 1: 9/25 [36%]; p=0.800) or with CR rate (60% vs 48%; p=0.462).

7) FN not resolving within 7 days was not considered a dose limiting toxicity (as well as some other grade 3+ AEs). However FN is a critical adverse event that requires

urgent high-level care and as such is treated as a DLT in many other dose escalation studies. Had this been considered a DLT, there may have been more patients enrolled in the initial dose level because 33% developed FN in the dose escalation cohort. Additionally, 7 (~18%) needed a dose reduction or interruption in dose expansion phase for neutropenia or FN. I recommend commenting on this study design decision in the discussion section and possibly in the limitations section.

Response: We appreciate the reviewer's comment, and agree that febrile neutropenia represents a critical AE that should be accounted for when defining DLTs. In the dose-escalation phase, 2 out the 6 patients experienced grade 3 febrile neutropenia, but these events occurred after the DLT observation period (cycle 1) and therefore did not meet the protocol-defined criteria for DLT that triggers cohort expansion. The discussion has been added to the Discussion section (Line 299-304).

Line 299-304: Febrile neutropenia occurred in 7 patients (18%) and was generally manageable with supportive care, including G-CSF. Although this event is typically considered a DLT in the majority of phase 1 trials, in this trial, the 2 cases in phase 1b occurred outside the DLT observation period (cycle 1), and therefore did not meet the protocol-defined criteria for DLT.

8) Some univariate methods to identify factors associated with higher and lower PFS have extravagant HR and 95% CI estimates, likely due to the small sample sizes compared. Recommend including this in the limitations and highlighting these

associations, while interesting, are exploratory, even when using penalized cox regression as there may be underlying selection biases among subgroups (e.g. patients not fit enough for previous HSCT are likely to have more dose interruptions with lipo-MIT and as such have reduced efficacy).

Response: We appreciate the reviewer's comment. Another reviewer also expressed concerns over this issue. Accordingly, subgroup analyses were only conducted for the primary outcome of complete response, and no longer for the survival outcomes, in the revised manuscript. The Discussion section of the manuscript has been revised to address this issue as a limitation (Line 342-345).

A numerically higher rate of dose modification was observed in patients with prior HSCT (60% [3/5]) compared with those without (34% [12/35]; p=0.345). Given the small sample size, the finding should be interpreted with caution (Line 289-292).

Line 289-292: A numerically higher rate of dose modification was observed in patients with prior ASCT (60% [3/5]) compared with those without (34% [12/35]), although the difference was not statistically significant (p=0.345).

Line 342-345: Second, the exploratory analyses of CR were conducted in a relatively small cohort and were not prespecified for confirmatory hypothesis testing. The limited number of patients and events may have resulted in imprecise estimates and increased susceptibility to bias across clinical subgroups.

MINOR COMMENTS

Methods

9) Line 368 – enrollment is misspelt (spelt as "enrolment")

Line 392 – enroll is misspelt (spelt as "enrol")

Response: We appreciate the reviewer's suggestion, and have revised the statement accordingly (Line 404-406, and 428).

Line 404-406: Multiplex immunohistochemistry was performed on formalin-fixed paraffin-embedded tumor samples obtained before enrollment in available patients.

Line 428: Dose-escalation phase followed the standard 3 + 3 design to enroll 6-12 patients.

Reviewer #5 (Remarks to the Author):

I co-reviewed this manuscript with one of the reviewers who provided the listed reports. This is part of the Nature Communications initiative to facilitate training in peer review and to provide appropriate recognition for Early Career Researchers who co-review manuscripts.

Response: We thank the reviewer and co-reviewer for their careful and constructive evaluation of our manuscript.

Reviewer #6 (Remarks to the Author):

I co-reviewed this manuscript with one of the reviewers who provided the listed reports. This is part of the Nature Communications initiative to facilitate training in peer review and to provide appropriate recognition for Early Career Researchers who co-review manuscripts.

Response: We thank the reviewer and co-reviewer for their careful and constructive evaluation of our manuscript.

Point-by-point responses

Reviewer #1 (Remarks to the Author):

I thank the authors for their careful and constructive responses to my previous comments. Overall, the revised manuscript has addressed most of my statistical concerns. I have one remaining statistical clarification and one minor presentation suggestion.

1. Given the importance of the amendment to the primary efficacy endpoint, I recommend that the following clarification or similar wording be explicitly included in the Methods or Statistical analysis section, rather than only being provided in the supplementary protocol amendment table:
The primary efficacy endpoint was clarified from ORR to best CR rate in a protocol amendment dated February 10, 2025, after completion of enrollment but before database lock and any efficacy analyses; the sample size assumptions and planned statistical test were based on CR rate and remained unchanged throughout the study.

Response: We thank the reviewer for this valuable comment. As recommended, the clarification regarding the amendment of the primary efficacy endpoint has now been explicitly added to the Statistical analysis section (Line 463-467).

Line 463-467: The primary efficacy endpoint was clarified from ORR to best CR rate in a protocol amendment dated February 10, 2025, after completion of enrollment but before database lock and any efficacy analyses; the sample size assumptions and planned statistical test were based on CR rate and remained unchanged throughout the study.

2. When reporting a one-sided 95% confidence interval in the main text, I suggest reporting it as a one-sided 95% lower confidence bound rather than as a confidence interval with an upper bound of 100%. For example, the current wording: The CR

rate was 53% (21/40; one-sided 95% CI 38%–100%; p<0.001) could be revised to: The CR rate was 53% (21/40; one-sided 95% lower confidence bound, 38%; one-sided p<0.001).

Response: We appreciate the reviewer's suggestion, and have revised the manuscript accordingly. The complete response rate is now reported using the one-sided 95% lower confidence bound (Line 77-78, 149-150, and 157).

Line 77-78: The prespecified primary efficacy endpoint was met, with a CR rate of 53% (21/40; one-sided 95% lower confidence bound, 38%)

Line 149-150: the best CR rate was 53% (21/40; one-sided 95% lower confidence bound, 38%; $p<0.001$)

Line 157: The CR rate was 55% (21/38; one-sided 95% lower confidence bound, 41%)

Reviewer #2 (Remarks to the Author):

No further suggestions.

Response: We thank the reviewer for the careful review and positive comments on our manuscript.

Reviewer #3 (Remarks to the Author):

The authors have responded to my comments. I have no further comments.

Response: We thank the reviewer for the careful review and positive comments on our manuscript.

Reviewer #5 (Remarks to the Author):

I co-reviewed this manuscript with one of the reviewers who provided the listed reports. This is part of the Nature Communications initiative to facilitate training in peer review and to provide appropriate recognition for Early Career Researchers who co-review manuscripts.

Response: We thank the reviewer for the careful review and positive comments on our manuscript.