STATISTICAL ANALYSIS PLAN

Marker Therapeutics, Inc.

MRKR-19-401-01

Protocol Title:	A Phase 2 Study of Donor-Derived Multi-Tumor-Associated Antigen-Specific T Cells (MT-401) Administered to Patients with Acute Myeloid Leukemia (AML) following Hematopoietic Stem Cell Transplantation
Protocol Version and Date:	Version 6.1; 22 November 2022
Sponsor:	Marker Therapeutics, Inc. 3200 Southwest Freeway, Suite 2240 Houston, TX 77027
Prepared By:	Brian Beus Precision for Medicine 6005 Hidden Valley Road, Suite 170 Carlsbad, CA 92011
Document Version and Date:	Version 7.0; 28 November 2022

1 STATISTICAL ANALYSIS PLAN APPROVAL

Sponsor:	Marker Therapeutics, Inc.
Clinical Protocol Number:	MRKR-19-401-01
Protocol Title:	A Phase 2 Study of Donor-Derived Multi-Tumor-Associated Antigen-Specific T Cells (MT-401) Administered to Patients with Acute Myeloid Leukemia (AML) following Hematopoietic Stem Cell Transplantation
Document File Name:	2022Nov28_MRKR-19-401-01_SAP_v7.0_.pdf
Document Version and Effective Date:	Version 7.0; 28 November 2022

Approved By:

DocuSigned by:
Mythili Konero
Signer Name: Mythili Konero
Signing Reason: I approve this document
Signing Time: 28-Nov-2022 | 1:14:28 PM PST
CB7092EAF3A04599A4F3EBEC28D09F6F

28-Nov-2022

Mythili Koneru, MD, PhD
Chief Medical Officer
Marker Therapeutics, Inc.

Date

DocuSigned by:
Karrie Wang
Signer Name: Karrie Wang
Signing Reason: I approve this document
Signing Time: 28-Nov-2022 | 1:28:44 PM PST
65C9EF5DFE80413483A0572EA400E341

28-Nov-2022

Karrie Wang
Vice President, Biostatistics
Marker Therapeutics, Inc.

Date

DocuSigned by:
Brian Beus
Signer Name: Brian Beus
Signing Reason: I am the author of this document
Signing Time: 28-Nov-2022 | 1:30:36 PM PST
9E683A073C334051A68D48084B8342E1

28-Nov-2022

Brian Beus
Principal Biostatistician
Precision for Medicine

Date

2 TABLE OF CONTENTS

TABLE OF TABLES

3 LIST OF ABBREVIATIONS

Table 1. List of Abbreviations

Abbreviation	Definition
ADSPAM	**A**dministration of **D**onor-Derived Multi-Tumor-Associated Antigen-**SP**ecific T Cells to Patients with **A**ML or **M**DS
AE	adverse event
AML	acute myeloid leukemia
ATC	Anatomical Therapeutic Chemical
BMI	body mass index
BOR	best overall response
CR	complete remission
CRi	complete remission with incomplete recovery
CR_{MRD-}	complete remission without minimal residual disease
CRS	cytokine release syndrome
CSR	clinical study report
CTCAE	Common Terminology Criteria for Adverse Events
DLT	dose limiting toxicity
DMC	Data Monitoring Committee
DOCR	duration of complete remission
DOR	duration of response
ECG	electrocardiogram
eCRF	electronic case report form
ELN	European LeukemiaNet
GRFS	graft-versus-host disease relapse-free survival
GVHD	graft-versus-host disease
HSCT	hematopoietic stem cell transplantation
ICH	International Council for Harmonisation
IERC	Independent Efficacy Review Committee
IRT	interactive response technology
IV	intravenous
KM	Kaplan-Meier
MedDRA	Medical Dictionary for Regulatory Activities
MFC	multiparameter flow cytometry
MLFS	morphologic leukemia-free state
MRD	minimal residual disease
MRD^+	patients with minimal residual disease
MRD^-	patients without minimal residual disease
NCI	National Cancer Institute

Abbreviation	Definition
ORR	overall response rate
OS	overall survival
PD	progressive disease
PET	positron-emission tomography
PFS	progression-free survival
PP	per-protocol
PR	partial remission
PRD	primary refractory disease
PRO	patient-reported outcome
Q1	25th percentile
Q3	75th percentile
QTcF	QT interval corrected using Fridericia's formula
RFS	relapse-free survival
RT-qPCR	quantitative reverse transcription polymerase chain reaction
SAE	serious adverse event
SAP	statistical analysis plan
SD	stable disease
SOC	standard of care
SRC	sponsor review committee
TEAE	treatment-emergent adverse event
WHODDE	World Health Organization Drug Dictionary Enhanced

4 INTRODUCTION

The purpose of this statistical analysis plan (SAP) is to provide comprehensive and detailed descriptions of the methods and presentation of data analyses proposed for Marker Therapeutics, Inc. Protocol MRKR-19-401-01 [A Phase 2 Study of Donor-Derived Multi-Tumor-Associated Antigen-Specific T Cells (MT-401) Administered to Patients with Acute Myeloid Leukemia (AML) following Hematopoietic Stem Cell Transplantation]. Descriptions of planned analyses are provided in order to avoid post hoc decisions that may affect the interpretation of the statistical analysis. The statistical methods applied in the design and planned analyses of this study are consistent with the International Council for Harmonisation (ICH) guideline *Statistical Principles for Clinical Trials* (E9) (1998).

This SAP will be finalized prior to data analysis at the interim analysis to provide full details to be presented in the clinical study report (CSR). Templates for tables, listings, and figures will be created in a separate document. Any changes between the statistical methods provided in the clinical study protocol and this SAP will be explained herein; any changes or deviations from this SAP relative to the final analysis will be fully documented in the CSR. Minor changes or deviations from the templates for tables, figures, and listings need not be documented in the CSR.

5 STUDY OBJECTIVES

5.1 Primary Study Objectives

Safety Lead-in (Cohorts I and II) & Cohorts III-V:

- To assess safety and tolerability of MT-401:
    - Manufactured using peptides produced by two different vendors (Safety Lead-in: Cohorts I and II)
    - At a higher dose than the dose used in the Phase 2 portion of the study (Cohort III)
    - Manufactured using an accelerated manufacturing process at two dose levels (Cohorts IV and V)

Phase 2 Adjuvant (Group 1):

- To compare relapse-free survival (RFS) for MT-401 (Arm A) vs standard of care (SOC; Arm B)

Phase 2 Active Disease (Group 2):

- Subgroup C (frank relapse): to estimate complete remission (CR) and duration of CR (DOCR)
- Subgroup D (MRD^+): to estimate complete remission (CR_{MRD}-) and duration of CR without minimal residual disease ($DOCR_{MRD}$-)

5.2 Secondary Study Objectives

Safety Lead-in (Cohorts I and II) & Cohorts III-V:

- To assess efficacy of MT-401:
    - Manufactured using peptides produced by two different vendors (Safety Lead-in: Cohorts I and II)
    - At a higher dose than the dose used in the Phase 2 portion of the study (Cohort III)
    - Manufactured using an accelerated manufacturing process at two dose levels (Cohorts IV and V)

Phase 2 Adjuvant (Group 1):

- To analyze overall survival (OS) for MT-401 (Arm A) and SOC (Arm B)
- To compare graft-versus-host disease RFS (GRFS) for MT-401 (Arm A) vs SOC (Arm B)

Phase 2 Active Disease (Group 2: Subgroups C and D):

- To evaluate overall response rate (ORR), duration of response (DOR), progression-free survival (PFS) and OS for MT-401 alone for Subgroup C patients
- To evaluate RFS, OS, and GRFS for patients who achieve CR following bridging therapy and Subgroup D patients

Phase 2 (Groups 1 and 2):

- To evaluate the safety and tolerability of administering donor-derived MT-401 to patients with AML post-HSCT

5.3 Exploratory Study Objectives

Safety Lead-in, Cohorts III-V, and Phase 2 (Groups 1 and 2):

- To examine the expansion, persistence, clonality and anti-tumor immune effects of the adoptively-transferred, donor-derived, MT-401, as well as the presence of epitope spreading in all groups
- To evaluate time to minimal residual disease (MRD) positivity in patients without MRD (MRD^-)
- To evaluate clearance of MRD after MT-401 infusion in patients with CR but persistent evidence of MRD

- Phase 2 (Group 2 Subgroup C only): to explore patient-reported outcomes (PROs)

6 INVESTIGATIONAL PLAN

6.1 Overall Study Design

This study is a Phase 2 multicenter study evaluating the safety and efficacy of MT-401 administered to patients with AML, who received their first allogeneic hematopoietic stem cell transplant (HSCT). Part 1 of the study includes a Safety Lead-in portion prior to the Phase 2 study and includes two cohorts (Cohort I and Cohort II) that will evaluate the safety and tolerability of three consecutive MT-401 doses of 50 x 10^6 cells manufactured with peptide pools obtained from two different vendors. Once Cohorts I and II in the Safety Lead-in are complete, the Phase 2 portion will open.

Phase 2 will initially evaluate the efficacy and safety of MT-401 at a flat dose of 100 x 10^6 cells either as adjuvant therapy in patients (Group 1) with no active disease (in CR and MRD^-) or in patients with active disease (Group 2, frank relapse or MRD^+).

Part 2 of the study includes Cohorts III, IV, and V which will run in parallel with Phase 2. Cohort III will evaluate MT-401 at a higher dose (200 x 10^6 million cells) than the dose used in the Phase 2 portion using the initial manufacturing process. If Cohort III data determines the 200 x 10^6 cell dose to be safe, the Phase 2 patients may be switched to receive the 200 x 10^6 cell dose, infused three times at intervals of at least 2 weeks.

Cohorts IV and V will evaluate MT-401 manufactured using an accelerated process; patients will receive three consecutive doses of MT-401 at fixed doses of 100 x 10^6 cells and 200 x 10^6 cells, respectively. Following completion of Cohorts IV and V, the Sponsor may change the Phase 2 dose to the dose deemed to be safe for MT-401 manufactured using the accelerated process.

A total of approximately 225-240 patients will be treated in this study as follows:

- Safety Lead-in: 6-12 patients
- Cohorts III, IV and V: 9-18 patients
- Phase 2: approximately 210 patients
    - ~150 patients in Group 1 (adjuvant therapy)
    - ~60 patients in Group 2 (active disease)

Enrollment will continue in Phase 2 to achieve approximately 210 evaluable patients randomized/treated with the final MT-401 dose and manufacturing process, thereby replacing any patients that have not been randomized/treated with the final Phase 2 product and dose, as determined by the Sponsor. Patients aged ≥18 years old undergoing or having relapse after their first allogeneic HSCT (matched sibling, matched unrelated donor, or haploidentical transplants) for AML are eligible.

Potential patients for the study may be screened/enrolled:

- Prior to their first allogeneic HSCT. Leukapheresis material, from the HSCT donor, for the production of MT-401 will be obtained as a separate collection from the HSCT or ~30 days after, based on donor status at time of collection.

or

- When experiencing their first relapse post-allogeneic transplant. The same HSCT donor must be available and agree to undergo leukapheresis for production of MT-401.

Patients eligible for the study will be placed into one of two groups:

- **Adjuvant (Group 1)**: Patients screened prior to their HSCT with CR without minimal residual disease (CR_{MRD}-) at 80 days post-transplant based on central testing will be randomized (1:1) in an unblinded fashion to:
  - MT-401 (Arm A)
  - SOC - observation (Arm B)

  Randomization will be stratified by pre-transplant MRD status (MRD^+ vs. MRD^- /Unknown) and cytogenetic risk (Unfavorable vs. Other). However, randomization will be performed only if MRD testing performed centrally at Baseline visit (~80 days post-transplant) is negative.

  Note: Patients whose product includes fewer cells or does not meet specifications may be treated with MT-401.

- **Active Disease: (Group 2**): Patients meeting the following criteria will be assigned to Subgroup C (frank relapse) or Subgroup D (MRD^+) in Group 2 and will receive MT-401:
  - Patients in Group 1 who experience relapse (patients with MRD [MRD^+] or frank relapse) prior to randomization
  - Patients in Arm B of Group 1 (SOC) who experience relapse (MRD^+ or frank relapse) after randomization (crossover patients)
  - Patients who do not consent prior to HSCT but are experiencing their first relapse (MRD^+ or frank relapse) and have the same donor available for manufacturing

  Note: Patients in Group 2 may receive bridging chemotherapy for ≤ 6 months prior to receiving MT-401. Additionally, patients that meet the eligibility criteria for Group 2 only can be used for the Safety Lead-in portion of the study.

Those patients who enter the study prior to HSCT must meet applicable inclusion/exclusion criteria prior to and following HSCT for study entry and must agree to provide bone marrow aspirate sample(s) to be held for testing.

Relapse is defined as bone marrow blasts ≥5%, or reappearance of blasts in the blood, or development of extramedullary disease, or development of MRD positivity after having a post-treatment MRD negative status.

Disease assessments will be performed by local and central laboratory facilities. Screening assessments for Group 1 patients can be a prior disease assessment completed within 45 days of screening as per SOC. For Group 1, pre-transplant testing of MRD should be performed locally

by any testing methodology and results recorded in eCRF. Post-transplant baseline MRD testing will be performed centrally and results must be available prior to randomization. MRD^+ will be defined as CR (including normalization of counts) with detection in blood, bone marrow or any other tissues by flow cytometry and/or molecular testing (eg, PCR, next generation sequencing).

The study will evaluate safety and efficacy endpoints in post-HSCT patients after the infusion of MT-401 using standard methods of assessment (peripheral blood and bone marrow evaluation, positron-emission tomography [PET] scan, etc.). Response will be evaluated using the European LeukemiaNet (ELN) recommendations and repeated MRD assessment. Historical data will inform the clinical relevance of the magnitude and durability of response in Group 2 and will supplement data from the concurrently randomized control arm in Group 1 (Arm B). A synthetic control arm may also be identified for Group 2C. If created, the specifications of this control arm and analysis methods to be used will be documented in a separate statistical analysis plan.

The Data Monitoring Committee (DMC) will review data after the first 45 RFS events in Group 1 have been observed to assess safety and evaluate estimates of RFS for possible Group 1 enrollment expansion. Additionally, the sponsor will review safety and efficacy data in Group 2 after the first 10 patients per subgroup have progressed or had the opportunity to be followed for 6 months.

6.2 Schedule of Assessments

For the complete schedule of assessments, refer to Section 1.2.1 (Table 1) of the clinical study protocol.

6.3 Treatment

6.3.1 Treatment Administered

MT-401 will be administered as a patient-specific product every 2 weeks for 3 doses. MT-401 will be given by intravenous (IV) infusion through either a peripheral or central line. The patients will be monitored for any safety concerns for 1 hour post-infusion as outpatients.

A dose of 50 x 10^6 cells (flat dosing) will be the dose explored in the Safety Lead-in portion for this study with a minus 1 (-1) dose level of 25 x 10^6 cells (flat dosing), should the need arise to de-escalate. The Safety Lead-in portion of the study, specifically, tests safety of MT-401 manufactured with peptides produced from different vendors (JPT/Almac).

Following completion of the Safety Lead-in portion the study will proceed to open Groups 1 and/or 2 of the Phase 2 study at the highest cleared dose of 100 x 10^6 cells (flat dosing) from the Phase 1 trial.

In parallel to the opening of Groups 1 and/or 2, Cohort III will test the flat dose equivalent of 200 x 10^6 cells. If the 200 x 10^6 cells dose being tested in Cohort III is noted to be safe, then the patients enrolled, but not yet treated, in Groups 1 and 2 may be treated at the 200 x 10^6 cells dose, unless the Sponsor decides to remain at the current dose of 100 x 10^6 cells.

Cohorts IV and V will evaluate the safety of MT-401 manufactured using the accelerated method at doses of 100 x 10^6 and 200 x 10^6 cells, respectively. A lower dose level of 50 x 10^6 cells (flat dosing) may be used should the need arise to de-escalate. At the conclusion of Cohorts IV and V, which evaluate the safety of the product manufactured using an accelerated process at two different doses, the Sponsor may change the Phase 2 dose to the dose deemed to be safe for MT-401 manufactured using the accelerated process.

6.3.2 Method of Assigning Patients to Treatment Groups

Patients in the Safety Lead-in (Cohorts I and II) and Cohorts III-V will not be randomized.

For Group 1, randomization codes will be prepared to allow the clinical study sites to enroll patients by using an interactive response technology (IRT) system. Treatment group assignments will be known to site personnel (ie, investigator, pharmacist, and study staff) and patients. After completion of required screening procedures, patients will be assigned, using the randomization schedule, to MT-401 vs observation (SOC) with a 1:1 allocation. The randomization schedule will be created with stratification factors of pre-transplant MRD status (MRD^+ vs. MRD^- /Unknown) and cytogenetic risk (Unfavorable vs. Other) using random block sizes. SOC patients in Group 1-Arm B who relapse after randomization may cross over to Group 2 to receive MT-401, if the subgroup is open and/or Sponsor agreement.

For Group 2, after the patient's eligibility is established and informed consent has been obtained, the patient will be enrolled, and a number will be assigned through the IRT. Since this is a single-arm cohort, all enrolled patients who meet eligibility criteria will be treated with MT-401.

Due to the nature of the administration of MT-401 along with the fact that the endpoints are objectively evaluated without the concern for bias, the study is not blinded.

6.4 Efficacy and Safety Variables

6.4.1 Efficacy Variables

6.4.1.1 Disease Assessment – Response Criteria in AML

Disease will be assessed by standard criteria at the time points listed in the schedule of assessments in Section 1.2.1 of the clinical study protocol. The ELN recommendations for AML (Döhner et al, 2017) are as follows in Table 2.

Table 2. Response Criteria in AML

Category	Definition
CR without minimal residual disease (CR_{MRD-})	If studied pretreatment, CR with negativity for a genetic marker by quantitative reverse transcription polymerase chain reaction (RT-qPCR), or CR with negativity by MFC by multiparameter flow cytometry (MFC)
Complete remission (CR)	Bone marrow blasts <5%; absence of circulating blasts and blasts with Auer rods; absence of extramedullary disease; absolute neutrophil count ≥1.0 x 10^9/L (1000/μL); platelet count ≥100 x 10^9/L (100,000/μL) Note: MRD^+ or unknown
CR with incomplete recovery (CRi)	All CR criteria except for residual neutropenia (<1.0 x 10^9/L [1000/μL]) or thrombocytopenia (<100 x 10^9/L [100,000/μL])
Morphologic leukemia-free state (MLFS)	Bone marrow blasts <5%; absence of blasts with Auer rods; absence of extramedullary disease; no hematologic recovery required
Partial remission (PR)	All hematologic criteria of CR; decrease of bone marrow blast percentage to 5 to 25 percent; and decrease of pretreatment bone marrow blast percentage by at least 50 percent
Treatment failure	
Primary refractory disease	No CR or CR_i after 2 courses of intensive induction treatment; excluding patients with death in aplasia or death due to indeterminate cause
Death in aplasia	Deaths occurring ≥7 days following completion of initial treatment while cytopenic; with an aplastic or hypoplastic bone marrow obtained within 7 days of death, without evidence of persistent leukemia
Death from indeterminate cause	Deaths occurring before completion of therapy, or <7 days following its completion; or deaths occurring ≥7 days following completion of initial therapy with no blasts in the blood, but no bone marrow examination available
Response criteria for clinical trials only	
Stable disease[a]	Absence of CR_{MRD-}, CR, CR_i, PR, MLFS; and criteria for PD not met
Progressive disease (PD)	Evidence for an increase in bone marrow blast percentage and/or increase of absolute blast counts in the blood: • >50% increase in marrow blasts over baseline (a minimum 15% point increase is required in cases with <30% blasts at baseline; or persistent marrow blast percentage of >70% over at least 3 months; without at least a 100% improvement in ANC to an absolute level (>0.5 x 10^9/L [500/μL], and/or platelet count to >50 x 10^9/L [50,000/μL] nontransfused); or • >50% increase in peripheral blasts (WBC X % blasts) to >25 X 10^9/L (>25,000/μL) (in the absence of differentiation syndrome); or • New extramedullary disease
Relapse	
Hematologic Relapse (after CR_{MRD-}, CR, CR_i)	Bone marrow blasts ≥5%; or reappearance of blasts in the blood; or development of extramedullary disease
Molecular relapse (after CR_{MRD-})	If studied pretreatment, reoccurrence of MRD as assessed by RT-qPCR or by MFC

[a] The period of stable disease should be at least 3 months to be considered SD.

6.4.1.2 Primary Efficacy Variables

Phase 2 Adjuvant (Group 1):

- RFS: defined as the time between the date of randomization and the date of first relapse per ELN recommendations or death by any cause, whichever occurs first. Relapse includes either hematologic and/or molecular relapse as defined in Table 2 at any disease assessment. In the event that no relapse or death is documented prior to analysis cutoff or the start of new anti-cancer therapy, RFS will be censored on the date of the last evaluable disease assessment demonstrating no relapse prior to analysis cutoff or the start of new therapy. For patients who remain alive and have no baseline disease assessment or have no post-randomization disease assessment, RFS will be censored on the date of randomization. In addition, if relapse or death occurs after extended time without disease assessments, RFS will be censored on the date of the last evaluable disease assessment prior to the missed disease assessments.

  For patients with relapse or death, a variable will be created to flag whether the earlier date of relapse or death occurs after extended time without disease assessments. The extended time is defined as below:

  - Time from randomization date to Day 0 visit + 12 weeks + assessment window of 2 weeks, if the latest assessment prior to the event (relapse or death, whichever is earlier) ≤ randomization date + 0.5*time from randomization date to Day 0 visit. For patients without Day 0 visit, an artificial Day 0 visit will be derived as randomization date + mean time from randomization date to Day 0 visit date in MT-401 arm, for analysis purpose only.

  - 24 weeks + assessment window of 2 weeks, if the latest assessment prior to the event (relapse or death, whichever is earlier) ≤ Day 0 visit + 6 weeks & > randomization date + 0.5*time from randomization date to Day 0 visit

  - 38 weeks + assessment window of 5 weeks, if the latest assessment prior to the event (relapse or death, whichever is earlier) ≤ Day 0 visit + 18 weeks & > Day 0 visit + 6 weeks

  - 78 weeks + assessment window of 5 weeks, if the latest assessment prior to the event (relapse or death, whichever is earlier) ≤ Day 0 visit + 37 weeks & > Day 0 visit + 18 weeks

  - 104 weeks + assessment window of 5 weeks, if the latest assessment prior to the event (relapse or death, whichever is earlier) ≤ Day 0 visit + 78 weeks & > Day 0 visit + 37 weeks

Phase 2 Active Disease (Group 2):

- CR: For Subgroup C, CR is defined as the proportion of patients who achieve a best response of CR_{MRD-} or CR using the ELN recommendations. For Subgroup D, CR rate is defined as the proportion of treated patients who achieve a best response of CR_{MRD-}. Best overall response (BOR) is defined as the best response designation recorded between the date of first dose and the date of objectively documented relapse per ELN recommendations or the date of subsequent therapy, whichever occurs first. For patients without documented relapse or subsequent therapy, all available response designations will contribute to the BOR assessment.

    - The incidence of converting MRD+ to MRD- was chosen as the primary endpoint for Subgroup D based on the following rational: (1) 2017 ELN recommendations (Döhner et al, 2017) introduced a new response category of CR MRD- recognizing MRD assessment provides a more reliable predictor of outcome than morphological assessment and MRD status has been consistently shown to be an independent prognostic marker for disease outcome (Short et al, 2020; Schuurhuis et al, 2018). (2) Given the nature of single arm design, a binary endpoint was considered a more appropriate primary endpoint than a time to event endpoint such as RFS or OS.

- DOCR: For Subgroup C, DOCR is defined as the time between the date of first documented response of CR_{MRD-} or CR to the date of the first documented relapse per ELN recommendations or death due to any cause, whichever occurs first. For Group 2 Subgroup D, duration of complete response is defined as the time between the date of first documented response of CR_{MRD-} to the date of the first documented relapse per ELN recommendations or death due to any cause, whichever occurs first. For patients who neither relapse nor die prior to analysis cutoff or the start of new anti-cancer therapy, DOCR will be censored on the date of their last evaluable disease assessment on or after the assessment of first documented CR and prior to analysis cutoff or start of new therapy.

6.4.1.3 Secondary Efficacy Variables

Safety Lead-in (Cohorts I and II) and Cohorts III-V:

- CR (Active disease patients): defined as above for Group 2

- DOCR (Active disease patients): defined as above for Group 2

Phase 2 Adjuvant (Group 1):

- OS: defined as the time between the date of randomization and the date of death. A patient who has not died will be censored on their last contact date ("last known alive date").

- Graft-versus-host disease RFS (GRFS): defined as the time from randomization to date of GVHD, relapse, or death, whichever occurs first. A GVHD event is defined as any grade

3-4 acute GVHD or chronic GVHD requiring systemic therapy. A patient who has not experienced any of these events prior to analysis cutoff or the start of new anti-cancer therapy, will be censored on the earlier date of their last GVHD assessment and evaluable disease assessment demonstrating no relapse prior to analysis cutoff or start of new therapy. For patients who remain alive and have no baseline disease assessment or have no recorded post-randomization disease assessment, GRFS will be censored on the date of randomization. In addition, if a GVHD event, relapse, or death occurs after extended time without GVHD assessments, GRFS will be censored on the date of the last evaluable GVHD assessment prior to the missed GVHD assessments or on the date of the last GVHD assessment prior to the missed GVHD assessments.

Phase 2 Active Disease (Group 2):

- ORR: defined as the proportion of treated patients who achieve a best response of CR_{MRD-}, CR, CRi, MLFS, or PR using the ELN recommendations and repeated MRD assessment. This endpoint only applies to Group 2 Subgroup C.

- DOR: defined as the time between the date of first documented response of CR_{MRD-}, CR, CRi, MLFS, or PR to the date of the first documented progression per ELN recommendations or death due to any cause, whichever occurs first. For patients who neither progress nor die prior to analysis cutoff or the start of new anti-cancer therapy, DOR will be censored on the date of their last evaluable disease assessment on or after the assessment of first documented CR_{MRD-}, CR, CRi, or PR and prior to analysis cutoff or start of new therapy. This endpoint only applies to Group 2 Subgroup C.

- PFS: defined as the time between the date of first dose and the date of documented progression or death by any cause, whichever occurs first. For patients who remain alive and whose disease has not progressed prior to analysis cutoff or the start of new anti-cancer therapy, PFS will be censored on the date of last evaluable disease assessment prior to analysis cutoff or start of new therapy. For patients who remain alive and have no baseline disease assessment or have no recorded post-dose disease assessment, PFS will be censored on the date of first dose of study medication. This endpoint only applies to Group 2 Subgroup C.

- OS: defined as above for Group 1 except starting at date of first dose of study medication rather than date of randomization.

- RFS (for patients who achieve CR following bridging therapy and for Group 2 Subgroup D patients): defined as above for Group 1 except starting at date of first dose of study medication rather than date of randomization

- GRFS (for patients who achieve CR following bridging therapy and for Group 2 Subgroup D patients): defined as above for Group 1 except starting at date of first dose of study medication rather than date of randomization

6.4.1.4 Other Efficacy Variables

Other efficacy endpoints include: markers of immune function, cell targeting, signaling pathways and disease status, including MRD.

Phase 2 Adjuvant (Group 1):

- Time to MRD positivity: defined as the time between the date of randomization and the date of first occurrence of MRD^+. In the event that no MRD^+ is documented prior to analysis cutoff or the start of new anti-cancer therapy, time to MRD positivity will be censored on the date of the last MRD status assessment demonstrating MRD^- prior to analysis cutoff or start of new therapy. For patients who remain alive and have no baseline MRD assessment or have no post-randomization MRD assessment, time to MRD positivity will be censored on the date of randomization.

Phase 2 Active Disease (Group 2 CR patients):

- Clearance of MRD: defined as the time between the date of first dose and the date of first occurrence of MRD^-. This endpoint will be assessed in patients whose best overall response is CR MRD-.

Phase 2 Active Disease (Group 2):

- PRO questionnaire: details regarding the PRO questionnaire and associated instructions will be provided in a separate document.

6.4.2 Description of Safety Variables

6.4.2.1 Dose Limiting Toxicities (DLTs)

A DLT is defined as one of the following AEs reported during the Safety Lead-in observation period (28 days from the day of the initial MT-401 infusion), if considered to be related to MT-401 and deemed clinically significant by the investigator; and fulfills any of the following criterion using National Cancer Institute Common Terminology Criteria for Adverse Events (NCI CTCAE) (Version 5.0) with the exceptions as noted below:

1) Grade 3-4 GVHD

2) Grade 3-4 toxicity felt to be related to, or resulting from cytokine release syndrome (CRS)

    - Exception: Grade 3 and 4 expected reactions seen with the use of T cell-based immunotherapy, such as (but not limited to) fever and hypotension not requiring pressor support, will not be considered DLTs

3) Grade 3-4 infusion reactions that are persistent beyond 72 hours despite optimal medical management

4) Any other Grade 3-4 toxicity that is ***not*** pre-existing and/or ***not*** due to the underlying malignancy or infection or treatment of disease lasting for extended period of time despite optimal supportive care (for example, nausea, vomiting, and diarrhea lasting > 3 days; fatigue lasting > 7 days)

    - Exception: Laboratory abnormality deemed not to be clinically significant

5) Grade 5 toxicity (that is, death)

6.4.2.2 Adverse Events

An AE is any untoward medical occurrence in a patient or clinical study patient, temporally associated with the use of MT-401, whether or not considered related to MT-401. An AE can therefore be any unfavorable and unintended sign (including an abnormal laboratory finding), symptom, or disease (new or exacerbated) temporally associated with the use of MT-401.

6.4.2.3 Laboratory Parameters

Clinical laboratory tests will be collected at the specified time points indicated in the schedule of assessments in Section 1.2.1 of the clinical study protocol. Laboratory tests to be collected (on days of MT-401 administration, labs should be performed prior to MT-401 administration) include hematology, serum chemistry, and coagulation tests.

6.4.2.4 Vital Signs

Vital signs will be performed at the specified time points indicated in the schedule of assessments in Section 1.2.1 of the clinical study protocol. Temperature, pulse rate, and blood pressure will be assessed, and if clinically indicated, respiratory rate will also be assessed.

6.4.2.5 Physical Examination

Physical examinations will be performed at the specified time points indicated in the schedule of assessments in Section 1.2.1 of the clinical study protocol. A complete physical examination will be performed as per institutional guidelines at baseline and will include, at a minimum, assessments of the Cardiovascular, Respiratory, Gastrointestinal and Neurological systems, and height (at Screening) and weight will also be measured and recorded. A brief physical examination will include, at a minimum, assessments of the Cardiovascular, Respiratory, and Gastrointestinal systems.

6.4.2.6 GVHD Assessments

A potential toxicity for patients receiving allogeneic T cells is GVHD. The risk that adoptively transferred leukemia-targeted T cells will cause GVHD is very low. GVHD will be assessed at the specified time points indicated in the schedule of assessments in Section 1.2.1 of the clinical study protocol and as per acute and chronic GVHD guidelines (Harris, et al 2016; Jagasia et al, 2015).

6.5 Data Quality Assurance

Report summaries will be generated using validated Base SAS® software, version 9.4 or higher, on a PC or server-based platform. Additional validated software may be used to generate analyses, as needed.

All SAS programs that create outputs or supporting analysis datasets will be validated by a second statistical programmer or biostatistician. At a minimum, validation of programs will consist of a review of the program log, review of output or dataset format and structure, and independent confirmatory programming to verify output results or dataset content. Additionally, all outputs will undergo a review by a senior level team member before finalization.

The content of the source data will be reviewed on an ongoing basis by project statistical programmers and statisticians. Data will be checked for missing values, invalid records, and extreme outliers through defensive programming applications, analysis-based edit checks, and other programmatic testing procedures. All findings will be forwarded to the project data manager for appropriate action and resolution.

7 STATISTICAL METHODS

7.1 General Methodology

Data will be analyzed by Precision for Medicine biostatistics personnel. Statistical analyses will be reported with tables, figures, and listings, presented in rich text format, and using recommended ICH numbering. Output specifications for all tables, figures, and listings will be in conformance with guidelines specified by the ICH in Appendix 7 of the *Electronic Common Technical Document Specification* (Apr 2003).

The primary analysis will include disease assessments associated with the first treatment regimen of MT-401 only. Re-treatment of MT-401 beyond planned study treatment (eg, re-starting MT-401 after relapse) will be considered as follow up anti-cancer therapy for the sake of efficacy analyses.

7.1.1 Reporting Conventions

Tables and figures will be summarized by study phase and cohort or treatment group. Tables summarizing demographics and other baseline characteristics will also include a column for all patients combined within study phase. Listings will be presented where appropriate and will be ordered by study phase, patient number, cohort or treatment group, and assessment or event date. For Cohort I-V and Group 2, the treatment group presented in tables, figures and listings will be based on the actual treatment received, unless otherwise noted. For Group 1, safety outputs will be based on the actual treatment received and other outputs will be based on the treatment a patient is randomized to.

In general, continuous variables will be summarized to indicate the study population sample size (N), number of patients with available data (n), mean, SD, median, 25^{th} (Q1) and 75^{th} (Q3)

quartiles, minimum, and maximum values. Categorical variables will be summarized by the population size (N), number of patients with available data (n), number of patients in each category, and the percentage of patients in each category. Unless otherwise noted, the denominator to determine the percentage of patients in each category will be based on the number of patients with available data. Select ordinal data may be summarized using both descriptive statistics and counts and percentages of patients in each category, as appropriate.

Non-zero percentages will be rounded to whole numbers. Rounding conventions for presentation of summary statistics will be based on the precision of the variable of summarization, as it is collected in its rawest form (ie, on the electronic case report form [eCRF] or as provided within an external file) and are outlined as follows:

- The mean and median will be rounded to one more decimal place than the precision of the variable of summarization, but no more than 2 decimal places;
- Measures of variability (eg, SD, SE) will be rounded to two more decimal places than the precision of the variable of summarization, but no more than 3 decimal places; and
- Minimum and maximum values will be presented using the same precision as the variable of summarization.

Other statistics (eg, CIs) will be presented using the same general rules outlined above, or assessed for the most appropriate presentation based on the underlying data.

Statistical significance testing will be two-sided and performed using α=0.05. P-values will be reported for all statistical tests, rounded to four decimal places. P-values less than 0.0001 will be displayed as "<0.0001"; p-values greater than 0.9999 will be displayed as ">0.9999".

7.1.2 Summarization by Visit

Data summarized by study visit will be based on the nominal, scheduled visit label as reported on the eCRF. Data collected at unscheduled visits will not be included in by-visit summaries, but will be considered when endpoint derivations potentially include multiple visits (eg, determination of baseline value, determination of worst post-baseline value, etc.). All data will be included in patient listings.

All efficacy data collected at unscheduled visits will be included in the derivation of efficacy endpoints.

7.1.3 Data Handling Rules

Unless otherwise noted, values reported as greater than or less than some quantifiable limit (eg, "< 1.0") will be summarized with the sign suppressed in summary tables and figures, using the numeric value reported. Data will display on patient listings to include the sign.

7.1.4 Standard Calculations

Where appropriate, the calculated study day of each assessment or event will be presented with the assessment or event date on patient data listings, where study day will be determined as:

- The assessment/event date minus the date of first dose of MT-401 for patients enrolled in Cohorts I-V or Group 2
- The assessment/event date minus the date of randomization for all patients randomized in Group 1.

Other variables requiring calculations will be derived using the following formulas:

- **Days:** A duration between two dates expressed in days will be calculated as: later date – earlier date + 1
- **Months:** A duration expressed in months will be calculated by dividing the duration in days by (365.25 / 12);
- **Years:** A duration expressed in years will be calculated by dividing the duration in days by 365.25;
- **Change from Baseline:** Change from baseline will be calculated as the post-baseline value minus the baseline value;
- **Percentage Change from Baseline:** Percentage change from baseline will be calculated as the change from baseline divided by the baseline value, multiplied by 100.

7.1.5 Definitions

Standard definitions include:

- Actual follow-up time: last date known alive or death date – date of first dose of MT-401 (or date of randomization for Group 1 patient) + 1
- Baseline: the last available measurement prior to the first dose of MT-401 for patients enrolled in Cohorts I-V or Group 2. For patients randomized in Group 1, baseline for efficacy endpoints will be the last available measurement on or before the date of randomization. For patients randomized in Group 1, baseline for safety endpoints will be the last available measurement prior to the first dose of MT-401 for patients randomized to MT-401 and the last available measurement on or before the date of randomization for patients randomized to SOC.
- Potential follow-up time: data cutoff date – date of first dose of MT-401 (or date of randomization for Group 1 patients) + 1

- Study Day 0: the date of first dose of MT-401 (or date of randomization for Group 1 patients)

- Treatment emergent adverse events (TEAEs): Adverse events (AEs) with onset after the first dose of MT-401 or existing events that worsened after the first dose of MT-401 during the study. For the Group 1 SOC arm, TEAEs include AEs with onset after randomization or existing events that worsened after randomization.

7.2 Analysis Populations

The analysis populations are defined as follows:

- All Treated Patients: Includes all patients who received at least one dose of MT-401 or were randomized to SOC.

- All Randomized Patients: Includes all patients in Group 1 who were randomized with final dose and manufacturing process and are MRD^- prior to randomization by central testing and local testing, if a local test was also performed. A patient randomized based on negative MRD result per central test but confirmed to be MRD^+ will be excluded from Group 1 analysis.

  When MRD testing occurs with different methodology (eg, flow cytometry and PCR), the patient is considered MRD+ if either test detects MRD. If adjudication occurs between central testing and local testing for the same type of testing (eg, flow by central testing and flow by local testing), the final adjudicated result will be taken into account for the confirmed MRD result. A patient confirmed to be MRD+ prior to randomization will be excluded from Group 1 and will be included in Group 2 analysis if treated with MT-401.

- Per-Protocol (PP) Population: Includes all treated patients, as defined above, who have no major protocol deviations. If there is <5% difference in the all randomized patient population and the PP population for Group 1, no analyses will be performed on the PP population.

7.3 Study Patients

7.3.1 Disposition of Patients

Patient disposition will be summarized for all enrolled patients by study phase and cohort or treatment group. Summaries will include the number and percentage of patients in each analysis population, completing all 3 doses, reason for not completing all doses, completing the study, and discontinuing the study early by the primary reason for discontinuation. Descriptive statistics for actual and potential follow-up times will be included.

7.3.2 Protocol Deviations

Important protocol violations will be summarized by study phase and cohort or treatment group and over all patients combined for all enrolled patients. Important protocol violations are protocol

deviations captured on-study that are deemed by the Sponsor to potentially impact the efficacy or safety conclusions of the study.

All important protocol violations will be determined and appropriately categorized prior to database lock. The number and percentage of patients with any important protocol violations as well as the number and percentage of patients with violations within each category will be presented.

7.4 Efficacy Evaluation

7.4.1 Datasets Analyzed

Efficacy analyses for Group 1 will be conducted in the all randomized patients population while efficacy analyses for Cohorts I-V and Group 2 will be conducted in the all treated patients population. Select efficacy analyses for Group 1 will be repeated in the PP population. A data listing of patients excluded from the PP Population, to include the reason for exclusion, will be presented.

7.4.2 Demographic and Other Baseline Characteristics

Demographic and baseline characteristics will include: age, sex, ethnicity, race, height, weight, and body mass index (BMI). Age will be calculated relative to date of informed consent, as follows:

- If the month and day portion of the informed consent date is prior to the month and day portion of the birthdate, age will be calculated as the year of informed consent minus the year of birth, minus one;
- If the month and day portion of the informed consent date is on or after the month and day portion of the birthdate, age will be calculated as the year of informed consent minus the year of birth.

BMI will be calculated as: weight (kg) / [height (cm) / 100]^2. Age, height, weight, and BMI will be summarized using descriptive statistics. Sex, age category (18-64 years, 65 years and above), ethnicity, and race will be summarized with the number and percentage of patients in each parameter category.

Disease history characteristics include: time since initial diagnosis (months), WHO classification, type of AML (secondary or treatment-related AML, de novo AML), disease status at screening, disease status at baseline, bone marrow blast percentage at baseline (for Cohort I-V and Group 2) complete remission prior to transplant (CR1/CR2/CR3; for Group 1 only), prior stem cell transplant (yes/no), time from transplant to randomization/first dose, time from first post-transplant relapse to first dose (for Cohort I-V and Group 2), history of post-transplant chronic GVHD (yes/no), maximum grade of chronic GVHD, history of post-transplant acute GVHD (yes/no), maximum grade of acute GVHD, type of transplant donor, conditioning regimen, MRD status pre- and post-transplant, genetic mutations identified, disease risk classification (favorable,

intermediate, unfavorable), present genetic abnormalities, bridging therapy (yes/no), number of lines of prior systemic therapy, prior cell therapy, and prior radiation therapy. Detailed summary of prior anti-cancer therapy will be provided for prior systemic therapy (excluding bridging therapy) and bridging therapy separately.

Demographic, baseline characteristics, and disease history characteristics will be summarized by study phase and cohort or treatment group for the all treated patients population. For Group 1, summaries will be repeated for the all randomized patients population and the PP population.

If an imbalance of more than 20% between the MT-401 and SOC arms is observed among the age, genetic mutations, transplant donor type, conditioning regimen, or complete remission status prior to transplant, the primary analysis may be modified to stratify for these factors.

Medical history data will be presented in patient data listings by study phase and patient.

7.4.3 Measurements of Treatment Compliance

Compliance to the study treatment regimen will be determined as the total dose received divided by the expected dose received, multiplied by 100. The total dose received will be calculated as the sum of actual dose received for each of the 3 planned doses. The expected dose received will be determined as the sum of the planned dose for each of the 3 doses (ie, 150 x 10^6 cells). Dosing compliance will be summarized using descriptive statistics, by study phase and cohort or treatment group, based on the all treated patients population. The number and percentages of patients who are < 66.6% compliant and ≥ 66.6% compliant will be summarized.

7.4.4 Primary Efficacy Endpoint Analysis Methods

Phase 2 Adjuvant (Group 1):

The primary analysis of RFS will be performed when 90 RFS events have been observed. RFS curves will be estimated using the Kaplan-Meier (KM) product limit method. The KM estimate of the median, 25th and 75th percentiles will be reported, and their corresponding 95% CIs will be computed. KM estimates of RFS rates at milestone time points (ie, 3 month intervals) will be reported along with two-sided 95% CIs using Greenwood's formula, provided the minimum follow-up in patients exceeds the time point to support stable estimation of the rate. The hazard ratio and corresponding 95% CI will be estimated using a Cox proportional hazards model including a factor for randomized arm and stratification factors of pre-transplant MRD status (MRD^+ vs. MRD^-/Unknown) and cytogenetic risk (Unfavorable vs. Other). The treatment effect on RFS will be evaluated using a stratified log-rank test. The significance level to be used in the RFS analysis at the final analysis will be based on α=0.05 (two-sided).

As noted in Section 7.4.2, if an imbalance of more than 20% between the MT-401 and SOC arms is observed among the age, genetic mutations, transplant donor type, conditioning regimen, or complete remission status prior to transplant, the primary analysis may be modified to stratify for these factors.

A landmark analysis will be performed including only patients who are relapse free, alive and have not withdrawn from the study by Day 0.

Phase 2 Active Disease (Group 2):

The CR rate and the corresponding 95% exact CI will be calculated. The number and percentages of patients in each best disease response category per ELN recommendations will also be presented. Results will be presented separately for Subgroups C and D. Data from a matched synthetic control arm may also be obtained and compared against the CR rate from Group 2 active disease patients as part of a separate statistical analysis plan.

The DOCR will be estimated using the KM product limit method and summarized similar to RFS as described above.

7.4.5 Secondary Endpoint Analysis Methods

Safety Lead-in (Cohorts I and II) and Cohorts III-V:

CR and DOCR will be summarized descriptively for active disease patients in the Safety Lead-in (Cohorts I and II) and Cohorts III-V similar to Group 2. Summaries will be presented by cohort and overall.

Phase 2 Adjuvant (Group 1):

The statistical methods for analyzing OS and GRFS for Group 1 patients are identical to those described for RFS.

Phase 2 Active Disease (Group 2):

The statistical methods for analyzing ORR for Group 2 are identical to those described for CR. For Group 2 active disease patients, DOR, PFS and OS will be estimated using the KM product limit method and summarized similar to RFS for Group 1 as described above. For Group 2 patients achieving CR following bridging therapy, RFS and GRFS will also be estimated using the KM product limit method as described above for Group 1.

7.4.6 Additional Efficacy Endpoint Analysis Methods

The analysis of markers of immune function, cell targeting, and signaling pathways will not be discussed in this SAP. If analyses of these endpoint are explored, they will be described in a separate analysis plan and may not be included in the primary CSR.

For Group 1 adjuvant patients, the time to MRD^{+} will be summarized similar to RFS, but with no statistical comparisons.

For Group 2 CR patients, the time to MRD^{-} will be summarized descriptively.

For Group 2 Subgroup D, among patients who had PCR detection of MRD at baseline, the percentage reduction from baseline will be summarized, as well as the number and percentage of patients whose disease level reduced by 50% or more.

A summary of PRO data will be presented using descriptive statistics for Group 2 only.

7.4.7 Statistical/Analytical Issues

7.4.7.1 Adjustments for Covariates

The planned primary analysis models will include adjustments for stratification factors of pre-transplant MRD status and cytogenetic risk. Subgroup analysis and/or analyses including additional factors may be conducted based on, but not limited to, the following covariates:

- Age (18-64 years, 65 years and above)
- Genetic mutations
- Transplant donor type (matched sibling, matched unrelated, haploidentical)
- Conditioning regimen (myeloablative, reduced intensity/nonmyeloablative)
- Complete remission status prior to transplant (CR1, CR2, CR3)

For time to event endpoints forest plots of the event-free rates at landmark times may be generated to display the outcomes in subgroups. Additionally, Kaplan-Meier plots with subgroup KM curves overlaid on the same plot may be presented. For binary endpoints, the rates within subgroups may be presented in forest plots.

Subgroup analyses are described in Section 7.4.7.7. In addition, data may be analyzed post-hoc to explore the effects of various factors. Such analyses will be labeled as post-hoc.

7.4.7.2 Handling of Dropouts or Missing Data

Patients with no post-baseline disease assessment will be included in the denominator for calculation of CR and ORR and will be treated as non-responders.

For time to event endpoints (RFS, OS, GRFS, PFS), patients with no baseline disease assessment or no post-baseline disease assessment will be censored using a censored value of 1 day.

No other imputations of missing data will be made.

7.4.7.3 Interim Analyses and Data Monitoring

A Sponsor review committee (SRC) will be convened to review the toxicities and endpoints related to the Safety Lead-in (Cohorts I and II), meet separately once Cohort III is complete and again once Cohorts IV-V are complete. The SRC will consist of at least 1-2 clinicians that includes the Medical Monitor and/or the Medical Director, along with 1 statistician.

Safety data will be reviewed at the end of the Safety Lead-in to determine if it is appropriate to move forward with the specific peptide manufacturer in the Phase 2 portion of the study. The evaluation of DLTs will be conducted in the DLT-evaluable set. Safety data will also be reviewed at the end of Cohort III to determine if it is appropriate to switch to a higher dose in the Phase 2 portion of the study. Additionally, safety data will also be reviewed at the end of Cohorts IV and V to determine if it is appropriate to switch to a dosing regimen with the MT-401 product manufactured using the accelerated process in the Phase 2 portion of the study. This decision will be made by the Sponsor based on multiple parameters, including the safety and biomarker data available.

A DMC will be formed for the Phase 2 portion of the study. The DMC will make recommendations on ongoing study conduct, including the accrual of additional patients to Group 1 based on the futility and efficacy criteria outlined below and any additional criteria established in the DMC charter. In addition, an Independent Efficacy Review Committee (IERC) may be formed to provide independent disease assessment. Separate charters will outline the details, including roles and responsibilities of these two committees.

Two formal interim analyses are planned during Phase 2, one for Group 1 and one for Group 2.

The interim analysis for Group 1 will occur when 45 RFS events have occurred. It is anticipated that this will occur approximately 30 months after the first patient randomized. At this analysis, the DMC will assess safety, futility, and potential expansion of study accrual. At the time of this analysis, enrollment of the remaining patients will continue, unless any safety concerns are identified.

The following options for the Group 1 interim analysis will be considered:

- Option 1: Terminate Group 1 due to futility if the RFS hazard ratio for MT-401:SOC is larger than 1.1 at the interim analysis. This futility stopping rule will be non-binding and can be overridden by the DMC and/or Marker.

- Option 2: Retain the planned target number for patient accrual and planned 90 RFS events if the conditional power (CP) calculated at the interim analysis is greater than 80% or less than 50% and the futility stopping rule is not met.

- Option 3: Increase the target number of planned accrual and RFS events if CP is between 50% and 80% and the futility stopping rule is not met. The target number of RFS events will be increased to achieve CP of 80% or increase by 50%, whichever is smaller. Target enrollment adjustments will be accomplished based on pooled event accumulation rate and a reasonable study duration projected via simulation.

The interim analysis for Group 2 will occur after the first approximately 10 patients in each subgroup have progressed or had the opportunity to be followed for 6 months. Clinical safety and efficacy will be reviewed comprehensively. Further enrollment for a subgroup may be terminated due to futility based on totality of the data. Patients treated in Cohorts IV-V with the same

manufacturing process as Group 2 may be included in Group 2 Subgroup C or Subgroup D analyses for decision making.

7.4.7.4 Multicenter Studies

This is a multicenter study, with approximately 20-25 study centers expected to participate. Efficacy data collected from all study centers will be pooled for data analysis. The effect of study center on the efficacy analysis results may be explored post-hoc, as needed.

7.4.7.5 Multiple Comparisons/Multiplicity

The primary endpoint of RFS for Group 1 will be tested at $\alpha = 0.05$ (two-sided) at the final analysis. The analysis of RFS at the interim analysis will be performed for the purposes of testing for futility and possible re-estimation of sample size as described in Section 7.4.7.3. Analysis at the interim and final analysis will be performed based on (Chen et al. 2004) in order to maintain the type 1 error at $\alpha = 0.05$ for RFS at the final analysis.

Secondary endpoints for Group 1, OS and GRFS, will be presented for descriptive purposes only and no adjustments will be made for multiplicity.

The primary endpoint of CR for Group 2 will be tested at $\alpha = 0.05$ (two-sided) at the final analysis. DOCR will be presented for descriptive purposes only and no adjustments will be made for multiplicity.

7.4.7.6 Use of an "Efficacy Subset" of Patients

The primary efficacy analysis for Group 1 will be performed on all randomized patients and the PP population will be utilized as a sensitivity analysis. The PP population will exclude patients received less than two-thirds of total planned dose or had other major protocol deviations.

7.4.7.7 Examination of Subgroups

Subgroup analyses may be conducted based on, but not limited to, the covariates defined in Section 7.4.7.1.

For time to event endpoints forest plots of the event-free rates at landmark times may be generated to display the outcomes in subgroups. Additionally, Kaplan-Meier plots with subgroup KM curves overlaid on the same plot may be presented. For binary endpoints, the rates within subgroups may be presented in forest plots.

7.5 Safety Evaluation

Safety analysis will be carried out for the all treated patients population, to include all patients who receive at least one dose of study drug. Patients who do not complete the study, for whatever reason, will have all available data up until the time of termination included in the analysis. For safety analysis presented by study visit, the baseline value will be defined as the last value reported prior to first study drug administration.

7.5.1 Extent of Exposure

Extent of exposure to study treatment will be summarized for the all treated patients population by study phase and cohort or treatment group. The number of infusions received, the total dose received (cells) and dose intensity will be summarized using descriptive statistics. The dose intensity is defined as the total dose received divided by number of infusions received.

7.5.2 Adverse Events

Treatment-emergent AEs will be summarized by study phase and cohort or treatment group. Events reported with a partial onset date (eg, month and year are reported but the day is missing) will be considered to be treatment-emergent if it cannot be confirmed that the event onset was prior to the first dose of study drug based on the available date entries.

Verbatim terms on case report forms will be mapped to preferred terms and system organ classes using the Medical Dictionary for Regulatory Activities (MedDRA, version 25.0 or higher).

Summaries that are displayed by system organ class and preferred terms will be ordered by alphabetic order of system organ class and by descending incidence of preferred term within each system organ class. Summaries displayed by preferred term only will be ordered by descending incidence of preferred term. Summaries of the following types will be presented:

- Overall summary of number of unique TEAEs and treatment-emergent serious adverse events (SAEs) and patient incidence of TEAEs meeting various criteria;
- Patient incidence of TEAEs by MedDRA system organ class (SOC), MedDRA preferred term and maximum CTCAE severity grade;
- Patient incidence of TEAEs by maximum CTCAE severity grade and MedDRA preferred term;
- Patient incidence of TEAEs related to study drug by maximum CTCAE severity grade and MedDRA preferred term;
- Patient incidence of treatment-emergent SAEs by maximum CTCAE severity grade and MedDRA preferred term;
- Patient incidence of treatment-emergent SAEs by SOC, MedDRA preferred term, and maximum CTCAE severity grade;
- Patient incidence of treatment-emergent SAEs related to study drug by maximum CTCAE severity grade and MedDRA preferred term;
- Patient incidence of Grade ≥ 3 TEAEs by maximum CTCAE severity grade and MedDRA preferred term; and

- Patient incidence of Grade ≥ 3 TEAEs related to study drug by maximum CTCAE severity grade and MedDRA preferred term.

At each level of summarization (eg, any AE, system organ class, and preferred term), patients experiencing more than one TEAE will be counted only once. In the summary of TEAEs by severity grade, patients will be counted once at the highest severity reported at each level of summarization.

Adverse event data will be presented in data listings by study phase, patient, cohort or treatment group, and event.

Some sensitivity analyses may be performed for Group 1 to summarize all TEAEs (ie, patient incidence of TEAEs by MedDRA preferred term) that occurred on or after date of randomization.

7.5.3 Deaths, Other Serious Adverse Events, and Other Significant Adverse Events

All deaths during the study, including the post treatment follow-up period, will be listed by patient, to include the primary cause of death. Serious AEs and other significant AEs, including those that led to withdrawal, interruption, or dose reduction of the study drug, will be provided in separate patient data listings.

The number and percentage of patient deaths, as well as primary cause of deaths will be summarized by study phase and cohort or treatment group.

All deaths within 30 and 60 days of the last dose of MT-401 will be summarized. For Group 1 SOC arm, deaths within 30 days and 60 days of randomization date plus average time from randomization to last dose in the MT-401 arm will also be summarized as reference.

7.5.4 Clinical Laboratory Evaluation

All descriptive summaries of laboratory results will be based on data collected by local laboratories and collected on the eCRF. All data will be included in by-patient data listings. Laboratory measurements identified as abnormal (ie, outside the normal range) will also be listed separately by study phase, patient, treatment group, laboratory test, and unit. In addition, normal ranges will be presented in a separate listing.

Clinical laboratory measurements, including hematology, serum chemistry, and coagulation, will be summarized by study phase and cohort or treatment group. Descriptive statistics will be presented for observed values and changes from baseline at each visit where parameters were scheduled to be collected per the clinical study protocol.

Where applicable, laboratory results for select parameters will be assigned a toxicity grade based on the U.S. Department of Health and Human Services *Common Terminology Criteria for Adverse Events (CTCAE)*, version 5.0 (27 Nov 2017). If the quantitative criteria for grading are equivalent for two grades and the differentiation is described by clinical interventions, the clinical intervention component will not be considered and the highest CTCAE grade will be assigned. Similarly, death related to AE (ie, Grade 5) cannot be determined with available laboratory-based

data collection and, thus, will not be summarized as a category. Laboratory parameters that include multiple sets of criteria for each direction (eg, separate criteria for potassium measures to assess hyperkalemia and hypokalemia) will be summarized separately to reflect each set of criteria.

Five-by-five contingency tables will be presented for lab tests where toxicity grading can be applied, to summarize the shift from the baseline grade to the worst post-baseline grade. Grades will be presented as none (Grade 0; ie, measurements did not meet any CTCAE criteria for Grades 1 through 4), mild (Grade 1), moderate (Grade 2), severe (Grade 3), or life-threatening (Grade 4). Summary results will include the count and percentage of patients within each shift category. In addition, a table will be presented to include the count and percentage of patients with at least one post-baseline Grade 3 or higher toxicity by laboratory parameter.

7.5.5 Vital Signs, Physical Findings, and Other Observations Related to Safety

7.5.5.1 Vital Signs

Vital sign parameter measurements will be summarized by study phase and cohort or treatment group. Descriptive statistics will be presented for results and change from baseline at each visit where parameters were scheduled to be collected.

7.5.5.2 Physical Examination

Results of the physical examination will be presented in patient data listings by study phase, patient, study visit, and body system.

7.5.5.3 GVHD Assessments

Shift from baseline in maximum overall grade for chronic GVHD will be summarized at each post-baseline visit and worst case post-baseline, and will include the count and percentage of patients within each shift category and treatment group.

7.5.5.4 Prior and Concomitant Medications

Medications will be coded using the World Health Organization Drug Dictionary Enhanced (WHODDE), March 2020 version. Medications entered on the eCRF will be mapped to Anatomic Therapeutic Chemical (ATC) drug class (level 4) and drug name.

Prior and concomitant medications will be summarized separately and the classification of prior/concomitant for each medication will be determined programmatically based on medication start and end dates. A prior medication is defined as any medication administered prior to the date of the first dose of study drug. A concomitant medication is defined as any medication administered on or after the date of the first dose of study drug. A medication may be defined as both prior and concomitant. If it cannot be determined whether a medication was received prior to the start of study drug dosing due to partial or missing medication start and/or end dates, it will be considered a prior medication. Likewise, if it cannot be determined whether a medication was received after the start of study drug dosing, it will be considered concomitant.

For both prior and concomitant medications summaries, the number and percentage of patients receiving any medication will be summarized by study phase and cohort or treatment group, as will the number and percentage receiving any medication by ATC drug class and generic drug name. Prior medications will also be summarized over all patients combined. Patients reporting use of more than one medication at each level of summarization (any medication received, ATC class, and generic drug name) will be counted only once. ATC class terms will be displayed alphabetically and by descending order of incidence of generic drug names within each ATC class. The classification of prior/concomitant/both for each medication will be presented on the listing of prior and concomitant medications.

7.6 Determination of Sample Size

Approximately 6-12 patients are expected to be enrolled in the Safety Lead-in portion of this study per a standard 3+3 design with 2 planned cohorts. In addition, approximately 9-18 patients (3-6 patients per cohort) are expected to be treated in the Cohorts III-V portion of this study as per a standard 3+3 design.

A total of 150 Group 1 patients will be randomized with a 1:1 allocation to receive MT-401 or to be under observation (SOC arm). Median RFS for the SOC arm is expected to be approximately 8 months. A clinically meaningful increase in median RFS for the MT-401 arm over SOC is thought to be 2 months or more. The primary analysis will be conducted when 90 RFS events have been observed. Ninety (90) RFS events provides 80% power with alpha=0.05 (two-sided) to detect a median RFS improvement of 6.5 months over the assumed SOC median RFS of 8 months (hazard ratio=0.55). Ninety RFS will be observed assuming uniform accrual of 150 patients (75 per arm) over a 50-month period and an additional 12 month minimum follow-up period. The DMC and Sponsor will review data from Phase 2 based on the planned interim analyses described in Section 7.4.7.3.

If the decision is made to dose patients in Phase 2 with any of the MT-401 dosing regimens evaluated in Cohorts III-V, enrollment will continue in Phase 2 Group 1 to achieve 150 evaluable patients randomized and approximately 75 patients treated with the final MT-401 dose and the intended manufacturing process, thereby replacing any patients that have not been treated with the final Phase 2 product, as determined by the Sponsor.

A total of approximately 60 Group 2 patients are planned to be enrolled and treated with MT-401, 30 patients in Subgroup C (frank relapse patients) and 30 patients in Subgroup D (MRD^{+}). After approximately 10 patients in each subgroup have progressed or had the opportunity to be followed for 6 months, clinical safety and efficacy will be reviewed comprehensively. Patients who develop CR_{MRD-} after bridging therapy and patients who relapse early (defined as less than 90 days ± 10 days and termed Group 2 Early Relapse patients) will be analyzed separately, and not be counted as part of the 60 patients in Group 2.

The sample size of 30 per subgroup is not based on a formal power calculation. An increase in the CR rate of greater than 10-15% compared to historical rates for standard of care would be considered clinically meaningful.

If the decision is made to dose patients in Phase 2 with any of the MT-401 dosing regimens evaluated in Cohorts III-V, enrollment will continue in Phase 2 Group 2 to achieve 60 evaluable patients treated with the final MT-401 dose and manufacturing process, thereby replacing any patients that have not been treated with the final Phase 2 product as determined by the Sponsor.

For historical control response rates ranging from 5% to 40%, the observed treatment effects (and associated 95% confidence intervals) needed to attain 80% and 90% power with a sample size of 30 per subgroup are provided in Table 3.

Table 3. Power[a] and 95% Confidence Intervals Relative to Historical Control Rates

Historical Control CR Rate	Observed Treatment Effect (80% Power) (95% CI)	Observed Treatment Effect (90% Power) (95% CI)
5%	21% (9%, 40%)	25% (11%, 44%)
10%	29% (14%, 48%)	32% (17%, 52%)
15%	36% (19%, 55%)	40% (22%, 59%)
20%	43% (25%, 62%)	47% (28%, 66%)
25%	50% (31%, 68%)	53% (34%, 72%)
30%	53% (34%, 71%)	57% (37%, 75%)
35%	60% (40%, 77%)	63% (44%, 80%)
40%	63% (43%, 79%)	66% (47%, 82%)

[a] based on a 1-sided exact test at the 0.05 alpha level.

7.7 Changes in the Conduct of the Study or Planned Analyses

There were no changes to the study conduct or planned analyses identified within the development of this SAP, relative to the descriptions provided within the clinical study protocol.

8 REFERENCE LIST

Döhner H, Estey E, Grimwade D, et al. Diagnosis and management of AML in adults: 2017 ELN recommendations from an international expert panel. Blood, 129(4), 424-447 (2017).

Harris AC, et al. International, Multicenter Standardization of Acute Graft-versus-Host Disease Clinical Data Collection: A Report from the Mount Sinai Acute GVHD International Consortium. Biol Blood Marrow Transplant 22, 4-10 (2016).

Jagasia M, et. al. National Institutes of Health Consensus Development Project on Criteria for Clinical Trials in Chronic Graft-versus-Host Disease: I. The 2014 Diagnosis and Staging Working Group Report. Biol Blood Marrow Transplant Vol 21, Issue 3, 389-401. e1 (2015).

Chen YHJ, DeMets DL, Lan KKG, Increasing the sample size when the unblinded interim result is promising, Statistics in medicine 23, 1023–1038 (2004).

Schuurhuis G et al. Minimal/measurable residual disease in AML: a consensus document from the European LeukemiaNet MRD Working Party. Blood, 141(12), 1275-1291 (2018).

Association of measurable residual disease with survival outcomes in patients with acute myeloid leukemia. JAMA Oncology, 6(12), 1890-1899 (2020).