1.0 STATISTICAL ANALYSIS PLAN

This section outlines the statistical analysis strategy and procedures for the study. Changes to exploratory or other non-confirmatory analyses made after the protocol has been finalized, but prior to final database lock, will be documented in a supplemental Statistical Analysis Plan (sSAP).

1.1 STATISTICAL ANALYSIS PLAN SUMMARY

Key elements of the statistical analysis plan (SAP) are summarized below; the comprehensive plan is provided in Sections 1.2 through 1.9.

Study Design Overview	The purpose of this Phase 1, single-site, modified "3+3" dose escalation trial was to assess the safety and tolerability of intratumoral PVSRIPO in unresectable (Stage IIIB, IIIC or IV per American Joint Commission of Cancer [AJCC] 7th Edition) anti-programmed cell death protein-1 (PD-1) resistant melanoma. Exploratory objectives included evaluation of overall response rate (ORR) and correlative analyses of immune response following intratumoral injection of PVSRIPO.
Treatment Assignment	Three subjects were enrolled into 4 cohorts (total n = 12 subjects) with increasing PVSRIPO dose based on the presence of dose-limiting toxicities (DLTs). PVSRIPO was administered intratumorally at $1x10^8$ median tissue culture infectious dose ($TCID_{50}$) per injection, with increasing PVSRIPO dose occurring as a result of increasing the total number of PVSRIPO injections administered over the course of the study, ie, the actual dose/injection was not increased; rather, the number of injections over time was increased. The number of injections/cohort was 1 in Cohort 0; 2 in Cohort 1 (into different lesions); 3 in Cohort 2 (into different lesions); and 3 injections into the same lesion for Cohort 3. Only one lesion was injected/visit such that each PVSRIPO injection was separated by 21 days. Decisions concerning dose adjustment for subsequent subjects were based upon the occurrence of DLTs during the first 3 weeks after each PVSRIPO administration.
Analysis Populations	Safety: All Subjects as Treated (ASaT) Efficacy: Full Analysis Set (FAS)
Primary Endpoint(s)	Safety: Dose-Limiting Toxicities (DLTs)
Key Exploratory Efficacy Endpoint	Overall response rate based on immune-related response criteria (irRC) (Wolchok et al, 2009), per site assessment
Statistical Methods for Key Efficacy Analyses	The efficacy objectives will be evaluated by the point estimate and 95% confidence interval (CI) for the best overall response (BOR) in the FAS using an exact binomial distribution.
Statistical Methods for Key Safety Analyses	Summary statistics, and 95% CI for the incidence rate of DLTs, as defined in Section 1.6.1.
Multiplicity	No multiplicity adjustment will be applied.
Sample Size and Power	Given the purpose of this study was to determine the recommended PVSRIPO dose level and dosing regimen for future studies, the sample size reflects the number of subjects needed, per this modified 3+3 trial design, to characterize the safety and initial anti-tumor activity of the recommended PVSRIPO dose. The sample size was not based on statistical considerations; therefore, no inference will be drawn from any statistical tests conducted.

1.2 RESPONSIBILITY FOR ANALYSES

The statistical analysis of the data obtained from this study will be the responsibility of the biostatistics department of the Sponsor. This trial was conducted as an open-label, non-randomized clinical trial (ie, subjects, investigators, and Sponsor personnel were aware of subject treatment assignments after each subject was enrolled and treatment was assigned).

1.3 HYPOTHESES/ESTIMATION

Primary hypothesis: Intratumoral injection of PVSRIPO into cutaneous, subcutaneous, or nodal lesions of subjects with unresectable melanoma will be feasible and safe at the dose levels tested.

1.4 ANALYSIS ENDPOINTS

Safety and efficacy endpoints to be evaluated are listed below.

1.4.1 SAFETY ENDPOINTS

Adverse events will be graded using National Cancer Institute (NCI) Common Terminology Criteria for Adverse Events (CTCAE) v4.0. The primary safety endpoint for this Phase 1 study is the proportion of subjects experiencing a DLT within each cohort. An adverse event (AE) is considered a DLT, if it meets the following criteria:

- Grade 4 or higher non-hematologic treatment-related adverse event (TEAE), with the exception of vitiligo.
- Grade 4 or higher hematologic TEAE

Additional safety endpoints include assessment of all TEAEs experienced by subjects who received at least one intratumoral injection of PVSRIPO, laboratory safety assessments, vital signs, reasons for treatment discontinuation, and physical examinations.

1.4.2 EXPLORATORY EFFICACY ENDPOINTS

Per protocol, all efficacy analyses are exploratory. The main efficacy endpoint is the ORR, defined as the proportion of subjects in the analysis population who have complete response (CR) or partial response (PR) at any time during the study using irRC (Wolchok et al, 2009).

Additional efficacy endpoints include:

- Duration of response (DOR), defined as the time from first response to disease progression in subjects who achieve a PR or better, based on assessment by irRC.
- Overall response rate and DOR based on response evaluation criteria in solid tumors version 1.1 (RECIST 1.1) (Eisenhauer et al, 2009).

1.5 ANALYSIS POPULATIONS

1.5.1 SAFETY ANALYSIS POPULATIONS

The ASaT population will include all patients who received at least one dose of PVSRIPO and will be used for the analysis of safety data. At least 1 laboratory or vital sign measurement obtained

after at least 1 dose of study treatment is required for inclusion in the analysis of each specific parameter in the ASaT population. To assess change from baseline, a baseline measurement is also required.

1.5.2 EFFICACY ANALYSIS POPULATIONS

The FAS population, which consists of subjects who receive at least 1 dose of PVSRIPO and with measurable disease at baseline, will serve as the primary population for the analyses of efficacy data.

1.5.3 SUBJECTS NOT ANALYZED

An additional cohort (Cohort 4) was planned to explore the safety and tolerability of treating multiple lesions during the same visit; however, no patients were enrolled in Cohort 4.

In addition, multi-lesional PVSRIPO retreatment of patients originally enrolled in Cohorts 0 through 3 was allowed; however, no patients from Cohorts 0 through 3 were retreated with PVSRIPO.

1.6 STATISTICAL METHODS

This section describes the statistical methods addressing the primary and exploratory objectives.

1.6.1 STATISTICAL METHODS FOR SAFETY ANALYSES

Safety and tolerability will be assessed by review of all relevant parameters including AEs, laboratory tests, and vital signs. Summary statistics (eg, counts, percentage, mean, standard deviation) will be provided for the safety endpoints as appropriate. The 95% CI for the incidence rate of injection site reactions and the incidence rate of Grade 3 or 4 AEs will be provided as appropriate.

Adverse Events

AEs will be coded using Medical Dictionary for Regulatory Activities (MedDRA®) Version 23.1, with each verbatim term coded to a lowest level term (LLT), which is mapped to a preferred term (PT), high level term (HLT), and a system organ class (SOC).

AE summaries will include only TEAEs, which are defined as AEs that are not present prior to the start of study medication, or AEs present before study medication that worsened after starting study medication. If a partially missing date or time of onset allows the possibility that an AE may be a TEAE, it will be assumed the event is a TEAE.

An overall summary of the number and percentage of subjects in each category will be presented, as well as an overall summary of the number of events in each category.

The Investigator will provide an assessment of the severity of each AE using CTCAE Version 4.0.

Note: "Severe" is a measure of intensity; it is not equivalent to "serious". Subjects will be counted and summarized under their highest severity within a SOC and PT.

If the relationship between the AE or serious adverse event (SAE) and the investigational product is determined to be possibly, probably, or definitely related, the event will be considered as related to the investigational product for the purposes of reporting. Otherwise, the AE or SAE will be considered 'unrelated.'

Each subject will be counted once within each preferred term. If a subject experienced more than one TEAE within a preferred term only the TEAE with the strongest relationship or the maximum intensity, as appropriate, will be included in the summaries of relationship and intensity. No inferential statistical tests will be performed.

For TEAE's, summary tables showing the incidence rates (frequencies and percentages) and number of events for the following categories will be prepared:

- Overall Summary of TEAEs
- TEAEs by SOC and PT
- TEAEs by PT in descending frequency
- PVSRIPO-Related TEAEs by SOC and PT
- PVSRIPO-Related TEAEs by PT in descending frequency
- Severe (Grade 3), Life Threatening (Grade 4) and Fatal (Grade 5) TEAEs by SOC and PT
- Severe (Grade 3), Life Threatening (Grade 4) and Fatal (Grade 5) PVSRIPO-Related TEAEs by SOC and PT
- Listing of SAEs
- Listing of DLTs

Definition of Dose Limiting Toxicities

For Cohorts 0 through 3, DLTs were identified from the time of each PVSRIPO injection until the end of a 21-day observation period following each injection. DLTs were defined as any of the below events that did not resolve to pre-treatment baseline or ≤ Grade 1 within 21 days or any toxicity indicated below that resolved within the 21-day period, but then recurred as Grade 4 or higher during that same 21-day period.

DLTs include:

- Any Grade 4 or higher non-hematologic toxicities probably, possibly, or definitely related to PVSRIPO, with the exception of vitiligo
- Any Grade 4 or higher hematologic toxicities probably, possibly, or definitely related to PVSRIPO.

1.6.2 STATISTICAL METHODS FOR EXPLORATORY EFFICACY ANALYSES

Efficacy will be evaluated in the FAS population. For the key efficacy endpoint of ORR based on irRC, the point estimate, and 95% CI (as determined by the upper and lower 97.5% one-sided

confidence bounds) will be provided using an exact binomial distribution. Subjects without response data will be counted as non-responders.

DOR will be summarized using Kaplan-Meier (KM) method in all responders.

An outline of the key efficacy analysis strategy is presented in Table 1.

Table 1 Primary Analysis Strategy for Efficacy Endpoints

Endpoint/Variable (Description, Time Point)	Statistical Method	Analysis Population	Missing Data Approach
Primary:			
Overall irRC response rate based on investigator assessment	Exact test of binomial parameter	FAS	Subjects with missing data are considered non-responders
Secondary:			
Overall RECIST 1.1 response rate based on investigator assessment	Exact methods for binomial parameter	FAS	Subjects with missingdata are considered non-responders
Response duration (RECIST 1.1)	Summary statistics using Kaplan-Meier method	All responders	Non-responders are excluded in analysis
Response duration (irRC)	Summary statistics using Kaplan-Meier method	All responders	Non-responders are excluded in analysis
Exploratory:			
Overall irRC response rate and RECIST 1.1 response rate based on investigator assessment without requiring confirmatory scans for (ir)CR/PR responses	Exact methods for binomial parameter	FAS	Subjects with missingdata are considered non-responders

CR = complete response; FAS = full analysis set; ir = immune related; irRC = immune-related response criteria; PR = partial response; RECIST = response evaluation criteria in solid tumors

Details of the exploratory biomarker analysis strategy will be documented in the sSAP.

1.6.3 DEMOGRAPHIC AND BASELINE CHARACTERISTICS

Baseline characteristics will be assessed using tables and/or graphs. No statistical hypothesis tests will be performed on these characteristics. The number and percentage of subjects screened, enrolled, primary reasons for screening failure, and discontinuation will be displayed. Demographic variables (eg, age, sex), baseline characteristics, primary and secondary diagnoses, and prior and concomitant therapies will be summarized either by descriptive statistics or categorical tables as appropriate.

1.7 MULTIPLICITY

The main efficacy objective of ORR per irRC in the FAS Population will be assessed by a point estimate and 95% CI. There are no formal hypotheses, and no adjustments for multiplicity will be applied. Evaluation of objectives should be interpreted with caution due to the potential for an increase in the false positive rate.

1.8 SAMPLE SIZE AND POWER CALCULATIONS

Given the purpose of this study was to determine the recommended dose and dosing regimen of PVSRIPO in anti-PD-1 resistant melanoma, the sample size was chosen to characterize the safety and initial anti-tumor activity of PVSRIPO. The sample size is not based on statistical considerations; therefore, no inference will be drawn from any statistical tests conducted.

1.9 RESPONSE CRITERIA

1.9.1 IMMUNE RELATED RESPONSE CRITERIA

For all patients who experience disease progression on study, the date noted for disease progression is the time of the scan where it is originally detected, and not the following date of the confirmatory scan. If a scan was not available, the date of disease progression was identified by the Investigator.

Definitions of measurable and non-measurable disease

Measurable disease: Neoplastic masses that can be precisely measured in 2 in-plane perpendicular diameters. Both the longest diameter and the longest perpendicular must be greater than or equal to 10 mm or 2 times the axial slice thickness, whichever is greater. Lymph nodes must have a short-axis line-length of ≥ 15 mm. Malignant lymph nodes must be measurable in 2 perpendicular diameters. Both the longest diameter and the longest perpendicular must be greater than or equal to 15 mm or 2 times the axial slice thickness. The quantitative endpoint will be defined as the product of the longest diameter with its longest perpendicular.

Non-measurable disease: Non-measurable lesions are those that are not suitable for quantitative assessment over time. These include:

1. Neoplastic masses that are too small to measure, because their longest uninterrupted diameter or longest perpendicular are less than 10 mm or two (2) times the axial slice thickness.

2. Neoplastic masses whose boundaries cannot be distinguished. This includes masses that cannot be demarcated from surrounding tissue because of inadequate contrast, masses with overly complex morphology, or those with highly heterogeneous tissue composition.

3. Other types of lesions that are confidently felt to represent neoplastic tissue but difficult to quantify in a reproducible manner. These include bone metastases, leptomeningeal metastases, malignant ascites, pleural/pericardial effusions, inflammatory breast disease, lymphangitis cutis/pulmonis, cystic lesions, ill-defined abdominal masses, etc.

For irRC, only target lesions selected at baseline and measurable new lesions are taken into account.

At the baseline tumor assessment, the sum of the products of the two largest perpendicular diameters (SPD) of all **index lesions** (5 lesions per organ, up to 10 visceral lesions and 5

cutaneous index lesions) is calculated.

At each subsequent tumor assessment, the SPD of the index lesions and of new, measurable lesions (≥ 5 X 5 mm; up to 5 new lesions per organ: 5 new cutaneous lesions and 10 visceral lesions) are added together to provide the total time-point **tumor burden**.

Overall response using irRC:

- Complete Response (irCR): Complete disappearance of all tumor lesions (whether measureable or not, and no new lesions). CR must be confirmed by repeated, consecutive assessments made no less than 4 weeks from the date first documented.
- Partial Response (irPR): ≥50% decrease in tumor burden compared with baseline in two observations at least 4 weeks apart.
- Stable Disease (irSD): Failure to meet criteria for irCR or irPR, in absence of irPD.
- Progressive Disease (irPD): At least 25% increase in SPD relative to nadir (minimum recorded tumor burden) Confirmation by a repeat, consecutive assessment no less than 4 weeks from the data first documented.
- Non-index lesions contribute to defining irCR (complete disappearance required).

Immune-related response criteria for the current protocol was adopted from Wolchok et al, 2009.

1.9.2 RESPONSE EVALUATION CRITERIA IN SOLID TUMORS (RECIST) 1.1

RECIST version 1.1 will be used in this study for assessment of tumor response (Eisenhauer et al, 2009).

1.9.3 ADDITIONAL EXPLORATORY RESPONSE EVALUATION CRITERIA

Because of the exploratory nature of the study, some subjects may not have had confirmatory scans prior to exposure to subsequent anti-cancer therapies. Therefore, both irRC and RECIST version 1.1 will be applied without the requirement for confirmatory scans for irCR/irPR and for CR/PR for irRC and RECIST version 1.1, respectively.

Summary statistics comparing the 2 methods of response criteria application (confirmed vs. unconfirmed response) may be calculated for qualitative comparisons.

2.0 REFERENCES

Eisenhauer E.A, Therasse P, Bogaerts J, Schwartz L.H, Sargent D, Ford R, Dancey J, ArbuckS, Gwyther S, Mooney M, Rubinstein L, Shankar L, Dodd L, Kaplan R, Lacombe D, Verweij J. New response evaluation criteria in solid tumours: Revised RECIST guideline (version 1.1). Eur J Cancer. 2009;45:228-47.

Wolchok JD, Hoos A, O'Day S, et al. Guidelines for the Evaluation of Immune Therapy Activity in Solid Tumors: Immune-Related Response Criteria. Clin Cancer Res. 2009;15:7412-20.