Supplementary Material

Methods

Sample size determination, enrollment continuation, and early termination

Selection of sample sizes for each Study Track varies and will be guided by multiple designs, including the Simon 2-stage, nested Simon-Fleming 4-stage, and single-stage designs [1-3]. Given the differences in prior immuno-oncology (I-O) therapy and existing care options for patients under each Track, different criteria will be used to determine the number of patients to be enrolled at each stage and the number of objective responses required to recommend continuation to the next stage or termination of a study treatment arm. Additional patients may be enrolled to compensate for patients who may drop out without being evaluable for a response. In addition to the number of responses, other aspects of clinical benefit—such as the duration of response and progression-free survival rate—safety, biomarkers, and the relative performance of patients in the other treatment combination arms within the same study will also be considered.

A limited number of patients previously treated in the I-O–naive Tracks will be permitted to be re-randomized to the I-O–experienced Tracks. The number of re-randomized patients permitted may change depending on the enrollment rates, rates of progression, and number of currently open treatment arms.

Non-small cell lung cancer

Selection of sample sizes for the Tracks will be guided by multiple designs: Tracks 1.b and 2, nested Simon-Fleming 4-stage design; Tracks 3 and 4, Simon 2-stage design; and Track 5, single-stage design [1-4]. Twenty patients will be treated in both Tracks 1.b (I-O naive; programmed death-ligand 1 [PD-L1] positive) and 2 (I-O naive; PD-L1 negative). If the number of responses is ≤ 3 of 20 and ≤ 2 of 20 in Tracks 1.b and 2, respectively, then the treatment arm will be terminated. Otherwise, enrollment will continue to stage 2, where an additional 15 patients will be enrolled in each of the Tracks, for a total of 35 patients in each. If the number of responses is ≤ 6 of 35 and ≤ 4 of 35, respectively, then the treatment arm will be terminated. If the number of responses is ≥ 7 of 35 and ≥ 5 of 35, respectively, then the treatment arm will proceed to stage 3. However, if the number of responses in stage 2 is ≥ 11 of 35 in Track 1.b and ≥ 9 of 35 in Track 2, then the treatment arm will be considered efficacious and proceed to further development.

If proceeding to stage 3, an additional 15 patients will be enrolled, for a total of 50 patients in Tracks 1.b and 2. If the number of responses is ≤ 11 of 50 and ≤ 8 of 50, respectively, then the treatment arm will be terminated. Otherwise, an additional 20 patients will be enrolled in both Tracks (for a total of 70 patients in each) and the treatment arm will proceed to stage 4. If the number of responses is ≤ 13 of 70 and ≤ 9 of 70 in Tracks 1.b and 2, respectively, then the treatment arm will be terminated. If the number of responses is ≥ 14 of 70 and ≥ 10 of 70, respectively, then the treatment arm will be considered efficacious and carried on for further clinical development. These designs provide 93% power and a 3% false-positive rate for Track 1.b, assuming historical and target objective response rates (ORRs) of 14% and 35%,

respectively, and 95% power and a 4.4% false-positive rate for Track 2, assuming historical and target ORRs of 10% and 30%, respectively.

In Track 3, an initial 12 I-O–experienced patients will be treated. If a response is observed in ≤ 1 patient, then the treatment arm will be deemed futile. Otherwise, an additional 23 patients will be enrolled, for a total of 35 patients. If ≤ 5 of 35 responses are observed, then the treatment arm will be terminated. If ≥ 6 responses are observed, then the treatment arm will proceed to further development. This design provides 90% power and a 10% false-positive rate, assuming a target ORR of 30%. Given the limited data available for I-O–experienced patients, the Track 3 design is based not on observed historical response rates but rather on the assumption that a response rate below 10% would not warrant further study.

In Track 4, I-O–naive patients will be stratified based on PD-L1 expression levels as assessed by their baseline tumor biopsy. A minimum of 19, 28, and 16 patients with < 1%, ≥ 1% and < 50%, and ≥ 50% PD-L1 expression, respectively, will be treated and evaluated for efficacy. The treatment arm will not be considered efficacious if ≤ 3 of 19, ≤ 11 of 28, and ≤ 11 of 16 responses are observed in the < 1%, ≥ 1% and < 50%, and ≥ 50% PD-L1 expression groups, respectively. Otherwise, the study will progress to stage 2 and enroll an additional 17, 13, and 9 patients, for a total of 36, 41, and 25 patients, respectively. If the total number of responses is < 10 of 36, < 20 of 41, and < 20 of 25, respectively, then the treatment arm will be considered futile. If > 10, > 20, and > 20 responses, respectively, are observed, then the treatment arm will be considered for further development. These designs provide 90% power and a 10% false-positive rate, assuming historical and target ORRs of 20% and 40% (< 1% PD-L1 expression),

40% and 60% (≥ 1% and < 50% PD-L1 expression), and 70% and 90% (≥ 50% PD-L1 expression), respectively.

In Track 5, a total of 35 I-O–experienced patients will be treated, regardless of PD-L1 expression status. If ≥ 8 responses are observed, then the treatment arm will be considered efficacious. Otherwise, the treatment arm will be closed. Patients previously treated in Track 4 or re-randomized patients from Track 5 can enroll in Track 5. This design provides 90% power and a 10% false-positive rate, assuming a null rate of 10% and a target ORR of 30%.

Gastric cancer

Selection of sample sizes for both Tracks 1 (I-O naive; stratified by PD-L1 status) and 2 (I-O experienced) will be guided by a Simon 2-stage (optimal) design [1, 5]. A minimum of 24 and 19 patients will be treated in the Track 1 PD-L1–positive and PD-L1–negative groups, respectively, and efficacy will be evaluated. If the number of responses is ≤ 8 of 24 and ≤ 3 of 19, respectively, then the treatment arm will be considered futile. Otherwise, the treatment arm will proceed to stage 2, where an additional 39 and 25 patients will be enrolled, for a total of 63 and 44 patients, respectively. If the total number of responses is ≤ 24 of 63 and ≤ 10 of 44, respectively, then the treatment arm will be terminated. If there are > 24 and > 10 responses, respectively, then the arm will proceed to further development.

In Track 2, an initial 21 patients will be treated and preliminary efficacy will be assessed. If the number of responses is ≤ 1 of 21, then the treatment arm will be terminated. Otherwise, stage 2 will be initiated, where an additional 20 patients will be enrolled. If the total number of

responses is ≤ 4 of 41, then the arm will be considered futile. If there are > 4 responses, then the treatment arm will proceed to further development.

These designs provide 90% power and a 5% false-positive rate, assuming historical and target ORRs of 30% and 50% for the Track 1 PD-L1–positive group and 15% and 35% for the Track 1 PD-L1–negative group, respectively, and a target response rate of 20% for Track 2. Given the limited data available for I-O–experienced patients, the Track 2 design is based not on observed historical response rates but rather on the assumption that a response rate below 5% would not warrant further study.

Renal cell carcinoma

Selection of sample sizes for both Tracks 1 (I-O naive) and 2 (I-O experienced) will be guided by a Simon 2-stage (optimal) design [1, 6]. A minimum of 24 and 21 patients will be treated in Tracks 1 and 2, respectively, and evaluated for efficacy. If the number of responses is ≤ 8 of 24 and ≤ 1 of 21, respectively, then the treatment arm will be considered futile and will be terminated. Otherwise, enrollment to stage 2 will continue; 39 and 20 additional patients will be enrolled in Tracks 1 and 2, for a total of 63 and 41 patients, respectively. If the total number of responses at the end of stage 2 is ≤ 24 of 63 and ≤ 4 of 41 in Tracks 1 and 2, respectively, then the arm will be terminated for futility. If there are > 24 and > 4 responses, respectively, then the arms will be carried on for further development.

These designs provide 90% power and a 5% false-positive rate, assuming historical and target ORRs of 30% and 50%, respectively, for Track 1 and a target response rate of 20% for Track 2.

Given the limited data available for I-O–experienced patients, the Track 2 design is based not on observed historical response rates but rather on the assumption that a response rate below 5% would not warrant further study.

Supplementary References

1. Simon R: Optimal two-stage designs for phase II clinical trials. Control Clin Trials 10:1-10, 1989

2. Fleming TR: One-sample multiple testing procedure for phase II clinical trials. Biometrics 38:143-151, 1982

3. Thall PF, Simon R: Practical Bayesian guidelines for phase IIB clinical trials. Biometrics 50:337-349, 1994

4. Fracasso PM, Freeman DJ, Simonsen K, et al: A phase 2 trial, fast real-time assessment of combination therapies in immuno-oncology, in patients with advanced non-small cell lung cancer (FRACTION-lung). Ann Oncol 27 (Suppl 6): Abstract 1295TiP, 2016

5. Aanur P, Gutierrez M, Kelly RJ, et al: FRACTION (Fast Real-time Assessment of Combination Therapies in Immuno-Oncology)—gastric cancer (GC): a randomized, open-label, adaptive, phase 2 study of nivolumab in combination with other immuno-oncology (IO) agents in patients with advanced GC. J Clin Oncol 35: Abstract TPS4137, 2017

6. Taitt CC, Motzer RJ, Choueiri TK, et al: FRACTION-RCC: a randomized, open-label, adaptive, phase 2 study of nivolumab in combination with other immuno-oncology agents in patients with advanced RCC. J Immunother Cancer 5 (Suppl 2): Abstract 87 P226, 2017