Serum alpha-fetoprotein levels and clinical outcomes in the phase 3 CELESTIAL study of cabozantinib versus placebo in patients with advanced hepatocellular carcinoma

Robin Kate Kelley,1 Tim Meyer,2 Lorenza Rimassa,3,4 Philippe Merle,5 Joong-Won Park,6 Thomas Yau,7 Stephen L. Chan,8 Jean-Frederic Blanc,9 Vincent C. Tam,10 Albert Tran,11 Vincenzo Dadduzio,12 David W. Markby,13 Rajesh Kaldate,13,a Ann-Lii Cheng,14 Anthony B. El-Khoueiry,15 Ghassan K. Abou-Alfa16,17

1UCSF Helen Diller Family Comprehensive Cancer Center, San Francisco, CA, USA
2Royal Free Hospital and UCL Cancer Institute, London, UK
3Humanitas Cancer Center, Humanitas Clinical and Research Center-IRCCS, Rozzano (Milan), Italy
4Department of Biomedical Sciences, Humanitas University, Pieve Emanuele (Milan), Italy
5Groupement Hospitalier Lyon Nord, Lyon, France
6National Cancer Center, Goyang, Republic of Korea
7Queen Mary Hospital, Hong Kong, China
8The Chinese University of Hong Kong, Shatin, Hong Kong, China
9Hôpital Haut-Lévêque, CHU Bordeaux, France
10Tom Baker Cancer Centre, University of Calgary, Calgary, AB, Canada
11Groupe Hospitalier L’Archet, Nice, France
12Medical Oncology Unit 1, Istituto Oncologico Veneto, IRCCS, Padova, Italy
13Exelixis, Inc., Alameda, CA, USA
14National Taiwan University Cancer Center, Taipei, Taiwan
15USC Norris Comprehensive Cancer Center, Los Angeles, CA, USA
16Memorial Sloan Kettering Cancer Center, New York, NY, USA
17Weill Medical College at Cornell University, New York, NY, USA

aNo longer an employee of Exelixis

Running title: Serum AFP levels and outcomes with cabozantinib in CELESTIAL

Keywords: Gastrointestinal cancers, biomarkers, small molecule agents, clinical trials, molecular oncology

Corresponding author: Robin Kate Kelley
Department of Medicine (Hematology/Oncology)
UCSF Helen Diller Family Comprehensive Cancer Center
1600 Divisadero St., MZ Bldg A
San Francisco, CA 94143-1770
Email: katie.kelley@ucsf.edu
Phone/fax: (415) 353-9888/(415) 353-9959

Tables/Figures: 2 tables, 4 figures
Disclosure of potential conflicts of interest:

Robin Kate Kelley reports institutional consulting or advisory roles for Agios, AstraZeneca, Bristol-Myers Squibb, and Target PharmaSolutions; noninstitutional consulting or advisory roles for Genentech/Roche, Gilead, Ipsen, and Target PharmaSolutions; and institutional research funding from Adaptimmune, Agios, AstraZeneca, Bayer, Bristol-Myers Squibb, EMD Serono, Exelixis, Lilly, MedImmune, Merck Sharp & Dohme, Novartis, and Taiho Pharmaceutical.

Tim Meyer reports grants from Bayer and BTG and personal fees (consultancy) from BMS, AstraZeneca, Tarveda, Ipsen, MSD, and Eisai.

Lorenza Rimassa reports honoraria from AstraZeneca, AbbVie, Gilead, and Roche; consulting/advisory role for Lilly, Bayer, Sirtex Medical, Italfarmaco, Sanofi, ArQule, Baxter, Ipsen, Exelixis, Amgen, Celgene, Incyte, and Eisai; and research funding from Exelixis, Beigene, AstraZeneca, Astellas, MSD, ArQule, Lilly, Agios, Ipsen, Eisai, and Incyte.

Philippe Merle reports advisory board activity for Bayer, Ipsen, Exelixis, Lilly, BMS, MSD, AstraZeneca, Eisai, and Roche.

Joong-Won Park declares no potential conflicts of interest.

Thomas Yau reports a consulting/advisory role for Exelixis and honoraria for Ipsen and Exelixis.

Stephen L. Chan reports clinical trial grants from Merck Sharp & Dohme (Asia) Ltd and personal fees (speaker and consulting/advisory role) for AstraZeneca Hong Kong Limited.

Jean-Frederic Blanc reports personal fees (board) from Bayer, Ipsen, Eisai, and Lilly.

Vincent C. Tam reports grants (research funding to institution) from Bayer and and personal fees (advisory board honorarium) from Eisai, Ipsen, and BMS.

Albert Tran reports a relationship with an Investigator.

Vincenzo Dadduzio reports personal fees from Ipsen, Bayer, and MSD.

David W. Markby reports salary and stock options as an employee of Exelixis, Inc.

Rajesh Kaldate reports salary and stock options as an employee of Exelixis, Inc.

Ann-Lii Cheng reports a consulting or advisory role for Bayer, Bristol-Myers Squibb, Eisai, Exelixis, MSD, Newbird Healthcare, Novartis and Ono; and speaker’s bureau activity for AstraZeneca, Bayer, and Eli Lilly.

Anthony B. El-Khoueiry reports honoraria from AstraZeneca, Bayer, Bristol-Myers Squibb, Genentech, and GlaxoSmithKline; consulting or advisory role for AstraZeneca, Bristol-Myers Squibb, and Genentech/Roche; speakers’ bureau activity for Merrimack; research funding from Astex Pharmaceuticals; and travel, accommodations, and expenses from AstraZeneca, Bayer, Bristol-Myers Squibb, Genentech, and GlaxoSmithKline.

Ghassan K. Abou-Alfa reports a consulting or advisory role for 3DMedcare, Agios, Alignmed, Amgen, Antengene, Aptus, Aslan, Astellas, AZ, Bayer, Beigene, Bioline, BMS, Boston Scientific, Bridgebio, CARsgen, Celgene, Casi, Cipla, CytomX, Daiichi, Debio, Delcath, Eisai, Exelixis, Genoscience, Halozyme, Hengrui, Incyte, Inovio, Ipsen, Jazz, Jansen, Kyowa Kirin, LAM, Lilly, Loxo, Merck, Mina, Novella, Onxeo, PCI Biotech, Pfizer, Pieris, QED, Redhill, Sanofi, Servier, Silenseed, Sillajen, Sobi, Targovax, Tekmira, Twoxar, Vicus, Yakult, and Yiviva; and research funding for ActaBiologica, Agios,
Array, AZ, Bayer, Beigene, BMS, Casi, Celgene, Exelixis, Genentech, Halozyme, Incyte, Lilly, Mabvax, Novartis, OncoQuest, Polaris Puma, QED, and Roche.

**Translational relevance:** Cabozantinib is approved for patients with advanced hepatocellular carcinoma (HCC) who have received prior sorafenib, based on the randomized phase 3 CELESTIAL study. With the recent expansion of treatment options in the second-line setting for HCC, there is an urgent need for biomarkers of response to help guide treatment decisions. High serum levels of alpha-fetoprotein (AFP) are associated with poor prognosis in patients with HCC, and studies suggest a correlation between on-treatment decrease in AFP and improved outcomes. In this exploratory analysis of the CELESTIAL study, we show that cabozantinib prolonged OS and PFS relative to placebo across a range of baseline AFP levels. On-treatment AFP response, defined as a decrease of ≥20% from baseline in serum AFP, was more common with cabozantinib than placebo and was associated with improved OS and PFS in the cabozantinib arm. Further analysis of AFP kinetics in large, prospective, randomized studies is warranted.
Abstract

Purpose

The phase 3 CELESTIAL study demonstrated improved overall survival (OS) and progression-free survival (PFS) with cabozantinib versus placebo in patients with previously treated, advanced hepatocellular carcinoma (HCC). We analyzed outcomes by baseline alpha-fetoprotein (AFP) and on-treatment AFP changes.

Experimental design

Serum AFP was measured every 8 weeks by blinded, centralized testing. Outcomes were analyzed by baseline AFP bifurcated at 400 ng/mL and by on-treatment AFP response (≥20% decrease from baseline at Week 8). The optimal cutoff for change in AFP at Week 8 was evaluated using maximally selected rank statistics.

Results

Median OS for cabozantinib versus placebo was 13.9 versus 10.3 months (HR, 0.81; 95% CI, 0.62–1.04) for patients with baseline AFP <400 ng/mL, and 8.5 versus 5.2 months (HR, 0.71; 95% CI, 0.54–0.94) for patients with baseline AFP ≥400 ng/mL. Week 8 AFP response rate was 50% for cabozantinib versus 13% for placebo. In the cabozantinib arm, median OS for patients with and without AFP response was 16.1 versus 9.1 months (HR, 0.61; 95% CI, 0.45–0.84). AFP response was independently associated with longer OS. The optimal cutoff for association with OS in the cabozantinib arm was ≤0% change in AFP at Week 8 ([AFP control; HR 0.50 [95% CI, 0.35–0.71]]. HRs for PFS were consistent with those for OS.

Conclusions

Cabozantinib improved outcomes versus placebo across a range of baseline AFP levels. On-treatment AFP response and control rates were higher with cabozantinib than placebo, and were associated with longer OS and PFS with cabozantinib.
Introduction

High serum levels of alpha-fetoprotein (AFP) are associated with poor prognosis in patients with hepatocellular carcinoma (HCC) across stages of disease. Studies have shown an association of pretreatment AFP level with tumor size, pathological grade, tumor stage, and survival. Elevated preoperative AFP has been associated with recurrence in patients undergoing surgical resection or transplant. In patients treated with transarterial chemoembolization or surgery, post-intervention AFP decreases are associated with improved outcomes, including longer time to progression or recurrence, while increases indicate disease progression. Retrospective studies of patients with HCC receiving systemic therapy also suggest an association between AFP decline on treatment and improved survival.

There is no consensus definition of AFP-based response or progression; criteria vary across studies, with thresholds of 20% and 50% change from baseline AFP frequently used. Furthermore, studies of AFP response and progression in advanced HCC have primarily included patients treated with chemotherapy or sorafenib, with limited data for new and emerging targeted systemic agents.

Tumors with high AFP expression may represent a distinct biological subtype of HCC, providing a basis for the observed prognostic effects of serum AFP. Gene expression profiling has identified three major molecular subtypes of HCC; one of these (the “S2” subtype) is characterized by elevated AFP and aggressive clinical features such as large tumor size, increased proliferation, and poor differentiation. Preclinical and clinical studies also suggest a correlation between elevated AFP levels and high vascular endothelial growth factor (VEGF) expression, suggesting that VEGF pathway inhibitors may be particularly effective for these tumors. Consistent with this mechanism, the monoclonal antibody ramucirumab which targets the VEGF receptor 2 (VEGFR2) isoform demonstrated improved survival in patients with baseline AFP ≥400 ng/mL, though it did not demonstrate a survival benefit in a study population without baseline AFP selection. Cabozantinib is an oral tyrosine kinase inhibitor whose targets include VEGF receptors, MET, and the TAM family of kinases (TYRO3, AXL, MER). Cabozantinib is approved for patients with advanced HCC who have previously been treated with sorafenib, based on outcomes from the pivotal phase 3 CELESTIAL study. In CELESTIAL, cabozantinib significantly prolonged overall survival (OS; hazard ratio [HR], 0.76; 95% confidence interval [CI], 0.63–0.92) and progression-free survival (PFS; HR, 0.44; 95% CI, 0.36–0.52) relative to placebo. Subgroup analyses of OS and PFS favored cabozantinib across subgroups based on patient demographics, clinical characteristics and biomarker levels, including baseline AFP levels ≥400 ng/mL and <400 ng/mL. Here, we describe exploratory analyses of outcomes in the phase 3 CELESTIAL study based on AFP levels at baseline and AFP changes during treatment.
Materials and methods

The study design and methods for CELESTIAL have been previously described. Briefly, 707 patients were randomized between September 2013 through June 2017 in a 2:1 ratio to receive either cabozantinib (60 mg once daily) or placebo. Patients must have received prior sorafenib and could have received up to two prior systemic regimens for HCC. Other key inclusion criteria were Child–Pugh class A liver function and Eastern Cooperative Oncology Group performance status (ECOG PS) of 0 or 1. The primary endpoint was OS; secondary endpoints were PFS and objective response rate (ORR). Tumor response and progression were assessed every 8 weeks by the investigator according to Response Evaluation Criteria in Solid Tumors (RECIST), version 1.1. Serum AFP levels were measured centrally (Covance Inc., Princeton, NJ, USA) at baseline, and every 8 weeks thereafter using a US Food and Drug Administration–approved chemiluminescence assay (Access AFP Immunoassay kit, Beckman Coulter) using a Beckman Coulter DXI 800 Access immunoassay analyzer (normal reference range, 0.4–300,000 ng/mL). Investigators and patients were blinded to treatment arm and to central AFP results.

Baseline AFP

Baseline characteristics and outcomes were analyzed for patients according to the baseline serum AFP level, using a cutoff of 400 ng/mL. This cutoff was based on prior studies demonstrating the prognostic value of this threshold. Outcomes included OS, PFS, tumor response, and safety. Additional analyses of OS and PFS were carried out using cutoffs of 20 ng/mL and 200 ng/mL, based on alternative cutoffs used in the literature.

AFP response

On-treatment AFP response was evaluated at Week 8 (week 9 day 1), which was also the timepoint for the first tumor assessment. AFP response was defined as ≥20% decrease from baseline in serum AFP at Week 8, in patients with baseline AFP ≥20 ng/ml and less than the upper limit of quantitation (300,000 ng/mL). This definition is consistent with previous studies. Outcomes including OS, PFS, tumor response, and safety were assessed according to AFP response. Additional analyses of OS and PFS were carried out in the same group of patients using alternative cutoffs based on review of the literature, including AFP response defined as a ≥50% decrease from baseline, and varying thresholds of AFP progression such as ≥20% or ≥50% increase.

Statistical analysis

Baseline characteristics in baseline AFP subgroups were compared using chi-squared tests in the case of two categorical factors, or ANOVA in the case of a categorical and a continuous factor.
Efficacy analyses included all randomized patients, and safety analyses included all patients who received at least one dose of study drug. OS and PFS were assessed using the Kaplan–Meier method. No adjustments for multiplicity were made for subgroup analyses. Confidence intervals are considered descriptive, and all HRs are unstratified. Survival analyses were adjusted for guarantee-time bias using the landmark method, which excluded patients with an event prior to Week 8.

To determine whether AFP response was independently associated with survival in the cabozantinib group, multivariable analyses were carried out using the Cox proportional hazard regression model to complement univariate analyses. The model also included the following baseline variables: baseline AFP level (<400 or ≥400 ng/mL), ECOG PS (0 or ≥1), macrovascular invasion (MVI; no or yes), extrahepatic spread (no or yes), age (<65 or ≥65 years), gender, and etiology (hepatitis B virus, hepatitis C virus, or other).

For analysis of the optimal AFP response cutoff, maximally selected rank statistics were used to determine the percent change in AFP from baseline to Week 8 that had the most significant association with OS. A rank statistic was calculated at each percent cutoff, and the statistics were then maximized using the method of Hothorn and Lausen.

To characterize the relationship between AFP response and radiographic response, a non-exact Spearman correlation test was performed between AFP percent change from baseline at Week 8 and percent change from baseline in the sum of diameters of target lesions at Week 8.

**Results**

**Patients**

The distribution of baseline AFP levels was similar between the cabozantinib and placebo treatment arms (Figure 1). Median baseline AFP was 154.7 ng/mL (interquartile range [IQR], 14.0–2988.9) for patients in the cabozantinib arm and 202.5 ng/mL (IQR, 10.2–5174.9) for patients in the placebo arm. Baseline characteristics and demographics according to baseline AFP level are shown in Table 1 and were generally balanced between the cabozantinib and placebo arms; however, some differences were noted between subgroups with baseline AFP levels <400 ng/mL versus ≥400 ng/mL. The proportion of patients with hepatitis B virus etiology was 33% and 45% for subgroups with baseline AFP levels <400 ng/mL and ≥400 ng/mL, respectively. A smaller proportion of patients had MVI in the subgroup with an AFP level <400 ng/mL relative to those with ≥400 ng/mL (24% versus 38%).
**Figure 1.** Logarithmic density plot of baseline AFP distribution in A, cabozantinib and B, placebo groups. Density refers to the probability distribution of AFP such that the area under the curve equals 1.

![Logarithmic density plot of baseline AFP distribution](image)

**Table 1.** Baseline characteristics according to baseline AFP

<table>
<thead>
<tr>
<th></th>
<th>AFP &lt;400 ng/mL</th>
<th></th>
<th>AFP ≥400 ng/mL</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Cabozantinib</td>
<td>Placebo</td>
<td>Cabozantinib</td>
<td>Placebo</td>
</tr>
<tr>
<td></td>
<td>(N=278)</td>
<td>(N=136)</td>
<td>(N=192)</td>
<td>(N=101)</td>
</tr>
<tr>
<td>Age,¹ median (range), years</td>
<td>64.0 (35–86)</td>
<td>65.5 (24–86)</td>
<td>64.0 (22–86)</td>
<td>62.0 (34–83)</td>
</tr>
<tr>
<td>Male, n (%)</td>
<td>235 (85)</td>
<td>117 (86)</td>
<td>144 (75)</td>
<td>85 (84)</td>
</tr>
<tr>
<td>Geographic region,b n (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asia</td>
<td>62 (22)</td>
<td>31 (23)</td>
<td>54 (28)</td>
<td>28 (28)</td>
</tr>
<tr>
<td>Europe</td>
<td>140 (50)</td>
<td>65 (48)</td>
<td>91 (47)</td>
<td>43 (43)</td>
</tr>
<tr>
<td>Pacific</td>
<td>10 (4)</td>
<td>6 (4)</td>
<td>5 (3)</td>
<td>5 (5)</td>
</tr>
<tr>
<td>North America</td>
<td>66 (24)</td>
<td>34 (25)</td>
<td>42 (22)</td>
<td>25 (25)</td>
</tr>
<tr>
<td>Race, n (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asian</td>
<td>86 (31)</td>
<td>42 (31)</td>
<td>73 (38)</td>
<td>40 (40)</td>
</tr>
<tr>
<td>White</td>
<td>162 (58)</td>
<td>80 (59)</td>
<td>102 (53)</td>
<td>50 (50)</td>
</tr>
<tr>
<td>Black</td>
<td>6 (2)</td>
<td>7 (5)</td>
<td>2 (1)</td>
<td>4 (4)</td>
</tr>
<tr>
<td>Other or not reported</td>
<td>24 (9)</td>
<td>7 (5)</td>
<td>15 (8)</td>
<td>7 (7)</td>
</tr>
<tr>
<td>ECOG performance status, n (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>147 (53)</td>
<td>85 (63)</td>
<td>98 (51)</td>
<td>46 (46)</td>
</tr>
<tr>
<td>1</td>
<td>131 (47)</td>
<td>51 (38)</td>
<td>93 (48)</td>
<td>55 (54)</td>
</tr>
<tr>
<td>Etiology of disease,² n (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HBV²</td>
<td>95 (34)</td>
<td>41 (30)</td>
<td>83 (43)</td>
<td>48 (48)</td>
</tr>
<tr>
<td>HCV</td>
<td>73 (26)</td>
<td>35 (26)</td>
<td>40 (21)</td>
<td>20 (20)</td>
</tr>
<tr>
<td>Dual HBV and HCV Infection³</td>
<td>8 (3)</td>
<td>3 (2)</td>
<td>0 (1)</td>
<td></td>
</tr>
<tr>
<td>Alcohol</td>
<td>69 (25)</td>
<td>21 (15)</td>
<td>43 (22)</td>
<td>18 (18)</td>
</tr>
<tr>
<td>Nonalcoholic steatohepatitis</td>
<td>27 (10)</td>
<td>15 (11)</td>
<td>16 (8)</td>
<td>8 (8)</td>
</tr>
<tr>
<td>Other or unknown</td>
<td>66 (24)</td>
<td>40 (29)</td>
<td>33 (17)</td>
<td>23 (23)</td>
</tr>
<tr>
<td>Extrahepatic spread of disease and/or macrovascular invasion, n (%)</td>
<td>237 (85)</td>
<td>113 (83)</td>
<td>161 (84)</td>
<td>87 (86)</td>
</tr>
<tr>
<td>Extrahepatic spread of disease</td>
<td>223 (80)</td>
<td>103 (76)</td>
<td>146 (76)</td>
<td>79 (78)</td>
</tr>
<tr>
<td>Macrovascular invasion³</td>
<td>64 (23)</td>
<td>36 (26)</td>
<td>65 (34)</td>
<td>45 (45)</td>
</tr>
</tbody>
</table>
Sites of disease, \( d \) %

<table>
<thead>
<tr>
<th></th>
<th>n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Liver</td>
<td>227 (82)</td>
</tr>
<tr>
<td>Bone</td>
<td>41 (15)</td>
</tr>
<tr>
<td>Visceral (excluding liver)</td>
<td>128 (46)</td>
</tr>
<tr>
<td>Lymph node</td>
<td>90 (32)</td>
</tr>
</tbody>
</table>

Number of prior systemic anticancer regimens for advanced HCC, n (%)

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>201 (72)</td>
<td>134 (70)</td>
</tr>
<tr>
<td>2</td>
<td>73 (26)</td>
<td>57 (30)</td>
</tr>
</tbody>
</table>

Chemoembolization for HCC, n (%)

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>113 (41)</td>
<td>90 (47)</td>
</tr>
</tbody>
</table>

Median total duration of prior sorafenib, months

|                | 1.6 | 1.7 | 1.5 | 1.7 |

Median time from disease progression to randomization, months

|                | 5.2 | 6.6 | 5.4 | 4.0 |

AFP, alpha-fetoprotein; ECOG, Eastern Cooperative Oncology Group; HBV, hepatitis B virus; HCC, hepatocellular carcinoma; HCV, hepatitis C virus.

\( a \) Baseline characteristics which were significantly different between the low and high baseline AFP subgroups of the pooled treatment arms (\( p < 0.05 \)).

\( b \) Asia includes Republic of Korea, Hong Kong, Taiwan, and Singapore. Pacific includes Australia and New Zealand.

\( c \) Etiology per case report form. Some patients had more than one disease etiology category.

\( d \) Investigator-assessed at baseline

Overall, 236 out of 470 (50%) patients in the cabozantinib group and 111 out of 237 (47%) patients in the placebo group were evaluable for AFP response at Week 8. Reasons for lack of evaluable AFP response at Week 8 are listed in Supplementary Table S1; the primary reasons for non-evaluability were baseline AFP <20 ng/mL (139 [30%] patients in the cabozantinib group and 77 [32%] patients in the placebo group) and discontinuation or death before Week 8 (59 [13%] patients in the cabozantinib group and 38 [16%] patients in the placebo group).

**Efficacy outcomes according to baseline AFP**

OS and PFS were improved with cabozantinib relative to placebo in both baseline AFP subgroups (Figure 2). For patients with baseline AFP levels <400 ng/mL, median OS was 13.9 months with cabozantinib versus 10.3 months with placebo (HR, 0.81; 95% CI, 0.62–1.04), and 8.5 months versus 5.2 months (HR, 0.71; 95% CI, 0.54–0.94) for patients with baseline AFP levels ≥400 ng/mL (Figure 2a, 2b). For patients with baseline AFP levels <400 ng/mL, median PFS was 5.5 months with cabozantinib versus 1.9 months with placebo (HR, 0.47; 95% CI, 0.37–0.60), and 3.9 months versus 1.9 months (HR, 0.42; 95% CI, 0.32–0.55) for patients with baseline AFP levels ≥400 ng/mL (Figure 2c, 2d). Subsequent anticancer therapy according to AFP subgroups is shown in Supplementary Table S2. For patients with baseline AFP <400 ng/mL, 26% and 35% of patients in the cabozantinib
and placebo groups went on to receive subsequent therapy; for AFP levels ≥400 ng/mL, this value was 23% for both treatment groups.

For baseline AFP <400 ng/mL, ORR was 5% (95% CI, 2.5–7.9) with cabozantinib and 0.7% (95% CI, 0.0–4.0) with placebo, and for baseline AFP ≥400 ng/mL, ORR was 3% (95% CI, 0.9–6.0) with cabozantinib versus 0% with placebo (Table 2).

**Figure 2.** Outcomes according to baseline AFP. A–B, overall survival and C–D, progression-free survival by baseline AFP.

<table>
<thead>
<tr>
<th>Baseline AFP</th>
<th>AFP responsea</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Baseline AFP</td>
</tr>
<tr>
<td></td>
<td>AFP &lt;400 ng/mL</td>
</tr>
<tr>
<td>ORR (95% CI), %</td>
<td>5.0 (2.5–7.9)</td>
</tr>
<tr>
<td>Best overall response, n (%)</td>
<td></td>
</tr>
<tr>
<td>Partial response</td>
<td>13 (5)</td>
</tr>
<tr>
<td>Stable disease</td>
<td>172 (62)</td>
</tr>
</tbody>
</table>

Table 2. Best overall tumor response
Progressive disease  50 (18)  71 (52)  48 (25)  60 (59)  17 (15)  34 (29)
Not evaluable/missing  43 (15)  10 (7)  29 (15)  17 (17)  1 (1)  2 (2)

AFP, alpha-fetoprotein; CI, confidence interval; ORR, objective response rate.
*≥20% decrease in AFP level from baseline at Week 8 in patients who had baseline AFP levels ≥20 ng/mL.
^All responses were partial responses.

Additional PFS and OS analyses used cutoffs of 20 ng/mL and 200 ng/mL for baseline AFP. The survival benefit with cabozantinib relative to placebo was similar using the alternative cutoffs, with the exception of OS for patients with <20 ng/mL baseline AFP which showed a HR of 0.97 (95% CI 0.67–1.40) (Supplementary Figure S1).

Efficacy outcomes according to AFP response (≥20% decrease from baseline)

Change from baseline in serum AFP for patients in the cabozantinib group and the placebo group at Week 8 are shown in Figure 3. AFP response (defined as ≥20% decrease from baseline) occurred in 50% of evaluable patients in the cabozantinib group compared with 13% in the placebo group.

Owing to the low rate of AFP response in the placebo group, analysis of outcomes by AFP response focused primarily on the cabozantinib group. Baseline characteristics according to AFP response in the cabozantinib arm are shown in Supplementary Table S3 and were generally balanced between subgroups.
In patients evaluable for AFP response (baseline AFP ≥20 ng/ml), OS and PFS were improved in patients who had an AFP response (defined as ≥20% decrease from baseline to Week 8) relative to those with no AFP response, irrespective of treatment. In the cabozantinib group, median OS for patients with an AFP response (N=117) and without an AFP response (N=119) was 16.1 months and 9.1 months (HR, 0.61; 95% CI, 0.45–0.84), while median PFS for these subgroups was 7.3 months and 4.0 months (HR, 0.55; 95% CI, 0.41–0.74) (Figure 4 A–B). For the subgroup of 139 patients (30%) in the cabozantinib group who were not evaluable for response analysis due to baseline AFP <20 ng/mL, median OS and PFS were 14.4 months and 5.6 months, respectively, in the cabozantinib group (Supplementary Figure S1).

The proportion of patients in the cabozantinib group who went on to receive at least one subsequent anticancer therapy was similar for patients with and without an AFP response (28% vs 27%) (Supplementary Table S2). In the placebo group, median OS was 11.3 months for patients with
an AFP response (N=14) and 7.2 months for patients without an AFP response (N=97) (HR, 0.79; 95% CI, 0.41–1.55), while median PFS for these subgroups was 3.8 months and 1.9 months (HR, 0.51; 95% CI, 0.27–0.96).

Landmark analyses of OS and PFS were performed to adjust for guarantee-time bias. For OS, results of the landmark and unadjusted analyses were identical because only patients who were alive at Week 8 were included in the unadjusted analysis. Landmark analysis of PFS at Week 8 was similar to the unadjusted analysis; median PFS for patients with and without an AFP response in the cabozantinib group was 7.4 and 5.4 months from randomization respectively (HR, 0.63; 95% CI, 0.46–0.87).

Additional analyses of the cabozantinib group explored the association of OS and PFS with AFP response (≥20% decrease from baseline to Week 8) in subgroups of patients with low baseline AFP (20 to <400 ng/mL) and high baseline AFP (>400 ng/mL) (Supplementary Figure S2). For both subgroups, outcomes favored patients with an AFP response versus those without an AFP response, with an HR for OS of 0.59 in the low baseline AFP subgroup and 0.69 in the high baseline AFP subgroup, and corresponding HRs for PFS of 0.47 and 0.69, respectively.

ORR for patients with and without an AFP response in the cabozantinib arm was 7% and 3% (Table 2). The rate of progressive disease as best response per RECIST version 1.1 in the AFP response subgroup was approximately half that of the AFP nonresponse subgroup (15% vs 29%).

**Alternative cutoffs for AFP response and AFP control**

Using an alternative cutoff of ≥50% decrease from baseline to define AFP response, HRs for OS and PFS were consistent with those for the ≥20% decrease cutoff (Supplementary Figure S2). In the cabozantinib group, 21 patients experienced an AFP response accompanied by a decrease in AFP level to <20 ng/mL at Week 8, compared with 3 patients in the placebo group. Median OS with cabozantinib was 20.4 months for patients with AFP reduction to <20 ng/mL versus 10.6 months for patients who did not achieve this threshold (HR 0.51, 95% CI 0.29–0.90), and median PFS was 14.6 months versus 5.4 months (HR 0.36, 95% CI 0.21–0.6).

Next, we conducted an exploratory analysis using maximally selected rank statistics to determine the optimal cutoff for percent change in AFP from baseline to Week 8 that provided the strongest association with OS. For patients with baseline AFP ≥20 ng/mL, the optimal cutoff was estimated as 0% change from baseline in AFP (Supplementary Figure S3); this cutoff grouped patients by those who had AFP control at Week 8 (a reduction or no change from baseline) and patients without AFP control (any increase from baseline). Using this cutoff, 61% (144/236) of evaluable patients in the cabozantinib group and 23% (26/111) of patients in the placebo group had AFP control at Week 8.
Median OS with cabozantinib was 17.0 months for patients with AFP control (N=144) and 8.1 months for patients without AFP control (N=92; HR, 0.50; 95% CI, 0.35–0.71; p<0.0001); median PFS was 7.3 and 3.7 months, respectively (HR 0.48; 95% CI, 0.34–0.67) (Figure 4 C–D).

**Figure 4.** Overall survival and progression-free survival in the cabozantinib group by AFP response (defined as ≥20% decrease in AFP level from baseline at Week 8; A, B) and AFP control (defined as reduction or no change from baseline at Week 8; C, D). Evaluable patients were those who had baseline AFP levels ≥20 ng/mL.

We also explored the impact of AFP progression on OS and PFS, with progression defined as an increase of ≥20% or ≥50% from baseline AFP level to Week 8. AFP progression was associated with shorter OS and PFS at both cutoffs, and HRs were similar for both cutoffs (Supplementary Figure S2).

The association of AFP change from baseline with OS and PFS was also evaluated using continuous analysis. Among patients evaluable for AFP response at Week 8 in the cabozantinib group, the percent increase in AFP from baseline was significantly associated with OS (HR, 1.18; 95% CI, 1.03–1.36; p=0.016) and PFS (HR 1.39; 95% CI 1.22–1.58; p<0.0001).
Multivariable analyses

In multivariable analyses, an AFP response defined as a ≥20% decrease from baseline (HR, 0.60; 
P=0.0002) was independently associated with longer OS in the cabozantinib group, as were baseline 
AFP level <400 ng/mL (HR, 0.74; P=0.02), ECOG PS 0 (HR, 0.66; P=0.002) and the absence of MVI (HR, 
0.68; P=0.007). AFP response (HR, 0.56; P=0.0002) and baseline AFP <400 ng/mL (HR, 0.63; P=0.004) 
were also associated with improved PFS in the cabozantinib group (Supplementary Table S4).

AFP response and radiographic response

The relationship between AFP kinetics and radiographic tumor response at Week 8 was investigated. 
The percent change in target lesion sum of diameters was positively correlated with percent change 
in AFP as assessed using a non-exact Spearman correlation test (rho=0.509, p<0.0001 in the pooled 
treatment groups) (Supplementary Figure S4). Among 376 patients in the cabozantinib group 
evaluable for tumor response at Week 8, 22 patients (6%) had a radiographic response, defined as 
≥30% reduction in target lesion sum of diameters from baseline. Using this definition and noting the 
small number of patients meeting the response criteria, radiographic response at Week 8 was not 
significantly associated with OS in the cabozantinib group; median OS from randomization was 16.0 
months for patients with a radiographic response (N=22) and 11.5 months for patients without a 
response (N=354) (HR, 0.64; 95% CI, 0.37-1.12; p=0.12).

Safety

For the subgroup of patients with AFP levels <400 ng/mL, median duration of exposure was 3.9 
months (range, 0.1–37.3) for cabozantinib and 2.1 months (range, 0.1–27.2) for placebo; median 
average daily dose was 35.5 mg for cabozantinib and 59.0 mg for placebo. For those with AFP levels 
≥400 ng/mL, median duration of exposure for cabozantinib and placebo was 3.7 months (range, 0.1– 
26.5) and 1.9 months (range, 0.0–13.5), respectively; median average daily dose was 36.3 mg and 
57.4 mg. The rate of all-cause grade 3/4 adverse events in the cabozantinib and placebo groups was 
70% and 38% for patients with baseline AFP levels <400 ng/mL, and 64% and 35% for patients with 
AFP levels ≥400 ng/mL. The rate of discontinuation due to treatment-related adverse events (TRAE) 
in the cabozantinib and placebo groups was similar for patients with AFP levels <400 ng/mL (15% vs 
3%) and ≥400 ng/mL (18% vs 3%).

Within the cabozantinib treatment arm, patients with an AFP response (≥20% decrease) had a higher 
median duration of exposure to the drug (5.7 months; range, 1.9–37.3) compared with those 
without an AFP response (3.7 months; range, 1.4–22.6); the median average daily dose of 
cabozantinib for these subgroups was 39.2 mg and 33.9 mg, respectively. All-cause grade 3 or 4 
adverse events occurred in 75% of patients with an AFP response and 69% of patients without an
AFP response, and the rate of discontinuation due to TRAEs was 13% for both groups. 

**Supplementary Table S5** lists grade 3/4 AEs occurring at ≥5.0% frequency in either treatment arm in the overall safety population, according to AFP subgroup.

**Discussion**

The phase 3 CELESTIAL study showed an improvement in OS and PFS with cabozantinib relative to placebo in patients with previously treated, advanced HCC. Results of the current analysis are consistent with those of the overall population; cabozantinib improved OS and PFS compared with placebo across a range of baseline AFP levels. On-treatment AFP response (≥20% decrease from baseline in serum AFP) or AFP control (reduction or no change from baseline) at Week 8 was more frequent in the cabozantinib arm versus placebo. In the cabozantinib arm, patients who achieved AFP response or control had improved OS and PFS relative to those who did not, while those with AFP progression had worse outcomes. The safety profile of cabozantinib according to the various AFP subgroups was consistent with that of the primary analysis.

High baseline AFP levels were associated with shorter median OS in both treatment arms, consistent with other phase 3 studies and with high baseline AFP levels as a negative prognostic indicator. Our results are similar to those reported in phase 3 studies of the multikinase inhibitors regorafenib and sorafenib, which showed a survival benefit relative to placebo across baseline AFP subgroups defined by cutoffs of 400 ng/mL and 200 ng/mL, respectively. In contrast, the phase 3 REACH study of the VEGFR2-targeted antibody ramucirumab did not show an OS benefit relative to placebo in the overall patient population, but subgroup analyses showed that patients with high AFP (≥400 ng/mL) had an OS benefit with ramucirumab, while those with low AFP (<400 ng/mL) did not. The ensuing REACH-2 study exclusively enrolled patients with baseline AFP levels ≥400 ng/mL and confirmed the OS benefit of ramucirumab relative to placebo in this patient population, suggesting increased dependence on VEGF pathway signaling in tumors with high AFP expression. Unlike ramucirumab, however, multikinase inhibitors demonstrate efficacy across a range of baseline AFP values suggesting that the inhibition of additional targets may contribute to antitumor activity across a broader range of tumor biology.

The association of on-treatment AFP response or control with improved survival in CELESTIAL is consistent with retrospective analyses of patients treated with targeted therapies including sorafenib, ramucirumab and regorafenib. Conversely, shorter survival in patients whose AFP levels increased during treatment has also been reported. High AFP levels are associated with advanced stages of HCC, and less differentiated, larger tumors; it is likely, therefore, that AFP levels may increase as the disease progresses. On the whole, these data suggest a potential role for
on-treatment serum AFP kinetics as a surrogate endpoint. In HCC, radiographic assessment methods such as RECIST version 1.1, and modified RECIST are challenged by the nodularity and heterogeneity of background cirrhotic liver, scarring or devascularization from prior treatment, or heterogeneity in the timing of injection of contrast dye and subsequent acquisition time. The addition of serum biomarkers to radiographic assessment of tumor response may help to address limitations of imaging in HCC; moreover, serum biomarkers have the potential to provide an early indication of treatment efficacy prior to radiographic assessment, and may be particularly useful for therapies such as TKIs which have a low radiographic response rate. Indeed, the low radiographic response rate at Week 8 in CELESTIAL likely accounted for the lack of a significant association with OS in the cabozantinib group, though median survival was longer for those with a radiographic response (N=22, median OS 16.0 months) versus those without (N=354, median OS 11.5 months).

It is important to note the lack of standardized cutoff values for AFP response, which have varied across studies. Cutoffs of at least 20% and 50% decrease from baseline are commonly used, though these have not been validated and the optimal biologic cutoff has not been established. In the current analysis, HRs for PFS and OS were almost identical with cutoffs of 20% and 50%, prompting further analysis to identify an optimal cutoff value. By using maximally ranked statistics, we estimated the optimal cutoff for change in AFP from baseline to Week 8 was 0% for OS; this essentially categorized on-treatment change in AFP into AFP control (reduction or no change) and no AFP control (any increase from baseline). Patients receiving cabozantinib who achieved AFP control by Week 8 had longer OS and PFS compared to those who did not achieve AFP control, suggesting that AFP control could help to inform treatment decisions.

There are multiple limitations inherent to this study, owing to the exploratory, retrospective nature of the analysis. Approximately 50% of patients in both treatment arms were unevaluable for AFP response; reasons included baseline AFP <20 ng/mL in 30% of patients, consistent with other cohort analyses, or lack of AFP assessment at Week 8 in an additional 20%. Several studies, including ours, have included only patients with ≥20 ng/mL baseline AFP in their response analyses, given that underlying viral hepatitis or other causes of hepatic inflammation may contribute to AFP elevation, particularly at lower levels. Therefore, it is important to note that the utility of AFP kinetics in this setting is presumably limited to patients with baseline AFP levels above the chosen threshold of 20 ng/mL for response analysis, which accounted for 70% of the CELESTIAL study population and represents the majority of patients with advanced HCC. Another important consideration is that AFP kinetics may be dependent upon therapeutic mechanism of action. AFP response, control, and progression kinetics warrant examination for association with clinical outcomes in patients treated with other systemic therapies, including immune checkpoint inhibition and combinations thereof.
A strength of this study is the independent, centralized testing of AFP using a standardized assay, without reporting back to investigators who were blinded to treatment arm as well as AFP results. Treatment decisions and response assessments therefore were performed independently of central AFP values, further strengthening the observed associations with clinical response.

Prospective studies evaluating changes in AFP on treatment as a surrogate endpoint for efficacy outcomes are lacking, and rigorous validation studies are needed. Future studies should seek to prospectively analyze AFP kinetics in large, randomized studies according to type of treatment, as has been done with biomarkers in other tumor types such as prostate cancer where the kinetics of prostate-specific antigen have an integral role in response assessment and treatment decisions.\textsuperscript{42} Such studies should be adequately powered for evaluation of appropriate AFP cutoffs for response and progression, as well as baseline threshold for evaluability.

In conclusion, our analysis shows improved outcomes with cabozantinib relative to placebo in patients with previously treated, advanced HCC across a range of baseline AFP levels. The on-treatment AFP response and control rates was higher with cabozantinib than with placebo, while the rate of AFP progression was higher for placebo-treated patients. In the cabozantinib group, AFP response and AFP control were associated with longer OS and PFS. Given the rapidly expanding treatment landscape in HCC, further investigation of AFP kinetics in patients treated with newly available therapies is warranted.

Acknowledgements

We thank the patients, their families, the investigators and site staff, and the study teams who participated in the CELESTIAL trial. This study was funded by Exelixis, Inc. Medical writing support was provided by Karen O’Leary and Michael Raffin of Fishawack Communications, Inc. (Conshohocken, PA, USA) and funded by Exelixis, Inc.

Author contributions

Conception and design: RKK, TM, ALC, ABE

Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): RKK, TM, LR, PM, JWP, TY, SLC, JFB, VCT, AT, VD, ALC, ABEK, GKAA

Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): RKK, DWM, RK

Writing, review, and/or revision of the manuscript: All authors

Study supervision: GKAA, RKK, TM, ALC, ABEK
References


