Interobserver Agreement Rates on Fibroblast Activation Protein Inhibitor–Directed Molecular Imaging and Therapy

Sebastian E. Serfling, MD,* Philipp E. Hartrampf, MD,* Yingjun Zhi, MD,† Takahiro Higuchi, MD, PhD,*§ Steven P. Rowe, MD, PhD,§ Lena Bundschuh, MSc,|| Markus Essler, MD,‡ Andreas K. Buck, MD,* Ralph Alexander Bundschuh, MD, PhD,|| and Rudolf A. Werner, MD*§

Objectives: Fibroblast activation protein (FAP) has emerged as a novel target for FAP inhibitor (FAPI)-directed molecular imaging and endoradiotherapy (ERT). We aimed to assess the interobserver agreement rates for interpretation of 68Ga-FAPI-4 PET/CT and decision for ERT.

Patients and Methods: A random order of 68Ga-FAPI-4 PET/CTs from 49 oncology patients were independent interpreted by 4 blinded readers. Per scan, visual assessment was performed, including overall scan impression, number of organ/lymph node (LN) metastases, and number of affected organs/LN regions. Moreover, a maximum of 3 target lesions, defined as largest in size and/or most intense, per organ compartment were identified, which allowed for an additional quantitative interobserver assessment of LN and organ lesions. To investigate potential reference tissues, quantification also included unaffected liver parenchyma and blood pool. Readers also had to indicate whether FAPI-directed ERT should be considered (based on intensity of uptake and widespread disease). Interobserver agreement rates were evaluated using intraclass correlation coefficients (ICC) and interpreted according to Cicchetti (0.4–0.59 indicating fair, 0.6–0.74 good agreement).

Results: On a visual basis, the agreement rate for an overall scan impression was fair (ICC, 0.42–0.79 CI, 0.27–0.57). The concordance rate for number of affected LN areas was also fair (ICC, 0.59–95 CI, 0.45–0.72), whereas the number of LN metastases, number of affected organs, and number of organ metastases achieved good agreement rates (ICC, ≥0.63). In a quantitative analysis, concordance rates for LN were good (ICC, 0.70; 0.45–0.88), but only fair for organ lesions (ICC, 0.43; 0.26–0.60). Concerning background tissues, ICCs were good for unaffected liver parenchyma (0.68; 0.54–0.79) and fair for blood pool (0.43; 0.29–0.58). When readers should decide on ERT, concordance rates were also fair (ICC, 0.60). In regards to background tissues, ICCs were good for unaffected liver parenchyma and blood pool. Readers also had to indicate whether FAPI-directed ERT should be considered (based on intensity of uptake and widespread disease). Interobserver agreement rates were evaluated using intraclass correlation coefficients (ICC).

Conclusions: For FAPI-directed molecular imaging and therapy, a fair to good interobserver agreement rate was achieved, supporting the adoption of this radiotracer for clinical routine and multicenter trials.

Key Words: fibroblast activation protein inhibitor, FAPI, molecular imaging, theranostics, endoradiotherapy, RADS

Overexpressed on cancer-associated fibroblasts, fibroblast activation protein (FAP) has been identified as a suitable target for molecular imaging,1 providing high contrast PET images along with reduced background accumulation when compared with the current reference standard FDG.2–5 Of note, FAP inhibitors (FAPIs) can also be labeled with 18F-emitters, and such theranostic approaches have demonstrated low toxicity profiles and increased radiotracer retention after administration of the therapeutic compound.6–8 FAPI-directed PET is needed to identify patients eligible for FAPI-targeted endoradiotherapy (ERT), which further emphasizes the importance of accurate scan interpretation.4–6 In this regard, a recent study reported on various diagnostic pitfalls, with false-positives due to scarring or degenerative lesions.7 Thus, similar to somatostatin receptor (SSTR) or prostate-specific membrane antigen (PSMA)–directed molecular imaging, the reliability of scan interpretation for FAPI-directed PET must be proven, which may allow for more widespread adoption in trials or in the clinic.8,9 In this study, we aimed to elucidate the interobserver agreement rates for interpretation of FAPI-directed PET scans on a visual and quantitative level for sites of disease, as well as unaffected organs for reference. In addition, we aimed to determine PET-based concordance rates for identifying patients eligible for ERT.

PATIENTS AND METHODS

Patient Population

In total, 49 patients afflicted with oncological disease were retrospectively evaluated. Table 1 displays patient’s characteristics. Parts of this cohort have been investigated in previous articles,10,11 although those studies did not assess interobserver agreement rates. All patients had signed written informed consent for conducted procedures. The local ethics committee waived the need for further approval due to the retrospective nature of this investigation (number 20210415 02).

FAPI-Directed PET/CT

PET/CTs were conducted using a Siemens Biograph mCT 64 or 128 (Siemens Healthineers, Erlangen, Germany). After administration of 136 ± 24 MBq 68Ga-FAPI-4 and an uptake time of 60 minutes, a full-body protocol (top of the skull to knees) was used. Helical CTs with or without intravenous contrast were acquired. PET emission data were then obtained using a 128 × 128 matrix in 3-dimensional (2–3 minutes emission time per bed position) and reconstructed

Received for publication November 29, 2021; revision accepted February 9, 2022. From the *Department of Nuclear Medicine, University Hospital Würzburg; †Department of Otorhinolaryngology, Plastic, Aesthetic and Reconstructive Head and Neck Surgery, Julius Maximilian University of Würzburg, Würzburg, Germany; ‡Faculty of Medicine, Dentistry and Pharmaceutical Sciences, Okayama University, Okayama, Japan; §The Russell H Morgan Department of Radiology and Radiological Science, Johns Hopkins School of Medicine, Baltimore, MD; and ||Department of Nuclear Medicine, University Hospital Bonn, Bonn, Germany. R.A.B. and R.A.W. equally contributed. This work was supported by the Okayama University “RECTOR” Program (T.H.). A KAKENHI grant (21K19450) has been provided for T.H. from the Japan Society for the Promotion of Science. Conflict of interest and sources of funding: none declared. Correspondence to: Ralph Alexander Bundschuh, Department of Nuclear Medicine, University Hospital Bonn, Venusberg-Campus 1, 53127 Bonn, Germany. E-mail: ralph.bundschuh@ukbonn.de. Copyright © 2022 The Author(s). Published by Wolters Kluwer Health, Inc. Copyright © 2022 The Author(s). Published by Wolters Kluwer Health, Inc. This is an open-access article distributed under the terms of the Creative Commons Attribution-Non Commercial-No Derivatives License 4.0 (CCBY-NC-ND), where it is permissible to download and share the work provided it is properly cited. The work cannot be changed in any way or used commercially without permission from the journal. ISSN: 0363-9762/22/4706-0512 DOI: 10.1097/RLU.0000000000004189
TABLE 1. Patients Characteristics

|                          | No. (%)                  |
|--------------------------|--------------------------|
| Female                   | 18/49 (36.7)             |
| Age, y                   | 63.9 ± 11.6              |
| Referred for             |                          |
| Staging                  | 30/49 (61.2)             |
| Restaging                | 19/49 (38.8)             |
| Diagnosis                |                          |
| Oral cavity tumor        | 15/49 (30.6)             |
| Head and neck cancer     | 10/49 (20.4)             |
| Pancreatic cancer        | 8/49 (16.3)              |
| Hepatocellular carcinoma | 6/49 (12.2)              |
| Neuroendocrine neoplasia | 3/49 (6.1)               |
| Lung carcinoma           | 2/49 (4.1)               |
| Sarcoma, adrenal carcinoma, colon carcinoma, GIST, SIFT | 1/49 (2), each |
| Previous therapies       |                          |
| Surgery                  | 16/49 (32.6)             |
| Chemotherapy             | 9/49 (18.4)              |
| Radiation therapy        | 2/49 (4.1)               |

For age, mean ± standard deviation is displayed. Percentages are indicated in paren.

GIST, gastrointestinal stromal tumor; SIFT, solitary fibrosis tumor.

iteratively as provided by the producer (Siemens Esoft; Siemens Healthineers, Erlangen, Germany). 11

Scan Interpretation

PET/CTs were analyzed using syngo.via (VB50; Siemens Healthcare Erlangen, Germany). Four readers (minimum 3 years of experience; range, 3–4 years) analyzed all scans independently and were blinded to the clinical status except for diagnosis, sex, age, and previous therapies. All readers also underwent a training session to gain experience with the workstation before scan interpretation.

Visual Interpretation

As described in 12, overall scan impression was assessed in a binary fashion. Scans were rated as positive when uptakes in disease sites were above background level. In addition, the number of organs affected, the number of organ metastases, the number of affected lymph node (LN) areas, and the number of LN metastases were indicated on a 5-point scale (ranging from 1 to 2–5 for each parameter). 12 FAPI uptake density was evaluated using a 4-point scale (none, 0; low, 1; intermediate, 2; or high, 3). 12 In this regard, 0 was defined as having equivalent uptake when compared with unaffected background tissue (eg, in the muscle, unaffected liver, or blood pool), whereas 3 indicated most intense uptake relative to background. In addition, all readers also had to decide whether FAPI-directed ERT should be considered, either based on intensity of uptake or widespread disease (WD).

Quantitative Assessment

Readers selected a maximum of 5 target lesions (TLs; defined as having the highest uptake and/or the largest in size on PET), using the predefined maximum intensity threshold by the software. No more than 3 lesions per organ compartment could be chosen (primary, LN, lung, skeleton, liver, and soft tissue). Target lesions not allocated to LN were subsumed under organ lesions (OLs). Volumes of interest were placed over all TL to assess the SUVmax. Target lesions identified by all readers were then used for further analysis. In addition, to investigate their potential as reference tissue for background corrections, additional volumes of interest were placed on unaffected liver tissue and blood pool (left ventricle, carefully avoiding the ventricular wall), also providing SUVmax.

Statistical Analysis

For evaluation of agreement rates, we calculated intraclass correlation coefficients (ICCs; including 95% confidence intervals [CIs]), by applying a mean-rating, single-measure, consistency model, and average Maxwell’s intrarater agreement. 12,13 The concordance rates were interpreted according to Cicchetti (poor agreement, ICC <0.4; fair, 0.4–0.59; good, 0.6–0.74; excellent, 0.75–1). 14 For statistical analysis, the following software tools were used: MedCalc statistical software (version 18.2.1; Med-Calc Software bvba) and R (version 3.6.1; R Core Team, 2019) with package irr (version 0.84.1) and boot (version 1.3–23).

RESULTS

Interobserver Agreement for Visual Scan Interpretation

On a visual assessment, the ICC was fair for an overall scan impression (0.42 [0.27–0.57]). For the number of affected organs (ICC, 0.63 [0.49–0.75]), number of organ metastases (ICC, 0.74 [0.64–0.90]), and number of LN metastases (ICC, 0.74 [0.64–0.83]), good interobserver agreement rates were achieved, with the last 2 items just missing excellent agreement. For the number of affected LN areas, interrater agreement was fair (ICC, 0.59 [0.45–0.72]; Table 2; Figure 1A).

Interobserver Agreement for Quantitative Assessment

Among all readers, 511 TLs were recorded as follows: LN (149/511 [29.2%]), primary (123/511 [24.1%]), liver (103/511 [20.2%]), skeleton (79/511 [15.5%]), soft tissue (31/511 [6%]), and lung (26/511 [5.1%]). The identical TLs were recorded by all observers in 208/511 (40.7%) of the cases. For SUVmax derived from LN, the agreement rate was good (ICC, 0.70 [0.48–0.88]) and fair for the remaining OL (ICC, 0.43 [0.26–0.60]). Comparable results were recorded for the reference organs, with unaffected liver achieving good (ICC, 0.68 [0.54–0.79]) and blood pool fair concordance rates (ICC, 0.43 [0.29–0.58]; Table 3; Figure 1B).

Decision for FAPI-Directed ERT

The agreement rate for ERT was fair for intensity (0.50 [0.37–0.65]) or WD alone (0.54 [0.41–0.69]) and demonstrated slightly higher ICC when both conditions were applicable (0.59 [0.46–0.73]; Table 4; Figure 1C), thereby just missing good concordance rates. These findings were further confirmed by a fair agreement rate for FAPI density on PET (ICC, 0.46 [0.31–0.61]; Table 2).

Decision for FAPI-Directed ERT

The agreement rate for ERT was fair for intensity (0.50 [0.37–0.65]) or WD alone (0.54 [0.41–0.69]) and demonstrated slightly higher ICC when both conditions were applicable (0.59 [0.46–0.73]; Table 4; Figure 1C), thereby just missing good concordance rates. These findings were further confirmed by a fair agreement rate for FAPI density on PET (ICC, 0.46 [0.31–0.61]; Table 2). Figure 2A shows a case where all readers agreed on conducting

TABLE 2. Overview of ICC for Visual Assessment

| Parameter                  | ICC            |
|----------------------------|----------------|
| Overall scan impression    | 0.42 (0.27–0.57) |
| Uptake density             | 0.46 (0.31–0.61) |
| No. affected organs        | 0.63 (0.49–0.75) |
| No. organ metastases       | 0.74 (0.64–0.90) |
| No. affected LN areas      | 0.59 (0.45–0.72) |
| No. LN metastases          | 0.74 (0.64–0.83) |

95% CIs are given in parens.
In recent years, FAPI-directed molecular imaging has been applied in multiple types of cancer, demonstrating that this imaging agent provides high contrast images, which can even outperform the current reference radiotracer, 18F-FDG. In addition, FAPI PET has also provided clinical value, for example, by triggering a TNM-based upstaging in more than half of cases or assisting in volumetric segmentation for radiotherapy planning. Nonetheless, interpretation of 68Ga-FAPI-4 PET/CT has multiple pitfalls, leading to false-positives in degenerative lesions, scarring, or in the uterus. As such, before an increased use of this PET agent in daily routine, a high interobserver rate should be proven. We observed an ICC of 0.74 for both LN and visceral lesions (Fig. 1A), thereby just missing excellent agreement rates. This is in line with previous investigations on 68Ga-labeled radiotracers for SSTR- or PSMA-directed imaging, which had comparable or slightly higher ICCs for LN or distant organ metastases (SSTR, ≥0.77; PSMA, ≥0.74). As a possible explanation for those discrepant ICC values, such previous studies focused on more homogeneous cohorts of enrolled patients, for example, gastrointestinal neuroendocrine neoplasms or prostate cancer, respectively. This is in contrast to our study, which included randomly chosen patients with different tumor entities. As such, the presented findings emphasize the potential of FAPI as a pantumor radiotracer, as multiple observers still have a fair to good agreement rate, even in difficult scenarios with a large variety of tumor types. Nonetheless, the acceptable concordance rate may be of importance for future studies using FAPI-directed PET/CT, standardized frameworks providing reliable methods for scan interpretation may further increase ICCs, preferably established in a consensus setting among a panel of experts in the field.

In addition to a visual assessment, we also investigated the interobserver repeatability for quantification of sites of disease. Again, visceral and LN lesions demonstrated fair to good interobserver agreement rates, in particular with LN-derived SUV. The observed variation among readers may be explained by intratumor lesion variability of FAP expression, as such a heterogeneity of fibroblast activation within a metastasis has also been observed in immunohistochemical analyses. Nonetheless, the acceptable concordance rate may be of importance for future studies using FAPI-directed PET/CT for identifying high-risk patients prone to early progressive disease under antitumor-specific therapies or to use the changes in

### TABLE 3. Overview of ICC for Investigated Quantitative Parameters

| Compartment          | ICC       |
|-----------------------|-----------|
| Organ lesions*        | 0.43 (0.26-0.60) |
| LN lesions            | 0.70 (0.48-0.88) |
| Blood pool            | 0.43 (0.29-0.58) |
| Unaffected liver      | 0.68 (0.54-0.79) |

*Includes target lesions of the primary, lung, skeleton, liver, and soft tissue. 95% CIs are given in parentheses.
uptake between baseline and follow-up scans to assess efficacy of various therapeutics. In addition to its potential use as a response assessment tool, quantification may be also of importance for treatment planning, for example, in radiation oncology or to assess the target retention for antifibrotic treatments. As such, given the ICCs we observed, the reader may have certainty that SUVs will not substantially vary among multiple readers for LN or distant organ metastases (Table 3). In addition, recent reports also demonstrated that tumor-to-background ratios (TBRs) on FAPI-PETs are significantly higher relative to TBR derived from 18F-FDG, for example, by using the SUV\textsubscript{max} of the liver or blood pool. As such, SUV\textsubscript{max} from those reference tissues were also investigated in the present analysis, yielding a higher ICC for unaffected liver parenchyma when compared with blood pool derived from the left ventricle. As a possible explanation, a recent study revealed a tight link between radiotracer accumulation in the myocardium and risk factors for cardiovascular disease in a large cohort of patients with cancer. Similar results of a large variety of uptake in oncology patients scanned with FAPI-PET have also been reported for the vasculature.

Several limitations have to be considered. Our findings should be confirmed in larger, prospective trials, preferably including more readers with different levels of experience for PET/CT interpretation. Moreover, interobserver agreement rates for FAPI-directed imaging and decision on ERT may have been biased by patients with low FAPI-avid tumor findings, but TL-based assessment still provided a high number of lesions. Nonetheless, the present study aimed to mimic a real-world scenario of whom not every patient is eligible for radiolabeled therapy or does per se have a high uptake in every

| Parameter | ICC |
|-----------|-----|
| Intensity | 0.50 (0.37–0.65) |
| WD | 0.54 (0.41–0.69) |
| Intensity + WD | 0.59 (0.46–0.73) |

95% CIs are given in parentheses.

In theranostic scenarios, recent reports demonstrated relatively long tumor retention on posttherapeutic scans, which may lead to an increased use of FAPI-targeted ERT. Of note, regardless if the β-emitter 90Y, 177Lu, or 153Sm was used, patient selection was primarily conducted with 68Ga-labeled FAPI PET/CT. In our analysis, the ICCs for FAPI-directed ERT were fair among all readers. Nonetheless, relative to concordance rates for visual and quantitative assessments, ICCs were slightly lower for deciding on FAPI-directed treatment (Fig. 1C). Fendler and coworkers also investigated the observer agreement for SSTR-directed imaging and decision for ERT in neuroendocrine neoplasms. Similar to results of our study, a larger variety of uptake in oncology patients scanned with FAPI-PET should then not only include recommendations on image interpretation, but also for selecting candidates for treatment.

Several limitations have to be considered. Our findings should be confirmed in larger, prospective trials, preferably including more readers with different levels of experience for PET/CT interpretation. Moreover, interobserver agreement rates for FAPI-directed imaging and decision on ERT may have been biased by patients with low FAPI-avid tumor findings, but TL-based assessment still provided a high number of lesions. Nonetheless, the present study aimed to mimic a real-world scenario of whom not every patient is eligible for radiolabeled therapy or does per se have a high uptake in every

![FIGURE 2. Patients imaged with 68Ga-FAPI-4 that were evaluated for ERT. A, Patient with diagnosis of a mixed neuroendocrine nonneuroendocrine neoplasm after resection of the primary in the duodenum. MIP, transaxial PET, CT, and PET/CT revealed intensive 68Ga-FAPI-4 uptake in multiple liver lesions. All readers considered this patient eligible for ERT, based on both intensity and WD. B, Patient with a carcinoma of the oropharynx. MP, transaxial PET, CT, and PET/CT revealed a FAPI-avid LN metastasis in the left cervical region. All readers agreed that FAPI-directed ERT would be feasible based on intensity of uptake, but not on WD.](https://www.nuclearmed.com)
site of disease. Last, clinical applications of FAPI-directed molecular imaging have been recently expanded, for example, to image systemic sclerosis,13 and thus, future studies may also investigate the interobserver agreement if FAPI-directed PET is used to image nonmalignant diseases.

CONCLUSIONS
Evaluating the interobserver rate for FAPI-directed imaging and therapy in a cohort with various tumor entities, a fair to good concordance rate was achieved on both visual and quantitative levels, including sites of disease and reference tissues. As such, this radiotracer may be appropriate for larger multicenter trials, for example, to select treatment candidates for FAPI-directed ERT.

ACKNOWLEDGMENTS
The authors thank Dr D. Muegge, Adelebsen for his statistical assistance.

REFERENCES
1. Galvani M, Bugiardi R, Ferrini D, et al. The relation between myocardial oxygen consumption and ischemia during exertion in patients with angina pectoris: effect of propranolol [in Italian]. G Ital Cardiol. 1985;15:1032–1038.
2. Chen H, Pang Y, Wu J, et al. Comparison of [18F]FDG PET/CT and [18F]FDG PET/CT for the diagnosis of primary and metastatic lesions in patients with various types of cancer. Eur J Nucl Med Mol Imaging. 2020;47:1820–1832.
3. Giesel FL, Kratochwil C, Schlimmer J, et al. Head-to-head intra-individual comparison of biodistribution and tumor uptake of [68Ga]FAPI and [18F]FDG PET/CT in cancer patients. Eur J Nucl Med Mol Imaging. 2021;48:4377–4385.
4. Ferdinandus J, Fragoso Costa P, Kessler L, et al. Initial clinical experience with [68Ga]-FAPI-46 radioligand therapy for advanced stage solid tumors: a case series of nine patients. J Nucl Med. 2021.
5. Baum RP, Schuchardt C, Singh A, et al. Feasibility, biodistribution, and preliminary dosimetry in peptide-targeted radionuclide therapy of diverse adenocarcinomas using [177Lu]-FAPI-228: first-in-humans results. J Nucl Med. 2022;63:415–423.
6. Pindil K, Schlimmer J, Staadinger F, et al. The role of fibroblast activation protein ligands in oncologic PET imaging. PET Clin. 2021;6:341–351.
7. Kessler L, Ferdinandus J, Hirmas N, et al. Pitfalls and common findings in [68Ga]-FAPI-PET—a pictorial analysis. J Nucl Med. 2021.
8. Fendler WP, Barrio M, Spick C, et al. [68Ga-DOTATATE PET/CT: interobserver agreement for prostate cancer assessment: an international multicenter prospective study. J Nucl Med. 2017;58:1617–1623.
9. Werner RA, Bundschuh RA, Bundschuh L, et al. Molecular imaging reporting and data systems (MI-RADS): a generalizable framework for targeted radiotracers with theranostic implications. Ann Nucl Med. 2018;32:512–522.
10. Röhrich M, Naumann P, Giesel FL, et al. Impact of [68Ga]-FAPI-PET/CT imaging on the therapeutic management of primary and recurrent pancreatic ductal adenocarcinomas. J Nucl Med. 2021;62:779–786.
11. Röhrich M, Syed M, Liew DP, et al. [68Ga]-FAPI-PET/CT improves diagnostic staging and radiotherapy planning of adenoid cystic carcinomas—imaging analysis and histological validation. Radiother Oncol. 2021;160:192–201.
12. Fendler WP, Calais J, Allen-Auerbach M, et al. [68Ga-PSMA-11 PET/CT: interobserver agreement for prostate cancer assessments: an international multicenter prospective study. J Nucl Med. 2017;58:1617–1623.
13. Maxwell AE. Comparing the classification of subjects by two independent judges. Br J Psychiatry. 1970;116:651–655.
14. Cicchetti DV. Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. Psychol Assess. 1994;6:284–290.
15. R Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2019. Available at: https://www.R-project.org/. Accessed November 1, 2021.
16. Gamer M, Lemon J, Singh IFP. irr: Various Coefficients of Interrater Reliability and Agreement. 2019. R package version 0.84.1. Available at: https://CRAN.R-project.org/package=irr. Accessed November 1, 2021.
17. Canty A, Ripley B. Bootstrap R (S-Plus) Functions. R package version 1.3-23. 2019. Available at: https://cran.r-project.org/web/packages/boot/citation.html. Accessed November 1, 2021.
18. Serfling E, Fleischig P, Lindner T, et al. (68Ga)-FAPI-PET/CT: tracer uptake in 28 different kinds of cancer. J Nucl Med. 2019;60:801–805.
19. Fendler WP, Calais J, Allen-Auerbach M, et al. [68Ga]-PSMA-11 PET/CT: interobserver agreement for prostate cancer assessments: an international multicenter prospective study. J Nucl Med. 2017;58:1617–1623.
20. Werner RA, Bundschuh RA, Bundschuh L, et al. Molecular imaging and standardized assessment instruments in psychology.