Round Robin Evaluation of MET Protein Expression in Lung Adenocarcinomas Improves Interobserver Concordance

Theresa A. Boyle, H. Lee Moffitt Cancer Center & Research Institute
Farah K. Khalil, University of South Florida
Mari Mino-Kenudson, Massachusetts General Hospital
Gabriel Sica, Emory University
Andre L. Moreira, New York University
Lynette M. Sholl, Brigham and Women’s Hospital
Mirna Z. Knight, Metropolitan Pathologists
Liping Zhang, Tongji University
James Saller, H. Lee Moffitt Cancer Center & Research Institute
Marileila Varella-Garcia, University of Colorado

Only first 10 authors above; see publication for full author list.

Journal Title: APPLIED IMMUNOHISTOCHEMISTRY & MOLECULAR MORPHOLOGY
Volume: Volume 28, Number 9
Publisher: Lippincott, Williams & Wilkins | 2020-10-01, Pages 669-677
Type of Work: Article | Final Publisher PDF
Publisher DOI: 10.1097/PAI.0000000000000810
Permanent URL: https://pid.emory.edu/ark:/25593/vssv8

Final published version: http://dx.doi.org/10.1097/PAI.0000000000000810

Copyright information:

© 2019 The Author(s). Published by Wolters Kluwer Health, Inc.

This is an Open Access work distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License (https://creativecommons.org/licenses/by-nc-nd/4.0/).

Accessed December 30, 2023 10:57 PM EST
Round Robin Evaluation of MET Protein Expression in Lung Adenocarcinomas Improves Interobserver Concordance

Theresa A. Boyle, MD, PhD,*** Farah K. Khalil, MD,‡‡ Mari Mino-Kenudson, MD,§ Gabriel L. Sica, MD,|| Andre L. Moreira, MD,¶ Lynette M. Sholl, MD,# Mirna Z. Knight, DO,‡‡‡†‡ Ling Zhang, MD,‡‡‡ James Saller, MD,‡† Mareilea Varella-Garcia, PhD,‡‡‡§§ Lynne D. Berry, PhD,¶¶¶ Heidi Chen, PhD,||||
Kim E. Ellison, MS,** Christopher J. Rivard, PhD,** Kelly Kugler, BA,¶¶
Igncio I. Wistuba, MD, PhD,### Junya Fujimoto, MD, PhD,##
David J. Kwiatkowski, MD, PhD,*** Paul A. Bunn, MD,††† Mark G. Kris, MD,‡‡‡‡
Eric B. Haura, MD,† and Fred R. Hirsch, MD, PhD,**
for Lung Cancer Mutation Consortium (LCMC) Pathologist Panel Group

Introduction: Overexpression of the mesenchymal-epithelial transition (MET) receptor, a receptor tyrosine kinase, can propel the growth of cancer cells and portends poor prognoses for patients with lung cancer. Evaluation of MET by immunohistochemistry is challenging, with MET protein overexpression varying from 20% to 80% between lung cancer cohorts. Clinical trials using MET protein expression to select patients have also reported a wide range of positivity rates and outcomes.

Materials and Methods: To overcome this variability, the Lung Cancer Mutation Consortium Pathologist Panel endeavored to standardize the evaluation of MET protein expression with “Round Robin” conferences. This panel used randomly selected Aperio-scanned formalin-fixed paraffin-embedded lung cancer specimens stained by MET immunohistochemistry for the Lung Cancer Mutation Consortium 2.0 study (N = 838). Seven pathologists in separate laboratories scored images of 5 initial cases and 2 subsequent rounds of 39 cases. The pathologists’ scores were compared for consistency using the intraclass correlation coefficient. Issues affecting reproducibility were discussed in Round Robin conferences between rounds, and steps were taken to improve scoring consistency, such as sharing reference materials and example images.

Results: The overall group intraclass correlation coefficient comparing the consistency of scoring improved from 0.50 (95% confidence interval, 0.37-0.64) for the first scoring round to 0.74 (95% confidence interval, 0.64-0.83) for the second round.

Discussion: We found that the consistency of MET immunohistochemistry scoring is improved by continuous training and communication between pathologists.

Key Words: mesenchymal-epithelial transition (MET) receptor, histology score (H-score), reproducibility, immunohistochemistry, lung cancer

(Appl Immunohistochem Mol Morphol 2020;28:669-677)

The mesenchymal-epithelial transition (MET) proto-oncogene on chromosome 7q21-31 encodes MET, a transmembrane tyrosine kinase receptor. After MET is activated by its ligand, hepatocyte growth factor (HGF), it initiates signaling through several downstream pathways to promote cellular proliferation, differentiation, angiogenesis, and survival. In lung cancers, increased MET or HGF expression can drive tumorigenesis and are associated with...
poor prognoses.\textsuperscript{1–4} Cigarette smoke is a factor that promotes c-MET addiction in lung cancer.\textsuperscript{5}

\textit{MET} genomic alterations, such as \textit{MET} amplification and/or \textit{MET} gene exon 14 (\textit{MET}ex14) splice-site mutations, can increase \textit{MET} protein expression.\textsuperscript{6–9} Investigational efforts continue to provide additional insights regarding diagnostic, prognostic, and therapeutic implications for \textit{MET} genomic alterations. These studies have refined our understanding that \textit{MET} genomic alterations are enriched in pulmonary sarcomatoid carcinomas.\textsuperscript{10} \textit{MET}ex14 splice-site mutations are also reported in other lung neoplasms, brain gliomas and tumors of unknown primary origin.\textsuperscript{11}

\textit{MET} protein expression may also be upregulated as a resistance mechanism in response to tyrosine kinase inhibitor therapy for epidermal growth factor receptor–(\textit{EGFR})–mutated lung adenocarcinomas.\textsuperscript{11,12} \textit{ERBB3} signaling resultant from \textit{MET} amplification is described as an underlying mechanism attributed to gefitinib resistance.\textsuperscript{13} Dual blockade of \textit{EGFR} and \textit{MET} may be a therapeutic strategy for tumors that harbor both an \textit{EGFR} mutation and \textit{MET} gene amplification, mutation, and/or increased expression or signaling.\textsuperscript{12–14}

Several \textit{MET} inhibitors, including monoclonal antibodies and small-molecule inhibitors, are under investigation in lung cancer clinical trials. Individuals whose tumors harbor a high \textit{MET} gene copy number and/or amplification and/or \textit{MET}ex14 skipping mutations have demonstrated therapeutic response to tyrosine kinase inhibitors such as crizotinib,\textsuperscript{6,9,15,16} cabozantinib,\textsuperscript{5,17} Onartuzumab (MetMab), a humanized monovalent monoclonal antibody that blocks HGF binding to \textit{MET}, was associated with improved survival in patients with advanced lung cancers and high \textit{MET} expression by immunohistochemistry (IHC) in a phase II trial, but subsequently failed to show benefit in a phase III study.\textsuperscript{18–20} Tivantinib (ARQ 197), a small molecule inhibitor of \textit{MET}, was studied in phase III clinical trial, but the overall trial failed to show benefit.\textsuperscript{21} However, subgroup analysis demonstrated that tivantinib improved survival in patients with \textit{MET} protein overexpression.\textsuperscript{21}

Thus, despite some setbacks, \textit{MET} inhibition has produced benefit in molecularly defined subsets of individuals with lung cancer, such as those with \textit{MET} protein overexpression (tivantinib trial subgroup analysis), high \textit{MET} gene amplification, and \textit{MET}ex14 skipping (case report collection and promising clinical trial data). \textit{MET} inhibition continues to be an area of investigation in lung cancers of all types.\textsuperscript{22–24} Crizotinib, cabozantinib, capmatinib, tepotinib, glesatinib, merestinib, and savolitinib are all in phase II trials as \textit{MET} inhibitors.\textsuperscript{25} Pilotumumab and ficituzumab are under investigation as monoclonal antibodies that target the \textit{MET} ligand, HGF.\textsuperscript{24,25}

These studies have highlighted the complexities with \textit{MET} as a potential biomarker. Analysis of only \textit{MET} gene amplification by fluorescence in-situ hybridization misses some patients with \textit{MET}ex14 splice-site mutations and/or protein overexpression who might respond to therapy and vice versa.\textsuperscript{26} To add to the complexity, increased HGF ligand can over-activate \textit{MET} signaling and may decrease \textit{MET} receptor protein due to feedback inhibition; yet, these patients may still respond to monoclonal antibodies that target HGF.\textsuperscript{27}

\textit{MET} protein demonstrates variable expression in the normal lung with heterogeneous expression in most lung tumors.\textsuperscript{28} The staining, scoring, and positivity cutoffs for \textit{MET} are not standardized and there has been a wide range of reported positivity rates for \textit{MET} expression from 20% to 80% in different studies.\textsuperscript{1,2,29} Given the intense investigation of \textit{MET} as a potential biomarker with predictive utility for improved outcomes in the context of the use of \textit{MET}-targeted therapies, it is important that pathologists refine the evaluation of cancer specimens for \textit{MET} expression.

This need for better standardization became apparent for our Lung Cancer Mutation Consortium 2.0 (LCMC 2.0) pathologist group when, at a monthly teleconference early in the LCMC study (February 2014), the \textit{MET} protein expression positivity rates for the first 150 LCMC specimens at 4 sites ranged from 27% to 83%. This wide range of positivity findings between sites propelled the initiation of a quality assurance effort with a series of “Round Robin” tests, a series of tests that are performed independently by separate laboratories several times with analyses of variance. In the LCMC 2.0 pathologist Round Robin effort, 7 pathologists performed a series of 3 rounds of evaluation of 83 total cases, with analysis and discussion between rounds. This endeavor improved the consistency of \textit{MET} scores between pathologists.

\section*{MATERIALS AND METHODS}

\subsection*{Population}

\textit{LCMC} 2.0 was a collective effort by 16 cancer centers across the United States to study genetic and protein biomarkers in patients with stage IV or recurrent lung adenocarcinomas.\textsuperscript{30} Eligibility requirements for this study were Eastern Cooperative Oncology Group performance status of 0, 1, or 2, expected survival of \textgreater 6 months, no prior treatment with targeted therapy, diagnosis of metastatic disease after May 1, 2012, and adequate tissue for molecular analyses. Of 1367 enrolled patients, 1009 patients were confirmed to have both a diagnosis of adenocarcinoma and adequate pathologic material for continuation in the biomarker study. Of 904 patients for whom at least 1 biomarker was assessed, 838 were successfully stained by \textit{MET} IHC. All sites obtained local Institutional Review Board approval and patient consent for participation in this study. The overall findings from \textit{LCMC} 2.0 are published separately.\textsuperscript{31}

\subsection*{\textit{MET} IHC and Digital Image Scanning}

Formalin-fixed paraffin-embedded (FFPE) tumor blocks were collected from patients enrolled in the \textit{LCMC} 2.0 biomarker study. Lung adenocarcinoma diagnosis and specimen adequacy were confirmed for each specimen by evaluation of hematoxylin and eosin–stained slides by 2 pathologists. Unstained slides from FFPE specimens were stained for \textit{MET} protein expression with the optimized prediluted CONFIRM anti-Total c-MET (SP44) rabbit monoclonal primary antibody by Ventana Medical
Systems (Tucson AZ; Catalog number 790-4430), as described in the Ventana product library (www.productlibrary.ventana.com, search “790-4430”). In short, FFPE tumor tissue blocks were cut at 4 μm thickness and placed onto positively charged slides, deparaffinized, placed on a Ventana Benchmark XT autostainer, hybridized with the primary SP44 antibody, analyzed with the Ventana ultraview universal DAB detection kit, stained with hematoxylin and bluing reagent, and cover slipped. The conditions for this general protocol were optimized at each participating LCMC 2.0 site.

Slides were scanned at ×40 power with an Aperio AT2 scanner at the University of Colorado Denver Biorepository. The Aperio AT2 utilizes an LED light source with cold white temperature (CCT of 5700K), for natural white illumination of the specimen. The AT2 also utilizes an ICC color correction profile to maintain digital slide image color as close as possible to what can be viewed under the microscope. The standard image format of the Aperio AT2 is SVS, which is a standard pyramid tile TIFF with JPEG2000 image compression. Compression quality was set to produce a 30:1 JPEG2000 image compression. The pathologists accessed images on the Spectrum Web Service, a shared biorepository imaging website hosted by the University of Colorado. Pathologists used ImageScope software downloaded on their on-site computers to view the images.

Scoring
Pathologists at multiple sites received specialized training and written guidelines for MET scoring before scoring. MET protein expression was evaluated for the percentage of tumor cells with no, faint, moderate, or strong staining. A cut point of at least 50% of tumor cells with moderate to strong MET staining for a final diagnosis of positive or negative was used as defined by the “original” MetMab clinical trial diagnostic criteria. Scores were also calculated with the histology score (H-score), a semi-quantitative calculation in which tumor cells with stronger staining are given more weight in the assessment of the overall protein expression of the tumor as follows:

\[
H\text{-score} = 1 \times (\% \text{ of faintly stained tumor cells}) + 2 \times (\% \text{ of moderately stained tumor cells}) + 3 \times (\% \text{ of strongly stained tumor cells}),
\]

Thus, the H-score ranged from 0 to 300.

Statistical Analyses
A nested random-effect model was used to evaluate the reproducibility of the MET protein expression technique. This statistic is based on the analysis of variance models and describes how strongly quantitative measurements between multiple observers resemble each other. The intraclass correlation coefficient (ICC) is calculated by the formula ICC = \( \frac{\sigma_{\text{inter}}^2}{\sigma_{\text{inter}}^2 + \sigma_{\text{intra}}^2} \), where \( \sigma_{\text{inter}}^2 \) is the variance between specimens and \( \sigma_{\text{intra}} \) is the pooled within-specimen variance. For each round of protein expression scoring, overall group and individual pathologists’ scores were compared with ICC analysis. The ICC has a range from 0 (no correlation) to 1 (perfect correlation). The interobserver agreement was interpreted as follows: <0.2, “poor”; 0.2 to 0.39, “fair”; 0.40 to 0.59, “moderate”; 0.60 to 0.79, “good”; and 0.80 to 1.00, “very good.”

RESULTS

Exploratory Round
An exploratory round of investigation was prompted by the discovery of a wide range of MET protein positivity rates, 27% to 83%, between LCMC 2.0 study sites. During a preliminary internal investigation, cases scored as “positive” (n = 66) at the University of Colorado were reviewed to identify “borderline positive” cases arbitrarily defined as those called “positive” (> 50%), but with ≤ 70% tumor cells with moderate to strong staining and/or an H-score ≤ 200. Overall, 39% (26/66) of positive cases fit into this “borderline positive” category. Of these 26 “borderline positive” cases, 5 cases were randomly selected for whole slide digital scanning and rescoring of the cases by 6 of the multi-institutional LCMC pathologists (Figs. 1A–E).

The percentages of tumor cells with moderate to strong staining for the 5 cases as scored by the 7 pathologists were compared (Fig. 2A). Only 1 case was concordantly determined to be positive by all 7 pathologists. Two cases were scored as “positive” with moderate to strong staining in ≥ 50% of tumor cells by 6 of the 7 pathologists; 1 case was scored as positive by 3 pathologists and negative by 4; and 1 case with an original score of 55% was scored as negative by the 6 additional pathologists. An H-score system with a cutoff of ≥ 150 for positivity showed a similar pattern of concordance for the same cases (Fig. 2B).

The LCMC pathologists discussed the results by teleconference and concluded that although it was still possible that differences in preanalytical factors, such as specimen processing and staining, could contribute to the variable positivity rates between sites, there was enough interobserver variability from pathologist scoring of the same images during this exploratory round to warrant further efforts to standardize scoring. It was also noted that MET heterogeneity inherently complicated scoring with most cases exhibiting a heterogeneous staining pattern of MET, that is, highly variable tumor-staining intensities in different tumor areas. All 7 pathologists agreed to score and discuss more cases as a “Round Robin” quality assurance project.

Round Robin 1
For Round Robin 1, 40 cases were included for scoring including 5 cases from the exploratory analysis and 35 additional randomly selected cases. One case was excluded from analysis due to inadequate tumor cellularity such that 39 cases were included in the final analysis. The average, lowest, and highest scores for each case are depicted in Figure 3. Analysis of the scores revealed a “moderate” resemblance of results between pathologists with an ICC of 0.50 (95% confidence interval, 0.37-0.64). The average positivity rate for MET based on the 50% cutoff was 59% (range, 36% to 82%) and the average H-score was 158 (range, 71 to 260). A teleconference was
held to discuss ways to improve the consistency of scoring. Suggestions included:
(1) Provide matched hematoxylin and eosin slide images for the next round of MET image scoring to facilitate tumor cell confirmation.

(2) Keep the MET-scoring training document open for reference during scoring.

(3) Provide example MET IHC images of Round 1 cases with concordant results between pathologists for “faint,” “moderate,” and “strong” staining (Figs. 4A–I).

FIGURE 1. Images of 5 lung adenocarcinoma specimens from 5 cases (A–E) stained by mesenchymal-epithelial transition immunohistochemistry are shown. These were selected and scored by 7 pathologists during an initial exploratory round of scoring (×20).
(4) Perform final scoring at low power (×10 objective).
(5) Self-adjust scoring higher or lower for consistency with the group positivity rate from Round Robin 1.

**Round Robin II**
For Round Robin 2, 40 cases not included in Round Robin 1 were randomly selected for scoring by the 7 pathologists.
LCMC pathologists; 1 of these cases was excluded from analysis due to poor image quality. The average, lowest, and highest scores for each case are depicted in Figure 5. Statistical analysis revealed “good” agreement with an ICC of 0.74 (95% confidence interval, 0.64-0.84). The average MET positivity rate was 66% (range, 46% to 92%) and the average H-score was 173 (range, 43 to 280).

The overall average MET H-score for all 78 cases from Round Robins I and II was 165 (H-score range, 43 to 280). The average H-score was <125 for 14 specimens, 125 to 175 for 35 specimens, and >175 for 29 specimens. A comparison of the individual pathologists’ ICCs demonstrated improved individual scoring consistency for all 7 pathologists between rounds, with improvement from “good” consistency with an average ICC of 0.64 (range, 0.43 to 0.76) for the first round to “very good” consistency with an average of 0.82 (range, 0.75 to 0.93) for the second round. During the pathologist teleconference to discuss the results, the ability to improve scoring consistency with communication was appreciated, but with the caveat that the heterogeneous staining pattern of MET remained a cause of inherent difficulty for scoring standardization.

**DISCUSSION**

Pathologists have the training and expertise to develop and refine the criteria to improve interobserver reproducibility of MET IHC evaluation. Scoring of protein expression on MET IHC-stained slides involves both the “art” of observing what is seen on the slide and the “science” of recording what is observed in a semiquantitative manner. There have been other efforts for standardization of IHC evaluation of other proteins with initiatives lead by pathologists to collectively determine reliable criteria and practical metrics for interpretative concordance among pathologists33; however, MET IHC evaluation has been notoriously challenging with an exceptionally high variation of positivity rates between studies.1,2,29

The collective experience of the LCMC 2.0 pathologists in the MET-scoring exercise has highlighted the importance of direct and continuous communication between pathologists when studies involve slide evaluation by multiple participants. With a Round Robin effort and good communication between the LCMC 2.0 pathologists, the overall group ICC improved from “moderate” (0.50) agreement for the first scoring round to “good” (0.74) agreement for the second round. Although the efforts to improve interobserver concordance in scoring were successful, standardized scoring of MET may continue to be inherently challenging due to its heterogeneous expression which complicates setting an accurate cutoff for positivity.28

This pathologist-driven effort to standardize scoring for MET has corroborated the conclusions of other studies.
that have emphasized the critical importance of communication among pathologists to achieve standardized scoring. Efforts to standardize scoring of other lung cancer receptor proteins, such as EGFR\cite{33} and ALK\cite{34} have also benefited from efforts to enhance direct communication between pathologists. Programmed death ligand-1 is another heterogeneous lung cancer protein with clinical relevance for which harmonization of scoring is in progress, such as with the programmed death ligand-1 Blueprint endeavor\cite{35}. The pathologist-initiated quality assurance project described here to standardize scoring of MET serves as a model for the use of “Round Robin” conferences to improve the evaluation and development of any potential biomarker.

In this study, scores from glass slides were not compared with scores from digital images due to the complexities associated with multiple pathologists performing both scores and the introduction of other variables between the scoring of the glass slides and digital images, such as training to improve the standardization of scoring. These complexities could complicate any analyses of differences between scores. However, a future study could be performed with a side-by-side comparison of the matched glass slides and digital images, with a defined wash out period between scoring rounds. Furthermore, an approach that incorporates machine learning to explore its potential as an adjunct tool for the standardization of MET scoring could be performed with this same set of cases. Previous studies have corroborated the effectiveness of machine learning-based scoring of IHC.\cite{36-40}

With clinical trials in progress, the evidence is growing for MET genomic biomarkers, such as MET gene amplification\cite{41} and METex14 splice-site mutations,\cite{9} to predict response to MET inhibitor therapy. Evaluation of MET protein expression and assessment of downstream signaling may play a role in determining which patients with MET genomic aberrations will most likely benefit from MET-targeted therapy.\cite{21} The practical implementation of standardized MET IHC evaluation for protein expression by pathologists sets the foundation for advancement of MET as a biomarker.

To date, only limited studies are available in the literature that investigates the correlation between MET IHC expression and MET genomic sequencing data.\cite{42,43} The utility of MET IHC will likely ultimately be dependent upon the extent to which protein expression is able to

**FIGURE 4.** The images represent examples of cases with consensus by the Lung Cancer Mutation Consortium (LCMC) 2.0 Pathologist Panel for faint mesenchymal-epithelial transition (MET) staining (A–C, ×40 magnification), moderate MET staining (D–F, ×40 magnification), and strong MET staining (G–I, ×40 magnification).
reliably complement next-generation sequencing findings. Future studies to refine the standardization of MET IHC scoring should be performed with an appropriate volume of cases, with routine peer monitoring to ensure quality control. Moreover, it is important to correlate MET IHC scoring with findings from genomic sequencing in the context of clinical outcomes with MET-targeted therapy.

ACKNOWLEDGMENTS

The authors thank Paul Fletcher and Daley Drucker, Moffitt Cancer Center, for editorial support. They did not receive any compensation beyond their regular salary. They also gratefully acknowledge Free to Breathe (Madison, WI) for funding to support this research. The authors also thank Andrea Ospyk, BS, and the University of Colorado at Denver Imaging Biorepository Shared Services (Colorado Lung SPORE Tissue Biobank and Biomarker Core P50CA058187 and University of Colorado Cancer Center Tissue Biobanking and Histology Shared Resource P30CA046934).

REFERENCES


FIGURE 5. The bar graph depicts the lowest, average, and highest histology scores (H-scores) for each of the 39 specimens that were scored by 7 pathologists during Round Robin 2. Error bars represent the SDs around the average score. MET indicates mesenchymal-epithelial transition.


