Published in final edited form as:

Am J Surg Pathol. 2012 December ; 36(12): 1747–1760. doi:10.1097/PAS.0b013e31825ea736.

Diagnostic Reproducibility of Hydatidiform Moles:

Ancillary Techniques (p57 Immunohistochemistry and Molecular Genotyping) Improve Morphologic Diagnosis for Both Recently Trained and Experienced Gynecologic Pathologists

Mamta Gupta, MBBS*, Russell Vang, MD*,†, Anna V. Yemelyanova, MD*, Robert J. Kurman, MD*,†,‡, Fanghong Rose Li, MD*, Emily C. Maambo, MD*, Kathleen M. Murphy, PhD*, Cheryl DeScipio, PhD*,†, Carol B. Thompson, MS, MBA§, and Brigitte M. Ronnett, MD*,†

*Department of Pathology, The Johns Hopkins University School of Medicine and Hospital, Baltimore, MD

†Department of Gynecology & Obstetrics, The Johns Hopkins University School of Medicine and Hospital, Baltimore, MD

‡Department of Oncology, The Johns Hopkins University School of Medicine and Hospital, Baltimore, MD

§Johns Hopkins Biostatistics Center, The Johns Hopkins Bloomberg School of Public Health, Baltimore, MD

Abstract

Distinction of hydatidiform moles from nonmolar specimens (NMs) and subclassification of hydatidiform moles as complete hydatidiform mole (CHM) and partial hydatidiform mole (PHM) are important for clinical practice and investigational studies; however, diagnosis based solely on morphology is affected by interobserver variability. Molecular genotyping can distinguish these entities by discerning androgenetic diploidy, diandric triploidy, and biparental diploidy to diagnose CHMs, PHMs, and NMs, respectively. Eighty genotyped cases (27 CHMs, 27 PHMs, 26 NMs) were selected from a series of 200 potentially molar specimens previously diagnosed using p57 immunohistochemistry and genotyping. Cases were classified by 6 pathologists (3 faculty level gynecologic pathologists and 3 fellows) on the basis of morphology, masked to p57 immunostaining and genotyping results, into 1 of 3 categories (CHM, PHM, or NM) during 2 diagnostic rounds; a third round incorporating p57 immunostaining results was also conducted. Consensus diagnoses (those rendered by 2 of 3 pathologists in each group) were also determined. Performance of experienced gynecologic pathologists versus fellow pathologists was compared, using genotyping results as the gold standard. Correct classification of CHMs ranged from 59% to

Correspondence: Brigitte M. Ronnett, MD, Department of Pathology, The Johns Hopkins Hospital, Weinberg 2242, 401N Broadway, Baltimore, MD 21231 (bronnnett@jhmi.edu).

Present address: Mamta Gupta, MBBS, Beth Israel Deaconess Medical Center and Harvard medical School, Boston, MA.

Present address: Rose Fanghong Li, MD, The First Affiliated Hospital of Nanjing Medical University, Nanjing Medical University, Nanjing, China.

Present address: Emily C. Maambo, MD, Medical Center Laboratory, Yakima, WA.

Present address: Kathleen M. Murphy, PhD, ProPath, Dallas, TX.

Conflicts of Interest and Source of Funding: The authors have disclosed that they have no significant relationships with, or financial interest in, any commercial companies pertaining to this article.

100%; there were no statistically significant differences in performance of faculty versus fellows in any round (*P*-values of 0.13, 0.67, and 0.54 for rounds 1 to 3, respectively). Correct classification of PHMs ranged from 26% to 93%, with statistically significantly better performance of faculty versus fellows in each round (*P*-values of 0.04, <0.01, and <0.01 for rounds 1 to 3, respectively). Correct classification of NMs ranged from 31% to 92%, with statistically significantly better performance of faculty only in round 2 (*P*-values of 1.0, <0.01, and 0.61 for rounds 1 to 3, respectively). Correct classification of all cases combined ranged from 51% to 75% by morphology and 70% to 80% with p57, with statistically significantly better performance of faculty only in round 2 (*P*-values of 0.69, <0.01, and 0.15 for rounds 1 to 3, respectively). p57 immunostaining significantly improved recognition of CHMs (*P*<0.01) and had high reproducibility (κ =0.93 to 0.96) but had no impact on distinction of PHMs and NMs. Genotyping provides a definitive diagnosis for the ~25% to 50% of cases that are misclassified by morphology, especially those that are also unresolved by p57 immunostaining.

Keywords

hydatidiform mole; reproducibility; p57 immunohistochemistry; molecular genotyping

Distinction of hydatidiform moles from nonmolar specimens (NMs) and subclassification of hydatidiform moles as complete hydatidiform mole (CHM) and partial hydatidiform mole (PHM) are important not only for clinical management but also for accurate ascertainment of the risk of persistent gestational trophoblastic disease (GTD). The risk of persistent GTD for CHMs (15% to 20%) is significantly higher than for PHMs (0.2% to 4%).¹⁻⁴ Despite the lower risk associated with PHMs, metastatic GTD and trophoblastic tumors coexistent with or subsequent to a diagnosis of PHM have been reported.⁵⁻⁸ Furthermore, distinction of PHMs from NMs is important for appropriate clinical management, as a diagnosis of PHM generates follow-up with serum β human chorionic gonadotropin (HCG) levels and contraception, which would be unnecessary for a diagnosis of NM and undesirable for infertility patients.

The 2 forms of hydatidiform moles, CHMs and PHMs, can usually be distinguished when morphologic features are well developed. Typical CHMs are characterized by enlarged edematous villi with moderate to marked circumferential trophoblastic hyperplasia, often with cytologic atypia, prominent central cistern formation, and trophoblastic inclusions.⁹⁻¹¹ In addition, an early form of CHM has been recognized. Early CHMs are characterized by a redundant bulbous villous growth pattern, hypercellular myxoid villous stroma, a labyrinthine network of villous stromal canaliculi, karyorrhectic debris within stroma, and at least focal trophoblastic hyperplasia on villi and the undersurface of the chorionic plate.¹² Both forms of CHMs are androgenetic diploid conceptions, arising by fertilization of an ovum that has lost its maternal DNA by either 1 sperm that duplicates (most commonly) or 2 sperm; thus they contain 2 paternal and no maternal chromosomal complements. Characteristic morphologic features of PHMs include the presence of 2 populations of villi (large, irregular, hydropic villi and small, immature, fibrotic villi), cisterns in some enlarged villi, markedly irregular villi with scalloped borders and trophoblastic inclusions, and generally mild circumferential trophoblastic hyperplasia.^{9,10,13-15} PHMs are diandric

triploid conceptions, arising by fertilization of an ovum by 2 sperm (most commonly) or 1 sperm that duplicates; thus they contain 2 paternal and 1 maternal chromosomal complement. Despite these differences, there is some degree of morphologic overlap between CHMs and PHMs, occasionally making subtyping problematic. In addition, there are several situations in which NMs can display some features that can cause misinterpretation as a hydatidiform mole. These include products of conception (POC) specimens with abnormal villous morphology (a nonmolar type of villous abnormality having some morphologic features suggestive of a PHM but lacking diandric triploidy, sometimes attributable to other genetic abnormalities such as trisomy),^{16–19} early nonmolar abortuses with prominent trophoblastic hyperplasia, and hydropic abortuses. Virtually all of these nonmolar entities are characterized by biparental diploidy (they contain 1 maternal and 1 paternal chromosomal complement), with the exception of the digynic triploid conceptions; these contain 2 maternal and 1 paternal chromosomal complement but do not exhibit molar features. In addition, mosaic/chimeric conceptions share some features with hydatidiform moles but are genetically complicated and distinct from typical moles.^{19–22}

Prior studies have demonstrated that there is marked variability in the diagnosis of hydatidiform moles based only on hematoxylin and eosin-stained (H&E) slides, even among experienced pathologists with specialized training.^{23–27} The value of ancillary techniques, including immunohistochemical analysis of p57 expression^{20,21,28–38} and molecular genotyping by polymerase chain reaction amplification of short tandem repeat loci,^{19,34,36,39–43} in improving the diagnosis of hydatidiform moles has been demonstrated in a number of recent studies. Furthermore, genotyping allows for distinction of CHMs, PHMs, and NMs from one another by specifically discerning the androgenetic diploidy, diandric triploidy, and biparental diploidy that define CHMs, PHMs, and NMs, respectively. Thus, in contrast to other ancillary techniques, genotyping has the ability to establish diagnostic truth. Prior studies have not used a definitive gold standard to determine the true accuracy and reproducibility of diagnosing hydatidiform moles. We recently reported a study of reproducibility of diagnosing hydatidiform moles by experienced gynecologic pathologists using genotyping results as the gold standard against which performance was measured.⁴⁴ The current analysis represents an expanded investigation of diagnostic reproducibility of hydatidiform moles in which the performances of experienced gynecologic pathologists and recently trained junior pathologists were compared to assess the impact of experience on diagnostic performance.

MATERIALS AND METHODS

A large prospective and consecutive series of potentially molar specimens encountered on the Gynecologic Pathology Service of the Johns Hopkins Hospital (Baltimore, MD) (comprising ~90% consultation cases), previously diagnosed using p57 immunostaining (monoclonal antibody from Neomarkers, Fremont, CA) and molecular genotyping (n=200) served as the pool of material from which cases were obtained. Molecular genotyping consisted of microdissection of separate foci of villous and decidual tissue in each case. DNA was subjected to polymerase chain reaction amplification of 9 polymorphic markers (microsatellites) from multiple (8) different chromosomes and the amelogenin locus (AmpFISTR Profiler kit; Applied Biosystems, Foster City, CA). The alleles at each locus

were identified for both the villous and decidual tissues, and the patterns were compared. Additional details of immunohistochemical and molecular analysis are provided elsewhere.^{19,34,43}

Eighty genotyped cases (1 H&E slide per case), including 27 CHMs, 27 PHMs, and 26 NMs, were selected by one of the authors (B.M.R.) familiar with the cases to assure the quality of the H&E-stained section and p57 immunostain. It is noteworthy that the H&E slide selected for each case was the most representative of the specimen, being derived from the block that was used to generate the slides for p57 immunostaining and genotyping for formal diagnosis of each case. Three faculty pathologists practicing in an active gynecologic pathology service, with varying experience in gynecologic pathology (5–31 years; A.V.Y., R.V., and R.J.K.), and 3 fellows in the Division of Gynecologic Pathology who had just recently completed training in anatomic and clinical pathology (M.G., F.R.L., and E.C.M.) evaluated the cases. Cases were classified by the reviewers on the basis of the H&E slides, masked to p57 immunostaining and genotyping results, into 1 of 3 categories—CHM, PHM, or NM—during 2 diagnostic rounds without training sessions, separated by 8 to 12 weeks. All observers used the histologic criteria they use in routine practice. A third diagnostic round incorporating p57 immunostaining results was conducted after an interval of at least 16 weeks from the second round, with a brief explanation of interpretation of p57 immunostaining conducted using a multiheaded microscope for all reviewers. In this round, reviewers read the H&E slide and p57 immunostain together to render a diagnosis and an interpretation of the stain. Briefly, to interpret immunohistochemical stains for p57, the presence or absence of nuclear positivity is assessed in villous stromal cells, cytotrophoblast, intermediate trophoblast, and maternal decidua. Reviewers were instructed that the p57 immunostain was to be interpreted as “negative” and satisfactory when maternal decidua and/or intermediate trophoblastic cells exhibited nuclear expression of p57 (serving as internal positive control in all cases, including CHMs) but villous stromal cells and cytotrophoblast were either entirely negative or demonstrated only limited expression (nuclear staining in <10% of these cell types). The p57 immunostain is interpreted as “positive” when the extent of staining is extensive or diffuse in these cell types. In addition to the typical diffusely positive result, 2 variants of positive staining can be encountered occasionally (<10% of cases in our large series): equivocal and discordant. The p57 immunostain is interpreted as “equivocal” when nuclear expression in both villous stromal cells and cytotrophoblast is in the focally positive range (10% but <50% of the villi in the stained section). The p57 immunostain is interpreted as “discordant” when there is any combination/admixture of negative and positive results for villous stromal cells and cytotrophoblast within individual villi, including positive staining in cytotrophoblast and negative staining in villous stromal cells or vice versa. A more detailed discussion of p57 expression patterns is provided elsewhere.¹⁹

Statistical Analysis

Diagnoses rendered by the individual pathologists for each case in each round were tabulated, and consensus diagnoses [those rendered by 2 of 3 pathologists in each group of reviewers (faculty and fellows)] were determined. Genotyping results were used as the gold standard (true) diagnosis for assessing diagnostic performance. Sensitivity and specificity

were calculated for each diagnostic category (CHM, PHM, and NM), and for all cases combined, (as determined by the genotyping results), for each individual pathologist and for consensus diagnoses for each round. Comparisons of correct classification rates between faculty and fellow pathologists, and between an H&E round and the p57 round, were performed using general linear models with generalized estimating equations to account for within-pathologist correlation.⁴⁵ The results of these comparisons are described with odds ratios (OR) and their 95% confidence intervals. For this study, the OR represents how much more or less the odds are that one group of observers (eg, faculty pathologists) has rendered a correct classification compared with another group of observers (eg, fellow pathologists). When the 95% confidence interval does not include the value 1.0, an OR is significant at $P < 0.05$. Interobserver and intraobserver agreement for diagnoses (rounds 1 and 2) and interobserver agreement for interpretation of the p57 stain (round 3) were determined using Cohen κ statistics. Statistical analyses were performed with the Stata software package (version 11.1; StataCorp LP, 2009).

RESULTS

Results comparing performance of faculty versus fellows are presented in Figures 1–5, with details of the statistical analysis provided in Tables 1–4. Using morphologic assessment alone (H&E rounds 1 and 2), sensitivity of a diagnosis of CHM ranged from 59% to 100% for individual pathologists and 70% to 96% by consensus (Figs. 1, 5); specificity ranged from 81% to 96% for individual pathologists and 89% to 98% by consensus. Sensitivity of a diagnosis of PHM ranged from 26% to 93% for individual pathologists and 41% to 78% by consensus (Figs. 2, 5); specificity ranged from 59% to 93% for individual pathologists and 74% to 91% by consensus. Sensitivity of a diagnosis of NM ranged from 31% to 92% for individual pathologists and 50% to 73% by consensus (Figs. 3, 5); specificity ranged from 63% to 100% for individual pathologists and 78% to 94% by consensus. Percent correct classification of all cases (H&E rounds) ranged from 51% to 75% for individual pathologists and 63% to 75% by consensus (Figs. 4, 5).

Comparison of performance of faculty versus fellows in correctly diagnosing CHMs demonstrated no statistically significant differences in any round (P -values of 0.13, 0.67, and 0.54 for rounds 1 to 3, respectively; Fig. 1; Table 1). Comparison of performance of faculty versus fellows in correctly diagnosing PHMs demonstrated statistically significant differences in each round (P -values of 0.04, <0.01 , and <0.01 for rounds 1 to 3, respectively; Fig. 2; Table 1). Comparison of performance of faculty versus fellows in correctly diagnosing NMs demonstrated a statistically significant difference only in round 2 (P -values of 1.0, <0.01 , and 0.61 for rounds 1 to 3, respectively; Fig. 3; Table 1). Comparison of performance of faculty versus fellows in correctly diagnosing all cases combined demonstrated a statistically significant difference only in round 2 (P -values of 0.69, <0.01 , and 0.15 for rounds 1 to 3, respectively; Fig. 4; Table 1). Comparison of performance of the consensus diagnoses rendered by faculty versus fellows in correctly diagnosing each entity (CHMs, PHMs, and NMs) and all cases combined demonstrated statistically significant differences for diagnosis of CHMs only in round 1 (performance of fellows better), for diagnosis of PHMs in all rounds (performance of faculty better), and for diagnosis of NMs

only in round 2 (performance of faculty better); no differences were observed for diagnosis of all cases combined in any round (see Fig. 5; Table 1 for details, including *P*-values).

In round 3, addition of the p57 immunostain improved sensitivity of a diagnosis of CHM to a range of 93% to 96% for individual pathologists and 96% by consensus (Figs. 1, 5); specificity ranged from 96% to 98% for individual pathologists and 96% to 98% by consensus. This improvement in sensitivity varied widely for individual pathologists, with little or no impact for 2 pathologists (1 faculty and 1 fellow pathologist) and notable improvement for the remaining 4 (% increase in sensitivity ranging from 11% to 34%). Thus, overall improvement in diagnosis of CHMs for all reviewers using the p57 stain relative to the H&E rounds was significant ($P<0.01$ for round 3 vs. 1 and $P<0.01$ for round 3 vs. 2; Table 2). Importantly, almost all variability between individual pathologists (whether faculty or fellows) in sensitivity of diagnosing CHMs was removed with the use of this ancillary diagnostic tool. There was no significant impact of the addition of this stain on diagnosis of PHMs and NMs (for PHMs, $P=0.81$ for round 3 vs. 1 and $P=0.17$ for round 3 vs. 2; for NMs, $P=0.07$ for round 3 vs. round 1 and $P=0.26$ for round 3 vs. round 2; Table 2). Sensitivity of a diagnosis of PHM maintained a wide range, from 48% to 85% for individual pathologists and 48% to 78% by consensus (Figs. 2, 5); specificity ranged from 70% to 94% for individual pathologists and 83% to 89% by consensus. Sensitivity of a diagnosis of NM also maintained a wide range, from 39% to 92% for individual pathologists and 69% to 81% by consensus (Figs. 3, 5); specificity ranged from 76% to 93% for individual pathologists and 76% to 93% by consensus. The 1 CHM that was not recognized as such by any observer in round 3 incorporating the p57 stain was a p57-positive androgenetic diploid CHM with a retained maternal copy of chromosome 11 (location of p57 gene).⁴⁶ This case was diagnosed by 5/6 reviewers as a CHM in both rounds 1 and 2 (outlier diagnoses were 1 NM and 1 PHM) but was interpreted as a PHM by 4/6 and as NM by 2/6 in round 3 on the basis of the diffusely positive p57 stain.

Interobserver variability for morphologic assessment ranged from 0.46 to 0.73 (moderate to good) for a diagnosis of CHM, 0.15 to 0.43 (poor to moderate) for a diagnosis of PHM, and 0.13 to 0.42 (poor to moderate) for a diagnosis of NM (Table 3). Furthermore, agreement was generally similar for faculty and fellows within each diagnostic category. In particular, agreement for diagnosis of CHMs was moderate for both faculty and fellows in round 1 and good for both groups in round 2. Agreement for diagnosis of both PHMs and NMs was similar for both faculty and fellows in round 1 (poor to fair) and only slightly different for faculty versus fellows in round 2 (moderate vs. fair, respectively). Intraobserver variability (round 1 vs. round 2) ranged from 0.37 to 0.67 (fair to good) overall, with similar ranges for faculty (0.44 to 0.67) and fellows (0.37 to 0.63) (Table 4).

Interobserver variability for interpretation of the p57 immunostain [including all variants of a positive interpretation (equivocal and discordant) as simply positive] was 0.96 (almost perfect) for faculty and 0.93 (almost perfect) for fellows. Positive cases most often were interpreted as diffusely positive, with only a minority of cases assessed as equivocal or discordant positive results. The 3 faculty reviewers subclassified the positive stains as equivocal positive in 0, 5, and 3 cases and as discordant in 0, 5, and 6 cases. The 3 fellow reviewers subclassified the positive stains as equivocal positive in 1, 4, and 0 cases and as

discordant in 1, 0, and 3 cases. Of 3 androgenetic/biparental mosaic cases, the discordant staining pattern was recognized by 3/6, 2/6, and 3/6 reviewers for each of these cases.

To determine the nature of the diagnostic errors, consensus diagnoses versus genotyping diagnoses were tabulated for each group of observers (faculty and fellows) in each round (Table 5). Misclassification was dominated by difficulties in distinguishing PHMs and NMs, with both overdiagnosis and underdiagnosis occurring for both faculty and fellow reviewers (12 to 13 cases for faculty and 14 to 20 cases for the fellows in the 2 H&E rounds). Misclassification of CHM versus PHM, with both overdiagnosis and underdiagnosis occurring for both faculty and fellow reviewers, also occurred but less frequently compared with the aforementioned difficulty (4 to 5 cases for faculty and 1 to 3 cases for the fellows in the 2 H&E rounds). Misclassification of CHM and NM was uncommon but did occur in a few cases, with both overdiagnosis and underdiagnosis for faculty (2 cases in the second H&E round) and only overdiagnosis for the fellows (3 to 4 cases in both H&E rounds). A number of cases in each round for each group of reviewers received no consensus diagnosis: 6 and 2 cases for faculty in rounds 1 and 2, respectively, and 4 and 3 cases for fellows in rounds 1 and 2, respectively. These included cases in all 3 diagnostic categories. Of the cases lacking consensus for the faculty in the 2 H&E rounds, only 1 case failed to achieve consensus in both rounds. Of the cases lacking consensus for the fellows in the 2 H&E rounds, only 1 case failed to achieve consensus in both rounds. It is noteworthy that there was no overlap in the cases not achieving consensus for faculty versus fellows in either round, meaning that the most problematic cases for the faculty versus the fellows were not the same ones. Representative examples of problematic cases are illustrated in Figures 6 and 7, with details for those cases provided in Table 6.

DISCUSSION

The current study confirms that diagnosis of hydatidiform moles suffers from imperfect reproducibility. In addition, by using a true gold standard diagnosis (that obtained by genotyping) it establishes that molar and NMs overall are incorrectly classified in a substantial number of cases. Correct classification of all specimens combined, based solely on review of H&E slides, ranged from 51% to 75% overall for all individual reviewers in both rounds and 63% to 75% by consensus. Even with addition of p57 immunohistochemistry, which is a highly reproducibly interpreted stain ($\kappa > 0.9$ regardless of experience), correct classification of all cases overall was 70% to 80%. This limited impact of p57 on correct classification of all cases combined is attributable to the expected finding that addition of p57 immunohistochemistry to the H&E evaluation improved the sensitivity of diagnosis of CHMs but not PHMs or NMs. Our study thus confirms that the main diagnostic problem concerns distinguishing PHMs from NMs. Nonetheless, distinction of CHMs from both PHMs (primarily) and NMs (occasionally) is also problematic, and failure to recognize all CHMs by morphology persists even for experienced gynecologic pathologists.

The current study represents an expansion of our recent study⁴⁴ and differs from that analysis in that it was specifically designed to address the role of experience (from several to many years of specialized practice focused on gynecologic pathology) in correctly

diagnosing molar and NMs by comparing the performance of experienced faculty gynecologic pathologists with that of junior pathologists who had only recently completed general training in anatomic and clinical pathology. Although a single-institution study using experienced gynecologic pathologists and trainees who work together has certain potential limitations and might not provide results that could be generalized across practices, there are several aspects of the study that are worth noting in this regard. First, there was notable variability in performance among the faculty pathologists, despite the fact that they have worked together for >5 years. For example, correct classification of CHMs ranged from 59% to 100% for faculty in the 2 H&E rounds, with 2 faculty members demonstrating essentially similar performance in both rounds (63% for 1 faculty member in both rounds and 96% to 100% for the other faculty member in the 2 rounds) and 1 showing notable improvement in the second round (59% to 78%). It is likely that these results reflect individual diagnostic practices (behaviors) related to inherent different tendencies in sensitivity and reproducibility; if the shared work setting were a major factor then diagnostic performance would be expected to be much more similar among the faculty members. Second, the fellows had not been trainees in our residency program but rather were trained at 3 different institutions in 3 different states. Third, the study began shortly after the fellows joined the fellowship program, and hence they had essentially no exposure to the algorithmic approach to diagnosing potentially molar cases used at our institution in the first round, and rather limited exposure by the second round (on the basis of their periodic rotation onto the consultation service). Fourth, no training sessions were conducted at any time during the study, so the fellows were not formally exposed to our diagnostic approaches or teaching methods in a focused manner. Thus, on the basis of this limited exposure to our training environment and the lack of preceding or intervening training sessions, their diagnostic skills in the 2 H&E rounds predominantly reflect their residency training rather than fellowship training. Therefore, the current study offers a comparison of general surgical pathologists with limited practice experience with experienced gynecologic pathologists. Nonetheless, conducting a follow-up multi-institutional study comparing experienced faculty with specialty training, general pathologists with varying levels of training, and trainees at different stages of the training process would be of value.

In this study setting, accuracy and reproducibility of a diagnosis of CHM appeared to be independent of having experience in gynecologic pathology, as there were no significant differences in performance between faculty and fellows in the H&E rounds, and interobserver agreement was similar for both groups of observers in these rounds. In contrast to the results for diagnosis of CHMs, our data suggest that experience might have some role in improving recognition of PHMs, as accuracy of a diagnosis of PHM was significantly better for faculty compared with fellows in all rounds. However, reasons for the difference in performance could not be ascertained. Interobserver agreement for diagnosis of PHMs by faculty was inconsistent across the H&E rounds, being poor in round 1 and inexplicably improved to moderate in round 2 despite a lack of an intervening training session or additional data for the cases. Experience did not appear to have a role in improving diagnosis of NMs, as accuracy and reproducibility of a diagnosis of NM were more similar than different for faculty versus fellows overall, with no significant difference in 2 of the 3 rounds and generally similar levels of interobserver agreement. For all diagnostic categories

combined, there was no difference in performance of faculty versus fellows in 2 of the 3 rounds when individual diagnoses were used and no difference in any of the rounds when consensus diagnoses were used, suggesting that experience did not significantly influence diagnostic accuracy overall.

This study was not designed to determine potential reasons for the case misclassification and diagnostic variability observed. There are plenty of studies in the literature establishing that diagnostic reproducibility varies for a wide variety of pathologic entities, and this is no different for hydatidiform moles. Prior studies have established the variable and suboptimal interobserver and intraobserver reproducibility of the diagnosis of hydatidiform moles when based only on routine morphologic assessment, even for experienced gynecologic/placental pathologists, but have not discerned specific reasons for this.^{23–27} In an attempt to identify reasons for diagnostic variability and errors, one would have to take all cases for which there was any misclassification and also those for which there was inconsistent classification across diagnostic rounds and ask each pathologist why those different diagnoses were rendered for each such example. A pathologist could speculate as to why a certain diagnosis was rendered but such assessment would be biased by this kind of retrospective attempt to explain an incorrect interpretation in the face of the correct diagnosis. It is not clear that this would actually determine the reason for the misclassification. In general, problems in classification of molar specimens and NMs can be attributed to several factors, including imperfect histologic criteria for diagnosing hydatidiform moles, variability in how pathologists apply diagnostic criteria, and the known variation in morphologic features dependent on the gestational age of the specimen. With the widespread use of routine first-trimester ultrasonography, the latter factor has become significant, as most POC specimens, including molar and nonmolar ones, are encountered at much earlier gestational ages when microscopic features are less well developed.⁴⁷

As discussed in our recent study,⁴⁴ all of the cases in these 2 studies were derived from our gynecologic pathology consultation service, so we acknowledge some degree of bias in this study set; however, these included a wide morphologic spectrum of cases, including typical hydatidiform moles with characteristic features (as illustrated in some of the figures in our previous study). Interestingly, this prospectively collected set has had, since its beginning, both a substantial number of hydatidiform moles and a greater proportion of CHMs than PHMs. The 200-case set from which these 80 study cases were extracted comprised 76 CHMs, 48 PHMs, 69 NMs, and 7 mosaic/chimeric conceptions. This has been a surprise to us, given that diagnosis of PHMs and NMs is more problematic than CHMs, and one would expect PHMs to outnumber CHMs in a set of consultation cases. Although early forms are expected to constitute a notable subset of CHMs (given that a substantial number of hydatidiform moles are encountered at relatively early gestational ages since the introduction of routine use of imaging studies early in pregnancy), they did not constitute the exclusive form of CHM represented in this set (as judged by subjective morphologic assessment, as genotyping is identical for all forms of CHMs). Thus, since this set of consultation cases included a sufficient number of CHMs, we believe that the cases in this study are not unduly biased toward unusually difficult cases.

One unique feature and significant advantage of our reproducibility study is that it measured performance against a true diagnostic standard, namely, the genotyping result. All prior studies assessing the accuracy/reproducibility of a diagnosis of hydatidiform moles with ancillary techniques have not included any type of definitive gold standard method to validate the diagnoses; therefore, it has not been possible to determine true diagnostic accuracy of morphologic assessment (H&E-stained slides). Although it has been observed that the addition of ploidy analysis improves both the interobserver and intraobserver reproducibility of diagnosis of hydatidiform moles,^{23,24} most ancillary techniques, including conventional cytogenetics (karyotype) and DNA ploidy analysis (flow cytometry, image analysis, and fluorescent in situ hybridization),^{16,20,21,23,24,30,48–57} have the limitation of not being able to establish maternal/parental contributions of chromosome complements and thus cannot absolutely determine the true diagnosis. For example, a diploid result by karyotyping, flow cytometry/image analysis, or fluorescent in situ hybridization cannot distinguish a CHM (androgenetic diploidy) from a diploid NM (biparental diploidy), and a triploid result cannot discern a PHM (diandric triploidy) from a triploid NM (digynic triploidy). Numerous studies have demonstrated the value of p57 immunohistochemistry for reliably distinguishing CHMs (absence of p57 expression due to lack of maternal DNA) from both PHMs and NMs (both positive due to the presence of a maternal chromosomal complement).^{20,21,28–38} However, this marker has the limitation of not being able to discern PHMs from NMs (the latter including both biparental diploid NMs and digynic triploid NMs) because all of these maintain p57 expression due to the presence of a maternal chromosomal complement. As this particular differential diagnosis is the one that most requires the assistance of an ancillary technique, another method is necessary to definitively distinguish these entities; genotyping is a technique that accomplishes this.

In this study, routine microscopic evaluation without use of ancillary techniques, even in the hands of gynecologic pathologists and even when a consensus diagnosis was used, demonstrated incorrect classification in at least 20% of cases. These results suggest that there are inherent limitations in the ability of morphologic assessment to provide accurate diagnosis in all cases, likely related to the known morphologic overlap of the entities and the lack of fully developed morphologic features in early examples. Given that the experienced pathologists in this study already have subspecialty training and focused practice in gynecologic pathology, it is unlikely that there is any way to improve diagnosis using traditional H&E assessment alone. Thus, the findings in this study support use of the algorithm that we previously developed for improving the diagnosis of hydatidiform moles, which combines p57 immunohistochemistry and molecular genotyping.^{19,34,42} Briefly, immunohistochemical staining for p57 is first performed. In the context of appropriate morphologic features, a negative p57 result establishes a diagnosis of CHM, and no further ancillary evaluation is required. If the p57 result is positive, then a CHM can be excluded (with rare exceptions⁴⁶), but molecular genotyping is required to definitively distinguish a PHM from an NM. When the molecular analysis reveals a diandric triploid or biparental diploid result, then a diagnosis of PHM or NM, respectively, can be rendered. We advocate use of this algorithm for all POCs with any suspicion for a hydatidiform mole, which includes either a clinical concern for a hydatidiform mole (eg, abnormally elevated β -HCG level, abnormal ultrasound findings, clinical diagnosis of “rule out mole,” etc.) or pathologic

concern because of some morphologic abnormality of the chorionic villi. The algorithm yields a definitive diagnosis in essentially all cases, with the exception of a few situations, some rare, in which diagnostic evaluation may be limited or confusing. These situations have been discussed in detail in our prior report.⁴⁴

Use of this algorithmic approach represents a compromise between traditional morphologic assessment, which has limitations, and genotyping of all cases, which is clearly more expensive. The p57 component of the algorithm should capture essentially all CHMs, the most important group to readily identify for clinical management purposes; this assessment is highly reproducible and is a technique that can be performed in most immunohistochemistry laboratories without the need for highly specialized equipment and expertise, such as that required for genotyping. As PHMs have a low but real risk of persistent GTD and are managed as molar pregnancies as per current guidelines (sometimes with abbreviated follow-up relative to management of CHMs),^{58,59} and because overdiagnosis of NMs as PHMs has implications for infertility patients, genotyping of all p57- positive potentially molar specimens as per the algorithm provides for definitive diagnosis of PHMs and NMs, allowing for refined management of these entities. In the setting of limited resources, use of ancillary techniques can be focused on identifying the entity with the greatest risk for persistent GTD, namely CHMs, by selectively applying only p57 immunohistochemical analysis to assist in diagnosing CHMs and foregoing genotyping for distinction of PHMs from NMs. For the latter situation, an equivocal diagnosis, such as “abnormal villous morphology, PHM cannot be excluded,” might need to be rendered. As the risk of persistent GTD in the case of PHMs is much closer to that of NMs than it is to that of CHMs, this may well be acceptable for routine practice, with the understanding that an equivocal diagnosis will potentially lead to clinical management as a PHM at least for some abbreviated time frame, and that this approach does have accompanying costs (clinic visit, multiple serum β -HCG levels, contraception), which might well rival the cost of genotyping. With such a limited approach, it also needs to be understood that an apparently unequivocal diagnosis of either PHM or NM established on the basis of morphologic assessment alone is not guaranteed to be accurate even when rendered by an experienced gynecologic pathologist. Therefore, given the established suboptimal performance of morphologic assessment alone, even for experienced gynecologic pathologists as demonstrated in this and prior studies, the most ideal method of correctly classifying all hydatidiform moles and NMs is a combined approach including correlation of morphologic features, p57 immunohistochemistry, and molecular genotyping.

References

1. Feltmate CM, Growdon WB, Wolfberg AJ, et al. Clinical characteristics of persistent gestational trophoblastic neoplasia after partial hydatidiform molar pregnancy. *J Reprod Med.* 2006; 51:902–906. [PubMed: 17165438]
2. Hancock BW, Nazir K, Everard JE. Persistent gestational trophoblastic neoplasia after partial hydatidiform mole incidence and outcome. *J Reprod Med.* 2006; 51:764–766. [PubMed: 17086803]
3. Sebire NJ, Fisher RA, Foskett M, et al. Risk of recurrent hydatidiform mole and subsequent pregnancy outcome following complete or partial hydatidiform molar pregnancy. *BJOG.* 2003; 110:22–26. [PubMed: 12504931]

4. Wielsma S, Kerkmeijer L, Bekkers R, et al. Persistent trophoblast disease following partial molar pregnancy. *Aust N Z J Obstet Gynaecol.* 2006; 46:119–123. [PubMed: 16638033]
5. Cheung AN, Khoo US, Lai CY, et al. Metastatic trophoblastic disease after an initial diagnosis of partial hydatidiform mole: genotyping and chromosome in situ hybridization analysis. *Cancer.* 2004; 100:1411–1417. [PubMed: 15042675]
6. Medeiros F, Callahan MJ, Elvin JA, et al. Intraplacental choriocarcinoma arising in a second trimester placenta with partial hydatidiform mole. *Int J Gynecol Pathol.* 2008; 27:247–251. [PubMed: 18317216]
7. Palmieri C, Fisher RA, Sebire NJ, et al. Placental site trophoblastic tumour arising from a partial hydatidiform mole. *Lancet.* 2005; 366:688. [PubMed: 16112308]
8. Seckl MJ, Fisher RA, Salerno G, et al. Choriocarcinoma and partial hydatidiform moles. *Lancet.* 2000; 356:36–39. [PubMed: 10892763]
9. Sebire NJ, Fisher RA, Rees HC. Histopathological diagnosis of partial and complete hydatidiform mole in the first trimester of pregnancy. *Pediatr Dev Pathol.* 2003; 6:69–77. [PubMed: 12469234]
10. Sebire NJ, Makrydimas G, Agnantis NJ, et al. Updated diagnostic criteria for partial and complete hydatidiform moles in early pregnancy. *Anticancer Res.* 2003; 23:1723–1728. [PubMed: 12820448]
11. Szulman AE, Surti U. The syndromes of hydatidiform mole. II. Morphologic evolution of the complete and partial mole. *Am J Obstet Gynecol.* 1978; 132:20–27. [PubMed: 696779]
12. Keep D, Zaragoza MV, Hassold T, et al. Very early complete hydatidiform mole. *Hum Pathol.* 1996; 27:708–713. [PubMed: 8698316]
13. Fukunaga M. Histopathologic study of partial hydatidiform moles and DNA triploid placentas. *Pathol Int.* 1994; 44:528–534. [PubMed: 7921197]
14. Fukunaga M. Early partial hydatidiform mole: prevalence, histopathology, DNA ploidy, and persistence rate. *Virchows Arch.* 2000; 437:180–184. [PubMed: 10993279]
15. Genest DR. Partial hydatidiform mole: clinicopathological features, differential diagnosis, ploidy and molecular studies, and gold standards for diagnosis. *Int J Gynecol Pathol.* 2001; 20:315–322. [PubMed: 11603213]
16. Chew SH, Perlman EJ, Williams R, et al. Morphology and DNA content analysis in the evaluation of first trimester placentas for partial hydatidiform mole (PHM). *Hum Pathol.* 2000; 31:914–924. [PubMed: 10987251]
17. Norris-Kirby A, Hagenkord JM, Kshirsagar M, et al. Abnormal villous morphology associated with triple trisomy of paternal origin. *J Mol Diagn.* 2010; 12:525–529. [PubMed: 20413680]
18. Redline RW, Hassold T, Zaragoza M. Determinants of villous trophoblastic hyperplasia in spontaneous abortions. *Mod Pathol.* 1998; 11:762–768. [PubMed: 9720505]
19. Ronnett BM, DeScipio C, Murphy KM. Hydatidiform moles: ancillary techniques to refine diagnosis. *Int J Gynecol Pathol.* 2011; 30:101–116. [PubMed: 21293291]
20. Hoffner L, Dunn J, Esposito N, et al. P57KIP2 immunostaining and molecular cytogenetics: combined approach aids in diagnosis of morphologically challenging cases with molar phenotype and in detecting androgenetic cell lines in mosaic/chimeric conceptions. *Hum Pathol.* 2008; 39:63–72. [PubMed: 17904617]
21. Hoffner L, Parks WT, Swerdlow SH, et al. Simultaneous detection of imprinted gene expression (p57(KIP2)) and molecular cytogenetics (FICTION) in the evaluation of molar pregnancies. *J Reprod Med.* 2010; 55:219–228. [PubMed: 20626178]
22. Kaiser-Rogers KA, McFadden DE, Livasy CA, et al. Androgenetic/biparental mosaicism causes placental mesenchymal dysplasia. *J Med Genet.* 2006; 43:187–192. [PubMed: 15908568]
23. Conran RM, Hitchcock CL, Popek EJ, et al. Diagnostic considerations in molar gestations. *Hum Pathol.* 1993; 24:41–48. [PubMed: 8418014]
24. Fukunaga M, Katabuchi H, Nagasaka T, et al. Interobserver and intraobserver variability in the diagnosis of hydatidiform mole. *Am J Surg Pathol.* 2005; 29:942–947. [PubMed: 15958860]
25. Howat AJ, Beck S, Fox H, et al. Can histopathologists reliably diagnose molar pregnancy? *J Clin Pathol.* 1993; 46:599–602. [PubMed: 8157742]

26. Javey H, Borazjani G, Behmard S, et al. Discrepancies in the histological diagnosis of hydatidiform mole. *Br J Obstet Gynaecol.* 1979; 86:480–483. [PubMed: 223623]
27. Messerli ML, Parmley T, Woodruff JD, et al. Inter- and intrapathologist variability in the diagnosis of gestational trophoblastic neoplasia. *Obstet Gynecol.* 1987; 69:622–626. [PubMed: 3029643]
28. Castrillon DH, Sun D, Weremowicz S, et al. Discrimination of complete hydatidiform mole from its mimics by immunohistochemistry of the paternally imprinted gene product p57KIP2. *Am J Surg Pathol.* 2001; 25:1225–1230. [PubMed: 11688455]
29. Chilosi M, Piazzola E, Lestani M, et al. Differential expression of p57kip2, a maternally imprinted cdk inhibitor, in normal human placenta and gestational trophoblastic disease. *Lab Invest.* 1998; 78:269–276. [PubMed: 9520940]
30. Crisp H, Burton JL, Stewart R, et al. Refining the diagnosis of hydatidiform mole: image ploidy analysis and p57KIP2 immunohistochemistry. *Histopathology.* 2003; 43:363–373. [PubMed: 14511255]
31. Fukunaga M. Immunohistochemical characterization of p57(KIP2) expression in early hydatidiform moles. *Hum Pathol.* 2002; 33:1188–1192. [PubMed: 12514787]
32. Fukunaga M. Immunohistochemical characterization of p57Kip2 expression in tetraploid hydropic placentas. *Arch Pathol Lab Med.* 2004; 128:897–900. [PubMed: 15270611]
33. Jun SY, Ro JY, Kim KR. p57kip2 is useful in the classification and differential diagnosis of complete and partial hydatidiform moles. *Histopathology.* 2003; 43:17–25. [PubMed: 12823708]
34. McConnell TG, Murphy KM, Hafez M, et al. Diagnosis and subclassification of hydatidiform moles using p57 immunohistochemistry and molecular genotyping: validation and prospective analysis in routine and consultation practice settings with development of an algorithmic approach. *Am J Surg Pathol.* 2009; 33:805–817. [PubMed: 19145201]
35. Merchant SH, Amin MB, Viswanatha DS, et al. p57KIP2 immunohistochemistry in early molar pregnancies: emphasis on its complementary role in the differential diagnosis of hydropic abortuses. *Hum Pathol.* 2005; 36:180–186. [PubMed: 15754295]
36. Popielek DA, Yee H, Mittal K, et al. Multiplex short tandem repeat DNA analysis confirms the accuracy of p57(KIP2) immunostaining in the diagnosis of complete hydatidiform mole. *Hum Pathol.* 2006; 37:1426–1434. [PubMed: 16949913]
37. Romaguera RL, Rodriguez MM, Bruce JH, et al. Molar gestations and hydropic abortions differentiated by p57 immunostaining. *Fetal Pediatr Pathol.* 2004; 23:181–190. [PubMed: 15768863]
38. Sarmadi S, Izadi-Mood N, Abbasi A, et al. p57KIP2 immunohistochemical expression: a useful diagnostic tool in discrimination between complete hydatidiform mole and its mimics. *Arch Gynecol Obstet.* 2011; 283:743–748. [PubMed: 20361201]
39. Bell KA, Van Deerlin V, Addya K, et al. Molecular genetic testing from paraffin-embedded tissue distinguishes nonmolar hydropic abortion from hydatidiform mole. *Mol Diagn.* 1999; 4:11–19. [PubMed: 10229770]
40. Bifulco C, Johnson C, Hao L, et al. Genotypic analysis of hydatidiform mole: an accurate and practical method of diagnosis. *Am J Surg Pathol.* 2008; 32:445–451. [PubMed: 18300805]
41. Lai CY, Chan KY, Khoo US, et al. Analysis of gestational trophoblastic disease by genotyping and chromosome in situ hybridization. *Mod Pathol.* 2004; 17:40–48. [PubMed: 14631372]
42. Murphy KM, Ronnett BM. Molecular analysis of hydatidiform moles: Utilizing p57 immunohistochemistry and molecular genotyping to refine morphologic diagnosis. *Pathol Case Rev.* 2010; 15:126–134.
43. Murphy KM, McConnell TG, Hafez MJ, et al. Molecular genotyping of hydatidiform moles: analytic validation of a multiplex short tandem repeat assay. *J Mol Diagn.* 2009; 11:598–605. [PubMed: 19815697]
44. Vang R, Gupta M, Wu LS, et al. Diagnostic reproducibility of hydatidiform moles: ancillary techniques (p57 immunohistochemistry and molecular genotyping) improve morphologic diagnosis. *Am J Surg Pathol.* 2012; 36:443–453. [PubMed: 22245958]
45. Diggle, P.; Heagerty, P.; Liang, KY., et al. *Analysis of Longitudinal Data.* New York: Oxford University Press; 2002.

46. McConnell TG, Norris-Kirby A, Hagenkord JM, et al. Complete hydatidiform mole with retained maternal chromosomes 6 and 11. *Am J Surg Pathol*. 2009; 33:1409–1415. [PubMed: 19542869]
47. Kerkmeijer LG, Massuger LF, Ten Kate-Booij MJ, et al. Earlier diagnosis and serum human chorionic gonadotropin regression in complete hydatidiform moles. *Obstet Gynecol*. 2009; 113:326–331. [PubMed: 19155902]
48. Chevillat JC, Greiner T, Robinson RA, et al. Ploidy analysis by flow cytometry and fluorescence in situ hybridization in hydropic placentas and gestational trophoblastic disease. *Hum Pathol*. 1995; 26:753–757. [PubMed: 7628847]
49. Chiang S, Fazlollahi L, Nguyen A, et al. Diagnosis of hydatidiform moles by polymorphic deletion probe fluorescence in situ hybridization. *J Mol Diagn*. 2011; 13:406–415. [PubMed: 21704275]
50. Fukunaga M, Endo Y, Ushigome S. Flow cytometric and clinicopathologic study of 197 hydatidiform moles with special reference to the significance of cytometric aneuploidy and literature review. *Cytometry*. 1995; 22:135–138. [PubMed: 7587744]
51. Gschwendtner A, Neher A, Kreczy A, et al. DNA ploidy determination of early molar pregnancies by image analysis: comparison to histologic classification. *Arch Pathol Lab Med*. 1998; 122:1000–1004. [PubMed: 9822129]
52. Kipp BR, Ketterling RP, Oberg TN, et al. Comparison of fluorescence in situ hybridization, p57 immunostaining, flow cytometry, and digital image analysis for diagnosing molar and nonmolar products of conception. *Am J Clin Pathol*. 2010; 133:196–204. [PubMed: 20093228]
53. Lage JM, Mark SD, Roberts DJ, et al. A flow cytometric study of 137 fresh hydropic placentas: correlation between types of hydatidiform moles and nuclear DNA ploidy. *Obstet Gynecol*. 1992; 79:403–410. [PubMed: 1371185]
54. LeGallo RD, Stelow EB, Ramirez NC, et al. Diagnosis of hydatidiform moles using p57 immunohistochemistry and HER2 fluorescent in situ hybridization. *Am J Clin Pathol*. 2008; 129:749–755. [PubMed: 18426735]
55. Lescoat D, Jouan H, Loeuillet-Olivo L, et al. Fluorescent in situ hybridization (FISH) on paraffin-embedded placental tissues as an adjunct for understanding the etiology of early spontaneous abortion. *Prenat Diagn*. 2005; 25:314–317. [PubMed: 15849799]
56. Sumithran E, Cheah PL, Susil BJ, et al. Problems in the histological assessment of hydatidiform moles: a study on consensus diagnosis and ploidy status by fluorescent in situ hybridisation. *Pathology*. 1996; 28:311–315. [PubMed: 9007948]
57. Yver M, Carles D, Bloch B, et al. Determination of DNA ploidy by fluorescence in situ hybridization (FISH) in hydatidiform moles: evaluation of FISH on isolated nuclei. *Hum Pathol*. 2004; 35:752–758. [PubMed: 15188143]
58. Berkowitz RS, Goldstein DP. Clinical practice. Molar pregnancy. *N Engl J Med*. 2009; 360:1639–1645. [PubMed: 19369669]
59. Soper JT, Mutch DG, Schink JC. Diagnosis and treatment of gestational trophoblastic disease: ACOG Practice Bulletin No. 53. *Gynecol Oncol*. 2004; 93:575–585. [PubMed: 15196847]

FIGURE 1.
Correct classification (sensitivity of diagnosis) of CHMs by faculty and fellow reviewers.

FIGURE 2.
Correct classification (sensitivity of diagnosis) of PHMs by faculty and fellow reviewers.

FIGURE 3.
Correct classification (sensitivity of diagnosis) of NMs by faculty and fellow reviewers.

FIGURE 4.

Correct classification of all cases combined by faculty and fellow reviewers.

FIGURE 5.

Correct classification by consensus diagnoses for each diagnostic category by faculty and fellow reviewers.

FIGURE 6.

Problematic cases (see Table 6 for details). A–D, Androgenetic diploid CHM, with negative p57 stain. This example was often misclassified by morphology as a PHM by both faculty and fellow reviewers (more often by faculty). It was recognized as a CHM by all reviewers with the assistance of the p57 stain. E–H, Androgenetic diploid CHM, with negative p57 stain. This example was sometimes misclassified by morphology as either an NM or PHM by faculty reviewers but was recognized as a CHM by the fellow reviewers. It was recognized as a CHM by all reviewers with the assistance of the p57 stain. I–L, Androgenetic diploid CHM, with negative p57 stain. This example was often misclassified by faculty reviewers as a PHM or NM (consistently so, with no consensus in either H&E round) but was recognized as a CHM by the fellow reviewers. It was recognized as a CHM by most reviewers with the assistance of the p57 stain. This is a rare example in which reviewers did not all agree on the interpretation of the p57 stain and did not all utilize the result to modify their morphologic diagnosis.

FIGURE 7.

Problematic cases (see Table 6 for details). A–D, Diandric triploid PHM, with positive p57 stain. This subtle example was most often misinterpreted as an NM, and the p57 stain did not influence the interpretation. E–H, Biparental diploid NM (with trisomies 16 and 21) with abnormal villous morphology and positive p57 stain. This example was often misinterpreted as a PHM, and the p57 stain did not influence the interpretation.

TABLE 1
Comparison of Faculty versus Fellow Diagnostic Performance by Round and Diagnostic Category (Depicted in Figs. 1–5)

Type of Diagnosis	Diagnostic Category	Round 1 (H&E)			Round 2 (H&E)			Round 3 (H&E+p57)		
		OR	(95% CI)	P	OR	(95% CI)	P	OR	(95% CI)	P
Individual pathologist*	CHM*	0.56	(0.27, 1.18)	0.13	0.85	(0.40, 1.80)	0.67	0.79	(0.37, 1.68)	0.54
Individual pathologist [†]	PHM [†]	1.88	(1.03, 3.46)	0.04	2.64	(1.49, 4.68)	<0.01	2.37	(1.30, 4.31)	<0.01
Individual pathologist [‡]	NM [‡]	1.00	(0.54, 1.85)	1.00	2.37	(1.33, 4.22)	<0.01	0.84	(0.44, 1.63)	0.61
Individual pathologist [§]	All cases [§]	1.08	(0.75, 1.54)	0.69	1.85	(1.32, 2.59)	<0.01	1.30	(0.91, 1.86)	0.15
Consensus//	CHM//	0.09	(0.01, 0.62)	0.01	0.17	(0.02, 1.69)	0.13	¶	¶	¶
Consensus//	PHM//	2.41	(1.01, 5.78)	0.05	3.45	(1.14, 10.41)	0.03	3.77	(1.44, 9.90)	0.01
Consensus//	NM//	1.00	(0.41, 2.46)	1.00	2.71	(1.13, 6.51)	0.03	0.54	(0.14, 2.07)	0.37
Consensus//	All cases//	0.89	(0.51, 1.55)	0.67	1.80	(0.98, 3.29)	0.06	1.44	(0.77, 2.79)	0.27

* See Figure 1.

[†] See Figure 2.

[‡] See Figure 3.

[§] See Figure 4.

// See Figure 5.

¶ Cannot be calculated due to identical high values for this category by consensus diagnosis for both sets of observers.

CI indicates confidence interval.

TABLE 2

Comparison of Diagnostic Performance Across all Pathologists Between p57 and H&E Rounds by Diagnostic Category (Depicted in Figs. 1–3)

Diagnostic Category	Round 3 vs. Round 1		Round 3 vs. Round 2	
	OR (95% CI)	<i>P</i>	OR (95% CI)	<i>P</i>
CHM [*]	4.85 (2.27, 10.42)	< 0.01	3.86 (1.81, 8.26)	< 0.01
PHM [†]	1.05 (0.69, 1.59)	0.81	1.32 (0.88, 1.99)	0.17
NM [‡]	1.49 (0.96, 2.31)	0.07	1.28 (0.83, 1.96)	0.26

^{*} See Figure 1.

[†] See Figure 2.

[‡] See Figure 3.

CI indicates confidence interval.

TABLE 3

Interobserver Agreement by Diagnostic Category and Experience Level in H&E Rounds

Diagnostic Category	κ (Interpretation)			
	Faculty Pathologists		Fellow Pathologists	
	Round 1	Round 2	Round 1	Round 2
CHM	0.59 (moderate)	0.73 (good)	0.46 (moderate)	0.61 (good)
PHM	0.15 (poor)	0.43 (moderate)	0.23 (fair)	0.31 (fair)
NM	0.13 (poor)	0.42 (moderate)	0.16 (poor)	0.36 (fair)

TABLE 4

Intraobserver Agreement (Between H&E Rounds 1 and 2) by Pathologist Across all Diagnostic Categories

	κ (Interpretation)	% Agreement
Faculty		
Pathologist 1	0.44 (moderate)	64%
Pathologist 2	0.66 (good)	79%
Pathologist 3	0.67 (good)	79%
Fellows		
Pathologist 4	0.53 (moderate)	69%
Pathologist 5	0.63 (good)	75%
Pathologist 6	0.37 (fair)	59%

TABLE 5
Diagnostic Performance of Faculty and Fellow Pathologists: Consensus versus Genotyping Diagnoses

Genotyping Diagnosis	Consensus diagnosis*				Total
	CHM	PHM	NM	None	
Faculty pathologists, round 1					
CHM	19	4	0	4	27
PHM	1	21	3	2	27
NM	0	10	16	0	26
Total	20	35	19	6	80
Faculty pathologists, round 2					
CHM	22	2	1	2	27
PHM	2	19	6	0	27
NM	1	6	19	0	26
Total	25	27	26	2	80
Fellow pathologists, round 1					
CHM	26	0	0	1	27
PHM	1	16	9	1	27
NM	3	5	16	2	26
Total	30	21	25	4	80
Fellow pathologists, round 2					
CHM	26	1	0	0	27
PHM	2	11	12	2	27
NM	4	8	13	1	26
Total	32	20	25	3	80

* Diagnosis rendered by 2 of 3 reviewers.

TABLE 6

Diagnostic Details for Case Examples (Figs. 6, 7)

Diagnoses/Interpretations *									
Figure	Faculty				Fellows				
	Round 1 (H&E)	Round 2 (H&E)	Round 3 (H&E+p57)	p57	Round 1 (H&E)	Round 2 (H&E)	Round 3 (H&E+p57)	p57	Genotyping Diagnosis
6A-D	CHM	PHM	CHM	Negative	CHM	CHM	CHM	Negative	CHM (androgenetic diploidy)
	PHM	CHM	CHM	Negative	PHM	PHM	CHM	Negative	
	PHM	PHM	CHM	Negative	NM	CHM	CHM	Negative	
6E-H	NM	NM	CHM	Negative	CHM	CHM	CHM	Negative	CHM (androgenetic diploidy)
	CHM	CHM	CHM	Negative	CHM	CHM	CHM	Negative	
	PHM	NM	CHM	Negative	CHM	CHM	CHM	Negative	
6I-L	PHM	PHM	CHM	Negative	CHM	CHM	CHM	Negative	CHM (androgenetic diploidy)
	CHM	CHM	CHM	Negative	CHM	CHM	NM	Positive	
	NM	NM	NM	Negative	CHM	NM	CHM	Negative	
7A-D	NM	NM	NM	Positive	NM	NM	PHM	Positive	PHM (diandric triploidy)
	PHM	NM	PHM	Positive	NM	NM	NM	Positive	
	NM	NM	NM	Positive	NM	NM	NM	Positive	
7E-H	PHM	NM	PHM	Positive	PHM	NM	NM	Positive	NM (biparental diploidy, trisomies 16 & 21)
	PHM	NM	PHM	Positive	CHM	NM	PHM	Positive	
	NM	NM	NM	Positive	NM	NM	NM	Positive	

* Individual diagnoses/interpretations listed for each group of observers in each round.