

# A validated protocol to quantify severity of male urogenital feminization using the MOUSE (Mouse objective urethral severity evaluation)

Ciro M. Amato<sup>1</sup> and Krista A. McCoy<sup>1</sup>

**BACKGROUND:** Congenital abnormalities vary in presentation, yet studies using model organisms tend to focus on occurrence rather than severity of the defect. Scoring severity of abnormalities in model systems allows explicit hypothesis testing during basic, translational, and reverse translational studies. We developed and validated a protocol to quantify severity of male urogenital feminization (hypospadias) in the mouse model. Hypospadias is one of the most common birth defects in the world.

**METHODS:** To induce genital feminization, pregnant mice were exposed to different concentrations of the antiandrogen vinclozolin. Genitalia were photographed at gestational age 18.5. A dichotomous scoring system to evaluate genital feminization was developed, and validated against histological measurements of urethral length. A training protocol was developed for novice scorers, and criteria were defined to evaluate precision and accuracy of scores.

**RESULTS:** Vinclozolin induced variation in hypospadias severity. Severity scores were tightly correlated with histologically determined urethral length and both techniques showed similar dose–response relationships. Novice observers were trained to precisely and accurately score hypospadias severity.

**CONCLUSION:** This standardized scoring system advances the mouse as a model to study urogenital development, and will facilitate research on the mechanisms driving genital feminization in males, and aid translational hypospadias research.

**M**orphological variation occurs because developmental and molecular mechanisms driving morphogenesis differ across genotypes and environments (1). Many traits, however, are studied in a binary context: present or absent; normal or abnormal. This framework limits our ability to build a complete, and integrated understanding of the mechanisms driving morphogenesis because developmental processes do not induce binary outcomes. The scoring system we develop and validate provides a fine scale quantification of the severity of a common genital birth defect so that the mechanisms driving differences in genital morphogenesis can be studied.

Congenital abnormalities induced by altered genital morphogenesis are surprisingly common in humans. For example, hypospadias is the second most common birth defect in the United States, occurring in almost 1% of males. In Denmark rates are as high as 4.6% (2). Hypospadias occurs when the urethra is shortened and opens along the shaft of the penis, rather than at the distal tip. It has a complex etiology and broad variation in its presentation (3). Hypospadias severity in newborns varies from requiring no intervention, to necessitating complete sex reassignment surgery and vaginoplasty (4).

In fact, the Hypospadias Objective Penile Evaluation (H.O.P.E.) scoring system is used, in humans, to evaluate the severity of several surgically correctable genitalia abnormalities (5) and facilitates evaluation of both preoperative options and postoperative outcomes. Determining the mechanisms driving the variation in genitalia abnormalities will advance our understanding of why they are occurring in humans, and aid development of therapies to reduce their occurrence and incidence. Such mechanistic studies are most efficiently conducted in established model systems that exhibit similar development and responsiveness to environmental perturbations. Mice are the accepted model for investigating urogenital development (6). As in humans, hypospadias severity varies in mice. However, current techniques to evaluate severity of genital feminization and hypospadias in the mouse require destructive techniques such as histology (7) or resin casting (8), which allow measurement of urethral length (which is shorter with hypospadias), but limit the use of tissues for genetic, physiological, or developmental studies.

Reproductive abnormalities in rodents increase in severity with mixtures of endocrine disrupting chemicals (9), and many studies suggest that the increase in incidence of hypospadias in humans is related to fetal endocrine disrupting chemicals exposure (10). Here we develop the Mouse Objective Urethral Severity Evaluation (M.O.U.S.E), as a standardized scoring system for the mouse that allows efficient and consistent comparisons of hypospadias severity to facilitate mechanistic studies, and comparisons across genetic and environmental manipulations. This scoring system is validated against urethral histology, the accepted endpoint used to evaluate hypospadias

<sup>1</sup>Department of Biology, East Carolina University, Greenville, North Carolina. Correspondence: [Ciro M. Amato \(amato.ciro27@gmail.com\)](mailto:amato.ciro27@gmail.com)

Received 29 March 2016; accepted 9 June 2016; advance online publication 7 September 2016. doi:10.1038/pr.2016.157

severity (11). We go further to show that scores are negatively associated with anogenital distance (AGD), a well accepted marker of masculinization. We provide a training protocol that includes guidelines for determining when observers can precisely and accurately score severity and show that scorers with different educational backgrounds can properly use the training protocol. M.O.U.S.E. provides a single measurement through which multiple laboratories can exchange information about how genetic manipulations or environmental treatments influence genital development.

## RESULTS

To develop M.O.U.S.E. as a standardized hypospadias severity scoring protocol, we needed to compare individuals with a range of hypospadias severities. To induce hypospadias at different severities, pregnant mouse dams ( $N = 3$ ) were exposed to one of four doses of the androgen antagonist vinclozolin (0, 100, 125, 150 mg/kg) every day for 4 d during the critical period for genitalia development (gestational age (GA) 13.5–16.5) (12). Genitalia of each pup (GA 18.5) was photographed in a standard position (see **Supplementary Methods** online), histologically processed, and evaluated based on predetermined landmarks (**Figure 1a,b**). Genital length scaled with urethral length at the same rate across all individuals ( $\chi^2 = 40.598$ ,  $P = 1.45e-07$ , slope ( $B_1$ ) = 0.3205). Across all treatments, males displayed 33% larger urethral lengths ( $\chi^2 = 108.47$ ,  $P = 2.2e-16$ ) than control (0 mg/kg vinclozolin) females. However, males from vinclozolin-treated groups exhibited a saturating dose-dependent shortening of urethral length ( $\chi^2 = 18.307$ ,  $P = 1.88e-05$ , **Figure 1c**) relative to control males.

Next, we created a dichotomous key with defined scoring criteria that distinguished among: normal (0), abnormal (1), feminized (2), and severely feminized (3) penis structure, based on comparing overall genital morphology and location of the urethral exit (**Figure 2** and **Supplementary Methods** online). We used the dichotomous key to evaluate pictures of genitalia generated in the dose response study. Analogous to urethral length, hypospadias severity using M.O.U.S.E scores showed a saturating dose-dependent increase in hypospadias severity between 0–150-mg/kg vinclozolin ( $\chi^2 = 12.783$ ,  $P = 2.858e-05$ , **Figure 3**).

To validate that M.O.U.S.E. accurately represents hypospadias severity, we regressed MOUSE severity scores to histologically determined urethral lengths and found a significant negative relationship; 97% of the variation in urethral length was explained by MOUSE hypospadias score ( $R^2 = 0.9718$ , slope =  $-1.5467$ ; **Figure 4a**). To further confirm that the M.O.U.S.E. scores provide valid information about masculinization, severity scores and AGD for male pups from each dam were separately averaged and then regressed against one another. Indeed, mean score was negatively related to mean AGD ( $R^2 = 0.6189$ ,  $P = 0.00146$ , **Figure 4b**). In fact, 62% of the variation in AGD could be explained by differences in hypospadias severity.

To determine if multiple people with diverse educational backgrounds could use the M.O.U.S.E scoring system, we

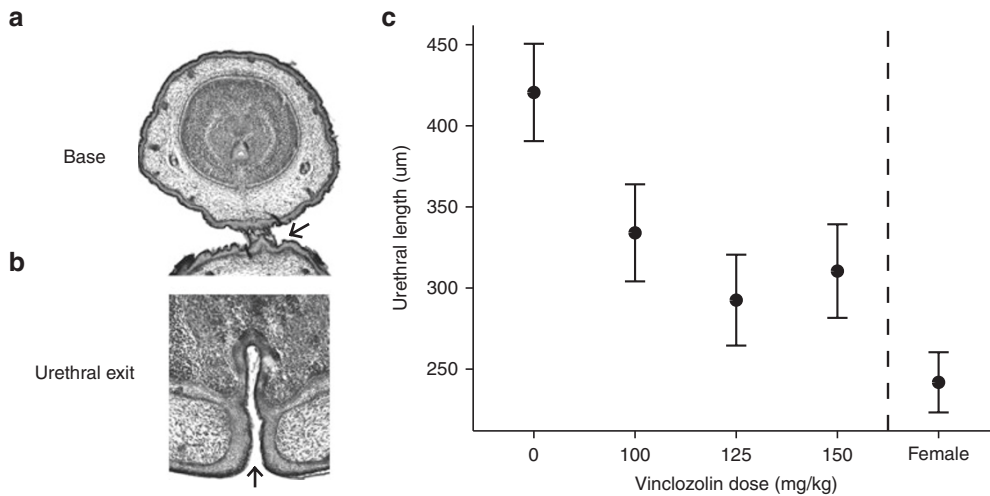
trained two novice scorers, with different educational experiences, (a biology graduate student and a high school English teacher) to score genitalia. Each person was trained for only 10 min prior to scoring. Important landmarks were indicated, and the trainer discussed the process of scoring using practice pictures, which were not included in the examination set (**Figure 2** and **Supplementary Methods** online).

Precision was evaluated in two ways. First, observers scored a single set of 24 sample photographs two times and a Pearson's correlation was run to determine if the two scoring attempts were consistent. A correlation coefficient of  $r = 0.8$  between the two attempts was required to be considered precise. Second, a paired  $t$ -test was run on the two scoring attempts to ensure that the difference in scores did not significantly deviate from zero (so that attempts were not different from one another). It is possible for scorers to be consistent but score differently between scoring sessions (e.g., consistently scoring lower in the second session). Quantifying precision in these two distinct ways is rigorous and necessary for standardized results. Accuracy was evaluated by comparing M.O.U.S.E scores to the histologically determined urethral length. To be considered accurate, urethral length and severity score had to show a significant negative relationship (urethral length decreases as severity increases) and have a  $R^2 \geq 0.95$  (see **Supplementary Methods** online and <http://thescholarship.ecu.edu/handle/10342/5650>, for more methodological detail, practice pictures and corresponding histological data).

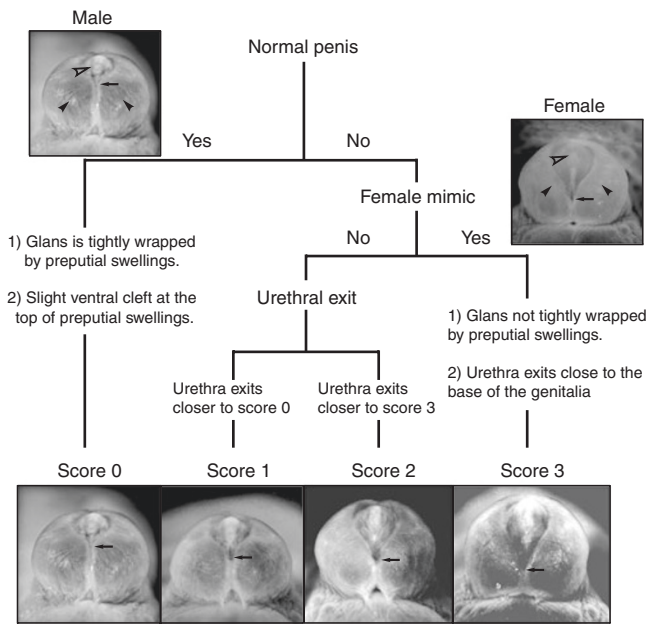
If the observers failed any precision or accuracy test, they were retrained on new practice pictures and underwent the scoring process a second time using new sample photographs. Both the graduate student and high school teacher required two training sessions and a total of 45 and 30 min (respectively) across both scoring sessions to meet the criteria described below (**Table 1**), at which point both individuals were considered capable of scoring hypospadias severity in future experiments. To facilitate standardized severity scoring across laboratories, we have provided a protocol for taking standardized photographs, a detailed training guide, genitalia photographs, and corresponding histological data in the **Supplementary Methods** online.

## DISCUSSION

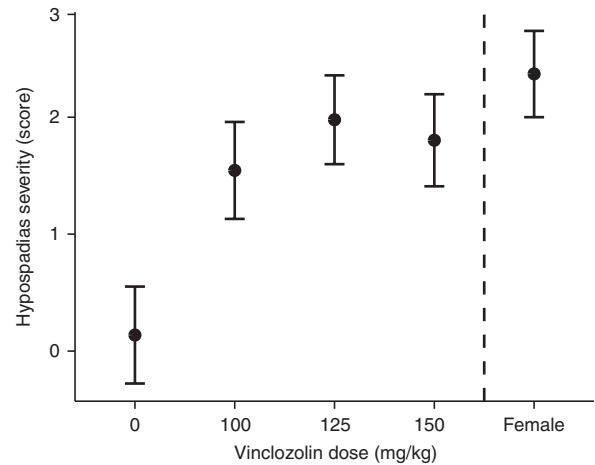
M.O.U.S.E. is an easy to use, standardized scoring system that allows accurate and precise quantification of hypospadias severity. All scoring is completed on photographs of the genitalia, and thus is nondestructive. The use of photographs to evaluate morphology is powerful for several reasons. First, pictures allow scorers to be blind to experimental treatment, pup sex, and dam. Second, the use of pictures frees the genitalia tissue to be used for mechanistic research. Individual samples can be dissected, photographed, and then processed in a variety of ways rather than having to be preserved for later randomization and scoring. Finally, photographs can be easily stored, shared, distributed, and scored multiple times to obtain average severity scores for each individual. Typical protein and mRNA preservation techniques require



**Figure 1.** Variation in hypospadias severity is induced by vinclozolin. (a,b) Histological images displaying the landmarks used for calculating urethral length. (a) Arrow indicates the first section where the penis is completely separated from the perineum and was recorded as the first section of the penis. (b) Urethral exit was counted as occurring on the first section where the urethral tube opens to the environment (arrow). (c) Urethral length exhibited a saturating dose response decreasing from 0–150 mg/kg ( $P = 1.88e-05$ ;  $N = 3$  dams). Error bars represent 95% confidence intervals.



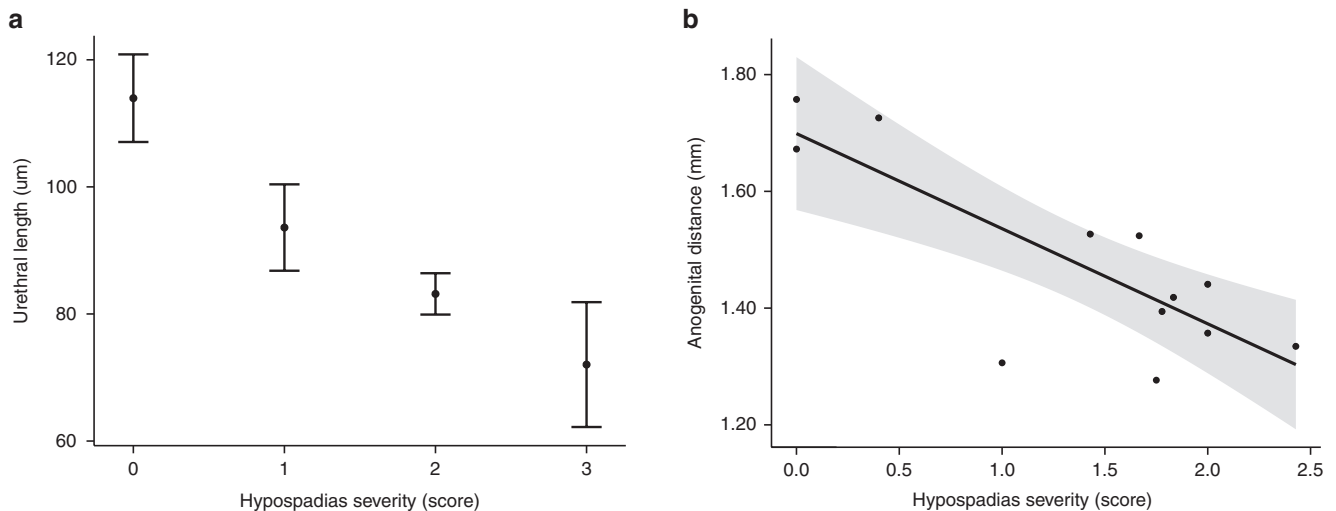
**Figure 2.** Evaluating hypospadias severity. Mouse objective urethral score evaluation (M.O.U.S.E.) is a dichotomous key that uses a stepwise method for scoring hypospadias severity in the mouse. The observer first asks if the penis is normal, defined as having a circular glans (hollow arrowhead) tightly wrapped by preputial swellings (solid arrowhead) and only a slight ventral cleft at the distal fusion of the preputial swellings (arrow). When the penis is considered normal, it is scored as zero. If the penis is not normal, the observer asks if the morphology mimics a female. In females the clitoral glans is not tightly wrapped by the preputial swellings, much of it is clearly visible, and the urethra exits below mid shaft (arrow in female picture). When the penis phenocopies a female genitalia, it is given a score of three. If the penis is neither normal nor completely feminized then it is scored as one, if the preputial swellings are wrapped around the glans of the penis forming an oval shaped glans but the urethral meatus is more proximal than normal. A score of two is recorded when the glans was not tightly wrapped, and the urethral exit occurs near mid shaft; see **Supplementary Methods** online for more details.



**Figure 3.** Scoring hypospadias severity. Hypospadias severity score scales significantly with vinclozolin dose ( $P = 2.858e-05$ ;  $N=3$  dams). Hypospadias severity exhibited a saturating dose response decreasing from 0–150 mg/kg similar to the pattern of urethral length (Figure 1). Females (scored blind) have the highest “hypospadias severity” when compared to males from each dose. Error bars represent 95% confidence intervals.

the tissues to be preserved in ways that change morphological structures and make multiple observations across time difficult.

Another advantage of scoring genitalia morphology using M.O.U.S.E. rather than relying on a single measure, such as position of the urethral opening (e.g., 2/3 of the way down the shaft), is that many aspects of morphology are integrated into a single M.O.U.S.E. score. Hypospadias is a disorder that affects more than urethral length. M.O.U.S.E. assesses: the extent of outgrowth of the preputial swellings, overall penis shape (relative to normal males and females), how tightly the preputial



**Figure 4.** Scoring validation. (a) Urethral length and M.O.U.S.E. scores are highly negatively correlated ( $R^2 = 0.97$ ), (b) Average dam score is negatively correlated with anogenital distance ( $R^2 = 0.6189$ ,  $P = 0.001$ ,  $N = 12$ ).  $N = 3$  dams/treatment; Score 0 ( $n = 12$  pups), Score 1 ( $n = 18$  pups), Score 2 ( $n = 37$  pups), and Score 3 ( $n = 5$  pups). Error bars and error envelope represent 95% confidence intervals.

**Table 1.** Members of the public and graduate students can be quickly trained to score hypospadias severity

Occupation	Trial	Time to complete	Precision (Pearson's $r$ )	Precision ( $P$ value)	Accuracy ( $R^2$ )
Graduate student	Trial 1	34 min	0.700 <sup>a</sup>	0.030 <sup>a</sup>	0.841 <sup>a</sup>
	Trial 2	14 min	0.860	0.714	0.991
High school teacher	Trial 1	18 min	0.850	0.001 <sup>a</sup>	0.929 <sup>a</sup>
	Trial 2	12 min	0.870	0.714	0.984

Each trainee required two trials to successfully score hypospadias severity.

<sup>a</sup>Indicates where trainee violated criteria.

swellings surround the glans, glans shape, as well as position and size of the urethral meatus. The increased resolution that M.O.U.S.E. provides for evaluating genitalia morphology will enhance basic and applied research.

#### Advancing the Mouse Model of Hypospadias

The M.O.U.S.E. scoring system provides a sensitive method to determine if penis morphology is affected even when incidence of hypospadias remains consistent across treatments. The increased sensitivity of M.O.U.S.E. will facilitate comparisons of hypospadias severity across studies using different genetic mutants and will help build a synthetic understanding of penile organogenesis and urethral tube closure. Developing a more synthetic knowledge of the molecular drivers of penile development in the mouse will facilitate our understanding of the mechanisms through which hypospadias occurs in humans.

This type of standardized scoring system is also essential to investigate the multifactorial nature of hypospadias. Evaluating the complex physiological and developmental processes involved in driving hypospadias severity is made possible when more discrete measurements are recorded and used to test explicit mechanistic hypotheses. M.O.U.S.E. scores can be quantitatively compared with measures of androgen levels,

gonadal function, AGD (Figure 4b), and other morphological or behavioral outcomes. With these types of data, we can integrate multiple morphological and physiological changes into a synthetic understanding of genital development and function.

M.O.U.S.E. will facilitate our understanding of the genetic and environmental drivers of hypospadias severity and advance translation of basic urogenital research (13). For example, human polymorphisms in several genes are significant risk factors for development of hypospadias (14–16). With M.O.U.S.E., we can conduct reverse translational work to determine which human gene mutations or combinations of mutations lead to more or less severe phenotypes in rodents, so that we can further study causal mechanisms. Furthermore, with M.O.U.S.E., researchers can test putative environmental factors associated with hypospadias in humans, and determine which induce the most severe hypospadias in the mouse. This approach will provide insight into what chemicals pregnant mothers should avoid. M.O.U.S.E. will also aid in development of therapies to reduce hypospadias severity. For example, previous epidemiological studies have shown that exposure to multivitamins generally (17) and folate specifically (18,19) reduces the risk hypospadias in newborns. M.O.U.S.E. will facilitate research identifying which nutritive supplements reduce the severity of hypospadias most dramatically.

The M.O.U.S.E. scoring system strengthens the mouse as model for evaluating the mechanisms driving congenital defects, and provides a means through which to evaluate sensitivity to specific chemicals and responses to therapeutic agents. Our work also advances the mouse as a model to study basic developmental and physiological processes that drive urogenital development.

#### METHODS

##### Mouse Maintenance and Treatment

All studies were approved by the East Carolina University Institutional Animal Care and Use Committee (AUP D-297). Eight-week-old CD-1 mice (Charles River Laboratories Raleigh, NC) were acclimated to 70–72 °F on a 12 h light–dark cycle with free access to food and



water (Purina ISOCHOW St. Hager City, WI) for at least 7 d prior to experimentation.

### Generating Variation in Hypospadias Severity

To generate differences in hypospadias severity, in the mouse, that phenocopies the variation seen in humans with hypospadias, we conducted a dose–response experiment fully replicated three times using vinclozolin (Sigma Aldrich, St. Louis, MO), a model antiandrogenic endocrine disrupting pesticide. Vinclozolin and its metabolites are known to competitively inhibit testosterone from binding to the androgen receptor and has been used to consistently induce hypospadias in previous studies (8,20,21). Acclimated nulliparous females were placed with male mice and checked every morning thereafter for vaginal plugs. Presence of a vaginal plug signified copulation occurred the previous night and that day at noon was considered GA 0.5. Pregnant dams ( $N = 3/\text{dose}$ ) were dosed on GA 13.5–16.5 with 0, 100, 125, or 150 mg/kg of vinclozolin. Tocopherol stripped corn oil (Millipore, Billerica, MA) was used as a vehicle and for vehicle control (0 mg/kg dose). During dosing, females were housed with same treatment dams or singly. To remove variation in developmental stage (age) caused by variation in day of birth (e.g., if parturition onset is affected by the treatments), females were humanly sacrificed on GA 18.5 and fetuses were removed. Sex was determined by examining the gonads (gonad morphology is not affected at the exposure time window—data not shown), and developmental stage of the embryo was verified by morphological evaluation (Theiler stage 26) (22). AGD was measured with a micrometer fitted on a Leica M80 stereoscope (Buffalo Grove, IL). Genitalia were removed, and photographed at 4 $\times$  zoom with a Leica M165FC stereoscope (Buffalo Grove, IL) using the z-stack add-on for Leica Application Suite software. Samples were then stored in 10% neutral buffered formalin (Fisher Scientific, Waltham, MA) until later histological processing. At GA 18.5 the genitalia are clearly sexually dimorphic, and hypospadias (shortened urethra which opens more proximally than normal) is observable both visually and histologically. To minimize the probability of misclassification of hypospadias severity, samples need to be photographed in a standard position. Pictures should be clear, especially where the urethra exits, and focus should be on the ventral aspect of the penis (for a full description see the **Supplementary Methods** online).

### Scoring Hypospadias Severity

When scoring, one observer, blind to treatment, scored all males in one attempt. All scoring was completed on photographs and not live or newly euthanized individuals. In practice, hypospadias severity is continuous and scores can technically lie between 1 and 2 or 2 and 3. In these situations, observers can use smaller score intervals (e.g., 1.5 or 2.5). Or, observers can provide the score that is best suited (e.g., closest integer score), and samples can be scored multiple times to obtain an average score per individual across attempts, which will more accurately estimate the true hypospadias score (we used this latter approach).

### Validating Scores

We validated the gross morphological scores of hypospadias severity by comparing them to histologically determined urethral lengths and AGD (an accepted biomarker of urogenital masculinization). Genitalia were dehydrated, infiltrated, and embedded vertically in paraffin wax, cross-sectioned at 10  $\mu\text{m}$ , and stained with hematoxylin and eosin. Each sample was sectioned entirely and data collected included the number of the section that contained the base of the penis, the urethral exit, and the tip of the penis (**Figure 1a,b**). The urethral length and total penis length were measured by counting sections from the base of the penis to exit of the urethra and the tip of the penis, respectively; then the total number of sections for each endpoint was multiplied by section thickness (10  $\mu\text{m}$ ). The base of the penis was defined as the first section not attached to perineal epidermis (**Figure 1a**). Urethral exit was the first section where the urethral epithelium opened to the environment, and adjacent epithelial walls were not touching (**Figure 1b**). The tip of the penis was identified as the last observable tissue on the slide.

Individuals are different sizes and thus have variable penis sizes, and the length of the urethra is dependent on penis size. To account

for this correlation, urethral length was corrected for penis size. To obtain the relationship between urethral length and penis size, we used a generalized linear random effects model, with dam treated as a random effect. Using urethra to penis length ratio is not an appropriate correction method due to the non-isometric relationship between penis size and urethral length at this developmental stage. To account for the allometry between urethral length ( $y$ ) and penis size ( $x$ ), we used a power law correction according to Equation 1 (23).

$$y/x^k = k \text{ (Equation 1)}$$

We used three approaches to validate our scoring system. First to insure that the scores we defined provided a fine enough scale to accurately classify individuals with different urethral lengths (the common marker for hypospadias) we visually compared the dose response generated from the M.O.U.S.E scores to the dose responses generated from the urethral lengths ( $\mu\text{m}$ ) determined histologically (**Figure 1c** and **Figure 3**). This allowed us to determine if a similar dose–response relationship was obtained with each data set. Second, to insure that each of our defined scores captured discrete nonoverlapping categories of urethral lengths, the histologically determined urethral lengths for individuals falling within each score severity category were averaged and regressed against score (**Figure 4a**). Here, if M.O.U.S.E scores capture true histological differences in urethral length, then score and mean urethral length should be highly related (**Figure 4a**). Third, we compared the M.O.U.S.E scores and AGDs to ensure that the hypospadias severity scores we generated were negatively associated. To test for a significant association between these measures, M.O.U.S.E scores and AGDs of male pups within each dam were averaged and regressed against one another.

### Validation of Our Training Protocol

To determine if multiple people with diverse educational backgrounds could use the standardized scoring system, we trained two individuals with different technical qualifications to score hypospadias severity using M.O.U.S.E. One individual was a graduate student in our laboratory who had no experience working on genitalia and the other was a high school English teacher with no formal scientific training. Each trainee was taught to recognize morphological landmarks, and score with a subset of photographs of mouse genitalia from the dose response experiment for which we had histologically validated urethral lengths (see **Supplementary Methods** online (<http://thescholarship.ecu.edu/handle/10342/5650>)). To assist researchers in learning how to follow the scoring protocol, we provide the detailed training guide, pictures, and corresponding histological data as **Supplementary Methods** online and **Supplementary Figure S1** online. Here, we detail the criteria used to evaluate whether a trainee is scoring properly and is ready to begin scoring experimental data.

After being trained, each scorer was provided with one set of 24 test pictures and asked to score each picture two times within the time span of 2 d. To be considered an adequate scorer, individuals had to show that their scores were both precise and accurate.

We determined precision in two ways. First, a Pearson's correlation between the two scoring attempts (for each scorer separately) was run to determine if the scorer was consistent in the way they scored single samples. A Pearson's correlation coefficient  $> 0.8$  was considered the threshold correlation coefficient. The second measure of precision was a paired  $t$ -test where we asked if the difference in scores obtained deviated significantly from zero. This tested whether the scores from the first and second scoring attempts were significantly different from one another and ruled out any potential positive or negative bias that cannot be detected in a Pearson's correlation (scores might be correlated but not the same). When the Pearson's correlation coefficient was  $> 0.8$ , and difference in the two scores did not significantly differ from zero the scoring was considered precise. If the individual showed acceptable internal consistency (precision) between scores, then we evaluated accuracy. If scores did not meet these criteria, scorers were retrained and provided with a second (different) set of test pictures.

Accuracy was evaluated by comparing the second (or last data set of) hypospadias scores, which is assumed to be the most accurate, to the histologically determined urethral length ( $\mu\text{m}$ ). If urethral length significantly negatively correlated with the observer M.O.U.S.E. score

(threshold  $R^2 = 0.95$  for averaged values) the trainee was deemed a trained precise and accurate hypospadias scorer. If the individual failed any of the precision or accuracy checkpoints they were retrained and asked to rescore a new set of 24 pictures, and undergo the process again. The trainee alternated between sets of randomly selected training pictures until they passed all the checkpoints (photographs, protocols, and histology data are provided in the **Supplemental Methods** online).

### Statistical Analysis

All data were analyzed using R statistical programming environment v. 3.1.2. Statistics used for scoring validation and within the training protocol are detailed above. Generalized linear mixed effects models (GLMM) from the lme4 R package (24) were used to analyze male and female urethral lengths as well as the dose effect on urethral length and hypospadias severity score in males. Dam was treated as a random effect. Stepwise likelihood ratio tests were used to evaluate the importance of parameter inclusion into the model. The effects package was used to extract the fitted values and confidence intervals from the GLMM (25).

### SUPPLEMENTARY MATERIAL

Supplementary material is linked to the online version of the paper at <http://www.nature.com/pr>

### ACKNOWLEDGMENTS

We thank J. Kelly and B. Blake for recording detailed notes and participating in the training protocol and A. Helms for providing helpful comments and edits. We would also like to thank M. McCoy for help with project development and data analysis.

### AUTHOR CONTRIBUTIONS

C.M.A. and K.A.M. conceived the project. C.M.A. conducted experimentation and tissue collection. C.M.A. and K.A.M. developed the M.O.U.S.E. protocol, analyzed the data, and wrote and edited the manuscript.

### STATEMENT OF FINANCIAL SUPPORT

This work was funded by East Carolina University's Division of Research and Graduate Studies and Thomas Harriot College of Arts and Sciences.

Disclosures: The authors declare no competing financial interests.

### REFERENCES

- Cooper WJ, Albertson RC. Quantification and variation in experimental studies of morphogenesis. *Dev Biol* 2008;321:295–302.
- Bergman Å, Heindel JJ, Jobling S, Kidd KA, Zoeller RT. State of the science of endocrine disrupting chemicals - 2012. *WHO* 2012:63–8.
- Kalfa N, Liu B, Klein O, et al. Mutations of CXorf6 are associated with a range of severities of hypospadias. *Eur J Endocrinol* 2008;159:453–8.
- Manzoni G, Bracka A, Palminteri E, Marrocco G. Hypospadias surgery: when, what and by whom? *BJU Int* 2004;94:1188–95.
- van der Toorn F, de Jong TP, de Gier RP, et al. Introducing the HOPE (Hypospadias Objective Penile Evaluation)-score: a validation study of an objective scoring system for evaluating cosmetic appearance in hypospadias patients. *J Pediatr Urol* 2013;9(6 Pt B):1006–16.
- Cunha GR, Sinclair A, Risbridger G, Hutson J, Baskin LS. Current understanding of hypospadias: relevance of animal models. *Nat Rev Urol* 2015;12:271–80.
- Yang JH, Menshenina J, Cunha GR, Place N, Baskin LS. Morphology of mouse external genitalia: implications for a role of estrogen in sexual dimorphism of the mouse genital tubercle. *J Urol* 2010;184(4 Suppl):1604–9.
- Buckley J, Willingham E, Agras K, Baskin LS. Embryonic exposure to the fungicide vinclozolin causes virilization of females and alteration of progesterone receptor expression in vivo: an experimental study in mice. *Environ Health* 2006;5:4.
- Rider CV, Furr J, Wilson VS, Gray LE Jr. A mixture of seven antiandrogens induces reproductive malformations in rats. *Int J Androl* 2008;31:249–62.
- Rocheleau CM, Romitti PA, Dennis LK. Pesticides and hypospadias: a meta-analysis. *J Pediatr Urol* 2009;5:17–24.
- Kim KS, Torres CR Jr, Yucel S, Raimondo K, Cunha GR, Baskin LS. Induction of hypospadias in a murine model by maternal exposure to synthetic estrogens. *Environ Res* 2004;94:267–75.
- Silversides DW, Price CA, Cooke GM. Effects of short-term exposure to hydroxyflutamide in utero on the development of the reproductive tract in male mice. *Can J Physiol Pharmacol* 1995;73:1582–8.
- Khoury MJ, Gwinn M, Yoon PW, Dowling N, Moore CA, Bradley L. The continuum of translation research in genomic medicine: how can we accelerate the appropriate integration of human genome discoveries into health care and disease prevention? *Genet Med* 2007;9:665–74.
- van der Zanden LF, van Rooij IA, Feitz WF, Franke B, Knoers NV, Roeleveld N. Aetiology of hypospadias: a systematic review of genes and environment. *Hum Reprod Update* 2012;18:260–83.
- Carmichael SL, Mohammed N, Ma C, et al. Diacylglycerol kinase K variants impact hypospadias in a California study population. *J Urol* 2013;189:305–11.
- Ogata T, Laporte J, Fukami M. MAMLD1 (CXorf6): a new gene involved in hypospadias. *Horm Res* 2009;71:245–52.
- Czeizel AE. The primary prevention of birth defects: multivitamins or folic acid? *Int J Med Sci* 2004;1:50–61.
- Mavrogenis S, Urban R, Czeizel AE, Ács N. Possible preventive effect of high doses of folic acid for isolated hypospadias: a national population-based case-control study. *Am J Med Genet A* 2014;164A:3108–14.
- Ormond G, Nieuwenhuijsen MJ, Nelson P, et al. Endocrine disruptors in the workplace, hair spray, folate supplementation, and risk of hypospadias: case-control study. *Environ Health Perspect* 2009;117:303–7.
- Kelce WR, Monosson E, Gamcsik MP, Laws SC, Gray LE Jr. Environmental hormone disruptors: evidence that vinclozolin developmental toxicity is mediated by antiandrogenic metabolites. *Toxicol Appl Pharmacol* 1994;126:276–85.
- Vilela ML, Willingham E, Buckley J, et al. Endocrine disruptors and hypospadias: role of genistein and the fungicide vinclozolin. *Urology* 2007;70:618–21.
- Theiler K. The house mouse- Atlas of embryonic development: Springer; 1989.
- Leonart J, Salat J, Torres GJ. Removing allometric effects of body size in morphological analysis. *J Theor Biol* 2000;205:85–93.
- Bates D, Machler M, Bolker BMM, and Walker SC. Fitting linear mixed-effects models using lme4. *R News* 2005:27–30.
- Fox J. Effect displays in R for generalised linear models. *J Stat Softw* 2003;8:1–19.