Reproductive outcomes predicted by phase imaging with computational specificity of spermatozoon ultrastructure

Mikhail E. Kandel[a,b,1], Marcello Rubessa[c,1], Yuchen R. He[a,b], Sierra Schreiber[a,c], Sasha Meyers[a,c], Luciana Matter Naves[c], Molly K. Sermersheim[c], G. Scott Sell[d], Michael J. Szewczyk[a], Nahil Sobh[a], Matthew B. Wheeler[a,c,e,1], and Gabriel Popescu[a,b,e,1,2]

[a]Beckman Institute for Advanced Science and Technology, The University of Illinois at Urbana–Champaign, Urbana, IL 61801; [b]Department of Electrical and Computer Engineering, The University of Illinois at Urbana–Champaign, Urbana, IL 61801; [c]Department of Animal Sciences, The University of Illinois at Urbana–Champaign, Urbana, IL 61801; [d]Edisto Research and Education Center, Clemson University, Blackville, SC 29817; and [e]Department of Bioengineering, The University of Illinois at Urbana–Champaign, Urbana, IL 61801

Edited by John A. Rogers, Northwestern University, Evanston, IL, and approved June 12, 2020 (received for review January 29, 2020)

The ability to evaluate sperm at the microscopic level, at high-throughput, would be useful for assisted reproductive technologies (ARTs), as it can allow specific selection of sperm cells for in vitro fertilization (IVF). The tradeoff between intrinsic imaging and external contrast agents is particularly acute in reproductive medicine. The use of fluorescence labels has enabled new cell-sorting strategies and given new insights into developmental biology. Nevertheless, using extrinsic contrast agents is often too invasive for routine clinical operation. Raising questions about cell viability, especially for single-cell selection, clinicians prefer intrinsic contrast in the form of phase-contrast, differential-interference contrast, or Hoffman modulation contrast. While such instruments are nondestructive, the resulting image suffers from a lack of specificity. In this work, we provide a template to circumvent the tradeoff between cell viability and specificity by combining high-sensitivity phase imaging with deep learning. In order to introduce specificity to label-free images, we trained a deep-convolutional neural network to perform semantic segmentation on quantitative phase maps. This approach, a form of phase imaging with computational specificity (PICS), allowed us to efficiently analyze thousands of sperm cells and identify correlations between dry-mass content and artificial-reproduction outcomes. Specifically, we found that the dry-mass content ratios between the head, midpiece, and tail of the cells can predict the percentages of success for zygote cleavage and embryo blastocyst formation.

assisted reproduction | quantitative phase imaging | phase imaging with computational specificity | machine learning | sperm

Males are responsible for more than 40% of the total infertile couples (1). The ability to evaluate sperm at the microscopic level would be useful for assisted reproductive technologies (ARTs), as it can allow specific selection of sperm cells for in vitro fertilization (IVF). For example, it has been shown that intracytoplasmic morphologically selected sperm injection (IMSI) improves the outcome of in vitro fertilization, as compared with the conventional intracytoplasmic sperm injection (ICSI) (2, 3). Previous studies suggested that the anomalies detected by IMSI are due to abnormalities in chromatin packaging (4). It is also important to note that the chemical reaction of the stain or label and the required fixative change the shape and size of spermatozoa (5–7). Due to the potential toxic effects of fluorophores, morphology of sperm cells is observed using intrinsic contrast imaging, such as differential interference contrast (DIC) (8), phase-contrast (9), or Hoffman modulation contrast (10) microscopy. These methods generate contrast from local variations in refractive index across the cell. As a result, these technologies do not require fluorescence tagging and are relatively harmless. However, the relationship between the intensity map generated by these methods and the properties of the cell (e.g., thickness, mass) are only qualitative.

Quantitative phase image (QPI) (11) is a label-free approach that has been gaining significant interest in the biomedical community (for a recent review, see ref. 12). QPI combines intrinsic contrast microscopy with holography and, as a result, can render precise information about the optical path-length map introduced by the specimen. This quantitative phase image is linearly related to the dry-mass density of the cell under investigation (13–15). As a result, QPI has been applied successfully to studying cell-mass transport and growth (16–20). Furthermore, acquiring QPI data over a third axis, e.g., axially, angularly, spectrally resolved, allows us to extract tomographic reconstructions of unlabeled cells (21–25). Due to these capabilities, recently, QPI has been applied to imaging sperm cells as well (26–29). In particular, several studies used QPI for measuring sperm motility (27, 30–33).

Significance

The high incidence of human male factor infertility suggests a need for examining new ways of evaluating sperm cells. We present an approach that combines label-free imaging and artificial intelligence to obtain nondestructive markers for reproductive outcomes. Our phase-imaging system reveals nanoscale morphological details from unlabeled cells. Deep learning, on the other hand, provides a structural specificity map segmenting with high accuracy the head, midpiece, and tail. Using these binary masks applied to the quantitative phase images, we measure precisely the dry-mass content of each component. Remarkably, we found that the dry-mass ratios represent intrinsic markers with predictive power for zygote cleavage and blastocyst development.

Author contributions: M.E.K., M.R., M.B.W., and G.P. designed research; M.E.K., M.R., Y.R.H., S.S., S.M., L.M.N., M.K.S., G.S.S., M.J.S., and N.S. performed research; M.R. and M.B.W. contributed new reagents/analytic tools; M.E.K., Y.R.H., and N.S. analyzed data; M.E.K., M.R., Y.R.H., M.B.W., and G.P. wrote the paper; and M.B.W. and G.P. managed the project.

Competing interest statement: G.P. has a financial interest in Phi Optics, Inc., a company developing quantitative phase-imaging technology for materials and life science applications.

This article is a PNAS Direct Submission.

[1]M.E.K., M.R., M.B.W., and G.P. contributed equally to this work.

[2]To whom correspondence may be addressed. Email: gpopescu@illinois.edu.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2001754117/-/DCSupplemental.

However, as an intrinsic contrast method, QPI also suffers from lack of molecular specificity. Our solution to this challenge has been to implement QPI as an add-on module to an existing microscope stand with fluorescence or bright-field capability (16). This way, the QPI and fluorescence (or staining) channels integrate seamlessly, providing both specific and quantitative information (for a recent implementation for pathology applications, see ref. 34). Of course, this solution does not avoid the fact that labels are needed to gain specificity, which brings in limitations due to fixation, photobleaching, and phototoxicity (35, 36).

Here, we show that using highly sensitive QPI in combination with deep learning allows us to identify subcellular compartments of unlabeled bovine spermatozoa. The deep-learning "semantic segmentation" model automatically segments the head, midpiece, and tail of individual cells. We used these predictions to measure the respective dry mass of each component. We found that the relative mass content of these components correlates with the zygote cleavage and embryo quality. The dry-mass ratios, i.e., head/midpiece, head/tail, midpiece/tail, can be used as intrinsic markers for reproductive outcomes. Because the QPI module can be attached to existing microscopes, we anticipate that our approach will be adopted broadly. This paper is structured as follows. We first present first the study design and instrumentation. Next, we illustrate the slide scanning capability of our automated spatial light-interference microscopy (SLIM) system (see also Movie S1). Then, we describe in detail how we optimized a U-Net–based deep-learning model for the semantic segmentation of individual cells. Once the deep-learning network is trained, we show that the inference can be performed very fast on thousands of cells across entire microscope slides. Finally, we identify the intrinsic dry mass-based markers that correlate with fertility outcomes.

Results

Imaging Spermatozoon Ultrastructure. Fig. 1*A* illustrates the principle of the proposed phase imaging with computational specificity (PICS). The semen from five bulls was processed for fertilization, and aliquots were fixed on microscope slides for imaging. To investigate the relationship between the ultrastructure of sperm cells and artificial-reproduction outcomes, we collected semen samples from frozen-thawed straws used for IVF for which the rates of zygote cleavage and embryo blastocyst development were known. The specimen preparation is detailed in *Materials and Methods*.

To image the unlabeled spermatozoa, we employed SLIM (Fig. 1*B*), which is a high-resolution, high-sensitivity phase imaging method developed in our laboratory (16, 24, 37). The SLIM system consists of a commercial module (CellVista SLIM Pro; Phi Optics., Inc) attached to the output port of an existing phase-contrast microscope (Z1 Axio Observer; Zeiss). SLIM augments traditional phase-contrast microscopy by acquiring four intensity images corresponding to different delays between the incident and scattered light. Combining the four frames, SLIM outputs in real time the quantitative phase map associated with the specimen. Due to the white light illumination, SLIM lacks speckles, which yields subnanometer path-length spatial sensitivity. The interference is created by waves propagating on a nearly identical path, which confers SLIM subangstrom path-length temporal sensitivity (37). Owing to these features and complete automation of data acquisition, SLIM can be used to study cell behavior of long periods of time and large fields of view across microscope slides and multiwell plates (18, 34). To demonstrate the potential for SLIM to be used in a high-throughput setting, we imaged a whole microscope slide prepared according to *Materials and Methods*. In Fig. 2*A*, we assembled a mosaic made from 27,000 tiles, covering an area of 7.1 cm^2 (Movie S1). The imaging time for the slide was ∼90 min (40×/0.75). The sensitivity in SLIM comes at the expense of a coherence-related halo-artifact present in phase-contrast geometries, which is removed by a Hilbert transform-based technique (Fig. 2*B*) (38, 39).

To illustrate the ability of our instrument to capture subtle cellular structures, we reconstruct a representative sperm cell from a series of through-focus measurements (z-stack). Fig. 2*C* shows the reconstructed cell tomography, which follows the procedure that we reported previously (24, 40). Various cellular

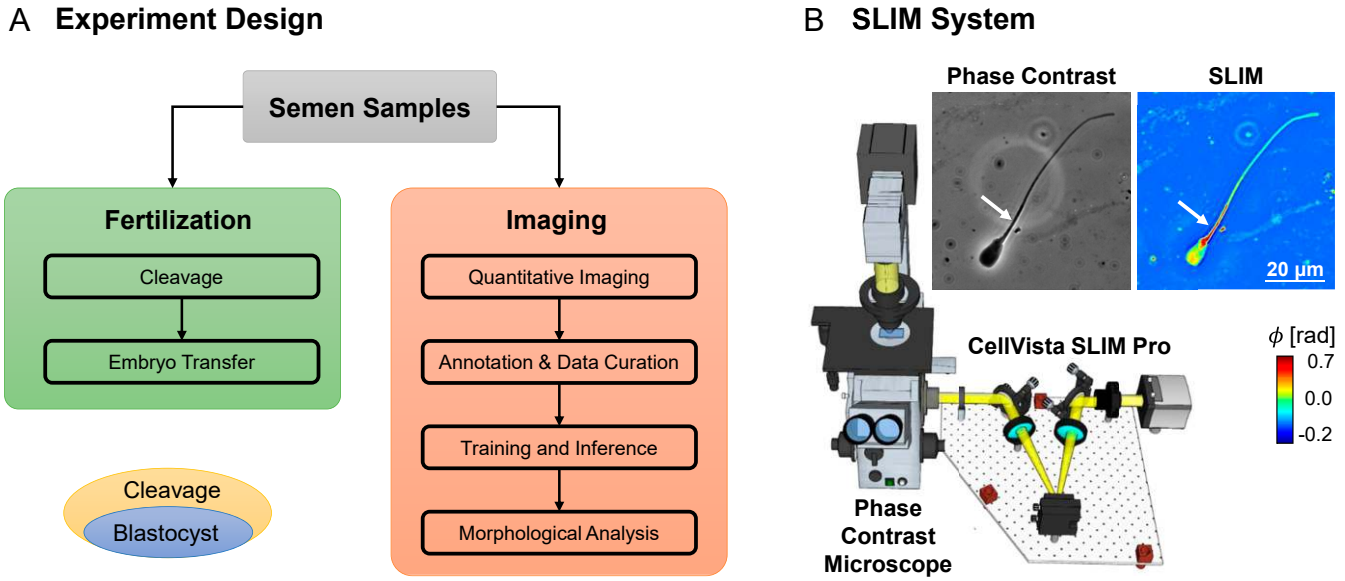


Fig. 1. Experiment design and SLIM. (*A*) To find the relationship between dry-mass and assisted-reproduction outcomes, semen samples from the same ejaculate were used for embryo transfer and quantitative phase imaging. For each animal, a fraction of the embryos underwent zygote cleavage (fertilization), from which a smaller fraction formed blastocysts. To assay cellular ultrastructure, we performed high-resolution quantitative phase imaging and developed a deep-convolutional neural network to perform automatic annotation. The results of this network were used to relate the morphology to the aforementioned outcomes. (*B*) The SLIM system upgrades a conventional phase-contrast microscope with the ability to measure optical path-length shifts. Compared with typical phase contrast, SLIM highlights the mitochondria-rich midpiece as a denser portion of the cell (white arrow).

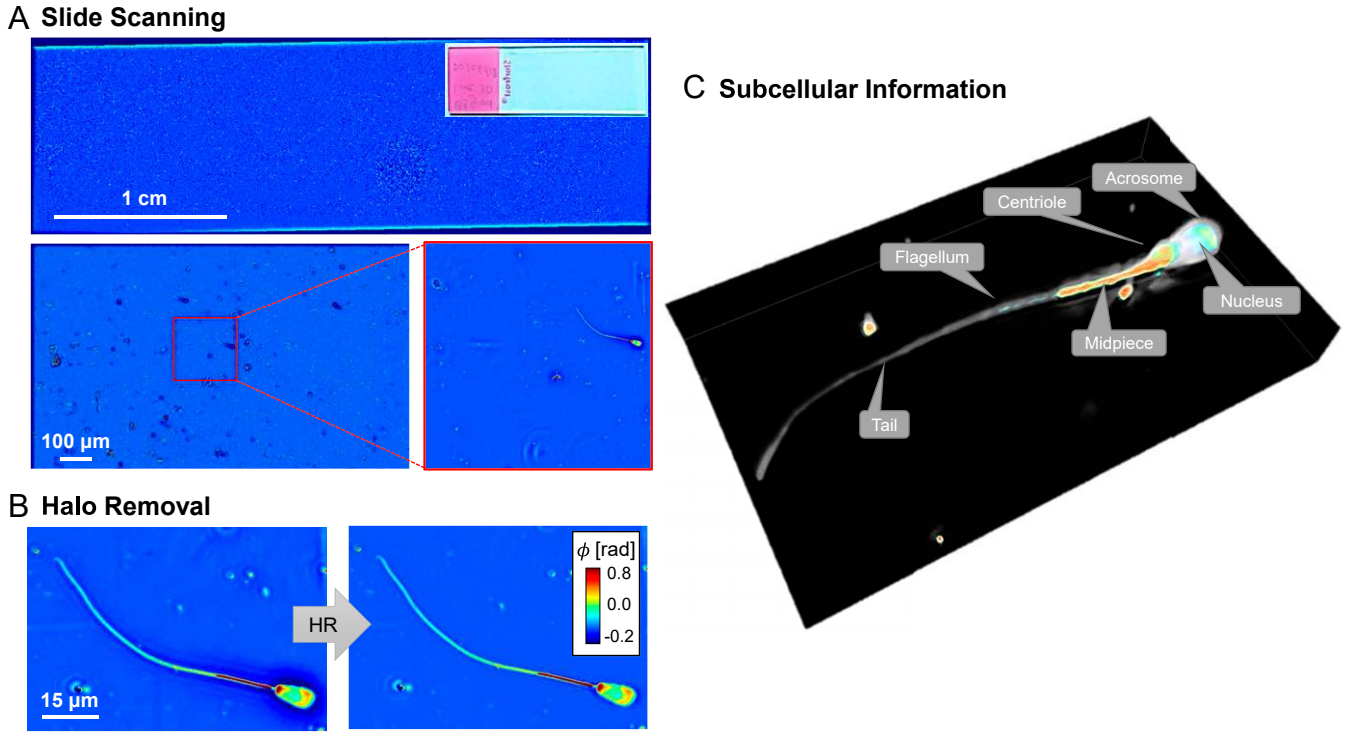


Fig. 2. SLIM can image sperm as a fully automated slide scanner, with thousands of samples on each slide. (*A*) A large number of samples in each slide motivates the use of automated segmentation techniques. (*B*) The superior sensitivity of SLIM images is, in part, due to the use of spatially and temporally broadband fields. The partially coherent illumination corrupts the low frequencies, evident as a halo glow surrounding the cell. The halos are corrected by solving a nonlinear inverse problem. (*C*) Tomographic rendering of a spermatozoon using SLIM. The mitochondria-rich midpiece appear as substantially higher in dry-mass density. Rendering of the tomogram was performed using Amira (Thermo Fisher) with the "physics" color map corresponding to high phase values and a grayscale color map corresponding to the lower-phase values in the nucleus and tail.

compartments are revealed with high resolution and contrast. We observe that the highest-density region of the sperm is the mitochondria-rich neck (or midpiece), which is connected to a denser centriole vault leading to the head (41). The posterior of the sperm consists of a flagellum followed by a less dense tail.

Deep Learning for Semantic Segmentation. To analyze microscope slides, we develop an artificial-intelligence (AI) system to label the pixels in the image as "head," "midpiece," "tail," or "background." The complete workflow and U-Net architecture are presented in Fig. 3. The challenges for developing such a system are twofold. First, the system must be trained on annotated images that are labor-intensive to generate and necessarily include bias from the individual annotators. As outlined in *Materials and Methods*, we developed a "bootstrapping" image-segmentation approach that ameliorates these concerns by annotating a subset of images that are then used to train the initial classification system. The results of this initial classification form a larger set of training images that can be quickly corrected for coarse defects and used to retrain the classification system. This approach, which produces a larger training set in a shorter period than annotating finely the entire training corpus, resembles a manual form of generative adversarial networks (42).

The second challenge is developing an efficient classification system. Our classification is based on deep-convolutional neural networks (43), which apply a sequence of nonlinear image filters that result in a transformed image. In contrast to previous techniques, such as the texture frameworks used in, for example ref. 44, deep-convolutional networks add pooling layers to efficiently integrate textural and contextual information while maintaining a large number of learnable free parameters through subsequent, hidden or "deep" layers. As we aim to incorporate both local textural and contextual information, we select a U-Net architecture where textural information is learned in the top layers and contextual information is learned at the bottom layers (45, 46).

The construction of the network is described in detail in *Materials and Methods*. As shown in *SI Appendix*, we achieve an F1-score above 0.85 after only 30 training epochs (47). This network is then used to annotate the sperm cells across all slides, resulting in a semantic map for every SLIM image. Additionally, we used the connected coordinate analysis (44) to generate instance segmentation, which separates individual sperm cells, enabling us to group the labels on a per-cell basis (48). This step also removes sperm cells that are stuck together.

Analysis of Cellular Compartments. Following the procedure outlined in ref. 18, the dry mass of the cellular structures can be integrated within the margin's given by our computationally generated labels to yield the total dry mass for the three cellular compartments. The dry mass for these elements can then be related to the known fertility outcomes for the sperm sample from which they were drawn.

We measured the dry-mass ratios between the head, midpiece, and tail, as the absolute dry-mass values showed no statistically significant correlations with outcomes. To investigate the cells that participate in fertilization, we excluded damaged or otherwise incomplete compartments by removing the bottom and top quarter of all detected objects. Interestingly, we found that the dry-mass ratios between the three components of interest are distributed relatively narrowly across the population (Fig. 4). We verify this observation by computing the percentage

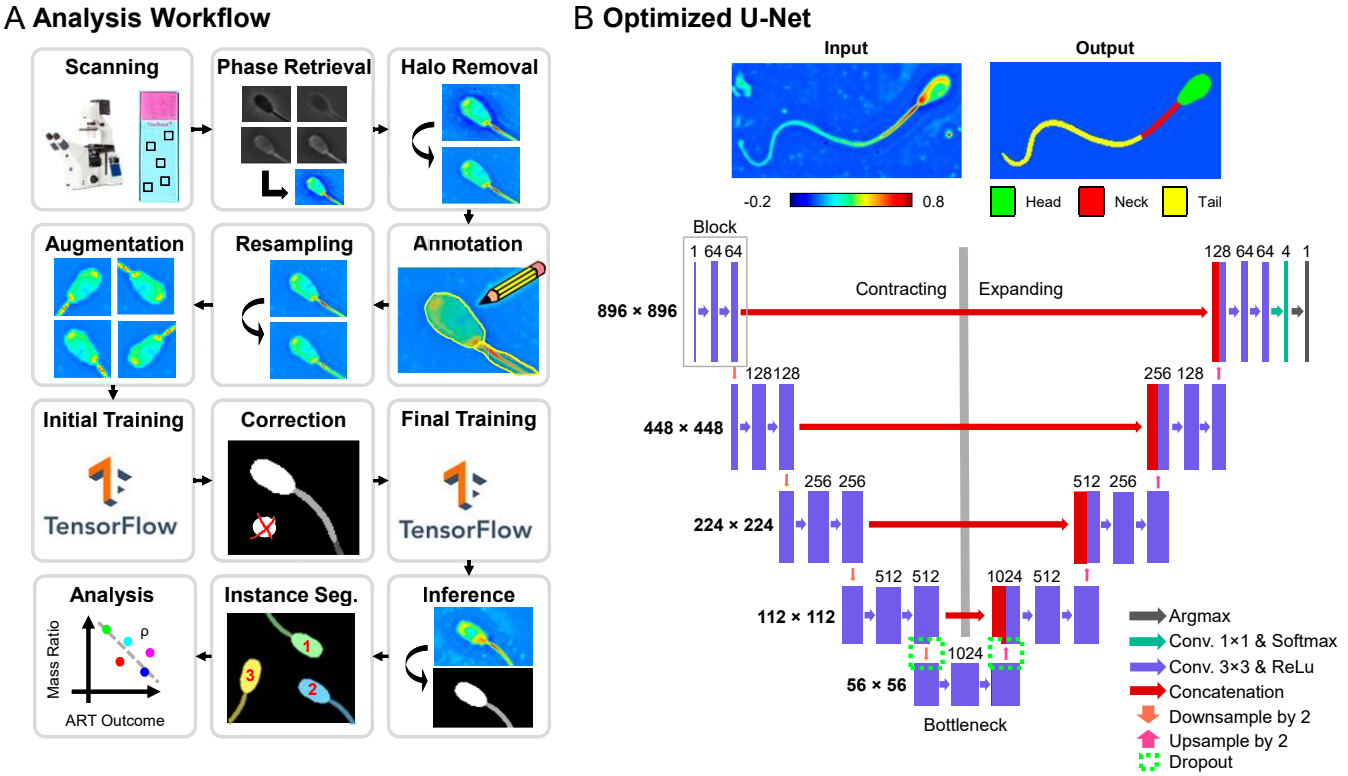


Fig. 3. Workflow for training a deep-convolutional neural network on quantitative phase images and inferring the semantic segmentation. (*A*) For each slide, we performed z-stacks, selecting the fields of view with intact sperm cells. We recover the phase from the four intensity images and perform halo removal to account for the partially coherent illumination. Next, we use ImageJ to manually segment the cells into the head, midpiece, and tail. We downsample the images to match the optical resolution and perform data augmentation by rotation and flipping. Training is performed using the TensorFlow backend for Keras. To boost the performance of our network, we correct for grossly defective segmentation using the Image Segmented App (part of MATLAB). This step is substantially faster than manually annotating every cell. A final training round is then performed using all data. By evaluating the network (inference) on all images, we obtain the segmentation results for all of the cells, which are grouped into individual sperm cells using connected-coordinate analysis. Finally, these results were used to determine the relationship between the dry mass of cellular ultrastructure and ART success rates. (*B*) Semantic segmentation converts phase maps into a binary mask corresponding to the head, midpiece, and tail. The U-Net architecture performs well on our data, as it contains a large receptive field, well suited for the rich, broadband images typically found in microscopy. The U-Net architecture consists of a series of nonlinear operations as outlined in *Materials and Methods*. In our implementation, we modify the training procedure for the network by introducing dropout at the bottom of the network as well as batch-normalization on all paths. The network results in a four-channel image with the probability for each of the four classes ("head," "midpiece," "tail," "background"), the largest of which assigned the label for the class.

difference between the first, Q_1, and third, Q_3, quartile points of the distributions,

$$\Delta Q = 100 \times \frac{|Q_3 - Q_1|}{\left[\frac{Q_3 + Q_1}{2}\right]}. \quad [1]$$

The quartile points, Q, are determined by finding $F_x(x) = \{0.25, 0.50, 0.75\}$, with $F_x(x)$ representing the cumulative probability, $F_x(x) = \Pr[X \leq x]$, of the dry-mass distribution, X.

The results in terms of percentages of zygotes cleaved and expanded blastocysts produced are summarized in Fig. 5*A*. Statistically significant correlations between dry-mass ratios and fertility outcomes are shown in bold. The graphs in Fig. 5 *B* and *C* show these findings in more detail. Essentially, cleavage outcomes correlate negatively with head/tail and midpiece/tail dry-mass ratios but not statistically significantly for head/midpiece ratios. These results suggest that a relatively more massive sperm tail is beneficial. However, when we evaluate the blastocyst development rate, it appears that a large head-to-midpiece value is desirable, while the other two ratios are only weakly correlated. This result appears to indicate that a denser head relative to the midpiece promotes blastocyst development. Note that this subgroup of spermatozoa that are associated with the embryo blastocyst development rate also has, with a high probability, larger tails (Fig. 5*B*).

Discussion

The relatively high incidence of male factor infertility may suggest a need for examining new ways of evaluating male gametes (49). In this context, we presented an approach that combines label-free imaging and artificial intelligence to obtain nondestructive markers for reproductive outcomes. Our phase imaging system, SLIM, reveals nanoscale morphological details from unlabeled cells. Deep learning, on the other hand, provides a semantic segmentation map, labeling with high accuracy the head, midpiece, and tail. Using these annotations applied to the quantitative phase images, we can precisely measure the dry-mass content of each component. Remarkably, we found that the dry-mass ratio represents intrinsic markers with predictive power for zygote cleavage and embryo blastocyst development.

The trend for relatively more massive head leading to poorer performance during early stages of embryo production could, in part, be attributed to the need for sperm cells to be capacitated before participating in fertilization (50). In other words, when a population of sperm cells appears to contain a relatively denser

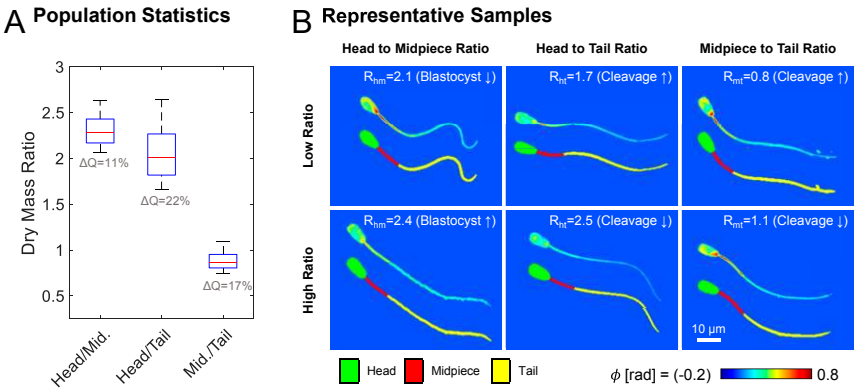
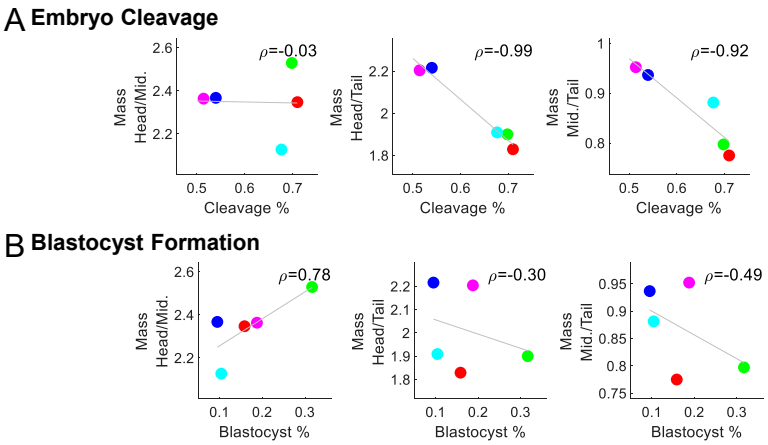


Fig. 4. Deep learning tracks subtle but significant differences in sperm morphology. (*A*) Histogram of the distribution of sperm dry-mass ratios shows that structural differences between sperm cells are relatively narrow, as evidenced by the percentage difference between the first and third quartile for the head to midpiece, head to tail, midpiece to tail are (only) 11, 24, and 17%, respectively. (*B*) Dry-mass maps of representative sperm cells along with semantic segmentation are labeled with their dry-mass ratios (head to midpiece [R_{hm}], head to tail [R_{ht}], midpiece to tail [R_{mt}]). Additionally, ↑ denotes an increase or ↓ represents a decrease in ART outcome, as determined in Fig. 5 and *SI Appendix*. These differences are especially difficult to visualize with conventional techniques as typical microscope images are not proportional to dry mass and the naked eye is unable to segment, integrate, and divide portions of an image (40×/0.75; SLIM).

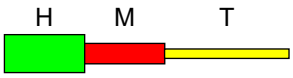
head, they are also more difficult to capacitate. Another interpretation of our findings is that more massive heads are harder to "push" with smaller tails, which is consistent with findings concerning head shape (51, 52). Likewise, it was observed that deer sperm with longer midpieces swim at an overall slower speed (53). Our results are compatible with this line of thought by confirming that such cells not only are swimming more slowly but also result in poorer fertility. This provides additional evidence for the hypothesis that faster swimmers are correlated with higher fertility (54). While it is known that swimming speed is important in cases of natural fertilization, our results add further weight that parameters affecting swimming speed are also important in artificial, in vitro scenarios. Another outcome of our work is to verify the conjecture that relative sizes, rather than absolute dimensions, are more relevant for fertility analysis (55).

Incorporating the findings from embryo-production rates, we observe two competing trends (Fig. 5 *B* and *C*). Having a head or midpiece with relatively more dry mass compared with the tail penalizes early stages of fertilization (zygote cleavage, negative trend), while having a larger head relative to midpiece is



C Summary of ART Outcomes

Embryo Cleavage	Pearson's R	P-Value
Head / Midpiece	-0.03	>0.36
Head / Tail	**-0.99**	**<0.01**
Midpiece / Tail	**-0.98**	**<0.01**
Blastocyst Formation		
Head / Midpiece	**+0.78**	**0.05**
Head / Tail	-0.30	0.31
Midpiece / Tail	-0.49	0.22



Cleavage
R_{ht}=H/T (lower is better)
R_{mt}=M/T (lower is better)

Blastocyst
R_{hm}=H/M (higher is better)

Fig. 5. Summary of outcomes. (*A* and *B*) Cleavage is strongly favored by a more massive tail (*A*), while blastocyst development is favored by a heavier head (*B*). (*C*) Summary across the five bulls for cleavage and blastocyst rates. H, head; Mid., midpiece; R_{hm}, head-to-midpiece ratio; R_{ht}, head-to-tail ratio; R_{mt}, midpiece-to-tail ratio; T, tail.

important for embryo development (blastocyst rate, positive trend). In bovine embryos, like in humans, the centriole is carried by the sperm and inherited after subsequent division (52). It has been hypothesized that damaged or otherwise defective centrioles can block an embryo from developing (56). In our images, the centriole structure is included in the dry mass of the head. It is not surprising, then, that the ratio of the head to the midpiece is positively related to the early stages of embryo development (as captured by embryo-production rates). Thus, our data suggest that for high cleavage rates, the head and midpiece should be light compared with the tail, while a larger head relative to the midpiece is preferable for blastocyst formation (Fig. 5*C*). We note that ratiometric differences between the sperm compartments are small, in absolute terms, but nonetheless significant for sperm selection. It is conceivable that our approach would be especially useful when selecting among seemingly healthy sperm, with no obvious defects. We note that the assay of these relatively small ratios is made possible by the high sensitivity of SLIM (57, 58). Importantly, in this work, we presented a template for automating the annotation of a large number of cells. By using a two-step training approach, we can expand a relatively small number of annotations with minimum effort, enabling the automated segmentation on entire microscope slides, while reducing practical problems such as bias between different annotators. The network training requires manual annotation, but we anticipate that this procedure needs to be performed only once per species. For example, moving to human studies will likely require a one-time retraining.

In a broader context, our methodology is in the same class with but distinct from very recent proposals to use neural networks in conjunctions with label-free phase images. In ref. 59, the authors recover synthetic histopathology stains by training on colocalized phase and hematoxylin/eosin images. A similar image-to-image translation estimation was used in ref. 60 to perform synthetic staining on sperm cells. While those proposals rely on the use of chemical-based markers to train the neural network, in this work, we avoid stains altogether and instead rely on SLIM's ability to directly observe cell ultrastructure.

IVF clinics have been using conventional phase-contrast microscopes for nondestructive observation, which does not allow for a detailed and quantitative evaluation of cellular compartments. We foresee that PICS can be implemented to these existing systems as an add-on. Thus, we envision that our system can be applied to individual sperm selection, which is performed on chemically slowed cells (61) or as a slide cytometer capable of automatically assessing sperm sample quality. Since transferring the AI code is straightforward, our methodology can likely be embraced at a large scale. The inference time, of ~70 ms per field of view is shorter than the image acquisition time. Thus, our technology may provide a solution for sperm classification and sorting in real time, as well as valuable insight into the relationships between cellular ultrastructure and fertility.

Materials and Methods

Semen Preparation. Postprocessing, semen samples were fixed with 4% paraformaldehyde (ThermoFisher Scientific) at room temperature. After ~30 min, fixed spermatozoa were spread on microscope slides and air-dried. Subsequently, slides were immersed for 1 h in deionized water, air-dried, and stored at 4 °C until morphometrical analyses.

Animals. Animal use was approved by the University of Illinois Institutional Animal Care and Use Committee (IACUC) under IACUC Protocol 19186. Healthy, nonpregnant, cycling Holstein cows (11 to 17 mo of age) were used as oocyte donors. All cattle were evaluated by transrectal palpation and ultrasonography before each procedure. No exogenous hormones were given to stimulate follicle production or synchronize the follicular cycle. All procedures were performed on animals with random estrous cycle.

Follicle Aspiration. Follicular aspiration was performed according to standard procedure (62). Briefly, all of the follicles with a diameter of 2 mm or more were aspirated via transrectal B-mode ultrasound (Ibex EVO II; E.I. Medical Imaging), from 5- to 10-MHz convex array transducer fitted into the intravaginal device (Ibex C9OPU-HD; E.I. Medical Imaging) and a stainless-steel needle guide. Follicular puncture was performed using a disposable 10-gauge × 3-mm hypodermic needle (Air-Tite Products Co., Inc.) connected to a 50-mL conical tube (Corning) via a silicon tube (0.8 m long; 2-mm inner diameter). Aspiration was performed using a vacuum pump (Watanabe Tecnologia Aplicada) with a negative pressure of between 60 and 80 mmHg. The collection medium was VIGRO Complete Flush (Vetoquinol). The aspirated material was immediately filtered through an EasyStrainer 100 μm (Greiner Bio-One) with ABT 360 complete flush medium (ABT 360). Cumulus oocyte complexes were classified as follows: 1) more than three layers of compact cumulus cells; 2) at least one layer of cumulus cells; 3) denuded; and 4) degenerated (cytoplasmic degeneration). After evaluation, only groups 1 and 2 were placed into maturation. Prior to in vitro maturation (IVM), cumulus oocytes complexes (COCs) were washed three times in M-199 (M7528) supplemented with 5% fetal bovine serum (FBS) (F2442) and 10 mg/mL penicillin streptomycin (Gibco Life Technologies). The COCs were maturated in IVM medium (described in *Media for In Vitro Embryo Production*), in an incubator, for 24 h at 38.5 °C, in 5% CO_2 in air.

Media for In Vitro Embryo Production. Unless otherwise stated, all reagents were purchased from Sigma-Aldrich. The IVM medium was in Medium 199 (M4530) supplemented with 15% FBS (F2442), 0.5 μg/mL FSH (F8174), 5 μg/mL LH (L5269), 1 μg/mL estradiol (E8875), and 50 μg/mL gentamycin (G1264). The IVF medium was Tyrode's modified medium (63) without glucose, supplemented with 95.6 USP/mL heparin (H3149), 30 μM penicillamine (P4875), 15 μM hypotaurine (H1384), 1 μM epinephrine (E4250), and 6 mg/mL bovine serum albumin (A6003). The intravaginal culture (IVC) medium consisted of synthetic oviduct fluid (SOF) medium (64), with 30 μL/mL essential amino acids, 10 μL/mL nonessential amino acids (Gibco Life Technologies), and 3% FBS.

Embryo Production. Embryos were produced as previously described (65). Briefly, in vitro-matured COCs were washed and transferred in 50-μL drops of IVF medium, covered with 4 mL of mineral oil (no. M8410; Sigma), and placed in the incubator at 38.5 °C in 5% CO_2 in air. The sperm samples were then processed via Bovipure discontinuous gradient (45 to 90%) (Nidacon Laboratories AB) (66). After processing, the sperm pellets were diluted with IVF medium and added to the fertilization drops at the concentration of 1×10^6 sperm per milliliter (final volume, 60 μL). Gametes were coincubated for 18 to 20 h at 38.5 °C, in 5% CO_2 in air, after which presumptive zygotes were mechanically pipetted to remove cumulus cells in hydroxyethyl piperazineethanesulfonic acid tissue culture medium with 5% FBS and then washed twice in the same medium. Presumptive zygotes were placed in 100-μL drops of SOF, where they were incubated in a humidified mixture of 5% CO_2, 6% O_2, and 89% N_2, at 38.5 °C (day 1). At day 3 of culture (72 h from the beginning of the culture), zygotes were evaluated for cleavage, and cleavage rate was calculated. The cleaved embryos were placed in fresh equilibrated SOF for an additional 4 d (96 h) of culture. At the end of 7 d of culture, the fertilized oocytes were evaluated and scored for quality on the basis of morphological criteria, and only grade 1 and 2 blastocysts (Bl) were considered in the evaluation of the final embryo rate (65, 67).

Construction of the Neural Network. Deep-convolutional neural networks consist of a series of pixel-wise nonlinear operations that remap the original data into another form such as a single number or whole images (43, 68). To annotate sperm cells, we develop a neural network that remaps the halo-removed SLIM image into a four-channel probability map that specifies the probability for the three cellular compartments as well as the background. The most probable label is taken to produce a single image that contains the annotated sperm.

Our architecture is based on U-Net (45), which benefits from a large receptive field enabling the network to learn textural (local) and contextual (surrounding) features (46).

The network transforms the image data through a series of layers with each layer operating on the results of subsequent layers. The U-Net architecture shown in Fig. 3*B* is divided into a "contracting" and "expanding" portion that is linked by a "bottleneck" at the bottom. Importantly, the U-Net architecture includes concatenation layers that effectively copy the output from previous levels to stabilize the training procedure.

Each "block" of the contracting path results in a smaller version of the input image that has been processed in a nonlinear way. The first block

consists of a filter bank of random initialized convolutional kernels ("conv"). In the first block, there are 64 three-by-three filters. The weights are updated at each training step so that the network output will converge to the training data. To introduce nonlinearity, the convolution is paired to a nonlinear activation function ("ReLu"), which removes negative outputs (69). The resulting image resembles a distorted version of the input, and the operation is repeated a second time with different weights. The values are propagated to the next block by max-pooling, i.e., selecting the maximum value of each two-by-two pixel neighborhood, which effectively down-samples the image by two.

Subsequent blocks decrease the image size while increasing the number of filters. After passing the bottleneck of the network, the image is up-sampled by a factor of 2, and a filter bank is applied. The results of that filter bank are combined with the matching filter bank operation from the contracting path, leading to improved numeric stability when training (Fig. 3*B*; 1,024 filter block). These operations are repeated, with the last block of the expansive path containing a one-by-one convolution applied to a "softmax" activation function (70). This operation reduces the number of channels in the output image to the four-channel probability map. A pixel-wise maximum argument of this four-channel probability map ("argmax") results in a one-channel image with a label for each of the four categories.

In *SI Appendix, Model for Real-Time Imaging*, we demonstrate that a similar architecture can be used for real-time operation.

Generation of Labels for the Neural-Network Training. The training data were annotated manually by individuals trained to identify the sperm head, midpiece, and tail. A fraction of the sperm images were manually segmented by three annotators using ImageJ. The final segmentations were verified by one of the annotators. For each sperm cell, we traced the sharp discontinuity between the background and the cell, separated by an abrupt change in phase shift. An alternative strategy is to annotate the ultrastructure using immunochemical stains (71) or genetic engineering (39, 71, 72), at the expense of potential artifacts, more complicated specimen preparation protocol, and difficulty in maintaining genetically modified livestock (73).

To improve the segmentation accuracy, we used a two-pass training procedure where an initial training round was corrected and used for a second, final round. Manual annotation for the second round is comparably fast, which also allows us to correct for debris and other forms of clearly defective segmentation (60). To obtain the dry mass, we ran inference using our newly trained network. The resulting semantic segmentation maps were applied to the phase images to compute the dry-mass content of each component. We note that by using a single neural network, rather than a group of annotators, we can compensate for differences between annotators. In total, training and inference were run on 20 slides.

Training the Neural Network. The network was trained from scratch by using an adaptive momentum optimizer (74) against the categorical cross-entropy loss,

$$E = -\frac{1}{r \times c} \cdot \sum_{r=1}^{h} \sum_{c=1}^{w} \sum_{k=1}^{4} [\delta(y[r][c] == k) \cdot \log(\hat{y}[r][c][k])], \quad [2]$$

where h and w represent the number of rows and columns in the image. δ is an indicator function, which evaluates to 1 if $y[r][c]$, the true label for that pixel (r, c) is k. This loss function takes the average of the negative log-likelihood for the target class across every pixel in one image. It penalizes small $\log\hat{y}[r][c][k]$, which is the predicted probability for the target class.

As shown in Fig. 3*B*, following the typical procedure for training a U-Net, we introduce dropout at the bottleneck (bottom) portion of the architecture. To improve performance as well as accelerate the training, we introduce batch-normalization between input layers in our network (75).

The training was performed using the Tensorflow backend for Keras (76, 77) on a workstation equipped with a GPU (GTX 1080; NVIDIA). The cross-entropy loss function was already implemented as a built-in loss function in Keras.

To reduce computing requirements, images were down-sampled to match the optical resolution (78). To account for the shift variance of convolutional neural networks, the training pairs were augmented using rotation, flipping, and translation.

In this work, we trained a total of three neural networks. An initial network was used to "bootstrap" the training by producing a coarser segmentation map. This coarse segmentation was used to train the second neural network which was used for data analysis. Finally, as discussed in *SI Appendix, Model for Real-Time Inference*, a third neural network was trained to demonstrate real-time imaging capabilities. The split in images between training, test, and validation is shown in *SI Appendix*, Table S3. A typical image contained between one and four sperm cells.

The model used for data analysis was trained with a learning rate of $5e^{-6}$ for 30 epochs. Due to memory limitations, the batch size is set to one, meaning the model will update its weights after seeing one image. In our case, after each epoch, the model weights were updated 3,296 times (with data augmentation).

Evaluating the Performance of the Neural Network. The learning curves for the final training round are shown in *SI Appendix*, Fig. S4*A*. Both the training and validation loss simultaneously converge to a minimum value, indicating a lack of training defects such as overfitting. As training was performed on thousands of images, we found little motivation to perform computationally expensive cross-validation. The number of images used is summarized in *SI Appendix*, Table S3.

The ability of the network to reproduce the manual annotation is shown in *SI Appendix*, Fig. S4*B*. The diagonal terms of the confusion matrix represent the number of pixels that are correctly classified for each class, while the off-diagonal terms indicate different types of errors. For a more intuitive presentation, we normalized the confusion matrix by the total number of pixels in the test set. The resulting F1 scores (47) are all above 0.85, indicating good performance.

We note that the performance metrics reported are a comparison to manual annotation, which invariably has error due to human factors. As the neural network, generally, averages out these discrepancies, we expect the real-world performance to exceed what we report.

Data Availability. Data are available upon reasonable request.

ACKNOWLEDGMENTS. This work was supported by NSF Grants CBET-0939511 STC (to G.P.), NRT-UtB 1735252 (to M.E.K.), R01GM129709 (to G.P.), 1R43GM133280-01 (to G.P.), R01 CA238191 (to G.P.). M.J.S. was supported by the Carle Neuroscience Institute Undergraduate Research Award. This work was partially supported by US Department of Agriculture Multistate Project W-4171 (M.B.W.) and the Ross Foundation (M.B.W. and M.R.).

1. Maxwell V. Meng, Kirsten L. Greene, Paul J. Turek, Surgery or assisted reproduction? A decision analysis of treatment costs in male infertility. *J. Urol.* **174**, 1926–1931, discussion 1931 (2005).
2. M. Nadalini, N. Tarozzi, V. Distratis, G. Scaravelli, A. Borini, Impact of intracytoplasmic morphologically selected sperm injection on assisted reproduction outcome: A review. *Reprod. Biomed. Online* **19** (suppl. 3), 45–55 (2009).
3. A. Berkovitz *et al*., How to improve IVF-ICSI outcome by sperm selection. *Reprod. Biomed. Online* **12**, 634–638 (2006).
4. P. Vanderzwalmen *et al*., *Intracytoplasmic Morphologically Selected Sperm Injection. In Vitro Fertilization*, (Springer, 2019).
5. S. Kondracki, A. Wysokińska, M. Kania, K. Górski, Application of two staining methods for sperm morphometric evaluation in domestic pigs. *J. Vet. Res. (Pulawy)* **61**, 345–349 (2017).
6. D. Banaszewska, K. Andraszek, M. Czubaszek, B. Biesiada-Drzazga, The effect of selected staining techniques on bull sperm morphometry. *Anim. Reprod. Sci.* **159**, 17–24 (2015).
7. L. Maree, S. S. du Plessis, R. Menkveld, G. van der Horst, Morphometric dimensions of the human sperm head depend on the staining method used. *Hum. Reprod.* **25**, 1369–1382 (2010).
8. G. Nomarski, Microinterféromètre différentiel à ondes polarisées. *J. Phys. Radium* **16**, 9S–13S (1955).
9. F. Zernike, How I discovered phase contrast. *Science* **121**, 345–349 (1955).
10. R. Hoffman, L. Gross, Modulation contrast microscope. *Appl. Opt.* **14**, 1169–1176 (1975).
11. G. Popescu, *Quantitative Phase Imaging of Cells and Tissues*, (McGraw-Hill, New York, NY, 2011).
12. Y. Park, C. Depeursinge, G. Popescu, Quantitative phase imaging in biomedicine. *Nat. Photonics* **12**, 578–589 (2018).
13. D. Zicha, E. Genot, G. A. Dunn, I. M. Kramer, TGFbeta1 induces a cell-cycle-dependent increase in motility of epithelial cells. *J. Cell Sci.* **112**, 447–454 (1999).
14. R. Barer, Interference microscopy and mass determination. *Nature* **169**, 366–367 (1952).
15. H. G. Davies, M. H. Wilkins, Interference microscopy and mass determination. *Nature* **169**, 541 (1952).
16. M. Mir *et al*., Optical measurement of cycle-dependent cell growth. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 13124–13129 (2011).
17. G. Popescu *et al*., Optical imaging of cell mass and growth dynamics. *Am. J. Physiol. Cell Physiol.* **295**, C538–C544 (2008).
18. M. E. Kandel *et al*., Cell-to-cell influence on growth in large populations. *Biomed. Opt. Express* **10**, 4664–4675 (2019).

19. K. Lee *et al*., Quantitative phase imaging techniques for the study of cell pathophysiology: From principles to applications. *Sensors (Basel)* **13**, 4170–4191 (2013).
20. M. Mir, A. Bergamaschi, B. S. Katzenellenbogen, G. Popescu, Highly sensitive quantitative imaging for monitoring single cancer cell growth kinetics and drug response. *PLoS One* **9**, e89000 (2014).
21. W. Choi *et al*., Tomographic phase microscopy. *Nat. Methods* **4**, 717–719 (2007).
22. F. Charrière *et al*., Cell refractive index tomography by digital holographic microscopy. *Opt. Lett.* **31**, 178–180 (2006).
23. F. Merola, P. Memmolo, L. Miccio, M. Mugnano, P. Ferraro, Phase contrast tomography at lab on chip scale by digital holography. *Methods* **136**, 108–115 (2018).
24. T. Kim *et al*., White-light diffraction tomography of unlabeled live cells. *Nat. Photonics* **8**, 256–263 (2014).
25. T. H. Nguyen, M. E. Kandel, M. Rubessa, M. B. Wheeler, G. Popescu, Gradient light interference microscopy for 3D imaging of unlabeled specimens. *Nat. Commun.* **8**, 210 (2017).
26. L. Miccio *et al*., "Detection and visualization improvement of Spermatozoa cells by digital holography". *Molecular Imaging III, Proceedings of the International Society of Optics and Photonics* **8089**, 1605–7422 (2011).
27. T. W. Su, L. Xue, A. Ozcan, High-throughput lensfree 3D tracking of human sperms reveals rare statistics of helical trajectories. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 16018–16022 (2012).
28. M. Haifler *et al*., Interferometric phase microscopy for label-free morphological evaluation of sperm cells. *Fertil. Steril.* **104**, 43–47.e2 (2015).
29. L. Liu *et al*., Topography and refractometry of sperm cells using spatial light interference microscopy. *J. Biomed. Opt.* **23**, 1–6 (2018).
30. P. Memmolo *et al*., Identification of bovine sperm head for morphometry analysis in quantitative phase-contrast holographic microscopy. *Opt. Express* **19**, 23215–23226 (2011).
31. G. Di Caprio *et al*., 4D tracking of clinical seminal samples for quantitative characterization of motility parameters. *Biomed. Opt. Express* **5**, 690–700 (2014).
32. G. Di Caprio *et al*., Holographic imaging of unlabelled sperm cells for semen analysis: A review. *J. Biophotonics* **8**, 779–789 (2015).
33. F. Merola *et al*., Digital holography as a method for 3D imaging and estimating the biovolume of motile cells. *Lab Chip* **13**, 4512–4516 (2013).
34. H. Majeed *et al*., Quantitative histopathology of stained tissues using color spatial light interference microscopy (cSLIM). *Sci. Rep.* **9**, 14679 (2019).
35. R. A. Hoebe *et al*., Controlled light-exposure microscopy reduces photobleaching and phototoxicity in fluorescence live-cell imaging. *Nat. Biotechnol.* **25**, 249–253 (2007).
36. D. A. Zacharias, J. D. Violin, A. C. Newton, R. Y. Tsien, Partitioning of lipid-modified monomeric GFPs into membrane microdomains of live cells. *Science* **296**, 913–916 (2002).
37. Z. Wang *et al*., Spatial light interference microscopy (SLIM). *Opt. Express* **19**, 1016–1026 (2011).
38. C. Edwards *et al*., Effects of spatial coherence in diffraction phase microscopy. *Opt. Express* **22**, 5133–5146 (2014).
39. M. E. Kandel, M. Fanous, C. Best-Popescu, G. Popescu, Real-time halo correction in phase contrast imaging. *Biomed. Opt. Express* **9**, 623–635 (2018).
40. M. E. Kandel *et al*., Three-dimensional intracellular transport in neuron bodies and neurites investigated by label-free dispersion-relation phase spectroscopy. *Cytom. Part A* **91**, 519–526 (2017).
41. P. Ounjai, K. D. Kim, P. V. Lishko, K. H. Downing, Three-dimensional structure of the bovine sperm connecting piece revealed by electron cryotomography. *Biol. Reprod.* **87**, 73 (2012).
42. A. Radford, L. Metz, S. Chintala, Unsupervised representation learning with deep convolutional generative adversarial networks. arXiv:1511.06434 (19 November 2015).
43. A. Krizhevsky, I. Sutskever, G. E. Hinton, Imagenet classification with deep convolutional neural networks. *Adv. Neural Inf. Process. Syst.* **25**, 1097–1105 (2012).
44. T. H. Nguyen *et al*., Automatic Gleason grading of prostate cancer using quantitative phase imaging and machine learning. *J. Biomed. Opt.* **22**, 36015 (2017).
45. O. Ronneberger, P. Fischer, T. Brox, *U-Net: Convolutional Networks for Biomedical Image Segmentation*, (Springer International Publishing, 2015), pp. 234–241.
46. M. E. Kandel *et al*, PICS: Phase imaging with computational specificity. arXiv: 2002.08361 (13 March 2020).
47. Y. Sasaki, The truth of the F-measure. https://www.cs.odu.edu/~mukka/cs795sum09dm/Lecturenotes/Day3/F-measure-YS-26Oct07.pdf. Accessed 20 April 2020.
48. R. M. Haralick, L. G. Shapiro, *Computer and Robot Vision*, (Addison-Wesley Reading, 1992).
49. A. Bartolacci *et al*., Abnormal sperm concentration and motility as well as advanced paternal age compromise early embryonic development but not pregnancy outcomes: A retrospective study of 1266 ICSI cycles. *J. Assist. Reprod. Genet.* **35**, 1897–1903 (2018).
50. S. S. Suarez, H. C. Ho, Hyperactivated motility in sperm. *Reprod. Domest. Anim.* **38**, 119–124 (2003).
51. A. F. P. Siqueira *et al*., Sperm traits on in vitro production (IVP) of bovine embryos: Too much of anything is good for nothing. *PLoS One* **13**, e0200273 (2018).
52. A. H. Sathananthan *et al*., Inheritance of sperm centrioles and centrosomes in bovine embryos. *Arch. Androl.* **38**, 37–48 (1997).
53. A. F. Malo *et al*., Sperm design and sperm function. *Biol. Lett.* **2**, 246–249 (2006).
54. M. Gomendio, E. R. Roldan, Implications of diversity in sperm size and function for sperm competition and fertility. *Int. J. Dev. Biol.* **52**, 439–447 (2008).
55. S. Humphries, J. P. Evans, L. W. Simmons, Sperm competition: Linking form to function. *BMC Evol. Biol.* **8**, 319 (2008).
56. I. Schneider, J. Ellenberg, Mysteries in embryonic development: How can errors arise so frequently at the beginning of mammalian life? *PLoS Biol.* **17**, e3000173 (2019).
57. M. Rubessa, S. N. Lotti, M. E. Kandel, G. Popescu, M. B. Wheeler, SLIM microscopy allows for visualization of DNA-containing liposomes designed for sperm-mediated gene transfer in cattle. *Mol. Biol. Rep.* **46**, 695–703 (2019).
58. M. Rubessa *et al*., Morphometric analysis of sperm used for IVP by three different separation methods with spatial light interference microscopy. *Syst Biol Reprod Med* **66**, 26–36 (2020).
59. Y. Rivenson *et al*., PhaseStain: The digital staining of label-free quantitative phase microscopy images using deep learning. *Light Sci. Appl.* **8**, 23 (2019).
60. Y. N. Nygate *et al*., Holographic virtual staining of individual biological cells. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 9223–9231 (2020).
61. B. Balaban *et al*., An alternative to PVP for slowing sperm prior to ICSI. *Hum. Reprod.* **18**, 1887–1889 (2003).
62. F. L. B. Cavalieri *et al*., Improvement of bovine in vitro embryo production by ovarian follicular wave synchronization prior to ovum pick-up. *Theriogenology* **117**, 57–60 (2018).
63. J. J. Parrish *et al*., Bovine in vitro fertilization with frozen-thawed semen. *Theriogenology* **25**, 591–600 (1986).
64. H. R. Tervit, D. G. Whittingham, L. E. Rowson, Successful culture in vitro of sheep and cattle ova. *J. Reprod. Fertil.* **30**, 493–497 (1972).
65. M. Rubessa, A. Ambrosi, D. Gonzalez-Pena, K. M. Polkoff, M. B. Wheeler, Non-invasive nuclear magnetic resonance analysis of male and female embryo metabolites during in vitro embryo culture. *Metabolomics* **14**, 113 (2018).
66. A. Sattar *et al*., The influence of gamete co-incubation length on the in vitro fertility and sex ratio of bovine bulls with different penetration speed. *Reprod. Domest. Anim.* **46**, 1090–1097 (2011).
67. I. Robertson, R. Nelson, "Certification and identification of the embryo" in *Manual of the International Embryo Transfer Society*, D. D. Stringfellow, S. M. Seidel, Eds. (International Embryo Transfer Society, Savoy, IL, 3rd Ed., 1998), pp. 103–134.
68. P. Isola, J. Zhu, T. Zhou, A. A. Efros, "Image-to-Image translation with conditional adversarial networks" in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, (IEEE, 2017), pp. 5967–5976.
69. V. Nair, G. E. Hinton, "Rectified linear units improve restricted boltzmann machines" in *Proceedings of the 27th International Conference on Machine Learning (ICML-10)*, (International Machine Learning Society, 2010), pp. 807–814.
70. C. Nwankpa, W. Ijomah, A. Gachagan, S. Marshall, Activation functions: Comparison of trends in practice and research for deep learning. arXiv:1811.03378 (8 November 2018).
71. Molecular Probes, "Probes for mitochondria—Section 12.2" in *The Molecular Probes Handbook*, (Molecular Probes, 11th Ed., 2010), pp. 503–515.
72. H. Hasuwa *et al*., Transgenic mouse sperm that have green acrosome and red mitochondria allow visualization of sperm and their acrosome reaction in vivo. *Exp. Anim.* **59**, 105–107 (2010).
73. A. L. Van Eenennaam, GMOs in animal agriculture: Time to consider both costs and benefits in regulatory evaluations. *J. Anim. Sci. Biotechnol.* **4**, 37 (2013).
74. D. P. Kingma, J. Ba, Adam: A method for stochastic optimization. *arXiv*:1412.6980 (30 January 2017).
75. S. Ioffe, C. Szegedy, Batch normalization: Accelerating deep network training by reducing internal covariate shift. arXiv:1502.03167 (2 March 2015).
76. F. Chollet, Data from "keras." GitHub. https://github.com/fchollet/keras. Accessed 10 September 2018.
77. M. Abadi *et al*., "Tensorflow: A system for large-scale machine learning" in *12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16)*, (USENIX, 2016), pp. 265–283.
78. R. Horstmeyer, R. Heintzmann, G. Popescu, L. Waller, C. Yang, Standardizing the resolution claims for coherent microscopy. *Nat. Photonics* **10**, 68–71 (2016).