Harnessing cytoplasmic particles movement of the human early embryo analysed by advanced imaging and artificial intelligence to predict development to blastocyst stage

Giovanni Coticchio^{1,*,#}, Giulia Fiorentino^{2,3,#}, Giovanna Nicora^{3,4}, Raffaella Sciajno¹, Federica Cavalera², Riccardo Bellazzi^{3,4}, Silvia Garagna^{2,3}, Andrea Borini¹, Maurizio Zuccotti^{2,3*}

1. 9.baby Family and Fertility Center, Bologna, Italy.
2. Department of Biology and Biotechnology ‘Lazzaro Spallanzani’, University of Pavia, Italy.
3. Centre for Health Technology, University of Pavia, Pavia, Italy.
4. Department of Electrical, Computer and Biomedical Engineering, University of Pavia, Pavia, Italy.

Contributed equally to the manuscript

(*) Correspondence

Giovanni Coticchio: 9.baby Family and Fertility Center, Via Dante, 15, 40125 - Bologna, Italy

e-mail: giovanni.coticchio@nove.baby

Maurizio Zuccotti: Laboratory of Developmental Biology, Department of Biology and Biotechnology ‘Lazzaro Spallanzani’, University of Pavia, Via Ferrata, 9, 27100 – Pavia, Italy.

e-mail: maurizio.zuccotti@unipv.it

Keywords: VF; embryo; Blastocyst; Cytoplasm; Artificial intelligence; artificial neuronal network

22 **Abstract**

23 **Research Question:** Progress in artificial intelligence (AI) and advanced image analysis offers unique
24 opportunities to develop novel embryo assessment approaches. In this study, we tested the hypothesis
25 that such technologies can extract and harness novel information derived from cytoplasmic movements
26 of the early human embryo to predict development to blastocyst.

27 **Design:** In a proof-of principle study, an artificial neural network (ANN) approach was undertaken to
28 assess retrospectively 230 human preimplantation embryos. After ICSI, embryos were subjected to time-
29 lapse monitoring for 44 hours. For comparison as a standard embryo assessment methodology, a single
30 senior embryologist assessed each embryo to predict development to blastocyst stage (BL) based on a
31 single picture frame taken at 42 hours of development. In the experimental approach, in embryos that
32 developed to blastocyst or destined to arrest (NoBL), cytoplasm movement velocity (CMV) was recorded
33 by time-lapse monitoring during the first 44 hours of culture and analysed with a Particle Image
34 Velocimetry (PIV) algorithm to extract quantitative information. Three main AI approaches, the k-
35 Nearest Neighbor (k-NN), the Long-Short Term Memory Neural Network (LSTM-NN) and the hybrid
36 ensemble classifier (HyEC) were employed to classify the two embryo classes.

37 **Results:** Blind operator assessment classified each embryo in terms of ability of development to
38 blastocyst, reaching a 75.4% accuracy, 76.5% sensitivity, 74.3% specificity, 74.3% precision and 75.4%
39 F1 score. After integration of results from AI models together with the blind operator classification, the
40 performance metrics improved significantly, with a 82.6% accuracy, 79.4% sensitivity, 85.7%
41 specificity, 84.4% precision and 81.8% F1 score.

42 **Conclusions:** The present study suggests the possibility to predict human blastocyst development at
43 early cleavage stages by detection of CMV and AI analysis. This indicates the importance of the
44 dynamics of the cytoplasm as a novel and valuable source of data to assess embryo viability.

45 **Introduction**

46 One of the long-term, and among the most important, goals of medically assisted reproduction (MAR) is
 47 the achievement, in the shortest possible period of time, of a single healthy term pregnancy following the
 48 transfer of a single embryo. To this end, embryos need to be assessed by non- or minimally-invasive
 49 methodologies and prioritized for transfer according to their developmental potential. Until the early
 50 2000's, in the vast majority of MAR programs, embryos were being cultured until day two or three post-
 51 insemination and selected for transfer according to static morphological criteria. More recently,
 52 significant paradigm shifts have emerged, based on the extension of embryo culture to day five or six to
 53 discriminate, by self-selection, embryos able to develop to the blastocyst stage. These blastocysts, on
 54 average with higher developmental potential compared with cleavage-stage embryos, can be used for
 55 embryo transfer with increased chances to achieve a viable pregnancy per transfer attempt (Glujovsky et
 56 al., 2016). To the aim of further increasing efficiency, i.e. to identify more precisely and prioritize for
 57 transfer the most viable embryo in a cohort, blastocysts can be assessed according to static ("The Istanbul
 58 consensus workshop on embryo assessment: proceedings of an expert meeting.", 2011) or
 59 dynamic (Gallego et al., 2019) morphological criteria, or even by adopting a strategy of preimplantation
 60 genetic screening for aneuploidies (PGT-A) (Griffin and Ogur, 2018). Collectively, indeed, these
 61 approaches based on blastocyst culture assure a higher treatment efficiency compared with day two or
 62 three embryo transfer. However, they cannot be considered ideal for several reasons. A major drawback
 63 of such methodologies lies in the culture to the blastocyst stage *per se*. In fact, blastocyst culture is more
 64 expensive and more demanding for the management of the IVF laboratory workflow. Even more
 65 importantly, if not mastered properly, blastocyst culture may impact on intrinsic embryo developmental
 66 potential, affecting overall treatment efficacy (Swain, 2019). This scenario clearly points towards the
 67 need for a more advanced, non-invasive embryo assessment methodology able to predict developmental
 68 potential at earlier stages, without necessarily resorting to extended culture.

69 Progress in the realm of artificial intelligence and in image analysis offers unique opportunities to
70 develop novel embryo assessment approaches. Recently, promising attempts were made to harness the
71 potential of deep learning to predict embryo implantation potential (Khosravi et al., 2019; Miyagi et al.,
72 2019; Tran et al., 2019). However, all such cases involved the use of static images or videos of blastocyst
73 stage embryos, leaving the question of prediction at earlier stages unanswered.

74 In this proof-of-principle study, we approached the option of advanced image analysis of early
75 development to predict ability of blastocyst formation. To this aim, sequential images collected by bright-
76 field time lapse microscopy (TLM) over less than two days of culture were subjected to particle image
77 velocimetry (PIV) analysis to detect the dynamics of cytoplasmic movements. The resulting data were
78 elaborated by neural network-based artificial intelligence (AI). The strategy to investigate cytoplasm
79 dynamics as a novel source of data to assess oocyte function and embryo viability finds justification in
80 studies carried out in the mouse model (Ajduk et al., 2011; Yi et al., 2013; Bui et al., 2017). Nevertheless,
81 movement of cytoplasmic particles has been until now totally neglected in attempts to predict human
82 embryo viability.

83 The study results suggest that three powerful technological and analytical tools, TLM, PIV and AI, can
84 be combined to extract predictive information on human embryo viability from cytoplasm dynamics
85 occurring at early developmental stages.

86 **Materials and Methods**

87 This retrospective proof-of-principle study included 230 embryos generated in ICSI cycles carried out
88 between October 2015 and May 2018. Approval for the study was obtained from the local Institutional
89 Review Board. Laboratory non-clinical data were used for research purposes only. Cases suitable for
90 analysis were selected according to the rule to have two sibling embryos from the same cohort (cycle),

91 of which one developed to blastocyst stage and one arrested to an earlier stage. In this fashion, we aimed
92 at limiting possible biases due to patient variability.

93 Diagnosis of infertility included various causes, including male factor, tubal factor and polycystic ovary
94 (PCO, but not PCOS) with or without chronic anovulation. Such inclusion criteria were chosen because
95 consistent with the expectation to have a relatively high number of embryos in a cohort and the ensuing
96 possibility to extend embryo culture to the blastocyst stage, a condition required by the study design.
97 Ovarian stimulation was carried out as previously described by Zacà et al. (Zacà et al., 2018).

98 Semen preparation and ICSI procedures

99 Semen selection for ICSI was performed by discontinuous PureSperm (Nidacon, Gothemborg, Sweden)
100 gradient (Borini et al., 2006). After preparation, sperm were evaluated for concentration, total and
101 progressive motility and morphology, according to WHO procedures (World Health Organization, 2010).

102 Embryo culture and time lapse image acquisition

103 Oocytes/embryo culture was carried out using the EmbryoScope equipment (Vitrolife, Goteborg,
104 Sweden), an integrated time-lapse technology-incubator system for embryo-culture, carried out in a
105 N₂/CO₂/O₂ (89:6:5, v/v) atmosphere at 37°C without control of humidity. Microinjected oocytes were
106 placed inside pre-equilibrated slides (EmbryoSlide; Vitrolife), each containing 12 droplets of 25 µl of
107 cleavage medium (Cook IVF, Australia) covered by 1.2 ml of mineral oil (SAGE, Biocare Europe, Rome,
108 Italy). On day 3, cleavage medium was replaced with blastocyst medium (Cook IVF, Australia) to extend
109 embryo culture until day 5, where appropriate (Lagalla 2017). Images were acquired over a period of 125
110 hours starting from the time of ICSI, with a 15 min interval between consecutive picture frames.

111 Operator-based prediction of development to blastocyst stage

112 The outcome of AI analysis was first comparatively assessed and then combined with standard operator-
 113 based prediction of development to blastocyst. Embryo assessment by standard morphology was
 114 preferred to alternative assessment approaches, i.e., morphokinetic algorithms, because static
 115 evaluation is routinely performed following the recommendations of an international consensus
 116 (“The Istanbul consensus workshop on embryo assessment: proceedings of an expert meeting.”, 2011)
 117 and remains the “standard of care” in clinical embryology. On the other hand, while having great
 118 potential, morphokinetic evaluation has not reached yet a similar degree of consensus, perhaps
 119 also as an effect of biases derived from different technological platforms and applicative protocols.
 120 On day two, at 42 hours post-ICSI a single embryologist with over 20 years of experience in clinical
 121 embryology assessed each embryo according to: a) blastomere number, size and mutual position; b)
 122 degree (percentage of total volume) of fragmentation; c) multinucleated blastomeres; d) cellular
 123 dysmorphisms (e.g., large vacuoles) (“The Istanbul consensus workshop on embryo assessment:
 124 proceedings of an expert meeting.”, 2011). Based on such morphological assessment performed statically
 125 (i.e., at a single time point), the operator formulated a binary prediction (yes/no) of development to
 126 blastocyst stage according to the recommendations of the Alpha and ESHRE Istanbul Consensus (“The
 127 Istanbul consensus workshop on embryo assessment: proceedings of an expert meeting.”, 2011).

128 Definition of blastocyst

129 At 116 hours post ICSI embryos were categorized as developed at the blastocyst stage if they showed at
 130 least the following minimal requirements: a) at least a Grade 1 blastocoel expansion; b) a discernible
 131 cluster of cells forming the inner cell mass (ICM); c) a wall of epithelial-like cells delimiting the
 132 blastocoel and enclosing the ICM (“The Istanbul consensus workshop on embryo assessment:
 133 proceedings of an expert meeting.”, 2011).

134 Cell_PIV analysis

135 Cell_PIV software (kindly provided by Dr. Shane Windsor), based on the Particle Image Velocimetry
 136 (PIV) algorithm, was used to detect embryo movement velocity (EMV) during the first two-cell division
 137 (44 hours) of BL and No-BL embryos. The movement velocity vectors, observed in the embryo, were
 138 calculated by cross correlation between the patterns of the pixels of adjacent video frames, extracting
 139 five features: (i) *Direction* (D), the mean of the inverse of the local standard deviation of the direction of
 140 the vectors; (ii) *Vorticity* (V), the mean vorticity of the vectors in each frame; (iii) *Hybrid* (H), a hybrid
 141 combination of the mean magnitude and inverse standard deviation data; (iv) *Meanmag* (M), the mean
 142 magnitude of the vectors in each frame; (v) *Summed* (S), the mean magnitude of the vectors summed
 143 over a set number of frames forward in time from the current point.

144 Statistics

145 Differences in BL and No-BL embryos cytoplasmic movements were evaluated using the Two-sample
 146 Kolmogorov-Smirnov test with the MATLAB software (R2018B). Data were considered significantly
 147 different when $p < 0.05$.

148 Artificial Intelligence

149 Both k-NN and LSTM-NN were implemented with MATLAB software (R2018B). In particular, the
 150 LSTM-NN was developed using the MATLAB Deep Learning Toolbox. A set of k-NN and LSTM-NN
 151 models, with empirically established parameters and different combination of features, was defined.
 152 Among them, four models (k-NN-1, k-NN-2, LSTM-NN-1, LSTM-NN-2) used for final comparison
 153 with the operator on the test set, were chosen based on the best results of the cross validation on the
 154 training set. The selected models were then trained on the entire training set. To evaluate the models
 155 generalization capabilities, a 10-fold cross validation on the entire dataset (230 embryos, 118 in NoBL

class and 112 in BL class) was also performed for the k-NN-1, k-NN2, LSTM-NN-1 and LSTM-NN-2 algorithms (Figure 5). For the k-NN models, we set the number of neighbours equal to four and the similarity between embryos was computed coupling the Dynamic Time Warping with the Kullback distance. The LSTM-NN, whose architecture is shown in Figure 3, was trained with the Adam optimization function for 100 epochs, with a learning rate of 0.1 and a batch size of 20. Results of the AI models on the test set were evaluated in terms of accuracy (BL-NoBL recognition rate), sensitivity (True-BL recognition rate) and specificity (True-NoBL recognition rate).

Results

Immediately after ICSI, each of the obtained 230 zygotes was transferred in a single Embryoscope well and cultured for a total of 125 hours (5.2 days). Of these, 112 reached the blastocyst stage (48.7%), whereas 118 arrested sometime earlier, but never before the second cleavage cycle.

Based on a single picture frame taken at 42 hours culture post ICSI, an expert operator blindly classified each of the 230 embryos - according to previously established morphological criteria (“Istanbul consensus workshop on embryo assessment: proceedings of an expert meeting.”, 2011) - as developmentally competent (BL) or incompetent (NoBL), reaching a 75.4% accuracy, 76.5% sensitivity and 74.3% specificity.

The time-lapse sequence recorded during 44 hours (175 frames), corresponding to completion of the second embryonic cell cycle division, was analysed using Particle Image Velocimetry software Cell_PIV (Figure 1A), with five different image processing features: *Direction*, *Vorticity*, *Hybrid*, *Meanmag* and *Summed* (see M&M).

177 Analysis of the first three parameters highlighted noisy and overlapping profiles of BL and NoBL embryo
178 movement velocity (EMV) (Figure 1B-D); instead, the *Meanmag* (Figure 1E) and, more significantly,
179 the *Summed* (Figure 1F) features identified two main and consistent time intervals with statistically
180 different EMVs.

181 Interestingly, when the *Summed* EMV temporal profiles of BL and NoBL embryos were correlated with
182 known key events occurring during the first 44 hours of development (Figure 2), the first time-interval
183 overlapped with pronuclear re-positioning and ended shortly before the time of the first embryonic cell
184 division. The second time-interval coincided with the second cleavage cycle, until the end of the time-
185 lapse frames analysed.

186 Whilst the morphodynamic patterns measured by Cell_PIV suggest, at least at these two time-intervals,
187 the potential to discriminate between the two embryo classes, their high cytoplasm movement velocity
188 (CMV) variability restricts their possible use for a single embryo classification in the clinical practice.
189 Nevertheless, this limitation was resolved by the subsequent AI analysis.

190 To this end, here we used two AI approaches, the k-Nearest Neighbor (k-NN) and the Long-Short Term
191 Memory Neural Network (LSTM-NN), whose algorithms have been exploited earlier in different
192 biomedical classification contexts to reveal hidden patterns from complex biological data and to support
193 the clinical decision process (He et al., 2019).

194 Out of 230 embryos analysed, 161 (70%) were randomly selected as training set of the AI algorithms,
195 whereas the remaining 69 (30%) were employed as test set. Both algorithms were fed with either the data
196 of a single Cell_PIV feature (i.e., M or S) or with a combination (i.e., M+S or D+V+H+M+S).

197 k-NN

198 The k-NN algorithm, largely used to solve time series classification problems (Yakovitz, 1987), classifies
199 a new sample based on its distance similarity with the training set. Since the computation of the distances
200 between two time series is frequently subject to slight time shifts, the k-NN on Cell_PIV temporal series
201 was preceded by the application of Dynamic Time Warping (DTW), a pre-processing technique that
202 produces the optimal alignment between sequences (Chaovalitwongse et al., 2007; Müller et al., 2007;
203 Geler et al., 2016; 2020; Tran et al., 2019).

204 Of the four k-NN implementations performed, the D+V+H+M+S (k-NN-2) showed the highest accuracy
205 (72.5%), the best balance between sensitivity (76.5%) and specificity (68.6%), the highest precision
206 (70.3%), and a F1 score equal to 73.2%; instead, the M+S (k-NN-1) approach gave a similar accuracy
207 (71.0%), lower specificity (60.0%) and precision (66.7%), but a higher sensitivity (82.3%) and F1 score
208 (73.7%). Implementation with a single feature (i.e., M or S) resulted in low accuracy (< 65%) and, thus,
209 this approach was abandoned.

210 LSTM-NN

211 LSTM-NN is a non-linear computational approach able to encode and store temporal information in its
212 layers and, through subsequent nodes, convert it into a classification output (Figure 3). For this reason,
213 LSTMs have been widely used to classify time series data (Karim et al., 2018). When trained on M+S
214 features (LSTM-NN-1), LSTM-NN showed a 71.0% accuracy, 76.5% sensitivity, 65.7% specificity,
215 68.4% precision and 72.2% F1 score; instead, the D+V+H+M+S combination (LSTM-NN-2) displayed
216 the same accuracy (71.0%), lower sensitivity (61.8%) higher specificity (80.0%), 75.0% precision and
217 67.7% F1 score. Implementation with a single feature (i.e., M or S) resulted in low accuracy (< 70%)
218 and, as before, this approach was abandoned.

219 Hybrid ensemble classifier

220 The results of the AI models (k-NN-1, k-NN-2, LSTM-NN-1 and LSTM-NN-2), showed a similar
221 classification accuracy to that reported by the operator (75.4%), although their comparison highlighted a
222 high classification variability for each single embryo (Figure 4A).

223 Next, we employed an ensemble classifier to integrate the results of the four AI models together with the
224 blind classification made by the expert operator, assuming, as decision rule, the agreement of at least 3
225 out of 5 classifiers. For its characteristics, this classifier was named *hybrid ensemble classifier* (HyEC).
226 The adoption of this strategy led to an improvement of all the performance metrics, with a final 82.6%
227 accuracy, 79.4% sensitivity, 85.7% specificity, 84.4% precision and 81.8% F1 score (Figure 4B).

228 In addition, the generalization capability of each of the four models and of the HyEC was evaluated by
229 10-fold cross validation on the entire dataset (230 embryos, 118 in NoBL class and 112 in BL class)
230 (Figure 5).

231

232 **Discussion**

233 The present study was designed to explore the use of a novel source of data to predict the ability of the
234 human embryo to develop to the blastocyst stage. This aim was pursued by harmonizing different
235 methodologies, i.e. TLM, advanced image analysis and AI computational and elaboration tools.
236 Collectively, the resulting data indicate that our approach can achieve a diagnostic accuracy classifiable
237 as “very good” (Šimundić, 2009). This is noteworthy, considering the proof-of-principle character of our
238 study and the small size of the data set analysed. However, such findings should be interpreted only in a
239 research context, lacking at present validation of their clinical applicability.

240 The current standard of embryo assessment in MAR relies on static morphological parameters (“The
241 Istanbul consensus workshop on embryo assessment: proceedings of an expert meeting.”, 2011). In the
242 large majority of clinics worldwide, embryos are ranked according to their presumed developmental
243 potential and prioritized for transfer at the cleavage (day three) or blastocyst (day five-six) stage. While
244 the two approaches generate comparable cumulative live birth rates after the transfer of fresh and
245 cryopreserved embryos (Glujovsky et al., 2016), self-selection derived from blastocyst culture offers the
246 advantage of a shorter time to pregnancy by reducing the number of embryos suitable for transfer.
247 However, because blastocyst culture is not immune from technical and management drawbacks, early
248 and robust prediction of development to blastocyst stage could significantly simplify and possibly
249 improve the ranking of embryos according to their developmental potential. As shown also by our study,
250 classical morphological criteria applied by an experienced embryologist have a measurable ability to
251 predict development to blastocyst stage. Nevertheless, a higher predictive power is required to make
252 short-term culture competitive with blastocyst culture. An approach similar to that adopted in our study
253 was used by Conaghan et al. (Conaghan et al., 2013). These authors reported that an algorithm applied
254 to TLM-based automatic annotation of embryo morphokinetics improves the ability of experienced
255 embryologists to predict development to the blastocyst stage. The potential of automatic morphokinetic
256 annotation and embryo selection at cleavage stages emerged also in another study based on the same
257 technology (Kieslinger et al., 2016). While these studies had the merit to introduce the concept of
258 combined operator-automated assessment, compared to our data it should be noted that in such cases
259 embryo assessment was extended to day three.

260 The ability of morphokinetic algorithms generated by TLM technology to predict development to
261 blastocyst and implantation has been extensively investigated as early as 2010 (Wong et al, 2010;
262 Meseguer et al., 2011), although without reaching a consensus (Armstrong et al., 2019). This has
263 prompted the need, recently tackled with the application of AI, to harness the immense potential of TLM

264 to generate morphokinetic data . An initial experience documenting the application of AI to oocyte and
 265 embryo scoring in clinical embryology was published as early as 2013 (Manna et al., 2013). Although
 266 relevant as a proof-of-principle and based on static images, this study was not followed by more
 267 extensive, independent investigations. The first major studies combining TLM technology and AI applied
 268 to human IVF were published only last year. Tran et al. (Tran et al., 2019) reported on a modality based
 269 on deep learning and raw data derived from TLM sequences of embryos developed to the blastocyst
 270 stage. In this research, the authors showed that the probability to achieve a clinical pregnancy can be
 271 automatically predicted with an AUC of 0.93. This study however requires independent confirmation,
 272 especially because its methodology appears largely unreported. In another study, Khosravi et
 273 al. (Khosravi et al., 2019) demonstrated that an AI platform trained on blastocyst images scored by
 274 experienced embryologists can predict blastocyst quality with an AUC >0.98 and the chances of
 275 pregnancy in a range between 13.8% and 66.3%, depending on blastocyst and patient characteristics. In
 276 addition, Rad and colleagues (Rad et al., 2019) reported that AI can predict embryo implantation based
 277 on the analysis of a single blastocyst image. A similar approach was adopted more recently also by
 278 VerMilyea and colleagues (VerMilyea et al., 2020), who described an improvement produced by AI of
 279 42% over routine embryologist assessment. These studies, together with others (Curchoe et al, 2019;
 280 Dirvanauskas et al., 2019; Kanakasabapathy et al., 2019), have stirred considerable interest in the field
 281 and will certainly be followed up by exciting developments.

282 Combining different technical and analytical tools, we conceived the present study to test the hypothesis
 283 that novel information derived from cytoplasmic movements occurring in the early embryo can be
 284 harnessed to develop predictive models of development.

285 Of note, focus on cytoplasmic movements as a source of data to predict human embryo development is
 286 a novelty in MAR. In fact, all embryo evaluation models, both static and dynamic, rely on large-scale

embryo morphological attributes (e.g. pronuclear position, cell number, degree of fragmentation, rate of blastocle expansion) and times of specific events (cleavages, compaction, start of blastocle formation), leaving cytoplasmic dynamics totally neglected. Therefore, in PIV analysis of cytoplasmic movements interpreted with the power of AI, we glimpsed a novel window of opportunity to improve early prediction of embryo developmental ability. Previous animal studies highlighted the significance of cytoplasmic dynamics for oocyte and embryo function. In a mouse model, Ajduk and colleagues (Ajduk et al., 2011) were the first to demonstrate that TLM associated with PIV can detect specific patterns of cytoplasmic flows occurring during fertilisation, which can be then used to predict embryo viability. Using the same image detection technology, we previously showed that mouse oocyte developmental competence, as assessed by chromatin rearrangement at the germinal vesicle stage, can be correctly predicted with a probability of 90% (Bui et al., 2017). Crucially, however, in this study we made more far-reaching and effective use of the data generated by TLM-PIV, by training an ANN. Therefore, we decided to exploit the excellent predictive power of this approach to assess embryo developmental competence *in vitro* in an MAR scenario. We recognise the limitations of our study, mainly represented by the small data set and the endpoint (blastocyst development), which is not the ultimate goal of infertility treatment. At the same time, however, we are encouraged by the significance and outcome of our study, which however should be intended as a proof-of principle. Firstly, compared with the AI studies of Tran et al. (Tran et al., 2019) and Khoshravi et al. (Khoshravi et al., 2019), we used a completely different and new source of data, i.e. analysis of cytoplasmic movement detected by PIV, as discussed above. This demonstrate that non-invasive bright field microscopy can potentially offer more than simply static morphology or morphokinetic information. In particular, our data suggest that sophisticated tracking of subcellular characteristics (discrete cytoplasmic particles) can add a “layer” of morphological information, qualitatively and quantitatively significant, never used or even considered so far for the analysis of human or animal cleavage-stage embryos. Secondly, our data suggest that AI has the potential to improve the

predictive ability of an experienced embryologist, achieving 82.6% accuracy, 79.4% sensitivity and 85.7% specificity to predict development to blastocyst. Human intervention may be seen as an obstacle to full automation. On the contrary we believe that valuable human skills should not be lost as an effect of technological development, but rather preserved and used. Thirdly, and particularly significant for the practice of MAR, our study suggests that blastocyst development can be predicted by day two, instead of day three as shown in other studies (Conaghan et al., 2013). Overall, however, two major limitations of our studies should be noted. The small sample size is compatible with the exploratory and preliminary nature of this investigation, but represents a weakness that prevents immediate application in a clinical scenario. Therefore, larger studies should be undertaken to confirm the validity of the present data and assure more stringent control of possible biases derived from clinical parameters. Indeed, we are in the process to set up a larger study focused on clinical endpoints. Not only will this require a larger study population and more participating clinics, but also an evolution of computational and IT tools in order to automate crucial steps of data extraction and elaboration, whose length and manual operativity represent an impediment to routine use of our model.

From a biological perspective, the study confirms the importance of early development for the correct unfolding of later stages. Interestingly, our approach could be further focused to the fertilization window, to assess the specific predictive power of this developmental segment. Consistent with this proposition, it is interesting to note that PIV patterns of competent and non-competent embryos diverge significantly during the interval of fertilisation comprised between pronuclear positioning and breakdown. Once again, this highlights the importance and predictive power of fertilisation events, as reported in recently published studies (Coticchio et al., 2018). Alternatively, the present approach could be extended to the cytoplasmic activity of human unfertilized oocytes, mirroring our previous mouse study (Bui et al., 2017).

From a more practical standpoint, should these preliminary data be confirmed, the possibility to limit embryo culture to two days would entail a reduction in costs and complexity of the IVF process compared with day five culture, while preserving a clinical efficiency comparable with blastocyst transfer.

Conclusions

Collectively, although not immediately translatable into clinical embryology practice, the present study suggests that movement of cytoplasmic particles of the human early embryo is a novel and valuable source of data to predict further development, after elaboration by advanced imaged analysis and AI. This opens new perspectives for non-invasive embryo assessment. Indeed, we are in the process to gather forces for a much larger study having live birth rate, as primary endpoint, and other parameters of clinical outcome as secondary endpoints.

Acknowledgments

This work was made possible thanks to support of 9.baby Family and Fertility Center, the Italian Ministry of Education, University and Research (MIUR) Dipartimenti di Eccellenza Program (2018–2022) to the Department of Biology and Biotechnology ‘L. Spallanzani’, University of Pavia, and a grant from the University of Pavia (FRG 2018). The authors also thank Merck-MilliQ Laboratory Water Solutions for support.

References

- Alpha Scientists in Reproductive Medicine and ESHRE Special Interest Group of Embryology, 2011. The Istanbul consensus workshop on embryo assessment: proceedings of an expert meeting., in: Presented at the Human reproduction (Oxford, England), pp. 1270–1283.
- Ajduk, A., Ilozue, T., Windsor, S., Yu, Y., Seres, K.B., Bomphrey, R.J., Tom, B.D., Swann, K., Thomas, A., Graham, C., Zernicka-Goetz, M., 2011. Rhythmic actomyosin-driven contractions induced by sperm entry predict mammalian embryo viability. *Nat Commun* 2, 417.
- Armstrong, S., Bhide, P., Jordan, V., Pacey, A., Marjoribanks, J., Farquhar, C., 2019. Time-lapse systems for embryo incubation and assessment in assisted reproduction. *Cochrane Database Syst Rev* 5, CD011320.
- Borini, A., Tarozzi, N., Bizzaro, D., Bonu, M.A., Fava, L., Flamigni, C., Coticchio, G., 2006. Sperm DNA fragmentation: paternal effect on early post-implantation embryo development in ART. *Hum Reprod* 21, 2876–81.
- Bui, T.T.H., Belli, M., Fassina, L., Vigone, G., Merico, V., Garagna, S., Zuccotti, M., 2017. Cytoplasmic movement profiles of mouse surrounding nucleolus and not-surrounding nucleolus antral oocytes during meiotic resumption. *Mol Reprod Dev* 84, 356–362.
- Conaghan, J., Chen, A.A., Willman, S.P., Ivani, K., Chenette, P.E., Boostanfar, R., Baker, V.L., Adamson, G.D., Abusief, M.E., Gvakharia, M., Loewke, K.E., Shen, S., 2013. Improving embryo selection using a computer-automated time-lapse image analysis test plus day 3 morphology: results from a prospective multicenter trial. *Fertil Steril* 100, 412–9.e5.

376 Coticchio, G., Mignini, R.M., Novara, P.V., Lain, M., De, P.E., Turchi, D., Fadini, R., Dal, C.M., 2018.
377 Focused time-lapse analysis reveals novel aspects of human fertilization and suggests new parameters of
378 embryo viability. *Hum Reprod* 33, 23–31.

379 Curchoe, C.L., Bormann, C.L., 2019. Artificial intelligence and machine learning for human
380 reproduction and embryology presented at ASRM and ESHRE 2018. *J Assist Reprod Genet* 36, 591–
381 600.

382 Dirvanauskas, D., Maskeliunas, R., Raudonis, V., Damasevicius, R., 2019. Embryo development stage
383 prediction algorithm for automated time lapse incubators. *Computer Methods and Programs in*
384 *Biomedicine* 177, 161–174.

385 Gallego, R.D., Remohí, J., Meseguer, M., 2019. Time-lapse imaging: the state of the art. *Biol Reprod*
386 101, 1146–1154.

387 Geler, Z., Kurbalija, V., Radovanović, M., & Ivanović, M., 2016. Comparison of different weighting
388 schemes for the kNN classifier on time-series data. *Knowl Inf Syst* 48, 331-378.

389 Geler, Z., Kurbalija, V., Ivanović, M., & Radovanović, M., 2020. Weighted kNN and constrained elastic
390 distances for time-series classification. *Expert Syst Appl* 162, 113829.

391 Glujovsky, D., Farquhar, C., Quinteiro, R.A.M., Alvarez, S.C.R., Blake, D., 2016. Cleavage stage versus
392 blastocyst stage embryo transfer in assisted reproductive technolog.. *Cochrane Database Syst Rev*
393 CD002118.

394 Griffin, D.K., Ogur, C., 2018. Chromosomal analysis in IVF: just how useful is it? *Reproduction* 156,
395 F29–F50.

396 He, J., Baxter, S.L., Xu, J., Xu, J., Zhou, X., Zhang, K., 2019. The practical implementation of artificial
397 intelligence technologies in medicine. *Nat Med* 25, 30–36.

398 Kanakasabapathy, M.K., Thirumalaraju, P., Bormann, C.L., Kandula, H., Dimitriadis, I., Souter, I.,
399 Yogesh, V., Kota Sai Pavan, S., Yarravarapu, D., Gupta, R., Pooniwala, R., Shafiee, H., 2019.
400 Development and evaluation of inexpensive automated deep learning-based imaging systems for
401 embryology. *Lab Chip* 19, 4139–4145.

402 Khosravi, P., Kazemi, E., Zhan, Q., Malmsten, J.E., Toschi, M., Zisimopoulos, P., Sigaras, A., Lavery,
403 S., Cooper, L.A.D., Hickman, C., Meseguer, M., Rosenwaks, Z., Elemento, O., Zaninovic, N.,
404 Hajirasouliha, I., 2019. Deep learning enables robust assessment and selection of human blastocysts after
405 in vitro fertilization. *NPJ Digit Med* 2, 21.

406 Kieslinger, D.C., De Gheselle, S., Lambalk, C.B., De Sutter, P., Kosteljik, E.H., Twisk, J.W.R., van
407 Rijswijk, J., Van den Abbeel, E., Vergouw, C.G., 2016. Embryo selection using time-lapse analysis
408 (Early Embryo Viability Assessment) in conjunction with standard morphology: a prospective two-
409 center pilot study. *Human Reproduction* 31, 2450–2457.

410 Manna, C., Nanni, L., Lumini, A., Pappalardo, S., 2013. Artificial intelligence techniques for embryo
411 and oocyte classification. *Reprod Biomed Online* 26, 42–9.

412 Meseguer, M., Herrero, J., Tejera, A., Hilligsøe, K.M., Ramsing, N.B., Remohí, J., 2011. The use of
413 morphokinetics as a predictor of embryo implantation. *Human Reproduction* 26, 2658–2671.

414 Miyagi, Y., Habara, T., Hirata, R., Hayashi, N., 2019. Feasibility of artificial intelligence for predicting
415 live birth without aneuploidy from a blastocyst image. *Reprod Med Biol* 18, 204–211.

416 Rad, R.M., Saeedi, P., Au, J., Havelock, J., 2019. Predicting Human Embryos' Implantation Outcome
417 from a Single Blastocyst Image. Conf Proc IEEE Eng Med Biol Soc 2019, 920–924. Šimundić, A.M.,
418 2009. Measures of Diagnostic Accuracy: Basic Definitions. EJIFCC 19, 203–11.

419 Swain, J.E., 2019. Controversies in ART: considerations and risks for uninterrupted embryo culture.
420 Reprod Biomed Online 39, 19–26.

421 Tran, D., Cooke, S., Illingworth, P.J., Gardner, D.K., 2019. Deep learning as a predictive tool for fetal
422 heart pregnancy following time-lapse incubation and blastocyst transfer. Hum Reprod 34, 1011–1018.

423 Tran, T. M., Le, X. M. T., Nguyen, H. T., & Huynh, V. N., 2019. A novel non-parametric method for
424 time series classification based on k-Nearest Neighbors and Dynamic Time Warping Barycenter
425 Averaging. Eng Appl Artif Intell 78, 173-185.

426 VerMilyea, M., Hall, J.M.M., Diakiw, S.M., Johnston, A., Nguyen, T., Perugini, D., Miller, A., Picou,
427 A., Murphy, A.P., Perugini, M., 2020. Development of an artificial intelligence-based assessment model
428 for prediction of embryo viability using static images captured by optical light microscopy during IVF.
429 Human Reproduction 35, 770–784.

430 Wong, C.C., Loewke, K.E., Bossert, N.L., Behr, B., De Jonge, C.J., Baer, T.M., Reijo-Pera, R.A., 2010.
431 Non-invasive imaging of human embryos before embryonic genome activation predicts development to
432 the blastocyst stage. Nature Biotechnology 28, 1115–1121.

433 World Health Organization. WHO laboratory manual for the examination and processing of human
434 semen. Fifth edition.

435 Yi, K., Rubinstein, B., Unruh, J.R., Guo, F., Slaughter, B.D., Li, R., 2013. Sequential actin-based pushing
436 forces drive meiosis I chromosome migration and symmetry breaking in oocytes. J. Cell Biol.

437 Zacà, C., Bazzocchi, A., Pennetta, F., Bonu, M.A., Coticchio, G., Borini, A., 2018. Cumulative live birth
438 rate in freeze-all cycles is comparable to that of a conventional embryo transfer policy at the cleavage
439 stage but superior at the blastocyst stage. *Fertil Steril* 110, 703–709.

440

441

442 **Figure Legends**

443 **Figure 1. Embryo movement velocity profiles of BL and NoBL embryos.** (A) Representative image
444 frame of an embryo at 32 hours after ICSI showing the Cell-PIV velocity vectors. The color and length
445 of the arrows (vectors) indicate the velocity module of the movements when comparing the previous
446 frame with that pictured. Colored vector scale bar: blue, lower velocity; purple, higher velocity. (B-F)
447 EMV profiles analysed using *Direction*, *Vorticity*, *Hybrid*, *Meanmag* or *Summed* Cell_PIV feature,
448 respectively.

449 **Figure 2. Cytological events occurring during the first two segmentation divisions.** The *Summed*
450 EMV profile of BL and NoBL embryos resulted significantly different during the time-intervals
451 corresponding to pronuclei re-allocation and the second embryonic division.

452 **Figure 3. LSTM Neural Network architecture.** The structure of the LSTM-NN is made of five main
453 layers. An input layer, whose number of nodes is equal to the number of features used; a bidirectional
454 LSTM layer, made with 3 nodes, whose function is to bring out hidden patterns along the time series; a
455 fully connected layer and a softmax layer, convert the temporal information stored in the LSTM layer in
456 a classification probability; an output layer provides the class with the highest probability.

457 **Figure 4. Improved embryo-quality classification with *hybrid ensemble classifier*.** (A) Test embryos
458 correctly (green) or incorrectly (red) classified by the AI models (k-NN-1, k-NN-2, LSTM-NN-1 or
459 LSTM-NN-2), by the operator or the *hybrid ensemble classifier* (HyEC). (B) Performance metrics
460 obtained by the operator alone or by the HyEC.

461 **Figure 5. 10-fold cross validation Mean and 95% Confidence Intervals of Accuracy, Specificity,**
462 **Sensitivity, Precision and F1 score, computed on the entire dataset (230 embryos, 118 in NoBL class and**
463 **112 in BL class) for the k-NN-1, k-NN-2, LSTM-NN-1, LSTM-NN-2 and HyEC classifiers.**

Figure 1

Figure 2

Figure 3

Figure 4

Figure 5