Assessing morphokinetic parameters via time lapse microscopy (TLM) to predict euploidy: are aneuploidy risk classification models universal?

Yael G. Kramer · Jason D. Kofinas · Katherine Melzer · Nicole Noyes ·
Caroline McCaffrey · Julia Buldo-Licciardi · David H. McCulloh ·
James A. Grifo

Received: 9 February 2014 / Accepted: 27 May 2014 / Published online: 25 June 2014
© Springer Science+Business Media New York 2014

Abstract

Purpose To determine if Aneuploidy Risk Classification Models are predictive of euploidy/aneuploidy amongst IVF facilities.

Methods We retrospectively applied key time lapse imaging events of embryos (Campbell et al.[5, 6]) to stratify embryos into 3 groups: low, medium and high risk of aneuploidy. The actual ploidy results (from array comparative genomic hybridization) were compared with expectations [5, 6]. Sources of variability in morphokinetic parameters were determined using Analysis of Variance (ANOVA).

Results The model failed to segregate euploid embryos from aneuploid embryos cultured at our facility. Further analysis indicated that the variability of embryos among patients was too great to allow selection of euploid embryos based on simple morphokinetic thresholds. Clinical selection of embryos based on morphokinetics alone is unlikely to identify euploid embryos accurately for transfer or yield higher rates of live delivery.

Capsule Morphokinetic parameters are insufficient to universally distinguish euploid from aneuploid embryos, predominantly due to the high degree of patient-to-patient variability and small inpatient variability for parameters previously associated with blastocyst development or embryo ploidy.

Y. G. Kramer (✉) · J. D. Kofinas · K. Melzer · N. Noyes ·
C. McCaffrey · J. Buldo-Licciardi · D. H. McCulloh · J. A. Grifo
NYU Fertility Center, NYU Langone School of Medicine, 660 First
Avenue, 5th floor, New York, NY 10016, USA
e-mail: yael.kramer@nyumc.org

Present Address:

K. Melzer
Genesis Fertility and Reproductive Medicine, 1355 84th Street,
Brooklyn, NY 11228, USA

Conclusions The use of non-invasive morphokinetics is unlikely to discriminate aneuploid from euploid embryos. Further, it does not approach the accuracy of preimplantation genetic screening with array comparative genomic hybridization.

Keywords Time-Lapse · Embryoscope · Preimplantation genetic screening · Morphokinetics · Aneuploidy risk models

Introduction

The goal of an IVF cycle is the delivery of a single healthy euploid fetus. Aneuploidy has been shown to be a cause of recurrent miscarriage [1] and the age related decline in fertility [2]. Analysis of morphokinetic parameters has been used to predict embryos that will attain the blastocyst stage [3], will implant [4] or are euploid [5, 6] (and Melzer et al., unpublished data). However, there are wide disparities in the morphokinetic events chosen in different facilities as to which important discriminatory events are crucial in selection of embryos with good prognosis over those with poor prognosis.

Campbell et al. (2013) [5, 6] presents an aneuploidy prediction model in which observed morphokinetic times were used to partition embryos into three groups corresponding to risk levels of aneuploidy (low, medium and high risk of aneuploidy). A significantly higher incidence of aneuploidy was seen in the high risk group (100 % of 12 embryos) and significantly lower incidence of aneuploidy in the low risk group (36 % aneuploidy of 36 embryos). It is important to note that even in the low risk group each embryo had a 36 % chance of being chromosomally abnormal.

Our study investigates morphokinetic parameters in embryos of known ploidy in order to determine if morphokinetic parameters can discriminate euploid from

aneuploid embryos. We have previously concluded that morphokinetic parameters are not particularly good discriminators of embryo ploidy (Melzer et al., unpublished data). In this manuscript, we investigated sources of variability for morphokinetic parameters, determining that the patient-to-patient variability of morphokinetic parameters was so great that it would make universal discrimination of ploidy by these parameters impossible.

Materials & methods

Patient population

Twenty five patients were undergoing aneuploidy screening for a variety of reasons: recurrent pregnancy loss ($n=8$), known translocation ($n=1$), family balancing ($n=2$), advanced maternal age ($n=11$) and unexplained (5). The mean age was 37.3 ± 3.9 y (27–43 y). These patients chose to undergo trophoctoderm (TE) biopsy due to various reasons such as aneuploidy screening (due to a history of recurrent pregnancy loss or infertility or advanced maternal age) [7], family balancing, or known parental translocation [8].

IVF stimulation

Stimulation protocols were tailored based on ovarian reserve, which varied considerably with some patients having clear evidence of diminished ovarian reserve with elevated day two FSH levels ($\text{max}=20$) and low AMH levels (<0.16) whereas others had normal serum testing. Patients received either FSH and/or human menopausal gonadotropins, and a GnRH antagonist was administered when a lead follicle reached 13 mm or if estradiol >1000 pg/mL. When at least 2 lead follicles reached ≥ 17 –18 mm, an ovulation trigger was administered using human chorionic gonadotropin ($n=19$) or GnRH agonist ($n=2$; leuprolide acetate, 0.4 cc=2 mg) or combination of both ($n=4$). Approximately 35 hours later, oocytes were collected via ultrasound-guided transvaginal aspiration.

Embryo culture, biopsy and vitrification

Following retrieval, oocytes were rinsed in hepes buffered media (LifeGlobal, Ontario, Canada) and immediately placed in 75 μ l droplets of pre-equilibrated Global for Fertilization media overlaid with mineral oil and incubated at 37 °C, 6%CO₂/air. Partner's sperm was collected on the day of retrieval or, if using frozen partner or donor sperm, thawed and processed to select motile sperm. Oocytes were inseminated at 4–6 h post retrieval or intracytoplasmic sperm injection (ICSI) performed when indicated for severe male factor or a prior history of failed or low fertilization. Fertilization was assessed ~18 h later (on day 1) by visualization of two

pronuclei (2PN). All embryos exhibiting 2 pronuclei (normal fertilization) were transferred to the Embryoscope™ incubator for culture and continuous TLM. Culture conditions as customary for Embryoscope™ culture consisted of individual embryo culture in 25 μ l drops of single-step medium (LifeGlobal) overlaid with 1.2 ml mineral oil (LifeGlobal) in embryoSlides (Unisense Fertilitech) at 37 °C, 6 % CO₂, 5%O₂ and 89%N₂. Embryos were removed from the TLM incubator for a brief period on day 3 and laser assisted zona-ablation performed (Cronus laser, Research Instruments; Falmouth, United Kingdom) to facilitate later biopsy of the trophectoderm once the embryo (s) had reached the blastocyst stage. Embryos were graded on days 5/6 according to standard morphologic criteria as described by Gardner and Lane [9]. Timing of key events characterized by the Campbell et al. “Risk for Aneuploidy” model [5, 6] were recorded: Initiation of compaction, initiation of blastulation, time to full blastocyst. Time of pronuclear envelope breakdown (Syngamy) was also recorded to be used as a normalization factor.

On Day 5 all blastocysts that had reached a sufficiently advanced stage of development (at least stage 2Cc) with a discernable inner cell mass (ICM) and trophectoderm (TE) were subjected to TE biopsy. Any embryos not suitable for biopsy were cultured to day 6 for possible biopsy. On the day of the biopsy a piece of the extruded trophectoderm, comprising several TE cells, was isolated through micromanipulation and incised using the laser. The biopsied cells were then placed in Eppendorf tubes, frozen in dry ice, and transported to Reprogenetics for chromosomal analysis. These cells were analyzed using 24-chromosome array aCGH as previously described [10, 11]. Embryos were cryopreserved using vitrification, being first equilibrated in media containing the lowest concentration of cryoprotectants (7.5 % ethylene glycol [EG] and 7.5 % dimethyl sulfoxide [DMSO]) to achieve the first level of dehydration. They were then placed in vitrification solution with cryoprotectants (15 % EG and 15 % DMSO; Vitrification Freeze Solutions for Embryo, Irvine Scientific, Santa Ana, CA). Embryos were then loaded onto Cryolocks (Cummings, GA) and immediately plunged directly into liquid nitrogen within 60–90 s.

Morphokinetic parameters

For each of the biopsied embryos, morphologic and kinetic variables were assessed prior to receiving the biopsy results. Images of the embryos were recorded every 20 min in seven focal planes. The timing of key morphokinetic events was recorded and reported as number of hours following insemination or ICSI. Recorded time points were selected based on previously described parameters [12–14] and included: onset of syngamy, appearance of a 2cell

cleavage furrow, attainment of the 2cell stage, length of division, 2–3 cell interval, 3 cell stage, 4 cell stage, 3–4 cell interval, 5 cell stage, 8 cell stage, start of compaction, duration of compaction, the start of cavitation, full blastocyst stage. Compaction was defined as the first time-point when blurred boundaries were visualized between cells and the overall diameter of the embryo decreases; cavitation as the first time-point at which a blastocoele was visible [13, 15]. Lastly, aneuploidy risk class was calculated as described by Campbell et al. (2013) [5, 6]. These criteria include time to initiation of compaction, initiation of blastulation, and time to full blastocyst.

Normalizing time morphokinetic parameters to time of syngamy

We chose to determine all of the morphokinetic times relative to the time of pronuclear envelope breakdown (syngamy) to avoid any variability in the time of sperm entry in oocytes inseminated by ICSI versus in vitro insemination. This was done by subtracting the time of syngamy from the time of the morphokinetic event. The resulting morphokinetic parameters had smaller standard deviations than the values measured from time of insemination, suggesting that this adjustment led to a more uniform method of measurement. These numbers were then analyzed using the Campbell model (with the average time of syngamy in their data subtracted from their criterion times).

Statistical analysis

Numbers of embryos in different groups were compared using Chi Squared. The ability to predict ploidy using morphokinetic parameters was assessed using the area under the Receiver Operator Characteristic (ROC) curve. Times and durations of stages were compared using Analysis of Variance.

Much of the statistical analysis was devoted to determining whether aneuploid and euploid embryos could be distinguished based upon the values of morphokinetic parameters (times and durations). We created criteria that we believe are necessary to be able to discriminate euploid from aneuploid embryos using morphokinetic parameters.

Criteria that must be satisfied to discriminate euploid from aneuploid Embryos

In considering whether morphokinetic parameters can be used to discriminate euploid embryos from aneuploid embryos, differences between euploid and aneuploid embryos must exist and must be sufficiently systematic that they continue

to be detected across all patients subjected to the morphokinetic analysis.

As a first level of consideration, we must consider whether a detectable difference exists between euploid and aneuploid embryos. We must consider whether differences can be observed within the embryos from each patient couple. As a second level of consideration, it is important that the differences between euploid and aneuploid embryos are large enough to remain distinguishable despite the greater variability inherently present when we consider embryos from different patient couples.

We have established several criteria that must be satisfied for morphokinetic parameters to be useful in discriminating euploid from aneuploid embryos. Their descriptions follow:

Criterion 1. A difference between euploid and aneuploid embryos must exist.

Differences between euploid and aneuploid embryos may take two different forms. Only one of these two criteria must be satisfied. The two different criteria (1A and 1B) are described as follows:

Criterion 1A. Differences in the mean value of the parameter

An example of such a difference could be when the time of a morphokinetic event occurs earlier in euploid embryos than it occurs in aneuploid embryos. The method for detecting such a difference would involve the use of a statistical test that distinguishes between the mean values M_{euploid} and $M_{\text{aneuploid}}$. We have used Analysis of Variance (ANOVA) to determine if significant differences are detected between the means of morphokinetic parameters from euploid embryo versus aneuploid embryos. The mean squared deviation within patients for all embryos (euploid and aneuploid) was compared with the mean squared deviation within patients for only euploid embryos.

Criterion 1B. Difference in the variability of the two parameters

An example of such a difference could be when the time of a morphokinetic event occurs at a particular time in euploid embryos but could be both earlier and later for aneuploid embryos. The method for detecting such a difference would involve the use of a statistical test that distinguishes between the standard deviations or the variances of the two sets of parameters: $\text{Var}_{\text{euploid}}$ and $\text{Var}_{\text{aneuploid}}$. We have used Analysis of Variance (ANOVA) to determine if significant differences are detected between the variances of morphokinetic parameters from euploid embryos versus aneuploid embryos. The mean squared deviation within patients for aneuploid embryos was compared with the mean squared deviation within patients for euploid embryos.

Criterion 2. Method must distinguish euploid from aneuploid embryos when applied across all patients

If a significant difference is detected, and a threshold (from Criterion 1A) or window (from Criterion 1B) is to be assigned to discriminate euploid from aneuploid embryos, then the difference must continue to discriminate euploid from aneuploid when applied across all patients. Euploid embryos must be similar to each other, and aneuploid embryos must be different from or more divergent than euploid embryos.

The variability of the morphokinetic parameter among patients should satisfy both Criterion 2A and 2B:

Criterion 2A The variability of the morphokinetic parameter for aneuploid embryos (or all embryos) among patients must be large compared to the variability of the same parameter for euploid embryos among patients. (The value of the parameter in euploid embryos should be characteristically more uniform than the variability of the parameter for aneuploid (or all embryos) among patients.) We have used ANOVA to compare the mean squared deviation for aneuploid embryos (or all embryos) among patients with the mean squared deviation for euploid embryos among patients. This comparison must be significant. If this is not significant, then the aneuploid embryos are not sufficiently different from euploid embryos to be universally distinguished from the euploid embryos.

Criterion 2B The variability of the parameter for euploid embryos among patients should not be large compared to the variability of the parameter for euploid embryos within patients. (The target value should not vary greatly from patient to patient). We have used ANOVA to compare the mean squared deviation for euploid embryos among patients with the mean squared deviation for euploid embryos within patients. This comparison must not be significant. When this is significant, the euploid embryos vary enough from patient to patient that universal criteria for selection are unlikely to be found.

Criterion 1A or 1B and Criteria 2A and 2B must be satisfied for discrimination to be effective and to universally and significantly enrich for euploidy.

Results

Embryos from 25 patients undergoing IVF cycles with PGS underwent culture. 360 oocytes were retrieved and fertilized

with routine insemination (n=18 patients) or intra cytoplasmic sperm injection/ICSI (n=7 patients). 235 embryos fertilized normally and were transferred to the Embryoscope™ incubator for further individual culture in a single-step media (LifeGlobal).

Of the 235 embryos, 149 blastocysts met criteria and were suitable for TE biopsy either on day 5 (n=108) or day 6 (n=41). Sixty-three (43 %) of the embryos were euploid, and 82 (57 %) embryos were aneuploid. Four embryos had chaotic profiles such that analysis was not possible and these were excluded from further consideration.

There was no difference in euploidy ($p>0.2$) based on day of biopsy: of the day 5 embryos, 49 (45 %) were euploid and 59 (55 %) were aneuploid. Of the day 6 embryos, 14 (34 %) were euploid and 27 (66 %) were aneuploid.

Morphokinetic events

Application of Campbell's Model to our data Campbell et al. [5, 6] created a mechanism for predicting euploidy based upon the values of two morphokinetic parameters: the time to blastocoele formation and the time to blastocyst expansion. We applied the criteria of Campbell et al. [5, 6] to our assembled morphokinetic parameters to determine if their criteria could predict the ploidy outcomes for our embryos. Using their criteria, we obtained predictions of 37 %, 69 % and 97 % aneuploidy in the low, medium and high risk groups, respectively (Fig. 1). The incidence of aneuploidy that we observed in these same three groups was 50.7 %, 60.3 % and 63.6 %. These values were significantly different than the Campbell expectations ($\chi^2=9.5$ with 2 d.f., $0.01<p<0.02$) and were not significantly different from a uniform expectation of 55.7 % in each group ($\chi^2=1.5$, with 2° of freedom, $p>0.25$). Therefore, the Campbell criteria were not able to provide significant enrichment in euploid embryos using

Fig. 1 Comparison of NYU results to published Campbell results

their criteria for our embryos monitored using the same morphokinetic parameters.

It should be noted however, that the timings of morphokinetic events were measured differently in our patients since standard in vitro insemination rather than ICSI was employed for 75 % of the resulting embryos in the data set that we analysed. Thus, we chose to determine all of the morphokinetic times relative to the time of pronuclear envelope breakdown (syngamy) to avoid any variability in the time of sperm entry in oocytes inseminated by ICSI versus in vitro insemination. 25 patients with 133 embryos had data for timing of syngamy. The remainder of the analysis focused on these embryos. The resulting morphokinetic parameters had smaller standard deviations than the values measured from time of insemination (data not presented), suggesting that this adjustment led to a more uniform method of measurement.

Upon adjustment of the Campbell criteria by subtraction of the average time of pronuclear envelope breakdown from the criterion time, we applied these corrected criteria to the adjusted morphokinetic parameters. Using their criteria, we obtained predictions of 37 %, 69 % and 97 % aneuploidy in the low, medium and high risk groups, respectively (Fig. 1). The incidence of aneuploidy that we observed in these same three groups was 54.5 %, 56.8 % and 66.6 %. Again, these values were significantly different than the Campbell expectations ($\chi^2=14.3$ with 2° of freedom; $p<0.001$) and were not significantly different from a uniform expectation of 56.4 % in each group ($\chi^2=0.62$, with 2° of freedom, $p>0.25$) (No Discrimination, Fig. 1).

Table 1 Areas under the curve for receiver operator characteristic (ROC) curves

Parameter	Area under ROC curve
Syngamy	0.529
2 cell furrow	0.539
2 cells	0.550
Length division	0.504
2 - 3 cell interval	0.520
3 cells	0.538
4 cells	0.533
3 - 4 cell interval	0.515
5 cells	0.536
8 cells	0.524
Starts compacting	0.565
Duration Compaction	0.674
Starts cavitation	0.535
Full blastocyst	0.562

Determination of the ploidy predictability within our own data

We examined our data for morphokinetic parameters in association with ploidy outcomes to find if any parameters were associated with ploidy diagnosis. The best test to determine if a morphokinetic parameter predicts ploidy is the determination of the area under the receiver operator characteristic curve (ROC curve). Areas under the curve (AUCs) in the neighborhood of 0.5 are indicative of a test that is not capable of distinguishing ploidy whereas AUCs approaching 1.0 are indicators of an excellent test. The only parameter that yielded an AUC indicative of any predictive value was the Duration of Compaction with an AUC of 0.674 (Table 1), an AUC comparable to the 0.72 value observed using the test paradigm described by Campbell et al. [5, 6] using their own data. All other AUCs were near 0.5 and were indicative of tests that were not capable of distinguishing ploidy (Fig. 2).

How could two such similar studies come to such divergent conclusions? This is what we hope to elucidate in the next sections.

Do systematic differences exist? Since we were able to discriminate euploid embryos from aneuploid embryos only poorly based upon morphokinetic parameters, we chose to examine the sources of variation in the morphokinetic parameters. This was done in the hope that understanding the source of the variability could help to explain why two data sets yield significantly different conclusions about parameters that predict ploidy.

Standard deviations of the morphokinetic parameters ranged between 11.6 % and 202 % of the mean values, indicating that the morphokinetic values may vary widely from embryo to embryo. (Table 2)

In order to determine how much of the variability was due to variability between euploid and aneuploid embryos, how much of the variability was due to variability in embryos

Fig. 2 Area Under the Curve for various time lapse time points

Table 2 Morphokinetic parameters and Results of Two Way Analysis of Variance (ANOVA) for patients, ploidy and interaction

Morphokinetic Parameters a	PB extr	PN abutt	Furrow	2 cells	Len div	Cc2	3 cell	4 cell	3cell- 4cell	5 cell	8 cell	Starts comp	Ends comp	Durat comp	Starts cav	blast
Mean±std ^b dev	-20±2.8	-12±2.9	2.2±0.39	2.8±1.0	0.7±0.98	11±2.4	14±2.1	15±1.8	1.0±2.0	27±4.5	33±6.1	50±12	59±11	13±7.6	72±9.1	84±10.0
F (pts) ^c	2.58	2.81	2.75	2.82	0.65	3.39	3.45	3.45	0.54	3.44	2.77	3.48	3.35	2.60	3.73	3.84
d.f	4/11	4/12	17/52	18/68	17/52	18/67	18/67	18/67	18/67	18/67	18/56	23/91	18/65	19/65	23/89	22/89
sig	NS	NS	Sig.	Sig	NS	Sig.	Sig.	Sig.	NS	Sig.	Sig.	Sig.	Sig.	Sig.	Sig.	Sig.
F (ploidy) ^d	.12	.07	.08	.22	.71	.01	.03	.02	.00	.03	.01	.07	.04	2.70	.01	.03
d.f.	1/11	1/12	1/52	1/68	1/52	1/67	1/67	1/67	1/67	1/67	1/56	1/91	1/65	1/65	1/91	1/89
sig	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS
F (intrct) ^e	0.27	0.01	0.11	0.08	.00	0.17	.16	.13	.01	.15	.17	.24	.14	.27	.22	.23
d.f.	4/11	4/12	17/52	18/68	17/52	18/67	18/67	18/67	18/67	18/67	18/56	23/91	18/65	19/65	23/91	22/89
sig	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS

^a Time of second polar body extrusion (PB extr). Time of initial juxtaposition and centralization of PN's (PN abutt). Time zygote undergoes first mitosis and beginnings of zygotic groove is seen (Furrow). Time of first embryo division to two distinct daughter cells (2 cells). Time interval between Furrow and 2 cells (Len div). Time interval between 2 cells 3 and cell (Cc2). Time of division into three distinct daughter cells (3 cell). Time of division into four distinct daughter cells (4 cell). Time interval between 3 cell to 4 cell (3cell-4cell). Time of division into five distinct daughter cells (5 cell). Time of division into eight distinct daughter cells (8 cell). Time when cells within embryo begin to adhere and lose an individual appearance (Starts comp) [5]. Time when embryo appears as indistinguishable mass of cells where cell boundaries are unclear (Ends Comp). Time interval between Starts comp and Ends comp (Dur comp). Earliest stage of a blastocyst with a small cavity visible (Starts cav) [5]. Time of formation of full blastocyst; when blastocoele filled embryo with <10% increase in diameter (blast) [5]

^b mean±standard deviation for morphokinetic parameters (presented in hours from syngamy). Negative values indicate morphokinetic parameters that preceded syngamy

^c F value for patients (mean squared deviation among patients/mean squared deviation within ploidy groups)

^d F value for ploidy (mean squared deviation among ploidy types (aneuploid and euploid)/mean squared deviation within ploidy groups)

^e F value for interaction (mean squared deviation for interaction between ploidy and patients/mean squared deviation within ploidy groups)

arising from within each patient couple and how much was due to the variability between embryos from different patient couples, we performed a two-way analysis of variance (ANOVA). The results of the two way ANOVA are displayed in Table 2.

The two way ANOVAs compared the mean squared deviation attributable to patients or to ploidy with the mean squared deviation within ploidy groups within patients and outcomes were quite consistent across all morphokinetic parameters. Significance was found for interpatient variability for nearly all morphokinetic parameters, indicating that morphokinetic parameters vary more from patient to patient than they do within patients (significant). Four morphokinetic parameters for early developmental events (PB extrusion, PN abuttal, length of division, and delay from the 3 to 4 cell stage) were not significant.

No significance was found for ploidy indicating that there was no significant difference in morphokinetic parameters comparing euploid to aneuploid embryos, taking into account the patient to patient variability. This indicated that there were no significant differences in the mean values of the morphokinetic parameters between ploidy groups. In addition, there was no significant interaction between patients and ploidy groups.

Consideration of criterion 1 for morphokinetic parameters

Tests for differences in the mean values of morphokinetic parameters. For all morphokinetic parameters examined (except for the duration of compaction), the F value comparing the mean squared deviation for all ploidies with the mean squared deviation within ploidy groups within patients was not significant. This indicated that variations in the means of the morphokinetic parameters were not associated with ploidy (Table 3 top). The F values for the time to the 4 cell stage and for the duration of compaction were significant. Therefore, there was no difference in timing between aneuploid and euploid embryos with the exception of these two times (Table 3).

Tests for differences in the variability of morphokinetic parameters. This comparison was made using the mean squared deviations within groups. For all morphokinetic parameters examined except 4, the variability of morphokinetic parameters for aneuploid embryos was not significantly greater than the variability of that morphokinetic parameter in euploid embryos (Table 3, bottom section). This indicates that there is no significant association between euploidy and any central tendency for the parameter. The four exceptions were the time of appearance of the first cleavage furrow, the time of attaining the 4 cell stage, completion of compaction and the time when cavitation began. This indicates that within each patient, it may be possible to discriminate euploid from

Table 3 ANOVA for comparisons of means (Criterion 1A) and comparisons of variances (Criterion 1B) within patients

Morphokinetic Parameters	PB extr	PN abutt	furrow	2 cells	Len Div	Cc2	3 cell	4 cell	3cell-4cell	5 cell	8 cell	Starts comp	Ends comp	Dur comp	Starts cav	blast
F w/in (all ploidies)/euploid ^a	0.83	0.83	1.45	0.514	0.47	0.77	0.83	1.88	1.09	1.02	1.05	0.89	1.13	2.03	1.25	1.48
d.f.	15/5	16/5	66/21	84/28	66/21	83/28	83/28	83/28	83/28	83/28	71/23	112/37	80/26	80/26	112/37	110/37
sig	NS	NS	NS	NS	NS	NS	NS	Sig.	NS	NS	NS	NS	NS	Sig.	NS	NS
F (w/in) aneuploid/euploid ^b	0.40	1.16	2.04	0.25	0.15	0.69	0.83	2.82	1.56	1.11	1.09	0.86	1.15	1.91	1.26	1.62
d.f.	6/5	7/5	31/21	40/28	31/21	39/28	39/28	39/28	39/28	39/28	33/23	54/37	39/26	39/26	54/37	52/37
sig	NS	NS	Sig.	NS	NS	NS	NS	Sig.	NS	NS	NS	NS	Sig.	Sig.	NS	NS

^a F value for comparison of means (Mean squared deviation of all ploidies within patients (aneuploid plus euploid)/mean squared deviation for euploid embryos within patients)

^b F value for comparison of variances (Mean squared deviation of aneuploid embryos within patients/mean squared deviation of euploid embryos within patients)

aneuploid embryos. In each case, euploid embryos tended to have times from syngamy to the morphological event that were more tightly clustered than the times for aneuploid embryos within that same patient (Table 3 bottom).

The first criterion is satisfied only for four parameters: the time to appearance of the first cleavage furrow, the time to the 4 cell stage, time to completion of compaction, and the duration of compaction. Two parameters, the time to the 4 cell stage and the duration of compaction were detectable both as a difference in the mean values of the parameter and as examples of aneuploid embryos being more variable than euploid embryos. The remaining two parameters, time to the first cleavage furrow and time to completion of compaction were examples in which aneuploid embryos were significantly more variable than euploid embryos, indicating that euploid embryos cluster around a value within the range of both euploid and aneuploid embryos.

Consideration of criteria 2A and 2B for morphokinetic parameters

Whereas several parameters satisfy the first criterion, it is necessary to determine if these parameters satisfy the additional criteria 2A and 2B for discrimination of euploid from aneuploid embryos.

Criterion 2A was investigated by two types of ANOVA. The first ANOVA compared the mean squared deviations for all embryos among patients with the mean squared deviations for euploid embryos among patients (Table 4 top). There were significant differences for the means of only time to the three cell stage and time to start of compaction. For the other means there was no significance.

The second ANOVA compared the mean squared deviations for aneuploid embryos among patients with the mean squared deviation for euploid embryos among patients (Table 4 bottom). There was a significant difference for the variances of only one of the morphokinetic parameters examined, the time to the 3 cell stage. For this stage the variability of aneuploid embryos was greater than the variability of euploid embryos.

Hence only two of the parameters satisfy Criterion 2A, one of them using both methods

Criterion 2B was investigated by comparing the mean squared deviations for euploid embryos among patients with the mean squared deviations for euploid embryos within patients (Table 5). Five parameters (4 cell, 8 cell, start of compaction, start of cavitation, and complete blastocyst) were significant, making these five parameters incapable of use for discriminating euploid from aneuploid embryos. All other parameters were not significant making them eligible for use in discriminating euploid from aneuploid embryos (Table 5).

Table 4 ANOVA for comparisons of means and variances (Criterion 2A) across all patients

Morphokinetic Parameters	PB extr	PN abut	furrow	2 cells	Len Div	Cc2	3 cell	4 cell	3cell-4cell	5 cell	8 cell	Starts comp	Ends comp	Dur comp	Starts cav	blast
F (among) All/euploid ^a	5.3	4.5	1.6	0.71	0.54	1.6	3.2	1.05	2.25	1.90	0.97	2.14	1.48	1.75	1.14	1.05
d.f.	4/3	4/3	17/13	17/15	17/13	17/15	17/15	17/15	17/15	17/15	17/15	22/20	17/14	18/15	22/20	22/20
Sig	NS	NS	NS	NS	NS	NS	Sig	NS	NS	NS	NS	Sig	NS	NS	NS	NS
F (among) aneuploid/euploid ^b	4.29	3.43	1.00	0.13	0.16	1.30	3.27	0.62	1.53	1.64	0.50	1.57	1.44	2.28	0.76	0.65
d.f.	4/3	4/3	17/13	17/15	17/13	17/15	17/15	17/15	17/15	17/15	17/14	22/20	17/14	17/15	22/20	22/20
sig	NS	NS	NS	NS	NS	NS	Sig.	NS	NS	NS	NS	NS	NS	NS	NS	NS

^a F value for comparison of means (Mean squared deviation of all ploidies among patients (aneuploid plus euploid)/mean squared deviation for euploid embryos among patients)

^b F value for comparison of variances (Mean squared deviation of aneuploid embryos among patients/mean squared deviation of euploid embryos among patients)

Table 5 ANOVA for comparison of euploid embryos among patients to euploid embryos within patients (Criterion 2B)

Morphokinetic Parameters	PB extr	PN abutt	Furrow	2 cells	Len Div	Cc2	3 cell	4 cell	3cell-4cell	5 cell	8 cell	Starts c0mp	Ends comp	Dur comp	Starts cav	blast
F (euploid) (among/within) ^a	0.91	0.61	2.11	0.91	0.79	1.01	0.52	2.16	0.88	1.15	2.49	1.95	1.37	1.91	2.92	3.43
d.f.	3/5	3/5	13/21	15/28	13/21	15/28	15/28	15/28	15/28	15/28	14/23	20/37	14/26	15/26	20/37	20/37
sig	NS	NS	NS	NS	NS	NS	NS	Sig.	NS	NS	Sig.	Sig.	NS	NS	Sig.	Sig.

^a F value for comparison of variability of morphokinetic parameters for euploid embryos (Mean squared deviation of euploid embryos among patients/mean squared deviation for euploid embryos within patients)

Table 6 summarizes the results of testing for criteria 1A, 1B, 2A and 2B. Note that none of the parameters fulfilled all of the criteria, Therefore we conclude that none of the parameters will function to discriminate euploid from aneuploid embryos (Table 6).

Discussion

When we applied the Campbell criteria to our morphokinetic data, there was no significant difference in the incidence of euploid (or aneuploid) embryos in any of the three groups (low, medium and high probability of aneuploidy).

Traditionally, numerous variables have been scrutinized in an attempt to select for embryos that will result in implantation and ongoing pregnancy. Classical embryo and blastocyst assessment, described by Gardner and Schoolcraft [16] has been the mainstay of embryo evaluation prior to endometrial transfer. Yet, qualification of embryos has suffered from inter/intra observer variation [17]. As such, time lapse imaging and morphokinetic assessment establishes standardization across clinical settings. In recent publications, morphogenetic time points have launched innovative prediction models in attempt to determine whether morphokinetic variables differ between aneuploid and euploid embryos. A model for ploidy classification while avoiding invasive selection of embryos by means of trophectoderm biopsy has been created [5, 6].

Classifying human preimplantation embryos according to their risk of aneuploidy, with the hopes of correlating risk models to clinical outcome [5, 6], was examined in our present study. When we applied the same criteria to our data, the parameters were unsuccessful in discriminating euploid from aneuploid embryos. This was consistent whether we used the time of insemination or the time of syngamy as the reference point from which events and criteria were timed. Nevertheless, we were able to establish one parameter (duration of compaction) that discriminated euploid from aneuploid embryos significantly, although poorly. It may be reasonable to believe that our criterion would not be capable of discriminating Campbell's euploid and aneuploid embryos. Any discriminator that enriches to only 64 % euploidy [5, 6] may be statistically significant; however, its clinical utility is in doubt when the probability of euploidy without selection is roughly 45 %. Therefore we feel compelled to explain how two different methods with significant ability to discriminate euploid from aneuploid embryos cannot be universally applied.

The failure of Campbell's criteria to discriminate euploid from aneuploid embryos in our laboratory suggests that Campbell's criteria are not universally applicable for use in selecting euploid embryos for transfer. Neither of the parameters in Campbell's model satisfied Criteria 2A (significant difference between aneuploid and euploid embryos among

Table 6 Overview of Satisfaction of Criteria for Discrimination^a

Criterion	PB extr	PN abutt	Furrow	2 cells	Len Div	Cc2	3 cell	4 cell	3cell-4cell	5 cell	8 cell	Starts comp	Ends comp	Dur comp	Starts cav	blast
1A ^b								Yes							Yes	
1B ^c			Yes					Yes					Yes	Yes		
2A ^d							Yes					Yes				
2B ^e	Yes	Yes	Yes	Yes	Yes	Yes	Yes		Yes	Yes			Yes	Yes		
Discrim? ^f	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No

^a This table summarizes the results of comparisons made in the preceding tables. Cells with “Yes” indicate criteria that were satisfied for discrimination of the parameter in the heading of that column. Cells that remain blank indicate criteria that were not satisfied for discrimination of the parameter in the heading of that column

^b Results of testing for criterion 1A in Table 3 (top sections)

^c Results of testing for criterion 1B in Table 3 (bottom sections)

^d Results of testing for criterion 2A in Table 4 (top and bottom sections)

^e Results of testing for criterion 2B in Table 5. Note that satisfaction of this criterion requires that the ANOVA not be significant.

^f Summarizes whether the parameters listed in the heading of each column is able to discriminate euploid from aneuploid embryos

patients) or 2B (uniformity of euploid embryos across all patients). Our observations that all morphokinetic parameters (besides duration of compaction) were very poor at discriminating euploid from aneuploid embryos provides further evidence supporting the notion that morphokinetic parameters are not capable of clearly distinguishing euploid from aneuploid embryos.

The observation that duration of compaction was the only parameter in our facility that was capable of providing enrichment for euploid embryos whereas different parameters were useful in Campbell's lab indicates that there may be differences, even slight, between our laboratories' culture methods or our patients' morphokinetics that contribute to this difference. However, our methods, dishes, media and TLM instrument (Embryoscope) were similar in most respects to these specific features in Campbell's laboratory (% CO₂ being the varying factor, 6 % CO₂ vs 5.5 % CO₂ used in Campbell's study). One distinct clinical difference between Campbell's patient treatment and ours was the different methods of insemination used, ICSI for all of Campbell's patients and standard insemination for most of our patients. We found that the use of syngamy as the reference point provided better consistency (smaller standard deviations) than the insemination time for morphokinetic parameters. We believe that the time of syngamy may be a better reference time point for future studies as well. Conversion of the morphokinetic parameters by normalization with the time of syngamy and similar adjustment to the criteria by subtracting the time of syngamy from the criterion presented by Campbell should have compensated for the differences in technique, and still the conclusions were similar: No discrimination of euploid from aneuploid embryos. The sustained difference in outcome despite the similarity of techniques suggests that the differences between our results are attributable to the patients. This

notion is supported statistically by the failure of this parameter to satisfy Criterion 2A (significant difference between aneuploid and euploid embryos among patients).

It has been pointed out in the literature that time ranges and sharp cutoffs may not correctly simulate embryo development [18]. Thus, the application of standard cutoffs may not be valid for all embryos and may be a significant source of error. In the case of ploidy, for a test to be universally valid in discriminating euploid embryos from aneuploid embryos, the testing parameters must vary more between euploid and aneuploid embryos than euploid embryos vary between patients. It may be possible to apply a test within a patient when euploid and aneuploid embryos are more variable than euploid embryos within a patient; however, application of the criteria requires that the values of euploid and aneuploid embryos within that patient must be known. Our observations that among patient variability was significant in the two way ANOVA whereas there was no significance attributable to ploidy indicates that the major source of variability in the data is patient to patient variability.

Examination of the variability of euploid embryos within patients to the variability of all embryos within patients or aneuploid embryos within patients yielded several candidate parameters that might have discriminated euploid embryos from aneuploid embryos. However, discrimination on the basis of within-patient variability means that we must establish the morphokinetic criteria for selection within each patient. Therefore, we would need to establish criteria with the knowledge of ploidy within each patient, requiring the performance of genetic screening within each patient. We have performed insufficient numbers of repeat patients to know whether our patient-to-patient variability is truly patient specific or is the same variability that we would see with cycle-to-cycle variability within the same patient. So, as it stands now,

genetic testing would be required for each cycle in order to establish criteria for discriminating ploidy of embryos. With the availability of the genetic screening data, morphokinetic predictions would be moot and doubtlessly, less accurate than the already available genetic screening data.

The observation that criterion 2A was not met for most morphokinetic parameters indicates that patient to patient variability for euploid embryos was too great or that patient to patient variability for aneuploid embryos was not great enough to permit discrimination. Similarly, the observation that criterion 2B was not satisfied (for most parameters) indicates that the within patient variability of euploid embryos was too small or the patient to patient variability for euploid embryos was too great for euploid embryos to be considered uniform enough that a universal criterion using single morphokinetic parameters could be used to discriminate them.

We conclude that among patient variability even for euploid embryos is so great that it will be impossible to use morphokinetic parameters to select euploid embryos using universal criteria. Further, we believe that within patient variability for good prognosis embryos will never be sufficiently small to provide discrimination of embryos according to ploidy without more information than is available at the time of morphokinetic analysis.

There remains the possibility that multiple parameters may be combined to create a multiparametric threshold or window. One might expect that multiple parameters would take advantage of the additivity of small enrichments using different criteria. However, ROC curves (Fig. 2) suggests that there is little if any enrichment available using the morphokinetic parameters investigated. The creation of a threshold via a multiparametric approach could yield a single predictor to which these same criteria could be applied to determine if the new predictor is capable of universal enrichment for euploid embryos. At this point in time, no such multiparametric threshold or window exists.

It is important to note that recent studies investigating morphokinetic parameters alone [3, 14] to select for embryos destined to become blastocysts included in their analysis embryos that arrested prior to the blastocyst stage. Their criteria predict the ability to form a blastocyst, but do not predict ploidy. Our current study only looked at embryos that became blastocysts and were amenable to biopsy, strikingly different embryo selection criteria. Further, expecting to see varied morphokinetic parameters due to aneuploidy may require the expression of the embryonic genome prior to any observable effects. Since expression of the embryonic genome is generally considered to occur around the time of 4–8 cells [19], it is not surprising that the morphokinetic parameters with the most promise of discriminating ploidy were events at or following this critical embryonic period.

Must euploid embryos have uniform morphokinetic measures? A traditional framework exists for overall predictive

timing of early embryo development based on cellular metabolic activity pathways and mitotic processes. Yet even among euploid embryos from the same patient, we see variability within the spectrum of accepted time frames. Aside from ploidy, which strictly looks at chromosomal number, underlying factors within gene expression driving an embryo through development are unfailingly ignored. Diversity in developmental rates has been attributed to expression of specific genes, *Ped* and its gene product *Qa-2* in the mouse [20] and *(HLA)-G* in the human [21] as well as other gene candidates, gene regulatory networks (GRN), transcripts and proteins [22]. Therefore, there may be ample genetic basis for the variability in morphokinetic parameters that may be ascribed to patients (patient to patient variability) and not just to ploidy.

A frequent cause of failed ART following embryo transfer is the unknowing transfer of aneuploid embryos [10]. We have shown that application of Campbell's criteria in our patient population does not predict those embryos that are chromosomally normal and may lead to the unknown transfer of aneuploid embryos.

Embryo profiling via preimplantation genetic screening (PGS), accurately analyzes all 24 chromosomes (autosomes 1–22, X, and Y) via comprehensive chromosomal screening (CCS), thereby screening embryos for chromosome abnormalities and allowing selection of euploid embryos for transfer. Estimates of the error rate with trophoctoderm biopsy and array CGH are less than 2 % [23]. With a false positive rate of 2 % and a False Negative rate of 2 %, the Area under the curve for an ROC curve would exceed 0.95. The AUC for Campbell's method in Campbell's patients (0.72) and for our use of duration of compaction in our patients (0.67) pale in comparison to the AUC for TE biopsy and aCGH (0.95). Therefore, we believe that the use of TE biopsy and aCGH is a remarkably better method of discriminating euploid and aneuploid embryos. Trophoctoderm (TE) biopsy particularly has allowed for an increased proportion of patients receiving single embryo transfer without compromising success rates [1, 24]. Morphokinetics used as a tool for embryo ploidy selection should provide prediction of euploid embryos with equal or greater certainty to TE biopsy; otherwise it is not useful as the standard of care. Further study should focus on determining how the additive knowledge of ploidy status and morphokinetic parameters will help to predict outcome rather than expecting morphokinetics to act as a substitute for PGS.

Summary

The evidence is insufficient to permit discrimination of embryo ploidy on the basis of morphokinetic values alone. The primary reason shown is that morphokinetic values varied more from patient to patient than they did between euploid and aneuploid embryos (2 way ANOVA). In addition,

differences between aneuploid and euploid embryos, while discernible within patients, were not large enough to be detected when the patient to patient variability was considered. Patients electing against PGS or clinics without access to PGS, will not increase birth outcomes utilizing currently available aneuploidy risk models described in the literature. Use of morphokinetic parameters alone to improve euploidy rates cannot provide results comparable to actual PGS. Ploidy of embryos assessed by screening of the trophectoderm to give full chromosome copy number is still the best predictor of an embryo with prognosis for a healthy outcome. Despite their poor ability to discriminate ploidy, morphokinetics may be useful as an adjunct to PGS in selecting those PGS-screened euploid embryos with the best chances of implantation and live birth; however, further study is needed.

References

- Grifo J, Hodes-Wertz B, Hsiao-Ling L, Amperloquio E, Clarke-Williams M, Adler A. Single thawed euploid embryo transfer improves IVF pregnancy, miscarriage and multiple gestation outcomes and has similar implantation rates as egg donation. *Journal of Assisted Reproduction and Genetics*. 2013. doi:10.1007/s10815-012-9929-1.
- Harton G, Munne S, Surrey M, Grifo J, Kaplan B, McCulloh D, Griffin D, Wells D. Diminished effect of maternal age on implantation after preimplantation genetic diagnosis with array comparative genomic hybridization.
- Conaghan J, Chen AA, Willman SP, Ivani K, Chenette PE, Boostanfar R, et al. Improving embryo selection using a computer-automated time-lapse image analysis test plus day 3 morphology: results from a prospective multicenter trial. *Fertility and Sterility*. 2013;100(2):412–9.e5.
- Chamayou S, Patrizio P, Storaci G, Tomaselli V, Alecci C, Ragolia C, et al. The use of morphokinetic parameters to select all embryos with full capacity to implant. *Journal of Assisted Reproduction and Genetics*. 2013;30:703–10.
- Campbell A, Fishel S, Bowman N, Duffy S, Sedler M, Hickman C. Modelling a risk classification of aneuploidy in human embryos using non-invasive morphokinetics. *Reproductive biomedicine online*. 2013;26:477–85.
- Campbell A, Fishel S, Bowman N, Duffy S, Sedler M, Thornton S. Retrospective analysis of outcomes after IVF using an aneuploidy risk model derived from time-lapse imaging without PGS. *Reproductive Biomedicine Online*. 2013;27:140–6.
- Rubio C, Buendía P, Rodrigo L, Mercader A, Mateu E, Peinado V, et al. Prognostic Factors for preimplantation genetic screening in repeated pregnancy loss. *Reproductive Biomedicine Online*. 2009;18(5):687–93.
- Huang CC, Chang LJ, Tsai YY, Hung CC, Fang MY, Su YN, Chen HF, Chen SU. A feasible strategy of preimplantation genetic diagnosis for carriers with chromosomal translocation: Using blastocyst biopsy and array comparative genomic hybridization.
- Gardner DK, Lane M. Culture and selection of viable blastocysts: a feasible proposition for human IVF? In: *Human reproduction update*. 3rd ed. 1997. p. 367–82.
- Gutierrez-Mateo C, Colls P, Sanchez-Garcia J, Escudero T, Prates R, Ketterson K, et al. Validation of microarray comparative genomic hybridization for comprehensive chromosome analysis of embryos. *Fertility and sterility*. 2011;95:953–8.
- Hodes-Wertz B, Grifo J, Ghadir S, Kaplan B, Laskin CA, Glassner M, et al. Idiopathic recurrent miscarriage is caused mostly by aneuploid embryos. *Fertility and Sterility*. 2012;98(3):675–80.
- Wong C, Loewke KE, Bossert NL, Behr B, De Jonge CJ, Baer TM, et al. Non-invasive imaging of human embryos before embryonic genome activation predicts development to the blastocyst stage. *Nature biotechnology*. 2010;28:1115–21.
- Kirkegaard K, Agerholm IE, Ingerslev HJ. Time-lapse monitoring as a tool for clinical embryo assessment. *Human reproduction*. 2012;27:1277–85.
- Meseguer M, Rubio I, Cruz M, Basile N, Marcos J, Requena A. Embryo incubation and selection in a time-lapse monitoring system improves pregnancy outcome compared with a standard incubator: a retrospective cohort study. *Fertility and sterility*. 2012;98(6):1481–9.
- Kirkegaard K, Hindkjaer J, Ingerslev HJ. Human embryonic development after blastomere removal: a time-lapse analysis. *Human reproduction*. 2012;27:97–105.
- Gardner DK, Schoolcraft WB. Human embryo viability: what determines developmental potential, and can it be assessed? *Journal of Assisted Reproduction and Genetics*. 1998;15(8):455–58.
- Patnot G, Devroe J, Debrock S, D'Hooghe TM, Spiessens C. Intra- and inter-observer analysis in the morphological assessment of early stage embryos. *Reproductive Biology and Endocrinology*. 2009;7:105.
- Herrero J, Meseguer M. Selection of high potential embryos using time-lapse imaging: the era of morphokinetics. *Fertility and Sterility*. 2013;99(4):1030–34.
- Braude P, Bolton V, Moore S. Human gene expression first occurs between the four- and eight-cell stages of preimplantation development. *Nature*. 1988;332:459–61.
- Levy R. Genetic regulation of preimplantation embryo survival. *International Review of Cytology*. 2001;210:1–37.
- Warner CM, Newmark JA, Comiskey M, De Fazio SR, O'Malley DM, Rajadhyaksha M, et al. Genetics and imaging to assess oocyte and preimplantation embryo health. *Reproduction Fertility and Development*. 2004;16(7):729–41.
- Cao W, Brenner CA, Alikani M, Cohen J, Warner CM. Search for a human homologue of the mouse *Ped* gene. *Molecular Human Reproduction*. 1999;5(6):541–47.
- Colls P, Escudero T, Fischer J, Cekleniak NA, Ben-Ozer S, Meyer B, et al. Validation of array comparative genome hybridization for diagnosis of translocations in preimplantation human embryos. *Reproductive BioMedicine Online*. 2012;24(6):621–9.
- Forman EJ, Hong KH, Ferry KM, Tao X, Taylor D, Levy B, et al. In vitro fertilization with single euploid blastocyst transfer: a randomized controlled trial. *Fertility and Sterility*. 2013;100(1):100–7.