Morphological and morphokinetic associations with aneuploidy

Bamford, Thomas; Barrie, Amy; Montgomery, Sue; Dhillon-Smith, Rima; Campbell, Alison; Easter, Christina; Coomarasamy, Arri

DOI:
10.1093/humupd/dmac022

Morphological and morphokinetic associations with aneuploidy: a systematic review and meta-analysis

Running title: Human embryo morphology, morphokinetics and ploidy

Authors: Thomas Bamford[a] MB. ChB, Amy Barrie[b] Ph.D, Sue Montgomery[a] Ph.D, Rima Dhillon-Smith Ph.D[c], Alison Campbell[d] Ph.D, Christina Easter[c] MSc, Arri Coomarasamy[c] M.D

a) CARE Fertility Manchester, United Kingdom, M14 5QH
b) CARE Fertility Chester, United Kingdom, CH2 1UL
c) Tommy's National Centre for Miscarriage Research, Institute of Metabolism and Systems Research, College of Medical and Dental Sciences, University of Birmingham, Edgbaston, United Kingdom
d) CARE Fertility, John Webster House, Nottingham, United Kingdom, NG8 6PZ

Corresponding Author: Dr Thomas Bamford, thomasbamford@nhs.net
ORCHID ID- 0000-0001-9048-3912
CARE Fertility Manchester,
109-112 Daisy Bank Rd
Manchester
M14 5QH

TABLE OF CONTENTS

ABSTRACT

BACKGROUND

A time lapse system (TLS) is utilised in some fertility clinics with the aim of predicting embryo viability and chance of live birth during IVF. It has been hypothesised that aneuploid embryos display altered morphokinetics as a consequence of their abnormal chromosome complement. Since aneuploidy is one of the fundamental reasons for IVF failure and miscarriage, attention has focused on utilising morphokinetics to develop models to non-invasively risk stratify embryos for ploidy status. This could avoid or reduce the costs associated with pre-implantation genetic testing for aneuploidy (PGT-A). Furthermore, TLS have provided an understanding of the true prevalence of other dysmorphisms. Hypothetically, the incorporation of morphological features into a model could act synergistically, improving a model's discriminative ability to predict ploidy status.

OBJECTIVE AND RATIONALE

The aim of this systematic review and meta-analysis was to investigate associations between ploidy status and morphokinetic or morphological features commonly denoted on a TLS. This

will determine the feasibility of a prediction model for euploidy and summarise the most useful prognostic markers to be included in model development.

SEARCH METHODS

Five separate searches were conducted in Medline, Embase, PubMed, and Cinahl from inception to 1st July 2021. Search terms and word variants included, among others, PGT-A, ploidy, morphokinetics and time lapse, and the latter were successively substituted for the following morphological parameters: fragmentation, multinucleation, abnormal cleavage and contraction. Studies were limited to human studies.

OUTCOMES

Overall, 58 studies were included incorporating over 40,000 embryos. All except one study had a moderate risk of bias in at least one domain when assessed by the quality in prognostic studies tool (QUIPS). Ten morphokinetic variables were significantly delayed in aneuploid embryos. When excluding studies using less reliable genetic technologies, the most notable variables were: time to 8 cells (t8, 1.13 hrs, 95% CI 0.21-2.05; three studies; n=742; I_2= 0%), t9 (2.27 hrs, 95% CI 0.5-4.03; two studies; n=671; I^2=33%), time to formation of a full blastocyst (tB, 1.99 hrs, 95% CI 0.15-3.81; four studies; n=1640; I^2=76%), and time to expanded blastocyst (tEB, 2.35 hrs, 95%CI 0.06-4.63; four studies; n=1640; I^2=83%). There is potentially some prognostic potential in the degree of fragmentation, multinucleation persisting to the 4-cell stage, and frequency of embryo contractions. Reverse cleavage was associated with euploidy in this meta-analysis, however this article argues that these are likely spurious results requiring further investigation. There was no association with direct unequal cleavage in an embryo that progressed to a blastocyst, or with multinucleation assessed on day 2 or at the 2-cell stage. However, owing to heterogenous results and poor-quality evidence,

associations between these morphological components needs to be investigated further before conclusions can be reliably drawn.

WIDER IMPLICATIONS

This first systematic review and meta-analysis of morphological and morphokinetic associations with ploidy status demonstrates the most useful morphokinetic variables, namely t8, t9 and tEB to be included in future model development. There is considerable variability within aneuploid and euploid embryos making definitively classifying them impossible, however, it is feasible that embryos could be prioritised for biopsy. Furthermore, these results support the mechanism by which algorithms for live birth may have predictive ability, suggesting aneuploidy causes delayed cytokinesis. We highlight significant heterogeneity in our results secondary to local conditions and diverse patient populations, therefore calling for future models to be robustly developed and tested in-house. If successful, such a model would constitute a meaningful breakthrough when accessing PGT-A is unsuitable for couples.

Key words: time-lapse, morphokinetics, ploidy, model, fragmentation, multinucleation, abnormal cleavage, contraction

Introduction

Pre-implantation embryo selection has historically relied upon morphological assessment using increasingly contested consensus guidelines (Gardner and Balaban 2016; Kemper et al. 2021; Gardner D. K. 1999; Alpha Scientists in Reproductive and Embryology 2011). Despite significant improvements since the inception of assisted reproduction, the average live birth rate in the UK remains low, at 32% per embryo transfer (for women < 35) (HFEA 2021). When one also considers the drive for single embryo transfers, advancing maternal age and higher associated aneuploid rates, the need for more advanced methods for assessing embryo viability is paramount.

A time lapse system (TLS) offers several advantages over static, basic morphological observations. This enclosed incubation system reduces the need to remove embryos from optimum atmospheric culture conditions by taking microscopic, multiplanar images at regular intervals. The retrospective analysis of these images allows the annotations of an embryo's developmental milestones (i.e., morphokinetics) to be compared to outcome variables, such as live birth or ploidy status. This allows embryos to be selected that display specific development patterns achieved at fixed times of development, for example blastocyst formation at 116hrs; usually recorded as hours post insemination (hpi). Unfortunately, due to poor quality evidence a Cochrane review was unable to conclude whether the use of a TLS increased live birth rates (Armstrong et al. 2019). In contrast, several large studies and randomised trials have reported improvements, therefore a TLS has become commonplace in many IVF laboratories worldwide (Pribenszky et al. 2017). A summary of definitions used for morphokinetic annotations and other morphological features denoted on a TLS can be found in Table I.

Aneuploidy is a major cause of implantation failure and miscarriage, however, there are barriers to accessing genetic testing. Aneuploidy arises from errors during mitosis or meiosis, such as non-disjunction. This increases with maternal age and therefore coincides with rapidly

declining success rates of IVF treatment in older women. For instance, in women under 35 years an average aneuploidy rate of 30-50% has been reported, increasing to 80% in women aged 42 years or older (Ata et al. 2012; Franasiak et al. 2014). Modern methods for pre-implantation genetic testing for aneuploidy (PGT-A) provide an accurate assessment of embryo chromosome complement using biopsy techniques in the majority of cases (Munné et al. 2017; Munné et al. 2019). For some patients, however, this technology may be inaccessible because it is prohibited by legislation, or they may deem it ethically inappropriate. They may also not have embryos suitable for biopsy. Moreover, PGT-A can cost over £3000 in the UK and in the USA it can be as high as $12,000, further limiting accessibility (Theobald et al. 2020). It is therefore not surprising that researchers have begun investigating methods to non-invasively detect aneuploidy.

It has been hypothesised that the morphokinetics of aneuploid embryos are delayed in comparison to euploid counterparts (Davies 2012; Campbell et al. 2013). Physiologically, this may be due to complex biochemical processes that occur when errors have been detected by the developing embryo (Coticchio, Barrie, et al. 2021). This results in slower cell division and is possibly a reason for the higher mitochondrial content seen in aneuploid embryos (Campbell et al. 2013; Ho et al. 2018). This has led to the development of several models using PGT-A and morphokinetic data aiming to risk-stratify embryos for euploidy (Campbell et al. 2013; Basile et al. 2014; Chawla et al. 2015; Mumusoglu et al. 2017; Del Carmen Nogales et al. 2017; Desai et al. 2018). At CARE Fertility a sophisticated time-lapse embryo selection model, "CAREmaps®", has been successfully developed that can predict an individual embryo's chance of resulting in a live birth. This was developed using a database of over 6000 transferred blastocysts with known live birth outcome data and has been shown to improve embryo selection (Fishel et al. 2018). Similar embryo selection algorithms have been developed by a variety of clinics internationally; some are commercially available (Petersen et al. 2016). It

remains unknown why embryos with higher scores should have better predicted outcomes; it would be sensible to hypothesise that the aetiology lies within delayed development as a sequela of chromosomal abnormalities. It would therefore prompt the assumption that if a TLS can identify embryos with the highest chance of live birth, it could be instrumented to enhance euploid embryo selection. This hypothesis is also supported by a recent meta-analysis that showed that the use of a TLS was associated with lower early miscarriage rates compared to traditional morphological assessment (Pribenszky et al. 2017). Other theories have also been suggested including partial compaction with or without cell extrusion or exclusion causing delayed cyto or karyokinesis, abnormal fertilisation, BMI, embryo sex, a failure of the embryo to undergo check points, and DNA repair mechanisms (Coticchio, Ezoe, et al. 2021; Coticchio, Barrie, et al. 2021; Bronet et al. 2015; Leary et al. 2015). It may therefore be feasible to utilise morphokinetics as a screening tool for ploidy status if this hypothesis becomes established by evidence.

The ability of morphokinetic models to predict ploidy status remains controversial and wide disparities exist in the morphokinetic events included in such models (Campbell et al. 2013; Kramer et al. 2014; Basile et al. 2014). This may be due to significant heterogeneity in study design and sample populations. For instance, the following have all been associated with altered morphokinetics: age, smoking status, biopsy techniques, stimulation protocols, insemination methods and culture conditions (Ciray et al. 2012; Muñoz et al. 2013; Lemmen et al. 2008; Bellver et al. 2013; Fréour et al. 2013; Kirkegaard et al. 2013).

Several morphological observations can be observed in greater detail when using a TLS, although historically there is limited correlation reported between ploidy status and these qualitative aspects (Magli et al. 2007; Minasi et al. 2016; Capalbo et al. 2014; Munné et al. 2017). In fact, several authors have identified that it is possible for aneuploid embryos to achieve good morphology scores (Munné 2006; Alfarawati et al. 2011; Fragouli et al. 2014).

Nonetheless, it must be taken into consideration that most studies investigating associations between morphology and ploidy status were undertaken using standard morphology assessments and not using a TLS. This results in an inability to identify dynamic changes occurring between check points. Furthermore, many of these studies utilised older, less reliable techniques such as fluorescence *in situ* hybridisation (FISH) and blastomere biopsy. This results in a higher chance of misclassifying mosaics or failing to detect aneuploidy due to the limited number of probes used (Fragouli and Wells 2011). We aim to investigate the association of various morphological components commonly observed on a TLS with ploidy status.

The first variable to be explored is fragmentation. Fragmentation is often considered during embryo selection owing to associations with embryo viability, but it remains one of the most enigmatic features identified in early development (Edwards et al. 1984; Puissant et al. 1987). Origins of these anucleated structures have been correlated with many factors including culture conditions, poor quality oocytes or spermatozoon, increased maternal age, oxidative stress and aneuploidy (Kim et al. 2018; Delimitreva et al. 2005; Fujimoto et al. 2011; Munne and Cohen 1998; Magli et al. 2007). It has even been associated with so called 'self-correction' mechanisms whereby an embryo extrudes sequestered chromosomes in order to become more genetically normal (Coticchio, Barrie, et al. 2021). Considering that the causation is poorly understood, association with ploidy status will be explored further in this review.

The second factor to be investigated is abnormal cleavage, the occurrence of which has become more apparent through a TLS yet causality remains unproven (Zhan et al. 2016; Athayde Wirka et al. 2014). The prevalence of these atypical cell divisions ranges from 4.4 to 26.1% and the implantation rates of these untested embryos has been found to be as low as 1.2%-17% (Barrie et al. 2017; Rubio et al. 2012; Ozbek et al. 2021). There is, therefore, a tendency to deselect these embryos (Desai et al. 2018; Zhan et al. 2016; Balakier et al. 2016; Hashimoto et al. 2016). Previous theories for aetiology include multipolar spindles, surplus centrosomes, quality of

spermatozoa and chromosome aberrations (Kalatova et al. 2015; Ozbek et al. 2021). Similarly, it has been speculated that abnormal cleavage may also be involved in the process of 'self-correction'. This is supported by the recent findings that abnormal cleavages are associated with partial compaction and the 'excluded phenotype' (Coticchio, Ezoe, et al. 2021). These excluded cells have also been shown to have a significantly higher abnormal chromosome content (Lagalla et al. 2017). We will assess the association between the most common types of abnormal cleavage and ploidy status: direct and reverse cleavage (Liu et al. 2014; Rubio et al. 2012).

Blastocyst contraction is the third feature to be examined that has been the focus of only a handful of studies. Physiologically this occurs through the inflow of liquid through aquaporin water channels and outflow through weak tight junctions (Watson et al. 2004; Marcos et al. 2015). The reason for it remains largely unknown, and it has been suggested that this process may assist in embryo hatching and has been associated with lower implantation rates (Marcos et al. 2015; Bodri et al. 2016; Niimura 2003). Hypothetically, this may be secondary to aneuploidy, therefore this will be investigated in this review.

Finally, multinucleation has been associated with poorer implantation outcomes and possibly aneuploidy (Kligman et al. 1996; Royen et al. 2003). This dysmorphism has been hypothesised to be the result of errors in nuclear replication without cytokinesis, nuclear fragmentation or defective DNA packaging and migration during anaphase (Pickering et al. 1995). It is therefore possible that this could be linked to aneuploidy as a consequence of errors occurring in chromosome segregation.

The aim of this systematic review and meta-analysis is to determine the most reliable morphokinetic prognostic factors for future model development and investigate associations between morphology and ploidy status. Specifically, the degree of fragmentation, presence of direct and reverse cleavage, blastocyst contractions and multinucleation will be investigated in

association with chromosomal status. Incorporating these morphological parameters may improve the discrimination of a morphokinetic model with regards to ploidy.

Methods

Registration

This review was prospectively registered with PROSPERO (ID number: CRD42021260795).

Data sources and search strategy

Five separate literature searches were conducted for potential prognostic factors and their associations with aneuploidy in concordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) guidelines (Moher et al. 2009). Electronic searches were conducted in MEDLINE, PubMed, EMBASE and CINAHL (from inception to 1st July 2021). Searches were conducted using the following MeSH key terms and word variants: 'pre-implantation genetic testing for aneuploidy (PGT-A)', OR 'pre-implantation genetic screening (PGS)', OR 'ploidy', OR 'aneuploid' AND 'morphokinetics', OR 'time-lapse'. For the four subsequent searches 'morphokinetics' and 'time-lapse' were successively substituted for: 'fragmentation', 'multinucleation', 'abnormal cleavage' and 'blastocyst contraction'. Similarly, word variants for each were included, such as 'trichotomous mitosis' for 'direct cleavage'.

Eligibility criteria

Studies were limited to human studies and included if the primary or secondary outcome was the ploidy status of biopsied embryos in relation to the presence of any of the prognostic factors under investigation. No language restrictions were applied. Manuscripts on mosaicism were included if they also provided data on aneuploid and euploid embryos. Exclusions include: polar body biopsy, those reporting clinical outcomes only, where the outcome was

translocations not aneuploidy, those that focussed on a subset of embryos with a particular morphological feature (such as abnormal cleavage in multinucleated embryos) or from a subgroup of patients (for example, endometriosis). The blastocyst contraction literature search aimed to determine association of embryo contraction kinetics (number or frequency), therefore studies investigating the rate or volume of expansion in relation to ploidy were excluded. Similarly, authors that correlated blastocyst expansion grading or morphology scores with ploidy status but not the individual prognostic factors being tested were excluded. Validation studies for a prognostic model already developed were excluded from the meta-analysis.

Study selection

Two reviewers initially screened all titles and abstracts independently for eligibility (T.B. and A.B.), and full length articles were then obtained and scrutinised. Any disputes were resolved by discussion with a third reviewer (S.M). Bibliographies of all relevant articles and review articles excluded were manually searched. Where more than 10 original articles met eligibility criteria, conference abstracts were subsequently excluded from the search. Otherwise, they were included due to a scarcity of published peer reviewed reports. Authors of all conference abstracts were contacted for additional information to assist with study selection, data extraction and quality assessments. Authors of original articles were contacted for further information where data presented was suboptimal. If data was not obtained or in a usable format, it was excluded from the meta-analysis but included in the systematic review.

Data extraction and study outcomes

Outcome and prognostic factor data were extracted independently by two reviewers into tables (T.B. and A.B.). The primary outcome extracted was the prevalence of aneuploid and euploid embryos for each potential prognostic factor assessed. This included the mean or median time

taken for both aneuploid and euploid embryos to achieve each morphokinetic variable. Data was also collected for: overall aneuploid rate, study design, primary outcome measured, number of patients and embryos included, TLS assessment period, PGT-A technique (including stage and type of biopsy), atmospheric culture conditions, infertility diagnosis and indication for PGT-A. Additionally, details of any model development, including attempts at model discrimination, calibration and validation, were recorded. Importantly, we collected data on potential study participant factors that could act as confounders including age, BMI, and stimulation drugs used. A recently published article by Barrie et al. (2021) described how age and BMI are the most important factors to control for in morphokinetic studies. Data were extracted only on those embryos with PGT-A results available.

Risk of bias and quality assessment

All articles meeting the selection criteria were quality assessed using the Quality in Prognosis Studies tool (QUIPS) (Grooten et al. 2019). It moves away from quantitative analysis of quality but rates the risk of bias in six domains (participation, attrition, prognostic factor measurement, outcome measurement, study confounding, and statistical analysis) as low, medium or high risk of bias (Higgins JPT 2021). The tool has been modified for use in this systematic review: an example can be found in Supplementary Table SI, including a summary of the bias domains and the criteria used to grade each category. Several items were removed from our adapted version of the tool. Firstly, the 'adequate participation', 'drop out' and 'attempt to collect information on participants who dropped out' prompting items were removed because they were less relevant to the study of embryos as research focuses on the retrospective analysis of existing PGT-A data sets. The original tool included a prompter within the confounding domain asking the reviewer to determine if the method used to measure confounding was reliable. This was removed because confounders for morphokinetics include readily available demographic data and standardised dosages.

There were a number of important factors to consider when undertaking the quality assessment. Firstly, if a particular study did not report on the proportion of embryos without PGT-A results, they were categorised as having a moderate risk of bias and if it was >5% they were deemed to have a high risk of bias. Using modern methods most genetic companies would now estimate that this occurs in up to approximately 2% of samples and this has been replicated in recent studies (Fiorentino et al. 2014; Neal et al. 2019; Tiegs et al. 2021). A significantly high proportion may lead to uncertainty regarding the validity of a study's conclusions and biopsy techniques. Secondly, if FISH was one of the genetic platforms used by a study it was considered a high risk of bias owing to the inaccuracies of this technique. Finally, due to the risk of inter-observer variability in morphological assessments of embryos, if there were no methods to account for internal validity then a publication was assessed as a moderate risk of bias. Similarly, if multinucleation was assessed as part of standard morphology assessment rather than the continuous observations enabled by a TLS, it was graded as a moderate risk of bias.

As per the Grading of Recommendations, Assessment, Development and Evaluation (GRADE) guidelines, publication bias was not assessed as less than 10 studies were included for each prognostic factor analysed, rendering the interpretation of funnel plots unreliable (Schünemann, 2013). The quality of reporting was assessed using the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) checklist according to published criteria (von Elm et al. 2007).

Data synthesis and analysis

Morphokinetics of aneuploid and euploid embryos were compared using a weighted mean difference analysis in concordance with the Meta-Analysis of Observational Studies in Epidemiology recommendations (Stroup, 2000). Where studies did not provide a SD value it was calculated from the 95% CI if the sample size was >100 (using methods recommended by

the Cochrane Handbook) (Higgins et al. 2021). Alternatively, data from studies describing only medians were summarised graphically by prognostic factor analysis using the interquartile range as a measure of dispersion.

The relationship between the prevalence of aneuploid embryos and percentage fragmentation has been presented on a line graph. Abnormal cleavage, embryo contraction and multinucleation data was dichotomised and meta-analysed with forest plots and corresponding calculated relative risks (RR). The results for abnormal cleavage were pooled and analyses were conducted for each type of abnormal cleavage to determine their relative contribution.

For all meta-analysed variables, heterogeneity was assessed using the I^2 statistic, whereby a result >50% was indicative of considerable heterogeneity. All analyses were undertaken using a random effects model by the Mantel-Haenszel method (DerSimonian and Laird 1986; Mantel and Haenszel 1959) using Review Manager (RevMan), version 5.4, The Cochrane Collaboration, (2020)

Sensitivity analysis

Sensitivity analyses were conducted based on the quality assessment derived from QUIPS. The use of random versus fixed effect models were also compared. Analyses will be restricted by excluding studies with a high risk of bias in any domain, an approach also taken by other authors using QUIPS (Taylor-Rowan et al. 2021).

Subgroup analyses

A subgroup analysis was performed by excluding studies using FISH, blastomere or day 3 biopsy to produce a more reliable effect estimate. During the last decade trophectoderm biopsy at the blastocyst stage has become the preferred method of testing; fewer embryos have mosaic results and there is less risk of damage and diminishing the live birth rates (Los et al. 2004; Staessen et al. 2004; Cohen et al. 2007; Goossens et al. 2008; Tarín et al. 1992). FISH also has several limitations, primarily, the impossibility to screen all chromosomes and the risk of

misdiagnosis is significant when multiple probes are used (DeUgarte et al. 2008; Ruangvutilert et al. 2000; Scriven and Bossuyt, 2010). These studies were not excluded as part of the screening process as they may still provide valuable prognostic information. None of the studies included in the prognostic factor graphs used FISH or blastomere biopsy.

Results

Search results

A total of 1557 studies were identified from the initial searches; 137 duplicates were removed, 1267 abstracts were screened, of which 123 were selected as being potentially relevant and 58 met selection criteria after screening of the full texts. Overall, 65 studies were excluded for the following main reasons: a different prognostic factor being investigated (n=15), describing an alternative outcome to ploidy status (n=11), measuring clinical outcomes only (n=10), or the study population was a subset of embryos with a particular characteristic (n=7). Figures 1-5 display the study screening process for each search and all exclusions are summarised in Supplementary Table SII (Moher et al. 2009). In total, 26 authors were contacted to attempt to identify missing information from their publication (Supplementary Table SIII). Two studies are 'awaiting classification' due to unanswered correspondence with the authors: this was intended to confirm whether their abstracts included different embryos than the later published articles (Desai, 2016; Lagalla, 2015). Responses were not received, therefore the publications with the most data were included in this systematic review (Lagalla et al. 2017; Desai et al. 2018). Responses were not received from 19 authors in total; this did not result in exclusion from the systematic review for any of these studies but exclusion from the quantitative analysis in 10. The remaining were contacted for supporting information only. In total, 58 studies were included in the narrative synthesis, 43 of which had results suitable for meta-analysis.

Characteristics of the included studies

Overall, 7,004 embryos that underwent PGT-A from at least 1,058 patients were included from 18 studies examining morphokinetics. Thirteen retrospective studies and one prospective cohort study provided morphokinetic data comparing euploid and aneuploid embryos, enabling their inclusion in the quantitative analysis (Chavez et al. 2012; Campbell et al. 2013; Basile et al. 2014; Chawla et al. 2015; Rienzi et al. 2015; Patel et al. 2016; Minasi et al. 2016; Mumusoglu et al. 2017; Zhang et al. 2017; Desai et al. 2018; Lee et al. 2019; Kimelman et al. 2019; Martin et al. 2021; Yang et al. 2014). Three studies were excluded from meta-analysis because two were validation studies (Kramer et al. 2014; Campbell et al, 2013) and one study presented data in an unusable format (Del Carmen Nogales et al. 2017). The included studies for morphokinetics were from eight different countries in total (USA, UK, Spain, Italy, India, Turkey, Taiwan and China). A summary of the characteristics of the included studies are reported in Tables II-VI.

Subsequently in this review we have considered morphological associations of aneuploidy. Firstly, fragmentation was assessed using data from 10,008 embryos from 1,842 patients, extracted from five studies (Magli et al. 2001; Ziebe et al. 2003; Delimitreva et al. 2005; Magli et al. 2007; Minasi et al. 2016). The remaining studies provided no raw data for interpretation or when provided it was in an unusable format (Moayeri et al. 2008; Chavez et al. 2012; Vera-Rodriguez et al. 2015). All were retrospective cohort studies apart from one publication which was a consecutive case series (Minasi et al. 2016).

Secondly, abnormal cleavage was assessed in relation to ploidy status, and this included 4,788 embryos from 1,100 patients from 10 retrospective cohort studies (Campbell et al. 2013; Rienzi et al. 2013; Vera-Rodriguez et al, 2015; Zhan et al. 2016; Lagalla et al. 2017; Zhang et al. 2017; Desai et al. 2018; Ho et al. 2018; McCoy et al. 2018; Ozbek et al. 2021). One study could not be included in the meta-analysis due to the limited provision of data (Davies 2012).

Thirdly, two cohort studies were meta-analysed to describe the relationship between embryo contraction and chromosome aberrations using data from 1,647 embryos from 460 patients (Vinals Gonzalez et al. 2018; Gazzo et al. 2020).

Finally, the presence of multinucleation was assessed in 18,676 embryos from 1,227 patients. Thirteen studies were included in the meta-analysis (Kligman et al. 1996; Magli et al. 2001; Agerholm et al. 2008; Ambroggio et al. 2011; Campbell et al. 2013; Mazur 2013; Munoz, 2014; Bayram, 2015; Balakier et al. 2016; Zhang et al. 2017; Hashimoto et al. 2016; Desai et al. 2018; Lee et al. 2019) and seven studies were included only as part of the narrative review (Scott, 2010; Davies, 2012; Melzer et al. 2013; Yilmaz et al. 2014; Li, 2015; Goodman, 2016; Del Carmen Nogales et al. 2017). Eleven are cohort studies and the remaining nine are conference abstracts (Davies, 2012; Mazur, 2013; Melzer et al. 2013; Munoz, 2014; Bayram, 2015; Li, 2015; Goodman, 2016; Del Carmen Nogales et al. 2017). The included manuscripts considering morphological prognostic factors were published from a broad range of countries (Tables II-VI).

Risk of bias and quality assessment results

Overall, the quality assessment of the eligible studies demonstrated a moderate risk of bias, whereby all but one study was scored with a moderate risk of bias in at least one domain. In total, only 17 out of 58 studies (29%) appropriately addressed confounding. Similarly, few authors adequately described participant characteristics or the selection criteria used (n=18/58, 31%). However, there was a low risk of bias for 'prognostic factor measurement' in most studies (n=45/58, 78%). The remaining studies had a moderate risk of bias within this category due to: unclear definitions of the prognostic factors (n=3), a lack of internal validation for the assessment of the morphological components (n=5), the use of standard morphology assessment at specific time points rather than the use of a TLS (n=4), or multiple methods used for prognostic factor measurement on the same cohort (n=1). Twenty-three studies were

considered a moderate risk of bias because they did not disclose the proportion of embryos with PGT-A results unavailable, and five studies had a high risk of bias since this was >5%. Furthermore, 11 studies were categorised as high risk of bias for the use of FISH and one study for 'statistics and reporting' as we consider their conclusions and results to be erroneous given the data presented (Davies, 2012). Finally, 17 studies within the 'statistics and reporting' domain were graded as a moderate risk of bias for: limited presentation of analytical strategy or data (n=10), poor modelling techniques and validation methods (n=5), and inappropriate statistical techniques (n=2). The results obtained using the QUIPS tool to determine the risk of bias are summarised in Supplementary Table SIV. The quality of reporting was assessed using STROBE (Supplementary Fig. S1).

Morphokinetics and ploidy

The following morphokinetic variables (Table I) were significantly delayed in aneuploid embryos: tPB2, t2, t4, t6, t7, t8, t9, tB, tEB, tHB. In contrast, tPNf, tM, tSB, cc3, S2, S3, and t_5-t_2 had no prognostic ability (Supplementary Fig. S2). Interestingly, euploid embryos were significantly delayed for cc2; however, this finding, t6 and tHB were no longer statistically significant when studies using FISH and/or blastomere biopsy were excluded. Additionally, t3, t5, and tPNf demonstrated significant differences exclusively in the subgroup analysis. The variables tPNf, t2, t3, t4, t5 were all delayed by up to 1 hour in aneuploid embryos. The variables tPB2 and t7 were delayed by > 1hour in aneuploid embryos, however, these results come from a subgroup analysis including only one study (1.3 hrs, 95% CI 0.88-1.72 and 1.8 hrs, 95% CI 0.34-3.26, respectively). The following variables were the most delayed in aneuploid embryos: t8 (1.13 hrs, 95% CI 0.21-2.05; three studies; n=742; I_2= 0%), t9 (2.27 hrs, 95% CI 0.5-4.03; two studies; n=671; I^2=33%), tB (1.99 hrs, 95% CI 0.15-3.81; four studies; n=1640; I^2=76%), and tEB (2.35 hrs, 95%CI 0.06-4.63; four studies; n=1640; I^2=83%) (Fig. 6).

On visual inspection, these results were concordant with the prognostic factor graphs, apart from t8. Similarly, tSC, which was analysed solely by a prognostic factor graph, resulted in inconsistent differences. The only study excluded from the meta-analysis that was not a validation study analysed morphokinetics per chromosomal abnormality and found that complex embryos had shorter cleavage times (Del Carmen Nogales et al. 2017). Finally, the sensitivity analyses did not change our conclusions with the exception of tSB that became significant using a fixed effects model (Supplementary Figs S3 and S4).

Fragmentation and ploidy

Fragmentation was associated with aneuploidy in six out of the eight included studies. The three most recent studies had the lowest risk of bias; two found no association and one found that a higher degree fragmentation was associated with aneuploidy (Minasi et al. 2016; Vera-Rodriguez et al. 2015; Chavez et al. 2012). Only four authors in total provided raw data that could be extracted into a line graph displaying a general trend of increasing prevalence of aneuploid embryos for increasing degrees of fragmentation (Supplementary Fig. S5).

Abnormal cleavage and ploidy

Pooled direct uneven cleavage, DUC1 and DUC2 (Table I), had no association with chromosomal normality (RR 1.09, 95%CI 0.83-1.44; RR 1.26, 0.98-1.61; RR 0.74 95% CI 0.26-2.1, respectively) (Supplementary Fig. S6). In contrast, reverse cleavage appears to provide some prognostic information specifically for euploidy (RR 1.36, 95% CI 1.14-1.63; five studies; n=3053; I_2=22%) (Fig. 7). There was a trend for more aneuploid embryos displaying DUC1 when studies using FISH and blastomere biopsy were excluded, however, this was not statistically significant (RR1.26, 95% CI 0.98-1.61; five studies, n=1917; I_2=27%) (Supplementary Fig. S6). Only one study was not included in the meta-analysis owing to limited provision of data, and this concluded that embryos exhibiting DUC1 were more likely

be aneuploid (57%, n=21) versus euploid (30%, n=44), p=0.01) (Davies 2012). Our findings were unchanged in a sensitivity analysis excluding studies with the highest risk of bias (Supplementary Fig. S7). A sensitivity analysis using a fixed effect model resulted in DUC2 being significantly more prevalent in aneuploid embryos relative to euploid (Supplementary Fig. S8).

Contractions and ploidy

Two studies examined the association between the presence of contractions and ploidy status and found that this observation was significantly more likely to occur in aneuploid embryos (RR 0.67, 95% CI 0.48-0.96; two studies, n=1,626, I_2=84%) (Fig. 8). These findings remained consistent in the sensitivity analysis (Supplementary Fig. S7).

Multinucleation and ploidy

No association with ploidy was found for embryos assessed on day 2 or at the 2-cell stage for multinucleation (RR 0.69, 95% CI 0.29-1.63, four studies, n=3650, I_2=0%; RR 0.82 95% CI 0.64-1.04, seven studies, n=2418, I_2=47%, respectively), however there may be prognostic potential in multinucleation persisting to the 4-cell stage (RR 0.52, 95% CI 0.29-0.91; six studies, n=1703, I_2=82%) (Supplementary Fig. S9 and Fig, 9, respectively). This remains uncertain since the subgroup analysis was insignificant, albeit trending towards an increased prevalence in aneuploid embryos (RR 0.56, 95% CI 0.28-1.14; four studies, n=1106, I_2=88%). Furthermore, 4-cell multinucleation was significantly associated with ploidy using a fixed rather than random effects model (Supplementary Fig. S8). Multinucleation on day 2 and at the 2-cell stage also had conflicting results in this sensitivity analysis; both were associated with aneuploidy but they remained insignificant in the subgroup analysis. Of the seven studies not included in the meta-analysis, two demonstrated association with ploidy when multinucleation was assessed during standard morphology assessments and one at the 4-cell stage (Melzer et

al. 2013; Yilmaz et al. 2014; Scott, 2010). In contrast**,** three studies reported no association with multinucleation when examined during daily morphology assessments (Davies, 2012; Goodman, 2016; Del Carmen Nogales et al. 2017) or at the 2-cell stage (Li, 2015). The findings of the main analysis were unchanged when excluding studies with the highest risk of bias (Supplementary Fig. S7).

Discussion

Key findings

Our study has found that aneuploid embryos are, on average, delayed by ≥ 1 hour in t8 and ≥ 2 hours in the morphokinetic variables t9 and tEB. Overall, in the weighted mean difference analysis, seven morphokinetic variables were significantly delayed in aneuploid embryos (tPB2, t2, t4, t7, t8, t9, tEB). Blastocysts displaying contractions are associated with aneuploidy and reverse cleavages are more prevalent in euploid embryos, although these results should be interpreted with caution and investigated further before any conclusions can be drawn. In addition, although not statistically significant, there is a trend towards aneuploid embryos displaying multinucleation persisting to the 4-cell stage. The trend between increasing percentage fragmentation and aneuploidy needs confirming in future studies owing to very low-quality evidence.

Morphokinetics and ploidy

Since the development of the Campbell model there has been a plethora of attempts to test and create models for ploidy status, each with significant limitations (Campbell et al. 2013). This original model has been tested by several authors; only Desai et al. (2018) was able to reliably risk stratify for aneuploidy (Kramer et al. 2014; Rienzi et al. 2015; Zhang et al. 2017; Desai et al. 2018). This may be because the morphokinetics of embryos are so sensitive to laboratory conditions that models may not be translatable between clinics or patient populations. Indeed,

this variability may also account for why some models incorporate early cleavage parameters (Chavez et al. 2012; Chawla et al. 2015; Patel et al. 2016; Del Carmen Nogales et al. 2017) and some late, blastulation variables (Campbell et al. 2013; Kramer et al. 2014; Desai et al. 2018; Lee et al. 2019; Martin et al. 2021). There are several common limitations to the published models, including the lack of control of confounders and the use of apparent validation by some authors, leading to model overestimation (Basile et al. 2014; Chawla et al. 2015; Del Carmen Nogales et al. 2017; Desai et al. 2018). In fact, confounding variables were overlooked in over 70% of the included studies (Supplementary Fig. S1). Four articles attempted to adjust for age, finding no association between age and morphokinetics (Rienzi et al. 2015; Mumusoglu et al. 2017; Desai et al. 2018; Martin et al. 2021). Conversely, BMI was found to be associated with delayed morphokinetics, while yet another such study demonstrated no such association (Martin et al. 2021; Mumusoglu et al. 2017). The effects of stimulation dosages were only assessed by three authors; two concluded there were no dose dependant differences, whereas one reported higher dosages were associated in delayed development kinetics (Campbell et al. 2013; Martin et al. 2021; Mumusoglu et al. 2017).

In comparison to t8, t9 and tEB, the variables tPB2, t2, t4, and t7 were less dramatically delayed in aneuploid embryos, therefore in the context of such wide CIs are less likely to reliably predict ploidy status. Of the most delayed variables, t8 and t9 had minimal heterogeneity (I_2=0% and 33%, respectively), whereas tB and tEB were substantially heterogenous (I_2=76% and 83%, respectively). The reasons for the heterogeneity are multifactorial including diverse patient populations, insufficient control for confounders, lack of standardisation of morphokinetic annotations, differences in laboratory and genetic testing techniques, and diverse embryo culture conditions. It must be highlighted that the results from tB and tEB are significantly heterogenous, therefore conclusions regarding these variables cannot reliably be drawn. That said, the heterogeneity for tEB is trending towards aneuploidy rather than traversing across the

line of no effect. Ordinarily we would be opposed to the meta-analysis of such heterogenous results, however the aim of this systematic review was not to provide a summary statistic to be translated directly into model development but to indicate potential prognostic markers for testing at local units. Whilst they are heterogenous, the results highlight the trend towards blastulation parameters predicting aneuploidy but, that said, we acknowledge that further research is needed to confirm our findings for tB and tEB.

The sensitivity analysis did not alter the results when studies with a high risk of bias were excluded, however tSB became significant with the use of a fixed effects model. This would indicate the need for more data to reliably conclude whether this variable could act as a prognostic marker. Interestingly, two morphokinetic studies were of higher quality and had comparable findings to our conclusions (Mumusoglu et al. 2017; Martin et al. 2021).

The association of day of blastocyst formation with aneuploid rates has been extensively studied, illustrating an increasing prevalence of aneuploidy from day 5 to day 7 blastocysts (Whitney et al. 2013; Minasi et al. 2016; Su et al. 2016; Kaing et al. 2018; Werland et al. 2017; Tiegs et al. 2019; Hernandez-Neito et al. 2019; McDaniel et al. 2021). Critics of time-lapse technology would argue that there is little to be gained from the study of cleavage parameters over the day of blastocyst formation using traditional monitoring. We argue that whilst day of blastocyst formation is a useful tool to counsel patients with limited access to time lapse, the accuracy and practicality that a TLS offers (for assessing readiness for biopsy whilst remaining in culture) is irreplaceable. Relying solely on traditional methods can lead to inaccuracies with the timing of blastulation in comparison to a TLS, where t0 is standardised to tPNf or time post insemination (hpi), allowing a more precise discrimination of a viable embryo despite slower development. The most successful morphokinetic logistic regression models for live birth are now much more complex than those using pre-defined thresholds, such as tSB <116 hrs or more traditional hierarchical models (Petersen et al. 2016; Zaninovic et al. 2017; Fishel et al.

2018). Time lapse therefore allows a statistical interpretation of embryo development whilst accounting for confounders that is not possible using traditional methods or univariate analysis. The variables more confidently associated with aneuploidy in this review are t8 and t9, factors that can only be considered through time lapse. It must also be considered that whilst tEB showed prognostic potential for ploidy, tSB and tB were not significantly associated, highlighting the precise nature of these associations rather than simply blastocyst formation. Finally, it has been suggested that there is some degree of multi-collinearity between cleavage and blastocyst kinetics, and this is illustrated by the fact that several authors have used earlier variables to predict blastocyst development (Wong et al. 2010; Cruz et al. 2012; Dal Canto et al. 2012; Hashimoto et al. 2012; Desai et al. 2014; Kirkegaard et al. 2013; Milewski et al. 2016). Therefore, this raises the question as to whether cleavage variables add prognostic value over the later blastulation parameters. Unfortunately, this has not been directly compared as published models either incorporate early cleavage parameters (Chavez et al. 2012; Chawla et al. 2015; Patel et al. 2016; Del Carmen Nogales et al. 2017) or blastulation variables (Campbell et al. 2013; Kramer et al. 2014; Desai et al. 2018; Lee et al. 2019; Martin et al. 2021). This would be an interesting question to drive future research, and care would need to be taken to not 'cherry-pick' variables to be included in prognostic model development however, as this can introduce significant bias outside the context of prognostic factor research (Riley et al. 2019).

More recently, artificial neural networks have demonstrated an impressive ability to evaluate images of pre-implantation embryos. Chavez-Badiola et al. (2020) developed a ranking system for ploidy status using this technology, with an impressive AUC of 0.70. Interestingly, two groups have investigated if there was an additive effect of using morphokinetic algorithms with artificial intelligence to improve diagnostic accuracy (Barnes et al. 2020; Huang et al. 2021). Barnes et al. (2020) demonstrated that both work synergistically to improve the AUC from

0.62 when solely image analysis is used to 0.76 (Barnes et al. 2020). Huang et al. (2021), similarly found the AUC increased from 0.57 to 0.77 with the addition of morphokinetics, age and full video analysis. This use of artificial intelligence in combination with morphokinetic models is a new direction of research that is evolving. Initial results appear promising and further studies are needed to demonstrate the application of this methodology. It would be beneficial for future work to include a prospective study design to validate these more complex models.

Morphological features and ploidy

It has been established that embryos with higher degrees of fragmentation have lower implantation rates; if the relationship suggested by our results is in fact true, the aetiology may be, in part, due to aneuploidy (Ziebe et al. 1997; Ebner et al. 2001). The quality of the evidence presented in all studies is poor, predominantly because of the use of unreliable genetic technologies (all used blastomere biopsy of intact cells and many adopted the use of FISH). Furthermore, the characteristics of the included patients are also extremely heterogenous. Some studies include couples with a good prognosis, in contrast to others focussing on patients with recurrent miscarriage or advanced maternal age, with no methods used to account for this (Tables II-VI). Notably, the fragmentation assessment method and timing were also inconsistent. This is important given that one author concluded that fragmentation was only associated with aneuploidy when assessed at the 7 and 8 cell stages and others when assessed at 48 hours (Magli et al. 2007; Ziebe et al. 2003). Other authors categorised fragmentation as 'high' or 'low'; these arbitrary thresholds make testing association more unreliable and to our knowledge there is no evidence to support such an approach (Vera-Rodriguez et al. 2015; Chavez et al. 2012). All considered, we cannot reliably conclude whether percentage fragmentation is associated with aneuploidy. There is, therefore, a need for future adequately

powered studies to examine fragmentation using time-lapse, next generation sequencing, and with adequate control of confounding.

Reverse cleavage has been associated with euploidy in our results, but it should be considered that these findings come from the contribution of one study and all other authors concluded that there was no significant difference (Ozbek et al. 2021). Whilst this was the largest study with 8% of embryos (n=78/1015) displaying reverse cleavage, the event rate remains low. For instance, we have calculated that for a power of 80% and a value of 0.05 for alpha you would need a sample size of 1617 embryos with at least 147 displaying reverse cleavage in order to find a difference when one truly exists. This is presuming a difference of 12% in the euploid rate between embryos displaying reverse cleavage and those that did not (extrapolated from the studies in this meta-analysis) and assuming a 1:10 ratio for the presence of this dysmorphism to normal cleavage. This illustrates a significant limitation of studies investigating dysmorphisms with such low prevalence. Ozbek et al. (2021) provide no explanation why embryos displaying reverse cleavage may have a higher incidence of euploidy, particularly in the context of the dramatically inferior live birth rates stated in their study when compared to normally cleaved euploid embryos (23% versus 56%). This association between reverse cleavage and inferior implantation rates has been replicated by several other authors, therefore we highly doubt that a relationship between euploidy and reverse cleavage truly exists (Barrie et al. 2017; Liu et al. 2014; Desai et al. 2018), particularly considering the underpowered nature of this study and the fact that reverse cleavage is often associated with compromised embryo development and quality. In fact, studies of bovine embryos have demonstrated an association with aneuploidy, strengthening the argument that these results are likely spurious (Magata et al. 2019). There have also been multiple factors independently associated with reverse cleavage, such as antagonist cycles, low progressive sperm motility and the use of ICSI (Liu et al. 2014).

While our main analyses indicate that direct uneven cleavage is not associated with ploidy, there is a significant limitation to the designs of the included studies. Aneuploid embryos may have been inadvertently excluded, either because only good quality embryos were biopsied or because a significant proportion (up to 87%) arrested in their development before biopsy (Zhan et al. 2016; Lagalla et al. 2017). It would be safer to conclude that embryos that have displayed direct cleavage that make it to the blastocyst stage could still be considered for biopsy or transfer: it has been demonstrated that they can result in live births, however, the patient must be warned of the increased likelihood of adverse outcome (Zhan et al. 2016; Fan et al. 2016; Ozbek et al. 2021). What causes these abnormal cleavages remains largely unknown, although it has previously been associated with the follicular environment of oocytes, poor-motility sperm and GnRH antagonists (Liu et al. 2014). Considering this, and the fixed effects sensitivity analysis that demonstrated DUC2 to be significantly associated with aneuploidy, further investigation is required to confirm or refute these findings.

Embryo contraction is a common phenomenon observed in a TLS (42% of embryos in the included studies), yet despite an understanding of the physiology, causality remains controversial. It has been hypothesised that contractions may assist in embryo hatching, although recent evidence does not support this theory (Gazzo et al. 2020). Future research should exclude studies that have undergone assisted hatching on day 3 as this has been related to altered frequency of contractions, a limitation of the included studies in the current analysis (Gazzo et al. 2020; Vinals Gonzalez et al. 2018). Embryos displaying contractions were more likely to be aneuploid, however this data comes from only two studies therefore further research is recommended to investigate this association.

The relationship between multinucleation at the 4-cell stage and ploidy is yet to be established given the significantly heterogenous results (I_2=88%) and contradictory findings in the subgroup and sensitivity analysis. It has been described how the presence of multinucleation

and associated aneuploidy can 'self-correct' by exclusion of cells during compaction or blastulation (Kligman et al. 1996; Ambroggio et al. 2011; Balakier et al. 2016; Desai et al. 2018). This complicates our understanding but may explain why only embryos displaying multinucleation at the 4-cell stage may be associated with aneuploidy and how healthy babies have been born from such embryos (Yilmaz et al. 2014; Meriano et al. 2004). Furthermore, multinucleation is only visible at interphase during conventional culture, therefore is likely to be underreported in the five included studies not utilising a TLS (Kligman et al. 1996; Magli et al. 2001; Agerholm et al. 2008; Scott, 2010; Ambroggio et al. 2011). In addition to aneuploidy, the presence of multinucleation has been related to the use of agonist down-regulation (perhaps associating it with poor ovarian reserve), high FSH dosages, high oestrogen levels and excessive oocyte numbers (Scott, 2010; De Cássia Savio Figueira et al. 2010; Desai et al. 2018). Despite this, across all the morphological studies only two manuscripts reporting the use of statistical modelling to adjust for age and no other confounders were considered (Minasi et al. 2016; Desai et al. 2018). In contrast to embryos displaying abnormal cleavage, there has been no difference demonstrated in the development of multinucleated embryos to expanded blastocyst, therefore our results are unlikely to be affected by arrested embryos (Goodman, 2016).

Strengths and limitations of this systematic review and meta-analysis

The findings of our study should be interpreted with caution due to an overall moderate risk of bias and significant heterogeneity of the included studies. Attempts have been made to control for sources of heterogeneity in our study design. This was primarily through subgroup analysis by excluding studies using older, unreliable technologies. In some variables the heterogeneity was calculated to be worse in the subgroup than in the main analysis, and this highlights the manifestation of other factors contributing to the diversified results. Heterogeneity may also exist in the way studies classified mosaics; this definition remains ambiguous in several studies.

This is important as mosaic embryos have previously been shown to have independent morphokinetic characteristics (Martin et al. 2021). It is also worth considering that whilst it is generally accepted that PGT-A biopsy results are concordant with the rest of the embryo in most cases, it is not absolute and sceptics exist (Victor et al. 2018; Esfandiari et al. 2016; Gleicher and Orvieto, 2017). There have been reports and suggested models of so called 'self-correction mechanisms' whereby mosaic embryos become more chromosomally normal as development progresses, although the existence of this phenomenon remains debatable (Capalbo and Rienzi, 2017; McCoy, 2017; Bolton et al. 2016; Munné et al. 2017; Coticchio, Barrie, et al. 2021).

Of the studies included in the meta-analysis of morphokinetic variables, all used ICSI, thus timing development from insemination apart from two groups, namely Lee et al. (2019) and Chavez et al. (2012). Unfortunately, t0 remains ambiguous in the study by Chavez et al. (2012) due to unanswered correspondence. Lee et al. (2019) used both standard IVF and ICSI for the included embryos, therefore this is a significant confounding factor to consider as they time conventional IVF embryos from the addition of spermatozoan to the oocyte; the accepted standard would be from tPNf. The exclusion of Chavez et al. (2012) for cc2 would make this variable not associated with ploidy status rather than associated with euploidy. That said, this study is not included in the subgroup analysis therefore the findings for this variable and S2 remain unchanged when considering the studies using the most reliable genetic technology.

While the conclusions drawn from this study are taken from data of over 40,000 embryos, the quality of evidence is low due to imprecision and large CIs. Only a limited number of studies tested each variable, leading to low event rates for some variables and the inclusion of only a handful of studies of those reporting usable data. As discussed previously, this is even more profound when the sample size of patients rather than embryos is considered. Furthermore, if the true population mean lies on the lower boundary of the 95% CI, we would be unable to

predict ploidy using tB and tEB. A final limitation is our inability to test for publication bias. That said, embryological studies tend to report a whole array of potential prognostic factors for ploidy per manuscript, therefore there is less risk of reporting only positive findings.

Our review does have multiple strengths, firstly the thorough methodological approach and comprehensive search of multiple variables and their association with ploidy status will be the first of its kind. Secondly, our meta-analyses of morphological and morphokinetic variables provide a strong argument for the local development of morphokinetic algorithms for ploidy and suggest those most likely to be included. Finally, we have provided an extensive critique of existing research and the quality of evidence in order to inform future prognostic methodologies.

Conclusion

In this first systematic review and meta-analysis of morphological and morphokinetic associations with ploidy, we have reported the most reliable prognostic markers to be t8, t9, and tEB. These results support the mechanism by which algorithms for live birth have predictive ability, suggesting that aneuploidy causes delayed cytokinesis. That said, we have demonstrated considerable variability within aneuploid and euploid embryos making definitively classifying them impossible. Time-lapse is, therefore, not suitable as a method to diagnose the ploidy status of pre-implantation embryos. Considering recent reports, it may be that morphokinetic algorithms can be used as a tool to risk stratify embryos for ploidy status, and more accurately by instrumenting artificial intelligence. Further research is needed to determine the suitability of machine learning for embryo assessment and selection.

Owing to the limited number of studies, heterogenous results and poor-quality evidence the suggested association between aneuploidy and multinucleation at the 4-cell stage, frequency of embryo contractions and fragmentation needs to be investigated further. Adequately powered studies should be conducted to test our hypothesis that reverse cleavage is not associated with

754 euploidy. We propose that incorporating associated morphological factors into a prognostic
755 model may work synergistically to improve euploid embryo selection. On the other hand,
756 multinucleation assessed on day 2 or at the 2-cell stage and direct unequal cleavage in an
757 embryo that progresses to a blastocyst do not appear to be associated with ploidy.
758 Differing clinical and laboratory practices and inadequate control for confounders in previous
759 research is most probably why TLS is rated as 'amber' by the UK regulatory body (HFEA
760 2021). There have been calls for multi-centre randomised controlled trials heard for many years
761 (Armstrong et al. 2019). Instead, we argue that since embryos are so significantly affected by
762 local conditions it may be more appropriate to robustly test models developed in-house.
763 While this review concludes that a TLS cannot be used to definitively diagnose ploidy status,
764 further research is needed to comprehend the potential of morphokinetic algorithms to prioritise
765 embryos for biopsy, or to use morphokinetics to select between euploid embryos. Therefore,
766 we will test this hypothesis in a cohort study at CARE Fertility using a morphokinetic dataset
767 of over 8,000 embryos with known PGT-A outcomes. This model will be trained, tested and
768 validated geographically and, if successful, a prospective study will determine its
769 discriminative ability. If successful, this has the potential to be a meaningful improvement for
770 patients, aiming to make more advanced and costly reproductive technologies more accessible.

771 **Data availability**

772 The data underlying this article are available in the article and in its online supplementary
773 material.

774 **Acknowledgements**

775 The first author would like to thank CARE Fertility for providing a platform for research and
776 clinical training through the creation of a clinical research fellow post. In particular, Mr Phillip
777 Lowe (Medical Director, CARE Manchester), Dr Victoria Sephton (CARE Group Deputy

Medical Director), and Professor Kingsland (CARE Group Clinical Director) for their continued support. The investigators would like to thank the following authors who supplied additional information about their research: Professor David McCulloh, Professor Magli, Dr Hakan Yarali, Dr Pavlo Mazur. Additionally, we would like to thank Dr Adam Devall for assistance in obtaining certain manuscripts.

Authors' roles

The study was conceived by Prof. Arri Coomarasamy and A.B. as part of a PhD programme of research undertaken by T.B at the University of Birmingham. The study protocol was designed by T.B., S.M., and supervised by Prof. Arri Coomarasamy. Study selection and extraction of data was performed by T.B. and A.B., followed by a quality assessment by T.B. All authors analysed and interpreted the data. T.B. drafted the first manuscript, this was subsequently approved by all authors before publication. C.E. provided statistical support and A. Campbell provided expertise within this field of interest.

Funding

T.B. receives funding obtained by CARE Fertility for Ph.D. tuition fees and stipend at the University of Birmingham

Conflict of interest

Alison Campbell is a minor shareholder at CARE Fertility. No other conflicts of interest exist. To note, CAREmaps® is a technology for which patients are charged extra for at CARE.

Figure legends

Figure 1 Flow diagram for study selection process for human embryo morphokinetics search.

Figure 2 Flow diagram for study selection process for human embryo fragmentation search.

Figure 3 Flow diagram for study selection process for human embryo abnormal cleavage search.

Figure 4 Flow diagram for study selection process for human embryo contraction search.

Figure 5 Flow diagram for study selection process for human embryo multinucleation search.

Figure 6 Weighted mean difference and prognostic factor analysis graphs of aneuploid vs. euploid human embryos for morphokinetic variables:

t8 **(A):** time from insemination to 8 cells (hpi)

t9 (**B):** time from insemination to 9 cells (hpi)

tB **(C):** time from insemination to the formation of a full blastocyst (hpi)

tEB (**D**): time from insemination to expanded blastocyst (hpi)

Figure 7 Relative risk of euploidy in a human embryo displaying reserve cleavage (RC).

Figure 8 Relative risk of euploidy in a human embryo displaying contractions.

Figure 9 Relative risk of a multinucleated (MN) human embryo being euploid when assessed at the 4- cell stage.

828 **References**

829 Agerholm IE, Hnida C, Cruger DG, Berg C, Bruun-Petersen G, Kolvraa S, and Ziebe S.
830 Nuclei size in relation to nuclear status and aneuploidy rate for 13 chromosomes in
831 donated four cells embryos. *J Assist Reprod Genet* 2008: **25**; 95-102.

832 Alfarawati S, Fragouli E, Colls P, Stevens J, Gutiérrez-Mateo C, Schoolcraft WB, Katz-Jaffe
833 MG, and Wells D. The relationship between blastocyst morphology, chromosomal
834 abnormality, and embryo gender. *Fertil Steril* 2011: **95**; 520-524.

835 Alpha Scientists in Reproductive M and Embryology ESIGo. The Istanbul consensus
836 workshop on embryo assessment: proceedings of an expert meeting†. *Human*
837 *Reproduction* 2011: **26**; 1270-1283.

838 Ambroggio J, Gindoff PR, Dayal MB, Khaldi R, Peak D, Frankfurter D, and Dubey AK.
839 Multinucleation of a sibling blastomere on day 2 suggests unsuitability for embryo
840 transfer in IVF-preimplantation genetic screening cycles. *Fertil Steril* 2011: **96**; 856-
841 859.

842 Armstrong S, Bhide P, Jordan V, Pacey A, Marjoribanks J, and Farquhar C. Time-lapse
843 systems for embryo incubation and assessment in assisted reproduction. *Cochrane*
844 *Database of Systematic Reviews* 2019.

845 Ata B, Kaplan B, Danzer H, Glassner M, Opsahl M, Tan SL, and Munné S. Array CGH
846 analysis shows that aneuploidy is not related to the number of embryos generated.
847 *Reprod Biomed Online* 2012: **24**; 614-620.

848 Athayde Wirka K, Chen AA, Conaghan J, Ivani K, Gvakharia M, Behr B, Suraj V, Tan L,
849 and Shen S. Atypical embryo phenotypes identified by time-lapse microscopy: high
850 prevalence and association with embryo development. *Fertil Steril* 2014: **101**; 1637-
851 1648.e1631-1635.

852 Balakier H, Sojecki A, Motamedi G, and Librach C. Impact of multinucleated blastomeres on
853 embryo developmental competence, morphokinetics, and aneuploidy. *Fertility and*
854 *sterility.* 2016: **106**; 608-614.e602.
855 Barnes J, Malmsten J, Zhan Q, Hajirasouliha I, Elemento O, Sierra J, Zaninovic N, and
856 Rosenwaks Z. Noninvasive detection of blastocyst ploidy (euploid vs. aneuploid)
857 using artificial intelligene (AI) with deep learning methods. *Fertility and Sterility*
858 2020: **114**; e76.
859 Barrie A, Homburg R, McDowell G, Brown J, Kingsland C, and Troup S. Preliminary
860 investigation of the prevalence and implantation potential of abnormal embryonic
861 phenotypes assessed using time-lapse imaging. *Reproductive BioMedicine Online*
862 2017: **34**; 455-462.
863 Barrie A, McDowell G, and Troup S. An investigation into the effect of potential
864 confounding patient and treatment parameters on human embryo morphokinetics.
865 *Fertility and Sterility* 2021: **115**; 1014-1022.
866 Basile N, Nogales Mdel C, Bronet F, Florensa M, Riqueiros M, Rodrigo L, Garcia-Velasco J,
867 and Meseguer M. Increasing the probability of selecting chromosomally normal
868 embryos by time-lapse morphokinetics analysis. *Fertil Steril* 2014: **101**; 699-704.
869 Bayram. A FM, Shunnar A., Divakaran J., Poukaram N. Does disappearance of
870 multinucleation from the 2-cell to 4-cell stage observed by time time-lapse correlate
871 chromosomal normality detected by PGS microarray? *Human reproduction.* 2015: **30**.
872 Bellver J, Mifsud A, Grau N, Privitera L, and Meseguer M. Similar morphokinetic patterns in
873 embryos derived from obese and normoweight infertile women: a time-lapse study.
874 *Hum Reprod* 2013: **28**; 794-800.

875 Bodri D, Sugimoto T, Yao Serna J, Kawachiya S, Kato R, and Matsumoto T. Blastocyst
876 collapse is not an independent predictor of reduced live birth: a time-lapse study.
877 *Fertil Steril* 2016: **105**; 1476-1483.e1473.
878 Bolton H, Graham SJL, Van der Aa N, Kumar P, Theunis K, Fernandez Gallardo E, Voet T,
879 and Zernicka-Goetz M. Mouse model of chromosome mosaicism reveals lineage-
880 specific depletion of aneuploid cells and normal developmental potential. *Nat*
881 *Commun* 2016: **7**; 11165.
882 Bronet F, Nogales MC, Martínez E, Ariza M, Rubio C, García-Velasco JA, and Meseguer M.
883 Is there a relationship between time-lapse parameters and embryo sex? *Fertil Steril*
884 2015: **103**; 396-401.e392.
885 Campbell A, Fishel S, Bowman N, Duffy S, Sedler M, and Hickman CF. Modelling a risk
886 classification of aneuploidy in human embryos using non-invasive morphokinetics.
887 *Reprod Biomed Online* 2013: **26**; 477-485.
888 Campbell A, Fishel S, Bowman N, Duffy S, Sedler M, and Thornton S. Retrospective
889 analysis of outcomes after IVF using an aneuploidy risk model derived from time-
890 lapse imaging without PGS. *Reprod Biomed Online* 2013: **27**; 140-146.
891 Capalbo A and Rienzi L. Mosaicism between trophectoderm and inner cell mass. *Fertility*
892 *and Sterility* 2017: **107**; 1098-1106.
893 Capalbo A, Rienzi L, Cimadomo D, Maggiulli R, Elliott T, Wright G, Nagy ZP, and Ubaldi
894 FM. Correlation between standard blastocyst morphology, euploidy and implantation:
895 an observational study in two centers involving 956 screened blastocysts. *Human*
896 *Reproduction* 2014: **29**; 1173-1181.
897 Chavez-Badiola A, Flores-Saiffe-Farías A, Mendizabal-Ruiz G, Drakeley AJ, and Cohen J.
898 Embryo Ranking Intelligent Classification Algorithm (ERICA): artificial intelligence

clinical assistant predicting embryo ploidy and implantation. *Reprod Biomed Online* 2020: **41**; 585-593.

Chavez SL, Loewke KE, Han J, Moussavi F, Colls P, Munne S, Behr B, and Reijo Pera RA. Dynamic blastomere behaviour reflects human embryo ploidy by the four-cell stage. *Nat Commun* 2012: **3**; 1251.

Chawla M, Fakih M, Shunnar A, Bayram A, Hellani A, Perumal V, Divakaran J, and Budak E. Morphokinetic analysis of cleavage stage embryos and its relationship to aneuploidy in a retrospective time-lapse imaging study. *J Assist Reprod Genet* 2015: **32**; 69-75.

Ciray HN, Aksoy T, Goktas C, Ozturk B, and Bahceci M. Time-lapse evaluation of human embryo development in single versus sequential culture media--a sibling oocyte study. *J Assist Reprod Genet* 2012: **29**; 891-900.

Cohen J, Wells D, and Munné S. Removal of 2 cells from cleavage stage embryos is likely to reduce the efficacy of chromosomal tests that are used to enhance implantation rates. *Fertil Steril* 2007: **87**; 496-503.

Coticchio G, Barrie A, Lagalla C, Borini A, Fishel S, Griffin D, and Campbell A. Plasticity of the human preimplantation embryo: developmental dogmas, variations on themes and self-correction. *Hum Reprod Update* 2021: **27**; 848-865.

Coticchio G, Ezoe K, Lagalla C, Shimazaki K, Ohata K, Ninomiya M, Wakabayashi N, Okimura T, Uchiyama K, Kato K*, et al.* Perturbations of morphogenesis at the compaction stage affect blastocyst implantation and live birth rates. *Human Reproduction* 2021: **36**; 918-928.

Cruz M, Garrido N, Herrero J, Pérez-Cano I, Muñoz M, and Meseguer M. Timing of cell division in human cleavage-stage embryos is linked with blastocyst formation and quality. *Reproductive BioMedicine Online* 2012: **25**; 371-381.

924 Davies S, Christopikou, D., Tsorva, E., Karagianni, T., Mastrominas, M., Handyside, A. H.
925 Altered morphokinetic parameters of embryos identified as aneuploid by single cell
926 array CGH analysis at the 8-cell stage. *Reproductive biomedicine online.* 2012: **24**;
927 S39-S40.
928 Davies S. TE, Christopikou D., Vogiatzi P., Karagianni A., Handyside A.H., Mastrominas M.
929 Morphokinetic assessment using time-lapse imaging allows more accurate
930 identification of embryos with multiple rather than single aneuploidies up to the 8-cell
931 stage. Abstract book of the 30th ESHRE Annual Meeting, Munich, Germany, 29 June
932 – 2 July 2014. . *Human Reproduction* 2014: **29**; i1-i389.
933 De Cássia Savio Figueira R, Souza Setti A, Paes De Almeida Ferreira Braga D, Iaconelli A,
934 Jr., and Borges E, Jr. Blastomere multinucleation: Contributing factors and effects on
935 embryo development and clinical outcome. *Hum Fertil (Camb)* 2010: **13**; 143-150.
936 Dal Canto M, Coticchio G, Mignini Renzini M, De Ponti E, Novara PV, Brambillasca F,
937 Comi R, and Fadini R. Cleavage kinetics analysis of human embryos predicts
938 development to blastocyst and implantation. *Reproductive BioMedicine Online* 2012:
939 **25**; 474-480.
940 Del Carmen Nogales M, Bronet F, Basile N, Martinez EM, Linan A, Rodrigo L, and
941 Meseguer M. Type of chromosome abnormality affects embryo morphology
942 dynamics. *Fertil Steril* 2017: **107**; 229-235 e222.
943 Delimitreva SM, Zhivkova RS, Vatev IT, and Toncheva DI. Chromosomal disorders and
944 nuclear and cell destruction in cleaving human embryos. *Int J Dev Biol* 2005: **49**; 409-
945 416.
946 DerSimonian R and Laird N. Meta-analysis in clinical trials. *Controlled Clinical Trials* 1986:
947 **7**; 177-188.

948 Desai N, Goldberg JM, Austin C, and Falcone T. Are cleavage anomalies, multinucleation, or
949 specific cell cycle kinetics observed with time-lapse imaging predictive of embryo
950 developmental capacity or ploidy? *Fertil Steril* 2018: **109**; 665-674.
951 Desai N. RP. Is there an increase in aneuploidy rate with delayed blastulation,
952 multinucleation or cleavage anomalies? *Fertility and sterility.* 2016: **106**; e145.
953 DeUgarte CM, Li M, Surrey M, Danzer H, Hill D, and DeCherney AH. Accuracy of FISH
954 analysis in predicting chromosomal status in patients undergoing preimplantation
955 genetic diagnosis. *Fertil Steril* 2008: **90**; 1049-1054.
956 Ebner T, Yaman C, Moser M, Sommergruber M, Pölz W, and Tews G. Embryo
957 fragmentation in vitro and its impact on treatment and pregnancy outcome. *Fertil*
958 *Steril* 2001: **76**; 281-285.
959 Edwards RG, Fishel SB, Cohen J, Fehilly CB, Purdy JM, Slater JM, Steptoe PC, and Webster
960 JM. Factors influencing the success of in vitro fertilization for alleviating human
961 infertility. *Journal of in Vitro Fertilization and Embryo Transfer* 1984: **1**; 3-23.
962 Esfandiari N, Bunnell ME, and Casper RF. Human embryo mosaicism: did we drop the ball
963 on chromosomal testing? *Journal of Assisted Reproduction and Genetics* 2016: **33**;
964 1439-1444.
965 Fan YL, Han SB, Wu LH, Wang YP, and Huang GN. Abnormally cleaving embryos are able
966 to produce live births: a time-lapse study. *J Assist Reprod Genet* 2016: **33**; 379-385.
967 Fiorentino F, Bono S, Biricik A, Nuccitelli A, Cotroneo E, Cottone G, Kokocinski F, Michel
968 C-E, Minasi MG, and Greco E. Application of next-generation sequencing technology
969 for comprehensive aneuploidy screening of blastocysts in clinical preimplantation
970 genetic screening cycles. *Human Reproduction* 2014: **29**; 2802-2813.
971 Fishel S, Campbell A, Montgomery S, Smith R, Nice L, Duffy S, Jenner L, Berrisford K,
972 Kellam L, Smith R*, et al.* Time-lapse imaging algorithms rank human preimplantation

973 embryos according to the probability of live birth. *Reprod Biomed Online* 2018: **37**;
974 304-313.
975 Fragouli E, Alfarawati S, Spath K, and Wells D. Morphological and cytogenetic assessment
976 of cleavage and blastocyst stage embryos. *Mol Hum Reprod* 2014: **20**; 117-126.
977 Fragouli E and Wells D. Aneuploidy in the human blastocyst. *Cytogenet Genome Res* 2011:
978 **133**; 149-159.
979 Franasiak JM, Forman EJ, Hong KH, Werner MD, Upham KM, Treff NR, and Scott RT, Jr.
980 The nature of aneuploidy with increasing age of the female partner: a review of
981 15,169 consecutive trophectoderm biopsies evaluated with comprehensive
982 chromosomal screening. *Fertil Steril* 2014: **101**; 656-663.e651.
983 Fréour T, Dessolle L, Lammers J, Lattes S, and Barrière P. Comparison of embryo
984 morphokinetics after in vitro fertilization-intracytoplasmic sperm injection in smoking
985 and nonsmoking women. *Fertil Steril* 2013: **99**; 1944-1950.
986 Fujimoto VY, Browne RW, Bloom MS, Sakkas D, and Alikani M. Pathogenesis,
987 developmental consequences, and clinical correlations of human embryo
988 fragmentation. *Fertility and Sterility* 2011: **95**; 1197-1204.
989 Gardner D. K. SWB. "In vitro culture of human blastocyst," in Towards Reproductive
990 Certainty: Infertility and Genetics Beyond In Mortimer D. (Carnforth: Parthenon
991 Press;) (ed). 1999.
992 Gardner DK and Balaban B. Assessment of human embryo development using morphological
993 criteria in an era of time-lapse, algorithms and 'OMICS': is looking good still
994 important? *Molecular Human Reproduction* 2016: **22**; 704-718.
995 Gazzo E, Pena F, Valdez F, Chung A, Velit M, Ascenzo M, and Escudero E. Blastocyst
996 contractions are strongly related with aneuploidy, lower implantation rates, and slow-
997 cleaving embryos: a time lapse study. *JBRA Assist Reprod* 2020: **24**; 77-81.

998 Gleicher N and Orvieto R. Is the hypothesis of preimplantation genetic screening (PGS) still
999 supportable? A review. *Journal of Ovarian Research* 2017: **10**; 21.
1000 Goodman L.R. GJM, Falcone T., Desai N., Rambhia P. Time-lapse imaging of
1001 multinucleated embryos and the association with aneuploidy determined by
1002 preimplantation genetic screening. *Fertility and sterility.* 2016: **106**; e151.
1003 Goossens V, De Rycke M, De Vos A, Staessen C, Michiels A, Verpoest W, Van Steirteghem
1004 A, Bertrand C, Liebaers I, Devroey P*, et al.* Diagnostic efficiency, embryonic
1005 development and clinical outcome after the biopsy of one or two blastomeres for
1006 preimplantation genetic diagnosis. *Hum Reprod* 2008: **23**; 481-492.
1007 Grooten WJA, Tseli E, Äng BO, Boersma K, Stålnacke B-M, Gerdle B, and Enthoven P.
1008 Elaborating on the assessment of the risk of bias in prognostic studies in pain
1009 rehabilitation using QUIPS—aspects of interrater agreement. *Diagnostic and*
1010 *Prognostic Research* 2019: **3**; 5.
1011 Hashimoto S, Kato N, Saeki K, and Morimoto Y. Selection of high-potential embryos by
1012 culture in poly(dimethylsiloxane) microwells and time-lapse imaging. *Fertility and*
1013 *Sterility* 2012: **97**; 332-337.
1014 Hashimoto S, Nakano T, Yamagata K, Inoue M, Morimoto Y, and Nakaoka Y.
1015 Multinucleation per se is not always sufficient as a marker of abnormality to decide
1016 against transferring human embryos. *Fertil Steril* 2016: **106**; 133-139 e136.
1017 Hernandez-Nieto C, Lee JA, Slifkin R, Sandler B, Copperman AB, and Flisser E. What is the
1018 reproductive potential of day 7 euploid embryos? *Hum Reprod* 2019: **34**; 1697-1706.
1019 HFEA. Fertility treatment 2019: trends and figures. 2021.
1020 HFEA. Treatment add-ons with limited evidence. 2021.

1021 Higgins JPT TJ, Chandler J, Cumpston M, Li T, Page MJ, Welch VA. Cochrane Handbook
1022 for Systematic Reviews of Interventions version 6.2 (updated February 2021).
1023 Cochrane. 2021.
1024 Ho JR, Arrach N, Rhodes-Long K, Salem W, McGinnis LK, Chung K, Bendikson KA,
1025 Paulson RJ, and Ahmady A. Blastulation timing is associated with differential
1026 mitochondrial content in euploid embryos. *J Assist Reprod Genet* 2018: **35**; 711-720.
1027 Huang B, Tan W, Li Z, and Jin L. An artificial intelligence model (euploid prediction
1028 algorithm) can predict embryo ploidy status based on time-lapse data. *Reprod Biol*
1029 *Endocrinol* 2021: **19**; 185-185.
1030 Kaing A, Kroener LL, Tassin R, Li M, Liu L, Buyalos R, Hubert G, and Shamonki M. Earlier
1031 day of blastocyst development is predictive of embryonic euploidy across all ages:
1032 essential data for physician decision-making and counseling patients. *Journal of*
1033 *Assisted Reproduction and Genetics* 2018: **35**; 119-125.
1034 Kalatova B, Jesenska R, Hlinka D, and Dudas M. Tripolar mitosis in human cells and
1035 embryos: occurrence, pathophysiology and medical implications. *Acta Histochem*
1036 2015: **117**; 111-125.
1037 Kemper JM, Liu Y, Afnan M, Hammond ER, Morbeck DE, and Mol BWJ. Should we look
1038 for a low-grade threshold for blastocyst transfer? A scoping review. *Reproductive*
1039 *BioMedicine Online* 2021: **42**; 709-716.
1040 Kim S-G, Kim Y-Y, Park J-Y, Kwak S-J, Yoo C-S, Park I-H, Sun H-G, Kim J-W, Lee K-H,
1041 Park H-D*, et al.* Early fragment removal on in vitro fertilization day 2 significantly
1042 improves the subsequent development and clinical outcomes of fragmented human
1043 embryos. *Clinical and experimental reproductive medicine* 2018: **45**; 122-128.
1044 Kimelman D, Confino R, Okeigwe I, Lambe-Steinmiller J, Confino E, Shulman LP, Zhang
1045 JX, and Pavone ME. Assessing the impact of delayed blastulation using time lapse

morphokinetics and preimplantation genetic testing in an IVF patient population. *J Assist Reprod Genet* 2019: **36**; 1561-1569.

Kirkegaard K, Kesmodel US, Hindkjær JJ, and Ingerslev HJ. Time-lapse parameters as predictors of blastocyst development and pregnancy outcome in embryos from good prognosis patients: a prospective cohort study. *Human Reproduction* 2013: **28**; 2643-2651.

Kirkegaard K, Hindkjaer JJ, and Ingerslev HJ. Effect of oxygen concentration on human embryo development evaluated by time-lapse monitoring. *Fertil Steril* 2013: **99**; 738-744.e734.

Kligman I, Benadiva C, Alikani M, and Munne S. The presence of multinucleated blastomeres in human embryos is correlated with chromosomal abnormalities. *Hum Reprod* 1996: **11**; 1492-1498.

Kramer YG, Kofinas JD, Melzer K, Noyes N, McCaffrey C, Buldo-Licciardi J, McCulloh DH, and Grifo JA. Assessing morphokinetic parameters via time lapse microscopy (TLM) to predict euploidy: are aneuploidy risk classification models universal? *J Assist Reprod Genet* 2014: **31**; 1231-1242.

Lagalla. Embryos with cell division aberrations monitored by timelapse imaging in a PGS program: Are they able to develop into euploid blastocysts? *Human reproduction.* 2015: **30**.

Lagalla C, Tarozzi N, Sciajno R, Wells D, Di Santo M, Nadalini M, Distratis V, and Borini A. Embryos with morphokinetic abnormalities may develop into euploid blastocysts. *Reprod Biomed Online* 2017: **34**; 137-146.

Leary C, Leese HJ, and Sturmey RG. Human embryos from overweight and obese women display phenotypic and metabolic abnormalities. *Hum Reprod* 2015: **30**; 122-132.

1070 Lee CI, Chen CH, Huang CC, Cheng EH, Chen HH, Ho ST, Lin PY, Lee MS, and Lee TH.
1071 Embryo morphokinetics is potentially associated with clinical outcomes of single-
1072 embryo transfers in preimplantation genetic testing for aneuploidy cycles. *Reprod*
1073 *Biomed Online* 2019: **39**; 569-579.
1074 Lemmen JG, Agerholm I, and Ziebe S. Kinetic markers of human embryo quality using time-
1075 lapse recordings of IVF/ICSI-fertilized oocytes. *Reprod Biomed Online* 2008: **17**;
1076 385-391.
1077 Li F. UM, Ayers J.W., Abuzeid M.I., Khan I. Euploidy is not predictable by time lapse
1078 imaging system with either blastomere or trophectoderm biopsies. *Fertility and*
1079 *sterility.* 2015: **104**; e314.
1080 Liu Y, Chapple V, Roberts P, and Matson P. Prevalence, consequence, and significance of
1081 reverse cleavage by human embryos viewed with the use of the Embryoscope time-
1082 lapse video system. *Fertil Steril* 2014: **102**; 1295-1300.e1292.
1083 Los FJ, Van Opstal D, and van den Berg C. The development of cytogenetically normal,
1084 abnormal and mosaic embryos: a theoretical model. *Hum Reprod Update* 2004: **10**;
1085 79-94.
1086 Magata F, Ideta A, Okubo H, Matsuda F, Urakawa M, and Oono Y. Growth potential of
1087 bovine embryos presenting abnormal cleavage observed through time lapse
1088 cinematography. *Theriogenology* 2019: **133**; 119-124.
1089 Magli MC, Gianaroli L, and Ferraretti AP. Chromosomal abnormalities in embryos. *Mol Cell*
1090 *Endocrinol* 2001: **183 Suppl 1**; S29-34.
1091 Magli MC, Gianaroli L, Ferraretti AP, Lappi M, Ruberti A, and Farfalli V. Embryo
1092 morphology and development are dependent on the chromosomal complement. *Fertil*
1093 *Steril* 2007: **87**; 534-541.

1094 Mantel N and Haenszel W. Statistical aspects of the analysis of data from retrospective
1095 studies of disease. *J Natl Cancer Inst* 1959: **22**; 719-748.
1096 Marcos J, Perez-Albala S, Mifsud A, Molla M, Landeras J, and Meseguer M. Collapse of
1097 blastocysts is strongly related to lower implantation success: a time-lapse study. *Hum*
1098 *Reprod* 2015: **30**; 2501-2508.
1099 Marcos J, Pérez-Albalá S, Mifsud A, Molla M, Landeras J, and Meseguer M. Collapse of
1100 blastocysts is strongly related to lower implantation success: a time-lapse study. *Hum*
1101 *Reprod* 2015: **30**; 2501-2508.
1102 Martin A, Rodrigo L, Beltran D, Meseguer M, Rubio C, Mercader A, and de Los Santos MJ.
1103 The morphokinetic signature of mosaic embryos: evidence in support of their own
1104 genetic identity. *Fertil Steril* 2021: **116**; 165-173.
1105 Mazur P, Nagornyy, V., Mykytenko, D., Semeniuk, L., Zukin, V. Multinucleation and
1106 aneuploidy, time-lapse imaging and aCGH may shed the light on causation. *Fertility*
1107 *and sterility.* 2013: **100**; S68.
1108 McCoy RC. Mosaicism in Preimplantation Human Embryos: When Chromosomal
1109 Abnormalities Are the Norm. *Trends Genet* 2017: **33**; 448-463.
1110 McCoy RC, Newnham LJ, Ottolini CS, Hoffmann ER, Chatzimeletiou K, Cornejo OE, Zhan
1111 Q, Zaninovic N, Rosenwaks Z, Petrov DA*, et al.* Tripolar chromosome segregation
1112 drives the association between maternal genotype at variants spanning PLK4 and
1113 aneuploidy in human preimplantation embryos. *Hum Mol Genet* 2018: **27**; 2573-2585.
1114 McDaniel KE, Awadalla MS, McGinnis LK, and Ahmady A. Transfer the best and biopsy the
1115 rest? Blastocyst euploidy rates differ by morphology and day of biopsy. *Archives of*
1116 *Gynecology and Obstetrics* 2021: **303**; 249-258.
1117 Melzer KE, Noyes N, Hodes-Wertz B, McCulloh D, Munne S, and Grifo JA. How well do
1118 morphokinetic (MK) parameters and time-lapse microscopy (TLM) predict euploidy?

a pilot study of TLM with trophectoderm (TE) biopsy with array comparative genomic hybridization (aCGH). *Fertility and sterility.* 2013: **100**; S209.

Meriano J, Clark C, Cadesky K, and Laskin CA. Binucleated and micronucleated blastomeres in embryos derived from human assisted reproduction cycles. *Reprod Biomed Online* 2004: **9**; 511-520.

Milewski R, Milewska AJ, Kuczyńska A, Stankiewicz B, and Kuczyński W. Do morphokinetic data sets inform pregnancy potential? *Journal of Assisted Reproduction and Genetics* 2016: **33**; 357-365.

Minasi MG, Colasante A, Riccio T, Ruberti A, Casciani V, Scarselli F, Spinella F, Fiorentino F, Varricchio MT, and Greco E. Correlation between aneuploidy, standard morphology evaluation and morphokinetic development in 1730 biopsied blastocysts: a consecutive case series study. *Hum Reprod* 2016: **31**; 2245-2254.

Moayeri SE, Allen RB, Brewster WR, Kim MH, Porto M, and Werlin LB. Day-3 embryo morphology predicts euploidy among older subjects. *Fertil Steril* 2008: **89**; 118-123.

Moher D, Liberati A, Tetzlaff J, Altman DG, and Group P. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *J Clin Epidemiol* 2009: **62**; 1006-1012.

Mumusoglu S, Ozbek IY, Sokmensuer LK, Polat M, Bozdag G, Papanikolaou E, and Yarali H. Duration of blastulation may be associated with ongoing pregnancy rate in single euploid blastocyst transfer cycles. *Reprod Biomed Online* 2017: **35**; 633-639.

Munné S. Chromosome abnormalities and their relationship to morphology and development of human embryos. *Reprod Biomed Online* 2006: **12**; 234-253.

Munné S, Blazek J, Large M, Martinez-Ortiz PA, Nisson H, Liu E, Tarozzi N, Borini A, Becker A, Zhang J*, et al.* Detailed investigation into the cytogenetic constitution and

pregnancy outcome of replacing mosaic blastocysts detected with the use of high-resolution next-generation sequencing. *Fertility and Sterility* 2017: **108**; 62-71.e68.

Munne S and Cohen J. Chromosome abnormalities in human embryos. *Hum Reprod Update* 1998: **4**; 842-855.

Munné S, Kaplan B, Frattarelli JL, Child T, Nakhuda G, Shamma FN, Silverberg K, Kalista T, Handyside AH, Katz-Jaffe M*, et al.* Preimplantation genetic testing for aneuploidy versus morphology as selection criteria for single frozen-thawed embryo transfer in good-prognosis patients: a multicenter randomized clinical trial. *Fertility and Sterility* 2019: **112**; 1071-1079.e1077.

Munoz E. OM, Perez M., Taboas E., Meseguer M., Aguilar J. Morphology dynamics of multinucleation affects embryo chromosome content. Abstract book of the 30 th ESHRE Annual Meeting, Munich, Germany, 29 June – 2 July 2014. *Human reproduction.* 2014: **29**; i1-i389.

Muñoz M, Cruz M, Humaidan P, Garrido N, Pérez-Cano I, and Meseguer M. The type of GnRH analogue used during controlled ovarian stimulation influences early embryo developmental kinetics: a time-lapse study. *Eur J Obstet Gynecol Reprod Biol* 2013: **168**; 167-172.

Neal SA, Sun L, Jalas C, Morin SJ, Molinaro TA, and Scott RT. When next-generation sequencing-based preimplantation genetic testing for aneuploidy (PGT-A) yields an inconclusive report: diagnostic results and clinical outcomes after re biopsy. *Journal of Assisted Reproduction and Genetics* 2019: **36**; 2103-2109.

Niimura S. Time-lapse videomicrographic analyses of contractions in mouse blastocysts. *J Reprod Dev* 2003: **49**; 413-423.

1166 Ozbek IY, Mumusoglu S, Polat M, Bozdag G, Sokmensuer LK, and Yarali H. Comparison of
1167 single euploid blastocyst transfer cycle outcome derived from embryos with normal or
1168 abnormal cleavage patterns. *Reprod Biomed Online* 2021: **42**; 892-900.
1169 Patel DV, Shah PB, Kotdawala AP, Herrero J, Rubio I, and Banker MR. Morphokinetic
1170 behavior of euploid and aneuploid embryos analyzed by time-lapse in embryoscope. *J*
1171 *Hum Reprod Sci* 2016: **9**; 112-118.
1172 Petersen BM, Boel M, Montag M, and Gardner DK. Development of a generally applicable
1173 morphokinetic algorithm capable of predicting the implantation potential of embryos
1174 transferred on Day 3. *Human reproduction (Oxford, England)* 2016: **31**; 2231-2244.
1175 Pickering SJ, Taylor A, Johnson MH, and Braude PR. An analysis of multinucleated
1176 blastomere formation in human embryos. *Hum Reprod* 1995: **10**; 1912-1922.
1177 Pribenszky C, Nilselid AM, and Montag M. Time-lapse culture with morphokinetic embryo
1178 selection improves pregnancy and live birth chances and reduces early pregnancy
1179 loss: a meta-analysis. *Reprod Biomed Online* 2017: **35**; 511-520.
1180 Puissant F, Van Rysselberge M, Barlow P, Deweze J, and Leroy F. Embryo scoring as a
1181 prognostic tool in IVF treatment. *Human Reproduction* 1987: **2**; 705-708.
1182 Rienzi L, Capalbo A, Stoppa M, Romano S, Maggiulli R, Albricci L, Scarica C, Farcomeni
1183 A, Vajta G, and Ubaldi FM. No evidence of association between blastocyst
1184 aneuploidy and morphokinetic assessment in a selected population of poor-prognosis
1185 patients: a longitudinal cohort study. *Reprod Biomed Online* 2015: **30**; 57-66.
1186 Rienzi S. BA, Capalbo L., Spizzichino C., Rubio F.M., Ubaldi F., Fiorentino. Human embryo
1187 aneuploidues cannot be predicted by morphokinetic assessement during the first
1188 phases of in vitro culture. *Human Reproduction. Abstract book,* 2013: **28**; i149-i206.
1189 Riley RD, van der Windt D, Croft P, and Moons KG. Prognosis research in healthcare:
1190 concepts, methods, and impact 2019. Oxford University Press.

1191 Royen EV, Mangelschots K, Vercruyssen M, Neubourg DD, Valkenburg M, Ryckaert G, and
1192 Gerris J. Multinucleation in cleavage stage embryos. *Human Reproduction* 2003: **18**;
1193 1062-1069.
1194 Ruangvutilert P, Delhanty JD, Rodeck CH, and Harper JC. Relative efficiency of FISH on
1195 metaphase and interphase nuclei from non-mosaic trisomic or triploid fibroblast
1196 cultures. *Prenat Diagn* 2000: **20**; 159-162.
1197 Rubio I, Kuhlmann R, Agerholm I, Kirk J, Herrero J, Escribá M-J, Bellver J, and Meseguer
1198 M. Limited implantation success of direct-cleaved human zygotes: a time-lapse study.
1199 *Fertility and Sterility* 2012: **98**; 1458-1463.
1200 Schünemann H BJ, Guyatt G, Oxman A, editors. GRADE handbook for grading quality of
1201 evidence and strength of recommendations. . The GRADE Working Group, 2013.
1202 2013.
1203 Scott L. FA, Kloos B., Davies D. The origin and consequences of Day-2 multinucleation of
1204 human embryos. *Human Reproduction. Conference: 26th Annual Meeting of the*
1205 *European Society of Human Reproduction and Embryology, ESHRE. Rome Italy.*
1206 *Conference Publication: (var.pagings). 25 (SUPPL. 1) (pp i195-i196), 2010. Date of*
1207 *Publication: June 2010.* 2010: **25**; i170-i210.
1208 Scriven PN and Bossuyt PM. Diagnostic accuracy: theoretical models for preimplantation
1209 genetic testing of a single nucleus using the fluorescence in situ hybridization
1210 technique. *Hum Reprod* 2010: **25**; 2622-2628.
1211 Staessen C, Platteau P, Van Assche E, Michiels A, Tournaye H, Camus M, Devroey P,
1212 Liebaers I, and Van Steirteghem A. Comparison of blastocyst transfer with or without
1213 preimplantation genetic diagnosis for aneuploidy screening in couples with advanced
1214 maternal age: a prospective randomized controlled trial. *Hum Reprod* 2004: **19**; 2849-
1215 2858.

1216 Stroup DF. Meta-analysis of Observational Studies in Epidemiology. *JAMA : the journal of*
1217 *the American Medical Association.* 2000: **283**; 2008.
1218 Su Y, Li J-J, Wang C, Haddad G, and Wang W-H. Aneuploidy analysis in day 7 human
1219 blastocysts produced by in vitro fertilization. *Reproductive Biology and*
1220 *Endocrinology* 2016: **14**; 1-7.
1221 Tarín JJ, Conaghan J, Winston RM, and Handyside AH. Human embryo biopsy on the 2nd
1222 day after insemination for preimplantation diagnosis: removal of a quarter of embryo
1223 retards cleavage. *Fertil Steril* 1992: **58**; 970-976.
1224 Taylor-Rowan M, Edwards S, Noel-Storr AH, McCleery J, Myint PK, Soiza R, Stewart C,
1225 Loke YK, and Quinn TJ. Anticholinergic burden (prognostic factor) for prediction of
1226 dementia or cognitive decline in older adults with no known cognitive syndrome.
1227 *Cochrane Database of Systematic Reviews* 2021.
1228 Theobald R, SenGupta S, and Harper J. The status of preimplantation genetic testing in the
1229 UK and USA. *Hum Reprod* 2020: **35**; 986-998.
1230 Tiegs AW, Tao X, Zhan Y, Whitehead C, Kim J, Hanson B, Osman E, Kim TJ, Patounakis
1231 G, Gutmann J*, et al.* A multicenter, prospective, blinded, nonselection study
1232 evaluating the predictive value of an aneuploid diagnosis using a targeted next-
1233 generation sequencing–based preimplantation genetic testing for aneuploidy assay and
1234 impact of biopsy. *Fertility and Sterility* 2021: **115**; 627-637.
1235 Vera-Rodriguez M, Chavez SL, Rubio C, Reijo Pera RA, and Simon C. Prediction model for
1236 aneuploidy in early human embryo development revealed by single-cell analysis. *Nat*
1237 *Commun* 2015: **6**; 7601.
1238 Victor AR, Griffin DK, Brake AJ, Tyndall JC, Murphy AE, Lepkowsky LT, Lal A, Zouves
1239 CG, Barnes FL, McCoy RC*, et al.* Assessment of aneuploidy concordance between

clinical trophectoderm biopsy and blastocyst. *Human Reproduction* 2018: **34**; 181-192.

Vinals Gonzalez X, Odia R, Cawood S, Gaunt M, Saab W, Seshadri S, and Serhal P. Contraction behaviour reduces embryo competence in high-quality euploid blastocysts. *J Assist Reprod Genet* 2018: **35**; 1509-1517.

von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, and Vandenbroucke JP. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *Lancet* 2007: **370**; 1453-1457.

Watson AJ, Natale DR, and Barcroft LC. Molecular regulation of blastocyst formation. *Anim Reprod Sci* 2004: **82-83**; 583-592.

Werland H, VerMilyea M, and Silverberg K. Culturing embryos to day 7. A viable option for IVF patients? *Fertility and Sterility* 2017: **108**; e161-e162.

Whitney J, Anderson R, and Schiewe M. Day 7 blastocyst euploidy and implantation rates warrant implentation for all programs using preimplantation genetic screening (PGS). *Fertility and Sterility* 2016: **106**; e146.

Wong CC, Loewke KE, Bossert NL, Behr B, De Jonge CJ, Baer TM, and Pera RAR. Non-invasive imaging of human embryos before embryonic genome activation predicts development to the blastocyst stage. *Nature Biotechnology* 2010: **28**; 1115-1121.

Yang Z, Zhang J, Salem SA, Liu X, Kuang Y, Salem RD, and Liu J. Selection of competent blastocysts for transfer by combining time-lapse monitoring and array CGH testing for patients undergoing preimplantation genetic screening: a prospective study with sibling oocytes. *BMC Med Genomics* 2014: **7**; 38.

1263 Yilmaz A, Zhang L, Zhang XY, Son WY, Holzer H, and Ao A. Chromosomal complement
1264 and clinical relevance of multinucleated embryos in PGD and PGS cycles. *Reprod*
1265 *Biomed Online* 2014: **28**; 380-387.
1266 Zaninovic N, Irani M, and Meseguer M. Assessment of embryo morphology and
1267 developmental dynamics by time-lapse microscopy: is there a relation to implantation
1268 and ploidy? *Fertility and Sterility* 2017: **108**; 722-729.
1269 Zhan Q, Ye Z, Clarke R, Rosenwaks Z, and Zaninovic N. Direct Unequal Cleavages: Embryo
1270 Developmental Competence, Genetic Constitution and Clinical Outcome. *PLoS One*
1271 2016: **11**; e0166398.
1272 Zhang J, Tao W, Liu H, Yu G, Li M, Ma S, and Wu K. Morphokinetic parameters from a
1273 time-lapse monitoring system cannot accurately predict the ploidy of embryos. *J*
1274 *Assist Reprod Genet* 2017: **34**; 1173-1178.
1275 Ziebe S, Lundin K, Loft A, Bergh C, Nyboe Andersen A, Selleskog U, Nielsen D, Grondahl
1276 C, Kim H, Arce JC*, et al.* FISH analysis for chromosomes 13, 16, 18, 21, 22, X and Y
1277 in all blastomeres of IVF pre-embryos from 144 randomly selected donated human
1278 oocytes and impact on pre-embryo morphology. *Hum Reprod* 2003: **18**; 2575-2581.
1279 Ziebe S, Petersen K, Lindenberg S, Andersen AG, Gabrielsen A, and Andersen AN. Embryo
1280 morphology or cleavage stage: how to select the best embryos for transfer after in-
1281 vitro fertilization. *Hum Reprod* 1997: **12**; 1545-1549.

1282