ASSOCIATE EDITOR’S COMMENT ON THE FAST-TRACK ARTICLE ‘DIAGNOSTIC EFFICIENCY, EMBRYONIC DEVELOPMENT AND CLINICAL OUTCOME AFTER THE BIOPSY OF ONE OR TWO BLASTOMERES FOR PREIMPLANTATION GENETIC DIAGNOSIS’ BY GOOSENS ET AL

What are the trade-offs between one-cell and two-cell biopsies of preimplantation embryos?

Catherine M.H. Combelles¹

Biology Department, Middlebury College, McCardell Bicentennial Hall 346, Middlebury, VT 05753, USA

¹Correspondence address. Tel: +1-802-443-5251; Fax: +1-802-443-2071; E-mail: ccombell@middlebury.edu

Keywords: preimplantation genetic diagnosis; preimplantation genetic screening; blastomere biopsy; diagnostic efficiency; clinical outcome

The fast changing landscape of reproductive technologies has experienced its share of controversies, among which is the genetic screening of preimplantation embryos. In principle, preimplantation genetic diagnosis (PGD), for known genetic afflictions, and preimplantation genetic screening (PGS), for aneuploidy detection, provide an option for fertile and infertile patients to detect any abnormal embryos and thus presumably augment implantation and take-home baby rates. In practice, PGD and PGS are not yet offered by all fertility centers with only 4–6% of IVF cycles combined with PGD in the USA (Baruch et al., 2007). Nonetheless, PGD is one of the fastest growing technologies in assisted reproduction (Sermon et al., 2007). Some experts even predict it to become one day a procedure ancillary to routine IVF cycles. Yet a divide characterizes the field, notably with respect to the potential clinical benefits of PGD/PGS. Regardless of which side of the divide a clinic may stand on, everyone agrees that intensive efforts should be placed on elucidating the true value of PGD/PGS. Only then will practitioners and patients be in a comfortable position to offer, select, or opt out of its use for treatment. While many facets of PGD remain investigational in nature, the technology already entered the clinical arena with even an aggressive use of it by some centers.

It is striking that despite the hundreds of IVF laboratories performing PGD/PGS worldwide and the 50 centers reporting to the ESHRE PGD Consortium (Sermon et al., 2007), the scientific literature appears somewhat dominated by a handful of groups. Together, these investigators have made significant strides. Nonetheless, a call must go out to others towards a more concerted involvement and scientific advancement of PGD. PGD/PGS is a complex procedure that includes a plethora of technological hurdles, from the biopsy to the processing of extracted blastomeres and the final diagnostic analysis. PGD can be used to detect an ever-increasing number of diseases, over a 100 different genetic conditions to date (Baruch et al., 2005). Ongoing efforts have also focused on the diagnostic procedures, with the testing of alternative and/or complementary approaches. FISH and PCR are widely employed with comparative genomic hybridization and spectral karyotyping representing some of the newer ones on the docket (Wells, 2004; Donoso et al., 2007). But grossly understudied is the methodology used for cell removal along with all of the critical laboratory procedures surrounding the biopsy itself. In this vein, experts have identified a number of pivotal discrepancies in the PGD lab procedure, among which are the cell removal technique, the fixation protocol, the selected set of chromosomal probes for FISH and the number of cells used for biopsy (De Vos and Van Steirteghem, 2001; Cohen and Munne, 2005; Cohen et al., 2007; Munné et al., 2007). The study published in this issue by Goossens et al. precisely addresses one of these long-standing questions as to whether more good than harm ensues from the biopsy of two cells versus one cell. This study is incredibly timely, especially in light of the most recently published article by...
Cohen et al. (2007) that has continued to fuel heated discussions in the field.

To undertake PGD/PGS, there is a minimal need for the removal of one cell, as one cannot currently circumvent the donation of a cell nucleus for the genetic evaluation of the embryo. However, after a few years of practice with single-cell biopsy, some groups began performing the routine biopsy of two cells (Donoso et al., 2007). There are several premises for doing so, among which is the practicality of providing a confirmed result with the analysis of two cells. This is particularly important in view of current limitations in diagnostic technologies with several studies reporting a range of error rates between 1 and 10% (ASRM Practice Committee, 2006). But even at its lowest level and in the hands of the most successful groups, a misdiagnosis is not an acceptable outcome for any patients undergoing PGD cycles. False-negatives result in the transfer of genetically inflicted embryos, and false-positives reduce the number of embryos available during a given cycle. When identified, errors are devastating to patients and all parties involved, so greatest efforts must be placed on ideally avoiding them altogether. For the highest standards of clinical care to be maintained, it is imperative that diagnosis be efficient and accurate (PGDIS, 2004; Thornhill et al., 2005). Sources of errors include technological and biological ones (Lewis et al., 2001). Errors may arise during the FISH fixation, staining and evaluation steps (Cohen et al., 2007; Munné et al., 2007). Biologically, the accuracy of diagnosis can be complicated by the existence of mosaicism in the early human embryo; indeed, a rather large proportion of embryos appear made up of blastomeres of different chromosomal constituencies (reviewed by Donoso et al., 2007). Given current technical limitations and confounding factors, increased accuracy may be gained from testing two cells.

Benefits aside, there may be developmental handicaps resulting from the removal of one-third to one-fourth of the cellular mass of an embryo. There is a clinical divide in the use of one-versus two-cell biopsy. The prominent centers in the USA largely advocate and practice the removal of a single cell, while several European groups routinely employ two-cell biopsy (Donoso et al., 2007). However, it is interesting to acknowledge that one-fourth of US clinics report on the biopsy of two cells (Baruch et al., 2007). In addition, groups in the Netherlands and Spain performed either one- or two-cell biopsy depending on the quality of Day 3 embryos (Donoso et al., 2007; Mastenbroek et al., 2007). Some centers may also resort to a combination of one- and two-cell removal, not by design but rather based on the success of cell removal at the time of biopsy. This lack of consensus on the number of cells to biopsy is also evident in published PGD practice guidelines (PGDIS, 2004; Thornhill et al., 2005).

So are two cells too many to remove? The time has come to settle the divide in PGD practices once and for all, and this is precisely where the study by Goossens et al. (2008) comes in. That said this study is not the first to address the issue. The literature is disseminated with evidence, or at times simply arguments, for and against two-cell biopsy. In a non-randomized design, Van De Velde et al. (2000) showed comparable pregnancy and implantation rates whether one or two cells were removed for FISH analysis. Beyond its retrospective nature and sample sizes of insufficient power, an inherent bias in embryo selection further weakened the study. The healthiest embryos were used for two-cell biopsy with only one cell removed from subquality embryos; therefore, any detrimental effects of two-cell biopsy may not be apparent due to the augmented developmental potential that the embryos started with. Recently, Cohen et al. (2007) drew analogies between cell loss during cryopreservation and embryo biopsy to argue that two-cell biopsy could reduce implantation potential by up to 20–30%. Irrespective of the biological validity of this prediction, the assumption that embryos recover similarly to freezing and thawing or blastomere biopsy is a dangerous one. Indeed, the causes and mechanisms for cell loss are distinct, and cryopreservation likely causes its own set of side effects.

Non-randomized studies reported a small benefit of two-cell over one-cell biopsy on diagnostic efficiency, with only FISH analysis considered to date (Emiliani et al., 2004; Michiels et al., 2006). The proportion of false-positives did not significantly differ with the FISH analysis of one or two cells (Baart et al., 2004, 2006; Michiels et al., 2006), and no definite conclusions can yet be drawn on the incidences of false-negatives with one- and two-cell biopsies (Emiliani et al., 2004; Michiels et al., 2006). The analysis of other diagnostic parameters (e.g. sensitivity, specificity and predictive values) failed to reveal significant differences according to the number of removed cells (Emiliani et al., 2004; Michiels et al., 2006). Taken together, recurring issues afflicted the design of studies actually testing the clinical and diagnostic effects of cell removal. Indeed, the one- and two-cell biopsy groups have not yet been randomly allocated irrespective of embryo quality at the time of biopsy. Sample sizes remain small, not adequately powered, and in most cases skewed to the two-cell biopsy group. Last but not least, all studies were retrospective in nature, thereby setting the stage for an eagerly awaited randomized controlled trial (RCT) on one- and two-cell biopsies. In a theoretical study on PCR diagnosis, Lewis et al. (2001) calculated that the probability to identify (and thus transfer) an unaffected embryo was increased with the biopsy of two cells when compared to one. Los et al. (2004) also modelled a reduction in FISH misdiagnosis with the removal of a second cell. While the potential to reduce error rates may be computed based on technical parameters and assumptions, one cannot predict from theory alone what the biological influences of cell removal will be on the developmental potential of the embryo.

The study by Goossens et al. (2008) aimed to test the potential influences of analyzing one or two blastomeres by either PCR or FISH on a clinically relevant set of primary outcomes, i.e. pregnancy outcomes and diagnostic efficiencies. Secondly, the investigators assessed diagnostic accuracy by PCR. As never reported to date, the authors undertook a prospective RCT at a single center, including both PGD and PGS patients with inclusion criteria for neither age nor history of recurrent miscarriage. There are several significant findings offered by this study, one of which pertains to the prospective evaluation of embryonic outcomes after the biopsy of one or two cells in embryos of comparable quality initially. Several commentaries
argue that the removal of two cells introduces significant handicaps, which may in turn irreversibly compromise the developmental competence of biopsied embryos (Magli et al., 2004; Cohen and Munné, 2005; Cohen et al., 2007; Munné et al., 2007). Goossens et al. (2008) are able to demonstrate that embryo quality, as based on morphology, declined \textit{in vitro} after the removal of two cells when compared with one. This finding could indicate that embryos become more stressed when two cells are removed, although additional detriment may also stem from culture conditions that are not yet up to par with those prevailing in the \textit{in vivo} environment. Drawing inferences about the exact influences of cell removal on embryo integrity is not that simple. Indeed, the authors poignantly identify and demonstrate that it is the quality of the embryos on Day 3 that predicts later (to Day 5 at least) developmental competence, rather than whether one or two cells were removed. This finding complements previous studies indicating that the quality of Day 3 embryos influences the subsequent success of PGD (Hardy et al., 1993). Goossens et al. (2008) thus demonstrate that as long as an embryo is of sufficient quality at the time of biopsy, it appears resilient to the removal of two cells, just as much as the ablation of a single cell. The only developmental assessment presented thus far has been embryo morphology, a measure that certainly does not hold as much predictive power as pregnancy outcomes. Interestingly, two-cell removal did not influence the implantation rate of transferred embryos, and the number of cells biopsied affected neither the availability nor the transfer rates of blastocysts. The best of primary outcome measures, i.e. live birth rates also did not differ depending on the number of cells biopsied, and thus in statistical terms, the incurred cost of removing two cells does not appear significant. But the clinical importance of any RCT finding should be addressed (Arce et al., 2005), and in this respect the authors show that for every 33 cycles, one live birth is accrued with one- over two-cell biopsy (Goossens et al., 2008).

Another strength to the study pertains to the follow-up clinical data including neonatal and post-natal reports, from which there is nothing to be alarmed about at this point. As for all ongoing studies on PGD populations, larger sample sizes are needed (PGDIS, 2004; Baruch et al., 2005; Thornhill et al., 2005; Sermon et al., 2007). The Developmental Origins of Adult Health and Disease supports the need for such monitoring (Gluckman et al., 2007). Such an invasive procedure (whether it be the removal of one or two cells) must be tested prospectively and rigorously for potential influences through later development and adulthood.

Based on the groundbreaking finding by Goossens et al. (2008), is it surprising that early human embryos display plasticity with a potential to make up for the absence of two cells? Historically, the use of PGD first proceeded based on the premise that in early developmental stages, all blastomeres are totipotent and that in turn every cell is an equal indicator of the whole. But this is a gross assumption that merits experimentation in human embryos. Simply put, we still do not know with any certainty whether and how the removal (conducted largely at random) of a nucleated cell harms the donor embryo. The biopsy procedure may disrupt any cellular and molecular polarities that potentially hold instructions to the developmental programme of the embryo. There is a striking and alarming paucity of literature on the subject of cell polarity and cell determination in the early mammalian embryo. In the early days of PGD, a couple of original studies began tackling the developmental potential of embryos post-biopsy (Hardy et al., 1990; Tarin et al., 1992). Since then, speculative and yet pertinent extrapolations have been made from model organisms to mammals (Edwards and Beard, 1997; Edwards, 2005). A case exists for pre-patterning in the early mammalian embryos, although with its own dose of controversy (Antczak and Van Blerkom, 1997; Zernicka-Goetz, 2002, 2006; Edwards, 2005; Gardner, 2006; Hiiragi and Solter, 2006; Hiiragi et al., 2006a,b).

An increased understanding of embryonic polarity should inform the practice of PGD, notably with respect to the selection of which cell (or two cells) to remove. The goal would be to remove representative cells (i.e. with respect to the genetic make-up of the embryo) while not disrupting irreversibly the polarity or placement of any early developmental determinants. How much cell fate determination a given blastomere has undergone by the 8-cell stage is another black box of early human embryonic development. There is a plea for some fundamental basic research on the early development of human embryos, particularly taking into consideration a current proposal for the co-existence of regulative development and pre-patterning in mouse preimplantation embryos (Zernicka-Goetz, 2002, 2006). Nonetheless, there are many preconceived notions about early human embryos, and virtually nothing is known on the true plasticity of the human embryo. For now, clinical studies such as the one published in this issue point towards a regulative ability of the remaining embryo, at least up to a threshold of two cells. The question thus remains as to whether a step-wise decline in embryo developmental competence exists when one or two cells are removed as opposed to none.

While the clinical study by Goossens et al. (2008) tackles the balance of benefits and risks between removing one versus two cells, the ultimate value of analyzing any number of cells when compared with none remains to be established (Shahine and Cedars, 2006; Twisk et al., 2007). If biopsy is to cause any harm to the embryo, the ‘good’ must be high enough to justify a little ‘bad’. Two adequately powered RCTs are most often cited, although not without their own share of disapprovals. Staessen et al. (2004) performed a first RCT with no demonstrated benefit of PGS over routine IVF cycles in patients of advanced maternal age. However, the study was criticized because of the removal of two cells from all ≥6-cell embryos (Cohen and Munné, 2005; Cohen et al., 2007; Munné et al., 2007). The argument put forth pertains to the lack of differences in clinical outcome that may simply reflect an actual benefit of PGS in face of the loss of two cells versus one cell. Such reasoning stands corrected with the report by Goossens et al. (2008). Most recently, Mastenbroek et al. (2007) show a compromised clinical outcome in women of advanced maternal age treated with PGS. This RCT was based on the routine biopsy of a single blastomere although the removal of a second cell was performed as needed, the prevalence of which was alas not...
reported in the published trial. Upon publication, this study was criticized for its use of potentially suboptimal biopsy protocols (Cohen and Grifo, 2007; Munné et al., 2007), thereby providing the impetus for testing various PGD methodologies. Even though critiques abound, discussions have delineated gold standards by which future RCTs on PGD efficacy and effectiveness should abide.

Goossens et al. (2008) report that the lack of diagnosis in PCR–PGD cycles was less likely to arise with two-versus one-cell biopsy. So while for live birth there is a slight benefit to one-cell removal, this is at the expense of a compromised efficiency in PCR diagnosis. Each type of biopsy method should thus be advocated based on the severity of the genetic disease and its reproductive risks as well as the diagnostic methodology and its proven efficiency in the hands of a given center. A choice to biopsy one or two cells will likely vary among centers with different technical proficiencies. This echoes with the sentiments from the group in New Jersey that PGD efficacy ultimately depends on ‘how good’ a center is (Munné et al., 2007). Diagnostically, removing a second cell provides a back-up system to any technical failures and procedural limitations that currently exist with PCR (Wells, 2004), notably the common lack of DNA amplification from a single cell (Goossens et al., 2008). The study by Goossens et al. (2008) thus facilitates the making of informed choices between the biopsy of one and two cells whether on a center, case and/or embryo basis.

In contrast to with PCR–PGD, there was no difference in the diagnostic efficiency of FISH–PGD with one-versus two-cell biopsy. Perhaps this simply reflects inequalities in technical capabilities using FISH versus PCR; FISH is the oldest of the PGD diagnostic method, and it may just be a matter of time before diagnosis by PCR be just as efficient whether one or two cells are removed. While a previous study by the same group (Michiels et al., 2006) described a significant increase in successful FISH diagnosis after the biopsy of two cells, the difference was only of 2.3%. Importantly, the two studies employed distinct approaches, and it is thus paramount to confirm findings using varied approaches (Twisk et al., 2005). Cohen et al. (2007) also report that the efficiency of detecting chromosomal errors increases with nine versus six probes. With the use of mostly 5–7 chromosomal pairs, Goossens et al. (2008) could have thus underestimated the number of errors detectable by FISH. Regardless, the investigators convincingly demonstrate that when using a common and useful set of chromosomal probes, there is no change in diagnostic efficiency between the one- and two-cell FISH–PGD groups.

As recommended for any PGD studies (PGDIS, 2004; Thornhill et al., 2005; Munné et al., 2007), Goossens et al. (2008) re-assessed non-transferred embryos, with a slightly higher misdiagnosis in the two-cell when compared with one-cell PCR–PGD groups. Although sample size was too small to draw statistical conclusions, this secondary outcome poses new questions. The fact that false-positives are higher in the two-cell group may simply reflect mosaicism during early cleavage stages. Logically, mosaicism only becomes detectable when two cells are analysed, and the analysis of two cells may inflate the number of Day 3 embryos assigned as genetically afflicted. To rationalize further the classification of an embryo as positive on Day 3 with yet a negative diagnosis on Day 5, it is possible that a specific genotype (notably a mosaic genotype) may not be detectable to sufficient efficiencies in whole blastocysts. After two additional days in culture, the full activation of the zygotic genome may conceivably enable embryos to self-correct. The repair of cellular damage (or even the elimination of aberrant cells altogether) could then account for diagnostic inaccuracies, notably false-positives. Likewise, it should be considered that in the case of false-negatives, new errors might arise after two more days in vitro, particularly under suboptimal culture conditions and at a pivotal developmental transition in the life of a mammalian embryo. Clearly, larger analyses of misdiagnosis promise to reveal areas of improvement, all in the grand goal to eliminate errors.

The study by Goossens et al. (2008) did not test the accuracy of FISH, but previous reports showed that in non-randomized, comparative studies on donated spare embryos, two-cell biopsy permitted a slight reduction, albeit non-significant, in false-positives (Baart et al., 2004, 2006; Michiels et al., 2006). The theoretical modeling by Los et al. (2004) also calculated higher rates of misdiagnosis in the one-cell when compared with the two-cell group. Additional prospective and randomized studies will lend further credence to the actual improvement in FISH diagnosis offered by two-cell biopsy. Interestingly, confirmation rates were further improved in the two-cell group when diagnosis was concordant on Day 3 (Baart et al., 2004, 2006). The study by Goossens et al. (2008) does not report on how congruent the diagnosis was between the two removed cells, however, previous studies provided notes of caution on a relatively high incidence of discordance (50–75% of cases) between the FISH analysis of two blastomeres from a single embryo (Michiels et al., 2006; Coulam et al., 2007). The few reports on two-cell biopsy indicate that it is mostly mosaicism that complicates the accuracy of PGD diagnosis (Baart et al., 2004, 2006).

Although the exact prevalence and biological relevance of mosaicism remains unknown, its existence is a certainty (Bielanska et al., 2002; Munné, 2006; Munné et al., 2006). Consequently, how could we ever be certain that the diagnosis of one or two cells is representative of the embryo? Can we comfortably advocate the culling of any embryos that are mosaic based on the random removal of two cells? Therefore, should we only perform two-cell biopsy so that any hints of mosaicism may be detected and avoided? What if mosaicism reflects a normal, perhaps strengthening process in unhealthy embryos? If so, may two-cell biopsy select out so-called developmentally incompetent embryos more often than necessary? It is arguable that embryos diagnosed as aberrant may develop into normal fetuses in utero. As aforementioned, the existence of false-positives may reflect a background level of rescuing ability in early embryos. However, the self-correcting potential of mammalian embryos remains unknown. Another poignant question pertains to the minimum number of cells that must be normal for an embryo to produce a viable pregnancy. Furthermore, Baart et al. (2006) proposed that the removal of two spatially adjacent and aberrant cells could effectively serve as a form of embryo therapy. Future research must thus
consider whether two-cell biopsy itself may assist some of the mosaic embryos. Also, with the removal of two cells proven not to be significantly detrimental, mosaicism can now be tested with an eye on prospectively and randomly examining not only its impact on diagnostic accuracy of PGD with one- or two-cell biopsy, but also its biological relevance to later development.

The work by Goossens et al. (2008) impacts all centers that offer PGD regardless of their current use of one- or two-cell biopsy, as it is never too late to learn what is true to do what is right. Best practice guidelines shall now be revised to reflect the novel data published herein on the diagnostic, embryonic and clinical trade-offs between one- versus two-cell biopsy. There is not yet a single protocol for PGD, and often times this limitation complicates and in the eyes of some even undermines previous RCTs. Reaching some commonalities in PGD protocols will strengthen the testing of PGS versus non-PGS cycles. Now that Goossens et al. (2008) addressed one of these fundamental technological variants, future studies should similarly test all aspects of the PGD procedures and analyses. For instance, RCT studies must address the potential influences of probe choices and fixation technique on clinical measures (Cohen et al., 2007). In this vein, a few methodological aspects (e.g. zona breaching and biopsy methods) are included in the ESHRE PGT Consortium database (Sermon et al., 2007). Once the field can identify and implement a congruent PGD protocol, future RCTs promise to tell once and for all, for whom, and how PGD may become a mainstay of ART. Based on prospective and randomized data rather than assumptions, biases, extrapolations or theoretical modeling, the study by Goossens et al. (2008) provides the field with a refreshing dataset. As a result, the longstanding debates surrounding one- versus two-cell biopsy can now evolve into new frontiers. The authors’ findings may influence other aspects of ART, notably the pressing need for reliable predictors of embryo quality. A combinatorial evaluation of embryos will most likely prove successful, and the removal of two cells may then permit diagnosis by multiple approaches without dangerously compromising the integrity of the embryo. Future efforts should also be placed on deriving embryo cohorts of highest quality, and thus presumably setting the stage for the most successful of genetic screening, be it from one or two cells.

References

Antczak M, Van Blerkom J. Oocyte influences on early development: the regulatory proteins leptin and STAT3 are polarized in mouse and human oocytes and differentially distributed within the cells of the preimplantation stage embryo. Mol Hum Reprod 1997;3:1067–1086.


Baart EB, Van Opstal D, Los FJ, Au Yeung DS, van Blerkom J. Oocyte influences on early development: the work by Goossens et al. (2008) impacts all centers that offer PGD regardless of their current use of one- or two-cell biopsy, as it is never too late to learn what is true to do what is right. Best practice guidelines shall now be revised to reflect the novel data published herein on the diagnostic, embryonic and clinical trade-offs between one- versus two-cell biopsy. There is not yet a single protocol for PGD, and often times this limitation complicates and in the eyes of some even undermines previous RCTs. Reaching some commonalities in PGD protocols will strengthen the testing of PGS versus non-PGS cycles. Now that Goossens et al. (2008) addressed one of these fundamental technological variants, future studies should similarly test all aspects of the PGD procedures and analyses. For instance, RCT studies must address the potential influences of probe choices and fixation technique on clinical measures (Cohen et al., 2007). In this vein, a few methodological aspects (e.g. zona breaching and biopsy methods) are included in the ESHRE PGT Consortium database (Sermon et al., 2007). Once the field can identify and implement a congruent PGD protocol, future RCTs promise to tell once and for all, for whom, and how PGD may become a mainstay of ART. Based on prospective and randomized data rather than assumptions, biases, extrapolations or theoretical modeling, the study by Goossens et al. (2008) provides the field with a refreshing dataset. As a result, the longstanding debates surrounding one- versus two-cell biopsy can now evolve into new frontiers. The authors’ findings may influence other aspects of ART, notably the pressing need for reliable predictors of embryo quality. A combinatorial evaluation of embryos will most likely prove successful, and the removal of two cells may then permit diagnosis by multiple approaches without dangerously compromising the integrity of the embryo. Future efforts should also be placed on deriving embryo cohorts of highest quality, and thus presumably setting the stage for the most successful of genetic screening, be it from one or two cells.


Submitted on November 20, 2007; resubmitted on November 20, 2007; accepted on November 20, 2007.