Running title: Raman detection of aneuploidy in embryo

Raman Profiling of Embryo Culture Medium to Identify Aneuploid and Euploid Embryos

Bo Liang¹ (PhD), Yuan Gao^{2,3} (PhD), Jiabao Xu (Msci)⁴, Yizhi Song⁴ (PhD), Liming Xuan⁵(MS), Ting Shi¹ (PhD), Ning Wang⁶ (PhD), Zhaoxu Hou⁶ (PhD), Yi-Lei Zhao^{1} (PhD), Wei E. Huang⁴ (PhD) and Zi-Jiang Chen^{2,3} (PhD, MD)*

1. State Key Laboratory of Microbial Metabolism, Joint International Research Laboratory of Metabolic and Developmental Sciences, School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, 800 Dongchuan Road, Shanghai 200240, China.

2. Center for Reproductive Medicine, Provincial Hospital Affiliated to Shandong University, Jinan, Shandong 250001, China.

3. The Key laboratory for Reproductive Endocrinology of Ministry of Education, Jinan, Shandong 250001, China.

4. Department of Engineering Science, University of Oxford, Parks Road, Oxford, OX1 3PJ, United Kingdom.

5. Basecare Medical Device Co., Ltd., 218 Xinghu Road, SIP., Suzhou, Jiangsu 215001, China.

6. Mathematical Institute, University of Oxford, Andrew Wiles Building, Radcliffe Observatory Quarter, Woodstock Rd, Oxford OX2 6GG

*Corresponding author:

Yi-Lei Zhao yileizhao@sjtu.edu.cn Tel: +86-21-34207190 (YZ)

Objective: To develop and validate Raman metabolic footprint analysis to determine chromosome euploidy and aneuploidy in IVF embryos

Design: Retrospective study.

Setting: Academic hospital.

Patient(s): Unselected assisted reproductive technology population.

Intervention(s): To establish the analysis protocol, spent embryo culture medium samples with known genetic outcomes from 87 human embryos were collected and measured by Raman spectroscopy. Individual Raman spectrum was analyzed to find biological components contributing to either euploidy or aneuploidy. To validate the protocol via machine learning algorithms, additional 1107 Raman spectra from 123 embryo culture medium (61 euploidy and 62 aneuploidy) were analyzed.

Main Outcome Measure(s): Raman-based footprint profiling of spent culture media and preimplantation genetic testing for aneuploidy (PGT-A) analysis.

Results: Mean-centered Raman spectra and principal component analysis (PCA) showed differences in the footprints of euploid and aneuploid embryos growing in culture medium. Significant differences in Raman bands associated with small RNAs and lipids were also observed. Stacking classification based on k-Nearest-Neighbor (kNN), Random Forests (RF) and Extreme Gradient Boosting (XGB) algorithms achieved an overall accuracy of 95.9% in correctly assigning either euploidy or aneuploidy based on Raman spectra, which is validated by PGT-A sequencing results.

Conclusions: This study suggests that chromosomal abnormalities in embryos should lead to the changes of metabolic footprints in embryo growth medium, which can be detected by Raman spectroscopy. The ploidy status of embryos was analyzed by Raman-based footprint profiling of spent culture media, which is consistent with PGT-A testing performed by next-generation sequencing (NGS).

Keywords: Raman microspectroscopy, machine learning, chromosome abnormality, IVF, preimplantation genetic testing for aneuploidy (PGT-A)

INTRODUCTION

Chromosome aneuploidy is one kind of chromosomal abnormalities in human characteristic of an abnormal number of chromosomes. The principal causes of aneuploidy in human oocytes and embryos are chromosome segregation errors (e.g. nondisjunction, anaphase lag, premature separation of sister chromatids). The vast majority of those abnormalities are lethal and can cause embryonic growth arrest before implantation or spontaneous abortion of the developing fetus during the first trimester of pregnancy (1, 2). According to NHS UK in 2010, the percentage of IVF treatments that resulted in a live birth was 32.2% for women under 35 and only 1.9% for women aged over 44 (3). Detection of such chromosomal abnormalities is therefore crucial to ensure healthy embryos created by *in vitro* fertilization (IVF) or intracytoplasmic sperm injection (ICSI) before transferring embryos back to the womb (4).

Current preimplantation genetic testing for aneuploidy (PGT-A) methods usually involve removing a single cell or a small number of cells from an embryo and then examined by genetic techniques such as single-cell whole-genome amplification (WGA) in combination with next-generation sequencing (NGS) technology or array comparative genomic hybridization (array CGH). Single cell NGS for diagnosis is powerful which shows a single sequence is accurate to 99.9% (5). Furthermore, Society of Obstetricians and Gynaecologists of Canada (SOGC) released a document on pre-implantation genetic diagnosis and screening techniques in 2017. However, single cell NGS is invasive and time-consuming which can be sophisticated and potentially cause a problem to the embryo if it is done improperly (6-8). Therefore, improved methods are in need to select the best embryos based on direct evidence showing normal chromosome structure.

One potential approach to evaluate embryonic reproductive competence using a rapid, noninvasive, sensitive, and clinically applicable platform is metabolic profiling of the embryo culture medium (9). By measuring metabolites remained in the culture media, metabolic footprints have been used to reveal the phenotype of mutations of yeast genes (10). Previous studies have shown that changes in the composition of culture medium correspond to differences in embryo quality and reproductive potential (11-13). In these attempts, the correlation between pyruvate uptake (14), glucose uptake (15), concentration of certain amino acids (11) and embryo viability was observed. However, many of these technologies are either expensive or slow. Moreover, these technologies arbitrarily pick up some metabolites from the complete array of small-molecule metabolites that constitutes the metabolome and reflects the

functional phenotype. Another technical challenge is that a small volume of culture medium (25 μ l) is typically used for embryo cultivation in which only 10-15 μ l is available for PGT-A.

Raman spectroscopy is a non-invasive and label-free tool to produce metabolic profiles of culture media. Raman spectra of embryo culture media usually contain rich and intrinsic biological information of nucleic acids, protein, carbohydrates and lipids (16), reflecting phenotypes and physiological states of cells. Raman spectroscopy has been shown to have a potential to identify the reproductive potential and/or the health of embryos (17). Due to the potential risk of laser damage to the embryo, many studies were performed on animal embryos instead of human embryos. Examples include application of Raman spectroscopy to identify the cataract in chick-embryo (18), the abnormal ossification of chick embryos caused by metals (19), ALVAC virus infected chicken embryos (20) and the existence of cytoplasmic polarization in the murine egg (21). For the reproductive potential assessment of human embryos, Raman spectroscopy was used for spent embryo medium and was proposed as a useful tool to complement current embryo assessment strategies (17, 22-26). Seli et al. used Raman spectra profile of embryo medium to predict the viability of individual embryos with a sensitivity of 86% and a specificity of 76.5% (25). The overall specificity for predicting a negative result was further increased to 80% in a later study (17). Among those studies, the relative ratios of sodium pyruvate/albumin and phenylalanine/albumin (26), and bands at 902 cm^{-1} and 943 cm^{-1} (24) extracted from Raman spectra could be potential indicators in the process of embryo screening (9, 12). Yet few studies have explored whether chromosomal abnormalities are associated with metabolic changes in embryo culture media.

The hypothesis is that the metabolic footprints of embryo culture media between normal and abnormal embryos are different, which are detectable by Raman spectroscopy coupled with traditional chemometric methods (principal component analysis) and advanced machine learning methods. Machine learning applies different mathematical algorithms to resolve the high-dimensional Raman dataset and enables automated and more accurate classification. Machine learning-Raman approaches have been used in many biological applications, especially for diagnosis purposes (27). In this study, we applied Raman spectroscopy to detect changes in metabolic footprints caused by chromosomal abnormalities in embryo culture media. This novel technique can potentially avoid transfer of aneuploid embryos and benefit patients from expensive and laborious assessment of aneuploidy risk (2, 6, 7, 9, 11-13, 17). Furthermore, by applying machine learning algorithms to build classification models based on Raman

spectra, Raman spectroscopy shows potential of establishing a diagnostic platform, providing rapid, non-invasive and affordable cytogenetic assessment of embryo health.

MATERIALS & METHODS

The ploidy status of embryos was analysed by Raman-based footprint profiling of spent culture media and compared to the results from PGT-A testing performed by NGS. All embryos in this study were cultured in the same media.

Ethics statement. This study was approved by the Reproductive Medicine Ethics Committee of Reproductive Hospital Affiliated to Shandong University. Written informed consent was obtained from all patients (Supplementary Information).

All patients participating in this study were treated at Reproductive Hospital affiliated to Shandong University. The age and ethnic group of the patients are listed in Supplementary Table S1 and S3.

Embryo cultivation and blastocyst development. On day 2 after fertilization, a micro-droplet dish was prepared with 10 of 25 μ L droplets of G-2 Plus medium covered with 3 mL OVOIL oil (Vitrolife, Ltd.). The samples were pre-equilibrated at 37 °C with 6% CO₂ for more than 6 hours. On day 3, the embryos were extensively washed in equilibrated G-2 PLUS droplets. They were then transferred into G-2 PLUS culture droplets, one embryo per droplet. In each case, 10 droplets were prepared. They were cultivated at 37 °C with 6% CO₂. On day 5 or 6, the blastocyst was frozen before transferring into patient's womb.

Sample collection. After removal of the embryos, 15 – 20 μ L of the spent embryo culture medium (G2; Vitrolife Sweden) was collected in labeled EP tubes and was immediately frozen and stored at – 80°C for Raman analysis. For genetic diagnosis, cells from the outermost of the blastocysts were sampled, and their genomic DNA was extracted and amplified by common commercial whole genome amplification (WGA) kits - PicoPLEXWGA Kit (Rubicon Genomics, U.S.A., version R30050). The WGA products were fragmented to an average size of 170 bp and end ligated with sequencing adaptors carrying identifiable barcode for Ion Proton platform (Basecare Ltd, China). Then the copy number variation of each embryo at day 5 or 6 was detected according to preimplantation genetic testing for aneuploidy (PGT-A) protocol by using next-generation sequencing (NGS) (28). All sequencing data was aligned to the human genome reference sequences (version NCBI Build37/hg19) by TMAP software. PCR

duplications were then removed by Picard software and uniquely mapped with reads. Reads with mapping quality score (MAPQ) ≥ 10 and length ≥ 35 bp were kept for further analysis. For each sample, all chromosomes were divided into segments with a 20 kbp window. The number of uniquely mapped reads and GC content in each window were determined. Windows containing no reads and windows with 'N's in the reference sequences were filtered and removed. The number of uniquely mapped reads in the remaining windows was corrected and normalized based on their GC content by LOESS regression. Finally, taking corresponding samples from normal embryo samples as controls, circular binary segmentation (CBS) algorithm was applied to segment the genome into regions with equal copy numbers, and screened out copy number variations (CNVs) with a mean of the log of the normalized read count less than -0.51 or more than 0.37, and a size more than 4 Mbp.

To establish Raman analysis protocol and biological interpretation, 87 spent embryo culture medium samples were collected. According to the PGT-A results (Supplementary Table S1), all of these samples were classified as either chromosome euploidy ($n = 54$) or chromosome aneuploidy ($n = 33$). Each sample was measured with 5 replicates in Raman spectroscopy which yielded 220 euploidy spectra and 165 aneuploidy spectra. To validate the protocol and machine learning algorithm, additional 123 spent culture medium samples were collected for Raman analysis. According to the PGT-A results (Supplementary Table S3), 61 of these samples were assigned to chromosome euploidy and 62 to chromosome aneuploidy. Each sample was measured in 9 replicates in Raman spectroscopy which yielded 549 euploidy spectra and 558 aneuploidy spectra. The replicate number was tested to achieve a better result of machine learning classification. In this study, samples with mosaic embryos were excluded, in order to assess Raman detection results.

Raman spectroscopy. Raman spectra were collected using a Raman spectroscopy system (BaseCare Raman 200, China) with a 785 nm laser and 320 mW of power. The size of the laser spot was 100 μm . Spent embryo culture medium samples from $-80\text{ }^{\circ}\text{C}$ freezer were thawed at room temperature ($25\text{ }^{\circ}\text{C}$) for 30 minutes before analysis (25). There was no observable difference in Raman spectra between the fresh samples and the samples stored at $-80\text{ }^{\circ}\text{C}$. To avoid sampling oil, a specifically designed sampler was carefully passed through the oil layer and directly sample 7 μL of the spent embryo culture medium, which then was dropped onto a quartz glass slide coated with a thin layer (100 nm) of gold film. Raman signals were obtained in standard mode using a charge-coupled device (CCD) camera with an integration time of 20 s. The spectra were acquired in a spectral range from 65 to 3200 cm^{-1} with a spectral resolution

of 1 cm⁻¹. The background control spectrum was subtracted from each sample spectrum. Five or nine replicates were taken for each aliquot.

The process of Raman analysis was 5 minutes per sample and the cost was \$20 per sample.

Data processing and analysis of Raman spectra. All spectra were pre-processed by subtracting the dark signal background from each spectrum. Spectral normalization was done by vector normalization of the fingerprint region from 600 – 1800 cm⁻¹ using LabSpec 6 (Horiba). Mean-centered Raman spectra were obtained by subtracting the means of the spectra from all samples of the individuals. Data analysis, statistics and visualization were done under R environment using in-house scripts. The region between 900 – 1500 cm⁻¹ was used for unsupervised PCA analysis. Quantification of metabolite concentration was done by integrating individual Raman bands and presented as box plots. The rectangle in the box plots represents the second and third quartiles with a line inside representing the median. The lower and upper quartiles are drawn as lines outside the box. Sample means were compared by using Welch's two sample t-test for unequal variances where * indicates statistical significance (* represents $p < 0.05$, ** represents $p < 0.01$ and *** represents $p < 0.001$).

Machine learning stacking classification. Machine learning models were implemented to classify embryo euploidy or aneuploidy based on original Raman spectra of embryo culture medium. A total of 1017 spectra (549 for euploidy and 558 for aneuploidy) was split into the train set and the test set with a ratio of 8:2. The train set was used to train a classification model and the test set was used to evaluate the model performance. Firstly, 3 machine learning algorithms were used separately to classify 1017 spectra (549 for euploidy and 558 for aneuploidy): k-Nearest Neighbors (kNN), Random Forests (RF) and Extreme Gradient Boosting (XGB). kNN is a non-parametric and simple algorithm based on calculating Euclidean distance and feature similarity. RF is an ensemble learning method by combining a number of decision trees. XGB is a boosting technique to achieve better classification results by combining several weak classifiers. Secondly, 3 models were stacked together to build a two-layer machine learning model for a better classification result. The prediction outcomes of the three models (kNN, RF and XGB) from the first layer were used as features for the second layer which employs XGB algorithm. Performance measures were computed as precision, sensitivity and F1 score for each class as well as an overall accuracy rate. Specificity was not used in the context as it is the true negative rate in complement of the sensitivity measure, therefore only one of them has been calculated. All models were constructed under a Python 3.3 environment.

$$\text{Precision} = \text{TP} / (\text{TP} + \text{FP})$$

$$\text{Sensitivity} = \text{TP} / (\text{TP} + \text{FN})$$

$$\text{F1 score} = 2 \times (\text{Sensitivity} \times \text{Precision}) / (\text{Sensitivity} + \text{Precision})$$

$$\text{Accuracy} = (\text{TP} + \text{TN}) / (\text{TP} + \text{TN} + \text{FP} + \text{FN})$$

where TP is the number of true positives, FP is the number of false positives, TN is the number of true negatives and FN is the number of false negatives.

RESULTS & DISCUSSION

Raman spectral analysis unveils metabolic differences in culture media of embryo with normal and abnormal chromosomes.

Figure 1a shows averaged Raman spectra of 87 spent culture medium samples from embryos with normal (euploidy) and abnormal (aneuploidy) chromosomes, which have been identified by PGT-A (Supplementary Table 1). Raman spectra of 33 aneuploidy (165 spectra) and 54 euploidy (270 spectra) were acquired. Compared with the media control, embryo culture media from euploidy and aneuploidy have significantly changed due to metabolic activities of embryos (Figure 1a). The fingerprint region from 600 – 1800 cm^{-1} (16) was used for analysis, which typically contains the most essential biological information (Figure 1a). It shows that Raman spectra of the aneuploidy samples have higher standard deviation/variation than those of the euploidy samples, due to the abnormal metabolism and cellular processes that result from the abnormal chromosomal content of the cells, with either extra or missing copies of chromosome (Supplementary Table S1 and Figure 1a). Abnormal metabolic profiling in aneuploidy samples could be caused by many possibilities, such as missing, additional and mutated chromosomes which change the diverse metabolic pathways. In the contrast, normal euploidy samples have normal and defined metabolic pathways.

Raman spectra from all samples were then averaged and subtracted from each sample at each wavenumber, and the mean values were assembled for each group to obtain mean-centered spectra (Figure 1b). Opposite directions of peak distribution in aneuploidy group and euploidy group were found in the region of 607 – 1790 cm^{-1} . It suggests that the changes in metabolic profiles of aneuploidy group and euploidy group are in different patterns. This result is in a good agreement with a previous study of embryos using mean-centered Raman spectra of embryo culture media (17).

Unsupervised principal component analysis (PCA) was then applied to all spectra to reduce the high-dimensional Raman dataset and transform data into appropriate variables that confer biological information. A spectral range of $900 - 1500 \text{ cm}^{-1}$ was used to minimize the effect of uneven baseline. Figure 1c illustrates the PCA results on dimension 1 and 3 visibly showing two clusters of aneuploidy and euploidy with some overlap of data points. Among the first batch of 87 embryos, 40 embryos developed to blastocyst stage on the fifth day, and 47 samples developed to blastocyst stage on the sixth day. So approximately the embryos from day 5 and day 6 were evenly distributed in this experiment. From figure 1c in the paper, the PCA results on dimension 1 and 3 visibly showed 54 euploid embryos and 33 aneuploid embryos with 13 overlap of misdiagnosed embryos. Of 54 euploid embryos, 46 Raman results were consistent with the PGS results and 8 Raman results (5 from day 5, 3 from day 6) were misdiagnosed according to PGS results (Table S2). Of 33 aneuploid embryos, 28 Raman results were consistent with PGS results and 5 Raman results (2 from day 5, 3 from day 6) were misdiagnosed according to PGS results (Table S2). Overall, the embryos that the Raman method misdiagnosed were the same embryos that develop into the blastocyst stage. The blastocyst of day 5 and day 6 that were diagnosed by Raman did not show obvious time-dependent results and were evenly distributed in the experiment.

This overlap can also be explained by the close similarity between human embryos as well as the limitation generated by a two-dimensional PCA plot. As one Raman spectrum contains over 1000 features, two principal components might be incapable to cover all variances. Therefore, we then sought to find biomarkers which can best separate the two groups and give diagnostic advices. As the difference between groups is most significant along dimension 3, a loading plot was made to assign each Raman wavenumber to indicate significant Raman bands along dimension 3 (Figure 1d). Three Raman zones were identified for describing the largest variances between two datasets at $967 - 1015 \text{ cm}^{-1}$, $1229 - 1295 \text{ cm}^{-1}$, and $1400 - 1430 \text{ cm}^{-1}$ (Figure 1d), some of which can be assigned to ribose-phosphate, -C-O- stretching (981 cm^{-1}), ring stretching (1001 cm^{-1}), U/C ring stretching (1249 cm^{-1}) and lipids CH_2 vibration (1420 cm^{-1}) (29, 30). Interestingly, the overall pattern of PCA loading at dimension 3 is similar to the Raman spectrum of small RNA (supplementary Figure S1).

In order to quantify the concentrations of these metabolites in embryo culture medium, each of the Raman bands was integrated and compared between groups. Figure 2 elucidates the quantification results in boxplots, showing significant differences between aneuploidy and euploidy groups in all three Raman positions ($p < 0.001$). The integration at the $967 - 1015$

cm⁻¹ was found to be higher in samples with normal chromosomes (mean = 1.61, s.d. = 0.03) compared to that with abnormal chromosomes (mean = 1.53, s.d. = 0.20) ($p = 1.1 \times 10^{-6}$). Quantification of the 1229 – 1295 cm⁻¹ on the other hand showed a higher content in medium of embryo aneuploidy (mean = 5.11, s.d. = 1.06) than that of euploidy (mean = 4.71, s.d. = 0.73) ($p = 2.7 \times 10^{-5}$). Lipid concentration, represented by the area at 1400 – 1430 cm⁻¹, was higher in the aneuploid group (mean = 1.91, s.d. = 0.51) compared to the euploid group (mean = 1.75, s.d. = 0.28) ($p = 0.00026$).

Quantification of all three Raman biomarkers reveals significant differences in the metabolism related to the variations of chromosomal abnormalities in embryos, which is consistent with a previous research showing the relationship of the metabolism, the cytogenetic composition, and the health of human embryos *in vitro* (31). Importantly, to our best knowledge, this is the first report to explore the effects of chromosomal abnormalities on metabolic changes in human embryo culture medium. Changes in gene copy number and perturbation of nuclear gene dosage can lead to altered regulation of transcription (32). Moreover, small RNA with size from 18 to 32 nucleotides could be secreted into embryo culture media and play essential roles in signaling, regulation and development of embryo (33, 34). Further study may include complementary metabolomics using traditional bulk chemistry analysis, for example mass spectrometry, to map with and consolidate the Raman results.

Machine learning models based on Raman spectra classify aneuploidy and euploidy samples with high performance.

Besides feature extraction to find informative bio-variables, classification of samples based on their spectra is often desirable for diagnostic purposes. An additional 549 euploid spectra and 558 aneuploid spectra were collected for classification purposes. The information of samples ID, replicate number and day of culturing was added as features to train machine learning models. Three models were first trained individually using k-Nearest Neighbors (kNN), Random Forests (RF) and Extreme Gradient Boosting (XGB). Then these three models were combined together to train a stacking classification model to achieve a better accuracy. The stacking model as well as the three classifiers were evaluated by calculating precision, sensitivity, F1 score and accuracy. Intuitively, precision is the ability of a classifier not labelling a negative sample as positive. Sensitivity is the ability of a classifier to find all the positive samples in one class. F1 score is a weighted mean of the precision and the sensitivity, where it reaches its best value 1 and worst 0. Because F1 score takes both false positives and false negatives into account, it is more useful to evaluate performance in one class.

Table 1 shows the results of kNN, RF and XGB, and the stacking analysis based on kNN, RF and XGB. The first model employed kNN algorithm which is a simple machine learning classifier based on Euclidean distances between individuals. With this classifier alone, 117 out of 125 chromosome euploidy (F1 score 0.951) and 93 out of 97 chromosome aneuploidy (F1 score 0.939) were assigned correctly. RF classifier is an ensemble learning technique by constructing a number of Decision Trees in which each tree gives a classification and the forest chooses the classification that is voted by most trees. RF classifier was tested and evaluated with a higher sensitivity of 79.4% towards classifying aneuploidy group (77 out of 97) than a sensitivity of 57.6% towards classifying euploidy group (72 out of 125). XGB is a boosting learning technique which is built on an ensemble of weak classifiers, optimized by an arbitrary differentiable loss function. The XGB model was able to correctly assign 116 out of 125 euploidy spectra and 88 out of 97 aneuploid spectra, with an F1 score of 92.8% and 90.7% respectively (Table 1). The stacking classification includes two layers with the first layer kNN, RF and XGB and the second layer XGB. It should be noted that RF was included in the stacking model despite its poor stand-alone performance because we hoped to combine algorithms based on different approaches i.e. numeric distance, binary decision and advanced boosting. The stacking model demonstrated a further boost of the performance and total accuracy rate compared with the three individual classifiers. It assigned correctly 121 out of 125 euploid spectra with an F1 score of 0.964 and 92 out of 97 aneuploid spectra with an F1 score of 0.953 (Table 1).

In terms of three models with a single classifier, kNN achieves a higher accuracy score of 94.6% compared to 67.1% by RF and 91.9% by XGB. Essentially, kNN outperforms other classifiers in the precision score classifying the euploid group and the sensitivity score classifying the aneuploid group. This is especially important because these two scores include the impact of misclassified aneuploid group. In practice, this will reduce the chance of misclassifying embryos with abnormal number of chromosomes as healthy embryos and being transferred back to the womb. Except for the kNN model, all other models achieved higher sensitivity assigning normal chromosomes, presumably due to high heterogeneity existed in abnormal chromosomes as in the PCA result. Compared to single classifiers, the stacking model has similar ability of correctly assigning two groups (Table 1). It also exceeds previous models in the F1 scores of both groups (0.964 and 0.953) with a total accuracy rate of 95.9%.

Previous metabolomics profiling using near infrared (NIR) spectroscopy for embryo viability selection has been shown unsuccessful in randomized controlled trials (35, 36). Presumably, it

could be caused by several technical reasons such as the sensitivity of Raman system, the quality of the spectra and, perhaps more importantly, the choice of the machine learning algorithm and insufficient training dataset. In addition, our study focused on a specific abnormality of embryo, aneuploidy, which can reduce complexity and heterogeneity. In this study, the quality of the spectra was good so that unsupervised chemometrics can be used to identify nucleic acids (e.g. RNA) contributing to the different metabolomic profiles of aneuploid and euploid embryos. With a relatively large training dataset and advanced algorithm of machine learning, the screening accuracy was improved. It is conceivable that in the future this approach could be more robust with high quality of Raman spectra, ever-improving and advanced algorithm, and an increasing size of training dataset.

By applying different machine learning algorithms to different datasets, the analysis of Raman spectra allows us to screen chromosomal quality based on changes in metabolites using small volume of culture medium (7 μ L). Moreover, with an increasing number of Raman spectra to construct the reference database, classification models can be improved to be more sensitive and robust. Within seconds after measuring a Raman spectrum of embryo culture medium, the embryo could be potentially classified as either healthy or unhealthy. This will bring advantages to feasibility in clinical practice complementary to current PGT-A method. Clinical implementation of this methodology, however, would require further study to confirm the current results (e.g. by a randomized controlled trial) and to test the robustness of the algorithm in terms of differences in culture media, gene copy numbers or other aspects of embryological practice.

CONCLUSIONS

To determine the effects of chromosomal abnormalities on embryo metabolism and developmental potential, full genomic, transcriptomic and proteomic analyses of each individual embryo are required. As clinical practice requires rapid screening with small volume of culture media (1-7 μ L), we developed a non-invasive and label-free strategy to use Raman spectroscopy to detect metabolic footprints of embryo culture medium. Analysis based on Raman spectra showed significant differences in metabolic profiles of embryo culture medium between aneuploidy and euploidy groups which were validated by PGT-A sequencing analysis. Machine learning models successfully classified spectra from two groups with high sensitivity and accuracy. Comparing to the current PGT-A methods, Raman measurements are cheap, easy

to use, fast and non-invasive. Moreover, the results can be obtained immediately with very small sample volumes. In summary, our data suggested that the non-invasive method of Raman microspectroscopy may be a useful tool to detect and predict the developmental potential of human embryos *in vitro* and to provide valuable information to the cytogenetic assessment of reproductive potential and embryo health. It provides considerable insights into embryo cytogenetic health and may improve embryo classification schemes and benefit pregnancy rates.

Conflicts of Interest: B.L. has nothing to disclose. Y.G. has nothing to disclose. J.X. has nothing to disclose. Y.S. has nothing to disclose. L.X. has nothing to disclose. T. S has nothing to disclose. N.W. has nothing to disclose. Z.H. has nothing to disclose. Z.J.C. has nothing to disclose. Y.L.Z has nothing to disclose. W.E.H. has nothing to disclose.

ACKNOWLEDGEMENT

This work was supported by the National Science Foundation of China (Grant number 21377085 and 31770070), SJTU Med-Eng fund (YG2016NS33), and HPC computing facility award. WEH acknowledges support from EPSRC (EP/M002403/1 and EP/M02833X/1) and NERC (NE/M002934/1) in the UK.

SUPPORTING INFORMATION

Patient and PGT-A results are included in supplementary Table 1, 2 and 3. Ethics form is in supplementary information.

REFERENCE

- Findikli N, Kahraman S, Saglam Y, Beyazyurek C, Sertyel S, Karlikaya G *et al.* Embryo aneuploidy screening for repeated implantation failure and unexplained recurrent miscarriage. *Reprod Biomed Online*. 2006;13:38-46.
- Wang Y, Chen Y, Tian F, Zhang J, Song Z, Wu Y *et al.* Maternal mosaicism is a significant contributor to discordant sex chromosomal aneuploidies associated with noninvasive prenatal testing. *Clin Chem*. 2014;60:251-9.
- <https://www.nhs.uk/conditions/ivf/>. Accessed on Aug 2 2018.

4. Demko ZP, Simon AL, McCoy RC, Petrov DA, Rabinowitz M. Effects of maternal age on euploidy rates in a large cohort of embryos analyzed with 24-chromosome single-nucleotide polymorphism-based preimplantation genetic screening. *Fertil and Steril*. 2016;105:1307-13.
5. Zhou S, Cheng D, Ouyang Q, Xie P, Lu C, Gong F *et al*. Prevalence and authenticity of de-novo segmental aneuploidy (>16 Mb) in human blastocysts as detected by next-generation sequencing. *Reprod Biomed Online*. 2018.
6. Chen M, Wei S, Hu J, Quan S. Can Comprehensive Chromosome Screening Technology Improve IVF/ICSI Outcomes? A Meta-Analysis. *PloS one*. 2015;10:e0140779.
7. Cimadomo D, Capalbo A, Ubaldi FM, Scarica C, Palagiano A, Canipari R *et al*. The Impact of Biopsy on Human Embryo Developmental Potential during Preimplantation Genetic Diagnosis. *BioMed Res Int*. 2016;2016:7193075.
8. Munne S, Weier HUG, Grifo J, Cohen J. Chromosome mosaicism in human embryos. *Biol Reprod*. 1994;51:373-9.
9. Botros L, Sakkas D, Seli E. Metabolomics and its application for non-invasive embryo assessment in IVF. *Mol Hum Reprod*. 2008;14:679-90.
10. Allen J, Davey HM, Broadhurst D, Heald JK, Rowland JJ, Oliver SG *et al*. High-throughput classification of yeast mutants for functional genomics using metabolic footprinting. *Nature Biotechnol*. 2003;21:692-6.
11. Brison DR, Houghton FD, Falconer D, Roberts SA, Hawkhead J, Humpherson PG *et al*. Identification of viable embryos in IVF by non-invasive measurement of amino acid turnover. *Hum Reprod*. 2004;19:2319-24.
12. Houghton FD, Hawkhead JA, Humpherson PG, Hogg JE, Balen AH, Rutherford AJ *et al*. Non-invasive amino acid turnover predicts human embryo developmental capacity. *Hum Reprod*. 2002;17:999-1005.
13. Zhao Q, Yin T, Peng J, Zou Y, Yang J, Shen A *et al*. Noninvasive metabolomic profiling of human embryo culture media using a simple spectroscopy adjunct to morphology for embryo assessment in in vitro fertilization (IVF). *Int J Mol Sci*. 2013;14:6556-70.
14. Conaghan J, Hardy K, Handyside AH, Winston RML, Leese HJ. Selection Criteria for Human Embryo Transfer - a Comparison of Pyruvate Uptake and Morphology. *J Assist Reprod Gen*. 1993;10:21-30.
15. Gardner DK, Lane M, Stevens J, Schoolcraft WB. Noninvasive assessment of human embryo nutrient consumption as a measure of developmental potential. *Fertil Steril*. 2001;76:1175-80.
16. Li MQ, Xu J, Romero-Gonzalez M, Banwart SA, Huang WE. Single cell Raman spectroscopy for cell sorting and imaging. *Curr Opin Biotechnol*. 2012;23:56-63.
17. Scott R, Seli E, Miller K, Sakkas D, Scott K, Burns DH. Noninvasive metabolomic profiling of human embryo culture media using Raman spectroscopy predicts embryonic reproductive potential: a prospective blinded pilot study. *Fertil Steril*. 2008;90:77-83.
18. Mizuno A, Nishigori H, Iwatsuru M. Glucocorticoid-Induced Cataract in Chick-Embryo Monitored by Raman-Spectroscopy. *Invest Ophth Vis Sci*. 1989;30:132-7.
19. Stahler AC, Monahan J, Markopoulos M, Dagher J, Baker J, Baker J *et al*. Investigating the abnormal ossification of tibiotarsi of chick embryos exposed to sublethal doses of platinum group metals through micro-Raman spectroscopy. *Abstr Pap Am Chem S*. 2012;244.
20. Misra AK, Kamemoto LE, Hu NJ, Dykes AC, Yu QG, Zinin PV *et al*. Micro-Raman Spectroscopy study of ALVAC virus infected chicken embryo cells. *Proc Spie*. 2011;8025.
21. Davidson B, Spears N, Murray A, Elfick A. The changing biochemical composition and organisation of the murine oocyte and early embryo as revealed by Raman spectroscopic mapping. *J Raman Spectrosc*. 2012;43:24-31.
22. Ding JY, Xu T, Tan XF, Jin H, Shao J, Li HB. Raman spectrum: A potential biomarker for embryo assessment during in vitro fertilization. *Exp Ther Med*. 2017;13:1789-92.

23. dos Santos EC, Martinho H, Annes K, da Silva T, Soares CA, Leite RF *et al.* Raman-based noninvasive metabolic profile evaluation of in vitro bovine embryos. *J Biomed Opt.* 2016;21.
24. Parlattan U, Basar G, Bavili N, Yumru H, Sag F, Bulgurcuoglu S *et al.* Embryo viability indexing using Raman spectroscopy of spent culture media. *Spectrosc Lett.* 2016;49:458-63.
25. Seli E, Sakkas D, Scott R, Kwok SC, Rosendahl SM, Burns DH. Noninvasive metabolomic profiling of embryo culture media using Raman and near-infrared spectroscopy correlates with reproductive potential of embryos in women undergoing in vitro fertilization. *Fertil Steril.* 2007;88:1350-7.
26. Shen AG, Peng J, Zhao QH, Su L, Wang XH, Hu JM *et al.* Accurate and noninvasive embryos screening during in vitro fertilization (IVF) assisted by Raman analysis of embryos culture medium. *Laser Phys Lett.* 2012;9:322-8.
27. Butler HJ, Ashton L, Bird B, Cinque G, Curtis K, Dorney J *et al.* Using Raman spectroscopy to characterize biological materials. *Nat Protoc.* 2016;11:664-87.
28. Zhang X, Liang B, Xu X, Zhou F, Kong L, Shen J *et al.* The comparison of the performance of four whole genome amplification kits on ion proton platform in copy number variation detection. *Biosci Rep.* 2017;37.
29. Hobro AJ, Rouhi M, Blanch EW, Conn GL. Raman and Raman optical activity (ROA) analysis of RNA structural motifs in Domain I of the EMCV IRES. *Nucleic Acids Res.* 2007;35:1169-77.
30. Wang Y, Huang WE, Cui L, Wagner M. Single cell stable isotope probing in microbiology using Raman microspectroscopy. *Curr Opin Biotechnol.* 2016;41:34-42.
31. Picton HM, Elder K, Houghton FD, Hawkhead JA, Rutherford AJ, Hogg JE *et al.* Association between amino acid turnover and chromosome aneuploidy during human preimplantation embryo development in vitro. *Mol Hum Reprod.* 2010;16:557-69.
32. Veitia RA, Bottani S, Birchler JA. Cellular reactions to gene dosage imbalance: genomic, transcriptomic and proteomic effects. *Trends Genet.* 2008;24:390-7.
33. Kropp J, Salih SM, Khatib H. Expression of microRNAs in bovine and human pre-implantation embryo culture media. *Front Genet.* 2014;5.
34. Suh N, Blelloch R. Small RNAs in early mammalian development: from gametes to gastrulation. *Dev.* 2011;138:1653-61.
35. Hardarson T, Ahlstrom A, Rogberg L, Botros L, Hillensjo T, Westlander G *et al.* Non-invasive metabolomic profiling of Day 2 and 5 embryo culture medium: a prospective randomized trial. *Hum Reprod.* 2012;27:89-96.
36. Vergouw CG, Heymans MW, Hardarson T, Sfontouris IA, Economou KA, Ahlstrom A *et al.* No evidence that embryo selection by near-infrared spectroscopy in addition to morphology is able to improve live birth rates: results from an individual patient data meta-analysis. *Hum Reprod.* 2014;29:455-61.

Table 1. Confusion matrix and performance evaluation of kNN, RF, XGB and stacking classification model for an independent test set of 222 Raman spectra. Models were trained from a training set of 885 Raman spectra. kNN: k-Nearest Neighbors, RF: Random Forests, XGB: Extreme Gradient Boosting. Stacking analysis is based on a first layer of kNN, RF and XGB and a second layer of XGB.

	Confusion Matrix			Performance evaluation			
	Actual class	Predicted Euploidy	Predicted Aneuploidy	Precision	Sensitivity	F1 score	Accuracy
kNN	Euploidy	117	8	96.7%	93.6%	0.951	94.6%
	Aneuploidy	4	93	92.1%	95.9%	0.939	
RF	Euploidy	72	53	78.2%	57.6%	0.664	67.1%
	Aneuploidy	20	77	59.2%	79.4%	0.678	
XGB	Euploidy	116	9	92.8%	92.8%	0.928	91.9%
	Aneuploidy	9	88	90.7%	90.7%	0.907	
Stacking	Euploidy	121	4	96.0%	96.8%	0.964	95.9%
	Aneuploidy	5	92	95.8%	94.9%	0.953	

503 **Figure captions:**

504

505

506 **Figure 1.** (a) Averaged Raman spectra of 87 spent embryo culture medium samples with
507 known PGT-A results, classified as either aneuploidy (n = 33, 165 spectra in total) or euploidy
508 (n = 54, 270 spectra in total, and culture medium background from 30 replicates. Shaded area
509 represents the standard deviation of the spread in all measurements. The relative intensities of
510 the spectra were manually separated to aid visual presentation. (b) Mean-centered Raman
511 spectra by subtracting means from all samples, showing differences between aneuploidy group
512 and euploidy group. (c) PCA plot of all Raman spectra showing clustering of either aneuploidy
513 or euploidy. (d) Raman wavenumber loading plots of contributions along dimension 3 of the
514 PCA, where * indicates wavenumbers responsible for the highest three contributions.

Figure 2. Boxplots of Raman band integration of aneuploidy and euploidy samples, showing significant differences at 975 – 993 cm^{-1} ($p = 1.1 \times 10^{-6}$), 1229 – 1295 cm^{-1} ($p = 2.7 \times 10^{-5}$), and 1400 – 1430 cm^{-1} ($p = 0.00026$), respectively.

521 **Supplementary figure**

524 **Supplementary Figure 1.** The PCA loading plot at dimension 3 (d3 PCA loading) shows a
525 similar spectral pattern to the Raman spectrum of small RNA.