

Defining Human Embryo Phenotypes by Cohort-Specific Prognostic Factors

Sunny H. Jun^{1,3}, Bokyoung Choi^{2,3}, Lora Shahine¹, Lynn M. Westphal¹, Barry Behr¹, Renee A. Reijo Pera^{1,3}, Wing H. Wong⁴, Mylene W. M. Yao^{1*}

1 Department of Obstetrics and Gynecology, Stanford University School of Medicine, Stanford, California, United States of America, **2** Department of Applied Physics, School of Humanities and Sciences, Stanford University, Stanford, California, United States of America, **3** Center for Human Embryo and Embryonic Stem Cell Research and Education, Institute for Stem Cell Biology & Regenerative Medicine, Stanford University, California, United States of America, **4** Department of Statistics, School of Humanities and Sciences, Stanford University, Stanford, California, United States of America

Abstract

Background: Hundreds of thousands of human embryos are cultured yearly at *in vitro* fertilization (IVF) centers worldwide, yet the vast majority fail to develop in culture or following transfer to the uterus. However, human embryo phenotypes have not been formally defined, and current criteria for embryo transfer largely focus on characteristics of individual embryos. We hypothesized that embryo cohort-specific variables describing sibling embryos as a group may predict developmental competence as measured by IVF cycle outcomes and serve to define human embryo phenotypes.

Methodology/Principal Findings: We retrieved data for all 1117 IVF cycles performed in 2005 at Stanford University Medical Center, and further analyzed clinical data from the 665 fresh IVF, non-donor cycles and their associated 4144 embryos. Thirty variables representing patient characteristics, clinical diagnoses, treatment protocol, and embryo parameters were analyzed in an unbiased manner by regression tree models, based on dichotomous pregnancy outcomes defined by positive serum β -human chorionic gonadotropin (β -hCG). IVF cycle outcomes were most accurately predicted at ~70% by four non-redundant, embryo cohort-specific variables that, remarkably, were more informative than any measures of individual, transferred embryos: Total number of embryos, number of 8-cell embryos, rate (percentage) of cleavage arrest in the cohort and day 3 follicle stimulating hormone (FSH) level. While three of these variables captured the effects of other significant variables, only the rate of cleavage arrest was independent of any known variables.

Conclusions/Significance: Our findings support defining human embryo phenotypes by non-redundant, prognostic variables that are specific to sibling embryos in a cohort.

Citation: Jun SH, Choi B, Shahine L, Westphal LM, Behr B, et al. (2008) Defining Human Embryo Phenotypes by Cohort-Specific Prognostic Factors. PLoS ONE 3(7): e2562. doi:10.1371/journal.pone.0002562

Editor: Sudhansu Kumar Dey, Vanderbilt University Medical Center, United States of America

Received: December 7, 2007; **Accepted:** May 22, 2008; **Published:** July 2, 2008

Copyright: © 2008 Jun et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: M.W.M.Y. received support from NIH HD057970 and HD01249; W.H.W. received support from NIH GM067250; R.A.R.P. received support from the March of Dimes. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: mylene.yao@stanford.edu

These authors contributed equally to this work.

Introduction

Developmental arrest of human embryos cultured *in vitro* is common and presents a major obstacle to achieving pregnancy through IVF, as well as a major obstacle to research in human embryonic stem cell (hESC) biology [1,2,3,4,5,6,7]. While the culture of embryos to blastocyst stage for subsequent transfer yields high pregnancy rates and minimizes the risk of multiple gestation, the availability of blastocysts is limited even in the best IVF clinics because of the high rates of attrition in *in vitro* embryo culture [1,3,4,8,9,10,11].

Although developmental defects such as cleavage arrest, polyploidy, and fragmentation are commonly encountered and have been used for scoring individual embryos in IVF [11,12,13,14,15,16,17,18,19,20,21], the lack of well-defined human embryo phenotypes has hindered translational research and mechanistic investigations. One key challenge to defining human

embryo phenotypes relates to the unclear and often highly interactive relationships amongst variables pertaining to patient characteristics, clinical infertility diagnoses, IVF treatment protocols, and observed embryo characteristics. Further, since any single couple may typically produce a few oocytes or embryos that are abnormal merely by chance, it is difficult to determine whether sibling embryos as a group, or an embryo cohort, is “normal”. (Note that “embryo cohort” refers to an embryo sibling group from the same couple within the same IVF treatment.) Nonetheless, we envision that the identification of cohort-specific parameters to define human embryo phenotypes is a necessary step towards translational investigations of molecular determinants of developmental competence. Thus, we sought to test the hypothesis that embryo cohort-specific variables have prognostic value in measuring IVF cycle outcomes by identifying non-redundant, prognostic variables in an unbiased manner using regression tree models.

Results

Of all 1117 IVF treatments performed at Stanford University in 2005, 822 were fresh IVF cycles that used the patients' own oocytes (Figure 1A). Based on our exclusion criteria, 157 cycles were excluded for a variety of medical and non-medical reasons (see results in *Supporting Information Text S1* for details). Clinical and embryology data on the remaining 665 cycles that satisfied inclusion and exclusion criteria, and their 4144 embryos, respectively, were analyzed to test the hypothesis that cohort-specific variables predict IVF cycle outcomes (Figure 1). Of those 4144 embryos, the number of blastomeres or cells on day 3 was recorded for 4002 embryos (96.6%). Overall, 38.8% had 8 cells, the developmentally appropriate cell number, while 18.2% of embryos had ≤ 4 cells, and 33.6% had 5–7 cells (Figure 2).

Prognostic Significance and Correlation of Variables

We systematically examined the association of each variable with IVF outcomes, and the correlation of each pair of variables. Pair-wise logistic regression tests confirmed many known prognostic variables, including female age, day 3 FSH, and the number of 8-cell embryos. However, in addition to these known prognostic variables, we observed that cohort-specific variables such as fertilization rate and the rate of cleavage arrest were also significantly associated with IVF cycle outcome ($p < 0.001$; Table 1). In contrast, except for male factor infertility ($p < 0.05$), none of the conventional clinical infertility diagnoses were significantly associated with IVF outcomes. Notably, despite a high degree of correlation between many variables and age or day 3 FSH level, which estimates ovarian aging, neither age nor day 3 FSH level was correlated to cohort-specific embryo parameters (see Table S1). Collectively, these results suggest that determinants other than age-related mechanisms and clinical diagnoses impact cohort-specific embryo developmental competence.

Thresholds of Non-redundant, Prognostic Variables Defining Human Embryo Cohort Phenotypes

Sequential Multiple Additive Regression Tree (MART[®]) and Classification and Regression Tree (CART) analyses of all 30 variables (listed in Table 1 and its legend) determined that IVF cycle outcomes were most accurately predicted at $\sim 70\%$ by using only four non-redundant variables: total number of embryos, rate of cleavage arrest in an embryo cohort, the number of 8-cell embryos in a cohort, and day 3 FSH level. Remarkably, these four variables all describe the embryo cohort rather than individual embryos, and were more informative than age, clinical diagnoses, or any measures of the transferred embryos. Interestingly, the total number of embryos, day 3 FSH, and the number of 8-cell embryos depended on and thus captured the effects of many other variables. In contrast, the rate of cleavage arrest was independent of any of those known variables. (Details on MART[®] and CART analyses are reported in Text S1, and Figure S1)

Of the prognostic thresholds identified, the most robust phenotypes are A1 and A2, and B1 and B2 (Table 2). Number of embryos < 6 or ≥ 6 is used by all 5 top CART models, defines all other phenotypes (B to F), and can be applied to all cases. Specifically, the phenotype defined by having fewer than 6 embryos, has an odds ratio of 3.9 for no pregnancy compared to cases with ≥ 6 embryos (95% Confidence Interval [CI], 2.8 to 5.5). Similarly, the next most robust phenotypes are defined by the number of embryos and cleavage arrest rate, such that for cases with ≥ 6 embryos, those with cleavage arrest rate $> 14.6\%$ are 3.0 times more likely to result in no pregnancy than those with cleavage arrest rate $\leq 14.6\%$ (95% CI, 1.9 to 4.9).

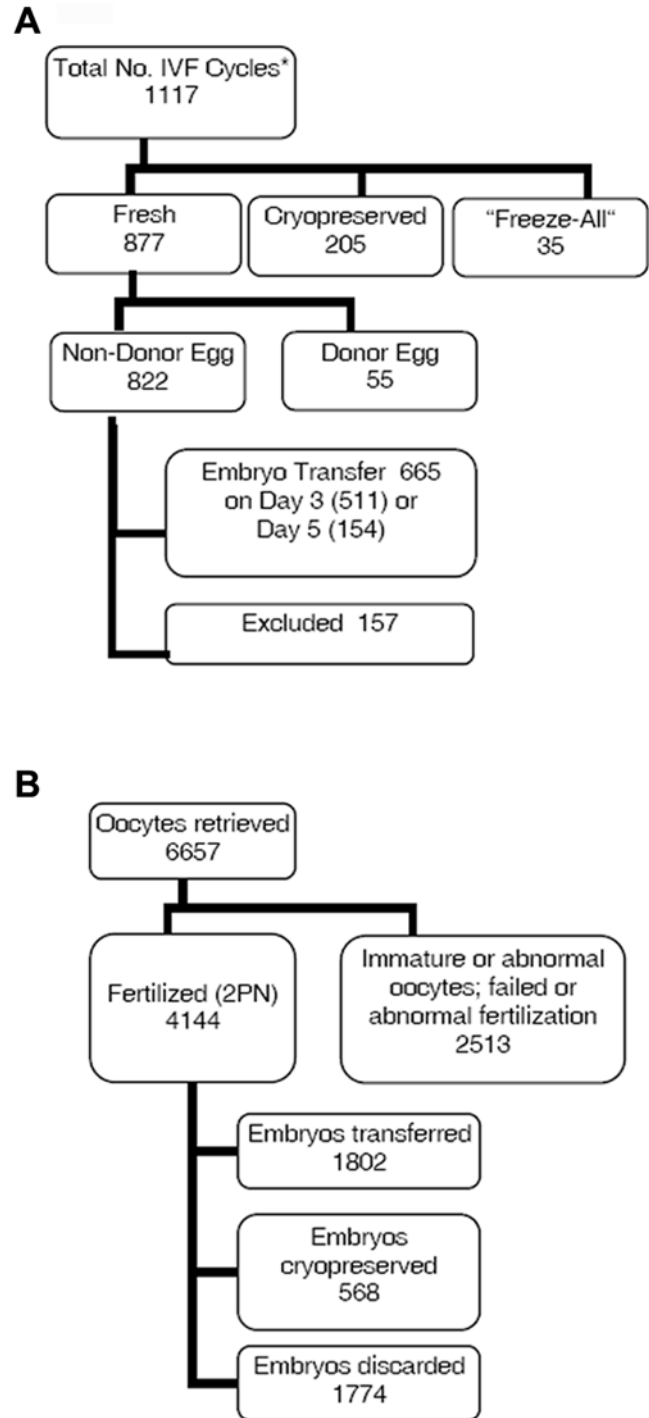


Figure 1. Source of data. A) IVF cycles performed in 2005. B) Utilization of oocytes and embryos in 665 fresh, non-donor IVF cycles. * All numbers in Panel A indicate the number of cycles and numbers in Panel B indicate the number of oocytes or embryos. Fresh cycles are defined by ovarian stimulation of gonadotropins and embryo transfer performed within the same cycle; cryopreserved cycles utilize embryos that were obtained and cryopreserved from a previous cycle; "freeze-all" are cycles in which ovarian stimulation was performed, but embryos were cryopreserved instead of being transferred back within the same cycle for medical or non-medical reasons. 157 cycles were removed from analysis for a variety of medical and non-medical reasons that did not result in fresh embryo transfer (see *S1 Text* for details). doi:10.1371/journal.pone.0002562.g001

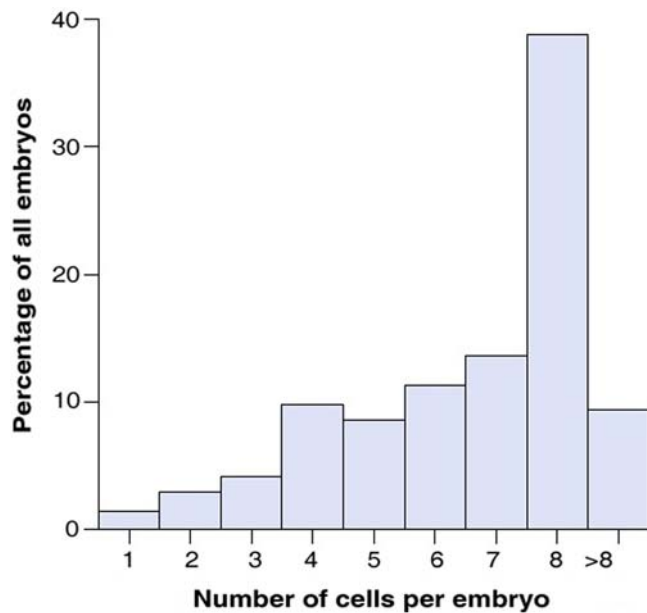


Figure 2. Distribution of all embryos from 665 fresh, non-donor IVF cases according to their cell number on Day 3.
doi:10.1371/journal.pone.0002562.g002

In contrast, the rest of the thresholds listed in Table 2 are used by only 1 CART model each, and is applicable to fewer cases. However, as some of those phenotypes describe very specific subset of cases and have odds ratios that are highly discriminatory, they may be extremely useful depending on the clinical or translational research context. For example, for cases with ≥ 6 embryos, having cleavage arrest rates of 14.6–52.8% and $>52.8\%$ increase the odds of no pregnancy by 2.6 (95% CI 1.6 to 4.3) and 10.6 (95% CI 3.2 to 49.6), respectively, when compared to cases with cleavage rates of $\leq 14.6\%$.

Discussion

Since the introduction of IVF in the 1970s, the major challenges of assisted reproductive technologies (ART) have been the high attrition rates of embryos cultured *in vitro* [1,3,4,8,9,10,11,22,23], the limited value of embryo morphology in predicting developmental competence [24,25,26], and finding criteria to help determine the number of embryos to transfer [22,23]. In addition, the benefit of aneuploidy screening by preimplantation genetic screening (PGS) has recently been refuted [27]. Thus, there is a need to reassess factors that determine human embryo quality.

Our findings represent a first step towards this goal by using regression tree models, MART[®] and CART, as unbiased methods to analyze IVF and embryo data. These methods allowed us to consider and control for a large number of variables, even if only a few of them have significant impact on outcomes. This feature is critical for the analysis of the highly interactive and multicollinear IVF and human embryo data, as arbitrary selection of variables may compromise completeness of data and introduce bias, while including all of them would cause the conventional multivariate regression to breakdown (see *SI Text*). Indeed, such application of CART analysis was taken by Guzik *et al.* to define semen parameters that predicted male infertility [28]. In our study, we further used MART[®], a more powerful statistical method that “boosts” or increases accuracy in the CART method [29,30,31,32].

We identified four non-redundant variables that predict outcomes in the current IVF cycle with $\sim 70\%$ accuracy. Most

Table 1. Association of each variable with pregnancy outcome.

Variables	Estimate*	S.E.	P-Value
Patient Characteristics and Clinical Diagnoses[†]			
Age	-0.10	0.02	2.16E-007
Maximum Day 3 FSH level	-0.08	0.03	1.70E-003
Gravidity	0.036	0.066	5.86E-001
Male Factor (infertility diagnosis)	0.50	0.24	3.71E-002
IVF Cycle Characteristics			
Microdose lupron (flare) protocol	-1.14	0.24	2.53E-006
Antagonist protocol	-0.74	0.19	9.98E-005
Performance of ICSI	-0.15	0.16	3.47E-001
No. of oocytes	0.08	0.01	1.58E-009
Embryo Cohort Parameters			
Fertilization rate	1.24	0.36	5.37E-004
No. of embryos	0.14	0.02	2.67E-012
Average cell no. of embryos	0.29	0.06	6.34E-006
No. of 8-cell embryos	0.26	0.04	2.88E-012
Percentage of 8-cell embryos	0.76	0.28	5.75E-003
Cleavage arrest rate [‡]	-1.28	0.35	2.76E-004
Average grade of embryos	-0.091	0.17	5.88E-001
Parameters of Transferred Embryos			
Day 5 embryo transfer [§]	1.40	0.19	7.51E-013
No. of embryos transferred	0.0058	0.053	9.12E-001
Average cell no. of embryos transferred	0.47	0.07	2.19E-010
Percentage of transferred embryos at the 8-cell stage	1.33	0.21	5.35E-010
No. of 8-cell embryos transferred	0.41	0.08	4.40E-008
No. of embryos with ≤ 4 cells transferred	-2.14	0.49	1.06E-005
Average grade of embryos transferred	-0.52	0.17	2.61E-003

*Positive and negative estimates indicate association with positive and negative pregnancy outcomes, respectively.

[†]Clinical infertility diagnoses that were not significantly associated with pregnancy outcome (p-value >0.05) were not listed: uterine factor, polycystic ovarian syndrome, endometriosis, tubal ligation, tubal disease, hydrosalpinges, unexplained infertility, and “other diagnoses”. Each IVF case may have more than one clinical infertility diagnosis.

[‡]Cleavage arrest rate is defined as the percentage of embryos with 4 or fewer cells on Day 3 of *in vitro* culture.

[§]Day 5 embryo transfer is arbitrarily listed under Parameters of Transferred Embryos. It can also be considered an Embryo Cohort Parameter, as it depends on the total number of embryos and the number of 8-cell embryos.

doi:10.1371/journal.pone.0002562.t001

remarkably, these variables—total number of embryos, cleavage arrest rate, number of 8-cell embryos, and day 3 FSH (in order of relative importance)—describe the entire embryo cohort, and are more predictive than any measures of the transferred embryos. In addition, we show that most prognostic information carried by highly interacting and multicollinear conventional variables such as age and clinical diagnoses, is captured by three of the four variables.

Previous reports mainly focused on the prognostic value of individual embryo scores, in which the relative weighting of score components was determined arbitrarily rather than by objective or statistical methods [15,18]. Further, although individual variables that were significantly related to IVF cycle outcomes were reported, there has been no attempt to compare their relative

Table 2. Prognostic thresholds defining cohort-specific phenotypes.

	Embryos (No.)*	Cleavage Arrest (%)*	8-cell embryo (No.)*	FSH (mIU/mL)*	Pregnancy -No. (%)†	No Pregnancy -No. (%) ‡	Applicable Cases-No. (%) §	No. Trees ¶	Reference Condition	Odds Ratio	95% Confidence Interval (C.I.)
A1**	≥6				177 (57.7)	130 (42.3)	307 (46.2)	5			
A2	<6				92 (25.7)	266 (74.3)	358 (53.8)	5	A1	3.9	(2.8, 5.5)
B1	≥6	≤14.6			112 (70.4)	47 (29.6)	159 (23.9)	4			
B2	≥6	>14.6			65 (43.9)	83 (56.1)	148 (22.3)	4	B1	3.0	(1.9, 4.9)
B3	≥6	14.6–52.8			62 (47.3)	69 (52.7)	131 (19.7)	1	B1	2.6	(1.6, 4.3)
B4	≥6	≤52.8			174 (60.0)	116 (40.0)	290 (43.6)	n/a			
B5	≥6	>52.8			3 (17.6)	14 (82.3)	17 (2.6)	1	B1	10.6	(3.2, 49.6)
									B4	6.7	(2.1, 30.9)
C1	≥6		≥2		157 (63.6)	90 (36.4)	247 (37.1)	1			
C2	≥6		<2		20 (33.3)	40 (66.7)	60 (9.2)	1	C1	3.5	(1.9, 6.4)
D1	≥6	>14.6	≥2		51 (53.1)	45 (46.9)	96 (14.4)	1			
D2	≥6	>14.6	<2		14 (26.9)	38 (73.1)	52 (7.8)	1	D1	3.0	(1.5, 6.5)
E1	≥6	>14.6	≥2	≤4.6	14 (82.4)	3 (17.6)	17 (2.6)	1			
E2	≥6	>14.6	≥2	>4.6	34 (46.9)	37 (53.1)	71 (12.2)	1	E1	4.8	(1.4, 23.4)

*Cohort phenotypes defined by thresholds of non-redundant prognostic variables. Each set of conditions (A–E) use “AND” as the operator where more than one condition is listed.

†No. of cases that satisfy the threshold conditions and have pregnancy outcome. This percentage is calculated by using the No. Applicable Cases as denominator. In general, conditions that discriminate between pregnancy and no pregnancy outcomes more highly are more robust and are expected to be more useful in both clinical management and translational research.

‡No. of cases that satisfy the threshold conditions and have no pregnancy outcome. This percentage is calculated by using the No. Applicable Cases as denominator.

§The No. Applicable Cases is the total number of cases that satisfy the threshold conditions. This percentage is calculated by using the total number of cycles (665) as the denominator. In general, the larger the number of applicable cases, the more useful the set of conditions are for clinical management and counseling. However, for the purpose of translational research, conditions that define a smaller number of cases may have more specific correlates on a molecular level.

¶No. Trees shows the number of CART trees that utilize each set of conditions. There are a total of 5 trees. (See Supplemental Results.) Increased utilization indicates “usefulness” or “robustness” of that particular set of conditions.

||Reference condition against which the Odds Ratio and 95% C.I. for having no pregnancy is calculated.

**Conditions A–E are listed from most robust and “useful” to least “useful” based on: the number of trees that utilize each set of conditions, the number of applicable cases, and the odds ratio and 95% C.I.

doi:10.1371/journal.pone.0002562.t002

prognostic value, or to identify redundancy amongst variables [12,14,17,21]. For example, age, serum FSH, number of oocytes and number of embryos were each reported to be significantly related to IVF outcomes [17]. However, as shown by our analyses, the prognostic value of age and number of oocytes was captured by three of the four non-redundant variables. Similarly, the total number of embryos and the number of 8-cell embryos have been advocated for use in selecting patients for blastocyst transfer in some IVF clinics to minimize the risk of having no embryos to transfer due to failed blastocyst development [21,33,34]. However, the prognostic value of these two variables has not been compared to that of others, and their ability to capture prognostic information from most other variables were not known.

Indeed, cleavage arrest rate is the only variable that is independent of the others, which suggests that it may be linked to biological mechanisms that are not currently recognized in the management of clinical infertility or hESC biology. Encountered in ~18% of human embryos cultured *in vitro* overall, its underlying defects are likely diverse, and may be due to suboptimal *in vitro* culture environment, biological mechanisms underlying infertility, a generally poor reproductive fitness of our species or all of these factors. Although cleavage arrest coincides with the maternal-embryonic transition during which maternal transcripts are degraded and the embryonic genome is activated [35], gene expression analyses of arrested single human embryos did not show failure in embryonic genome activation, and no specific molecular defects have been identified [8,11].

Our study has some limitations. Although we took advantage of the power of regression tree models to analyze a very comprehensive range of variables, we did not include cryopreservation of sibling embryos and assisted hatching as variables. In addition, it would also be valuable to analyze blastocyst development rate of sibling embryos, because this variable has been shown to correlate with positive pregnancy outcomes [36]. Those variables are now being investigated in a larger study that encompasses four years of data. As the goal of this current study was to explore new paradigms in human embryo development in IVF, and not to arrive at recommendations to change clinical practice, we used positive serum hCG status as the surrogate outcome measure to identify nonredundant predictors of IVF cycles in which at least one embryo attaches to the endometrium and secretes hCG, from those in which no embryo attaches. In the future, we will use later endpoints, such as clinical pregnancy or live birth, to address clinical questions.

In spite of over 30 years of ART, many challenges remain. Ongoing and future investigations may incorporate approaches common to genetics and developmental biology, in order to reassess defective human embryo development in terms of phenotypes that can be diagnosed, defined, and translated into improved clinical practices. Collectively, our results indicate that embryos from a cohort share as yet undefined genetic or epigenetic determinants of developmental competence, which is consistent with the greater increase in implantation relative to pregnancy rates conferred by blastocyst transfer [37]. The concept

of cohort-specific determinants suggest a paradigm shift from strictly focusing research efforts on selecting the “best” embryos to identifying methods that would improve the quality of the entire cohort. In addition, it raises the question of whether quality of the entire cohort is intrinsic due to the shared origins of the embryos, or if it is merely a result of group culture *in vitro*, especially since the benefits of group culture have been reported in animal and human embryos [38,39,40,41]. While embryo-specific parameters may help to identify embryos that would maximize the immediate pregnancy outcome for each couple, in the long term, understanding cohort-specific parameters is critical in counseling patients, improving treatment, and ultimately in developing mechanism-specific and more customized treatments.

We reason that well-defined criteria for embryo cohort phenotypes in selecting abnormal embryos for molecular analyses would maximize the chance of finding non-random genetic or epigenetic molecular defects that are consistent in an embryo cohort. For example, we are applying our findings to analyze arrested embryos from embryo cohorts in which the number of embryos are ≥ 6 and cleavage arrest rate is $>52.8\%$ (see Condition B5 in Table 2). Overall, $\sim 2.5\%$ of fresh, non-donor IVF cases (or ~ 17 cases per year, at our center) are expected to fulfill these criteria. This approach should allow for objective interpretation and comparison of data both internally and amongst research groups.

We are also applying this research strategy to investigate predictors of pregnancy outcomes in subsequent IVF cycles to contrast couple- versus embryo cohort-specific prognostics variables. More importantly, new hypotheses that are generated by this investigation can be further tested as additional years of data become available. For example, our findings indicate that a low day 3 FSH (<4.6 mIU/mL) confer high pregnancy rates in a very small and specific subset of patients (see Condition E in Table 2), and offer new perspectives on this controversial entity. While abnormally high levels of day 3 FSH have been associated with ovarian aging, poor ovarian response in IVF, and poor IVF cycle outcomes, many studies have cautioned against its use in clinical management due to its low sensitivity, especially in women under 40 [42,43,44,45]. However, the clinical utility of this test may be improved by determining appropriate thresholds and conditions [46].

Similar to the implications for ART, our results also raise questions about the effects of cohort-specific determinants on the success rate of hESC line derivation, the quality of hESC lines, and most importantly, embryo cohort selection for hESC line derivation, or oocyte cohort selection for somatic cell nuclear transfer. Currently, most scientific reports on successful derivation of hESC lines do not include information on embryo cohort characteristics, clinical information or IVF outcomes of sibling embryos. Our findings suggest that correlation of clinical IVF data and hESC line characteristics may provide valuable insight that would move both the fields of reproductive medicine and hESC research forward. We envision that dissection of human embryo phenotypes and their corresponding molecular correlates is not only a necessary step towards improving the treatment of clinical infertility, but will also contribute significantly to research efforts in the hESC field.

Materials and Methods

Data Collection, Inclusion and Exclusion Criteria

Data related to clinical diagnoses, IVF treatment protocol and monitoring, embryology data and treatment outcomes for all IVF cycles performed between January 1, 2005 and December 31, 2005 at Stanford University Medical Center were retrieved from

BabySentryPro (BabySentry Ltd, Limassol, Cyprus), a widely used fertility database management system, or obtained from medical and embryology records as necessary. Retrospective data collection, de-identification, and analysis were performed according to a Stanford University Institutional Review Board-approved protocol. The inclusion criteria for data analysis were fresh, stimulated, non-donor oocyte IVF cycles. We excluded cycles that did not result in embryo transfer for any reason, cycles performed for women aged over 45, and those performed for preimplantation genetic screening.

Assessment of Embryo Development

Our standard clinical protocols for ART treatment, fertilization, embryo culture, embryo assessment, cryopreservation criteria, and clinical outcomes are described in methods in *SI Text*. The normal progression of human embryo development *in vitro* is characterized by the appearance of 2 pronuclei at 16–20 hours after insemination as evidence of fertilization on Day 1, with Day 0 as the day of oocyte retrieval. By late Day 1, embryo development has reached the 2-cell stage, followed by the 4-cell and 8-cell stages on Days 2 and 3, respectively. On Days 4 and 5, embryo development is characterized by the establishment of the morula and blastocyst stages, respectively. All embryos were available for evaluation on Day 3. The day of embryo transfer was determined by the number of blastomeres on Day 3. In general, if 4 or more 8-cell embryos were present, we would recommend extended embryo culture until Day 5, when blastocyst transfer, which has been associated with higher pregnancy rates, would be performed. If fewer than four 8-cell embryos were present, embryo transfer would be performed on Day 3.

Patient, IVF Cycle, and Embryo Parameters

We analyzed 30 variables for association with IVF treatment outcomes, as listed in Table 1, under four main categories: patient characteristics and clinical diagnoses, IVF cycle characteristics, embryo cohort parameters, and parameters of transferred embryos. The cleavage arrest rate was defined as the percentage of embryos within a cohort with 4 or fewer cells on Day 3 of *in vitro* culture. All other variables were self-explanatory.

Statistical Analysis

Since some patients underwent more than one IVF cycle during the study period, the analyses were performed based on treatment cycles rather than patients. Statistical analyses were performed based on the dichotomous outcomes of no pregnancy, as defined by negative serum β -hCG, and pregnancy, as defined by positive serum β -hCG, and included biochemical pregnancy, clinical pregnancy, spontaneous abortion, and ectopic pregnancy. We performed pair-wise logistic regression of each variable to the outcome and determined the Pearson correlation coefficient between each pair of continuous variables.

For the main analyses, boosted classification trees were constructed by MART[®] to identify non-redundant prognostic variables, which were then further analyzed by CART to identify thresholds that would define them as categorical variables. MART[®] is a robust method used to identify interactive structure of variables that are predictive of outcomes [29,30,31,32]. The use of cross-validation and boosting in parameter selection and model assessment in MART[®] also preserve parsimony and prevent overfitting [31]. In the MART[®] tree constructions, the whole data set is divided into 10 subsets to achieve 10 fold cross validation for model assessment. The same 10 fold cross validation was repeated 1000 times to perform a robust prediction rate estimation and identify tree models with the highest prediction rates in the CART. While MART[®] is powerful in selecting non-redundant prognostic

variables from a large set of highly interactive variables, CART analysis results in simple algorithms, and more easily understood “decision trees”, that are used in the medical literature [28]. Thus non-redundant, prognostic variables identified by MART[®] to confer prediction were analyzed by CART to further define prognostic thresholds.

Supporting Information

Text S1 This SI file contains details pertaining to methods, results, and statistical analyses which may be of interest to certain readers. It also contains an SI Table 2.

Found at: doi:10.1371/journal.pone.0002562.s001 (0.10 MB DOC)

Table S1 Correlation between each pair of variables.

Found at: doi:10.1371/journal.pone.0002562.s002 (0.08 MB DOC)

References

- Behr B (1999) Blastocyst culture and transfer. *Hum Reprod* 14: 5–6.
- Cowan CA, Klimanskaya I, McMahon J, Atienza J, Witmyer J, et al. (2004) Derivation of embryonic stem-cell lines from human blastocysts. *N Engl J Med* 350: 1353–1356.
- Gardner DK, Lane M, Schoolcraft WB (2000) Culture and transfer of viable blastocysts: a feasible proposition for human IVF. *Hum Reprod* 15 Suppl 6: 9–23.
- Milki AA, Hinckley MD, Fisch JD, Dasig D, Behr B (2000) Comparison of blastocyst transfer with day 3 embryo transfer in similar patient populations. *Fertil Steril* 73: 126–129.
- Chen H, Qian K, Hu J, Liu D, Lu W, et al. (2005) The derivation of two additional human embryonic stem cell lines from day 3 embryos with low morphological scores. *Hum Reprod* 20: 2201–2206.
- Strelchenko N, Verlinsky O, Kukharencov V, Verlinsky Y (2004) Morula-derived human embryonic stem cells. *Reprod Biomed Online* 9: 623–629.
- Mitalipova M, Calhoun J, Shin S, Wminger D, Schulz T, et al. (2003) Human embryonic stem cell lines derived from discarded embryos. *Stem Cells* 21: 521–526.
- Artley JK, Braude PR, Johnson MH (1992) Gene activity and cleavage arrest in human pre-embryos. *Hum Reprod* 7: 1014–1021.
- Barratt CL, St John JC, Afnan M (2004) Clinical challenges in providing embryos for stem-cell initiatives. *Lancet* 364: 115–118.
- Hardy K, Spanos S, Becker D, Iannelli P, Winston RM, et al. (2001) From cell death to embryo arrest: mathematical models of human preimplantation embryo development. *Proc Natl Acad Sci U S A* 98: 1655–1660.
- Dobson AT, Raja R, Abeyta MJ, Taylor T, Shen S, et al. (2004) The unique transcriptome through day 3 of human preimplantation development. *Hum Mol Genet* 13: 1461–1470.
- Alikani M, Cohen J, Tomkin G, Garrisi GJ, Mack C, et al. (1999) Human embryo fragmentation in vitro and its implications for pregnancy and implantation. *Fertil Steril* 71: 836–842.
- Ebner T, Moser M, Sommergruber M, Tews G (2003) Selection based on morphological assessment of oocytes and embryos at different stages of preimplantation development: a review. *Hum Reprod Update* 9: 251–262.
- Ebner T, Yaman C, Moser M, Sommergruber M, Polz W, et al. (2001) Embryo fragmentation in vitro and its impact on treatment and pregnancy outcome. *Fertil Steril* 76: 281–285.
- Giorgetti C, Terriou P, Auquier P, Hans E, Spach JL, et al. (1995) Embryo score to predict implantation after in-vitro fertilization: based on 957 single embryo transfers. *Hum Reprod* 10: 2427–2431.
- Jun SH, O’Leary T, Jackson KV, Racowsky C (2006) Benefit of intracytoplasmic sperm injection in patients with a high incidence of triploidy in a prior in vitro fertilization cycle. *Fertil Steril* 86: 825–829.
- Keltz MD, Skorupski JC, Bradley K, Stein D (2006) Predictors of embryo fragmentation and outcome after fragment removal in in vitro fertilization. *Fertil Steril* 86: 321–324.
- Lan KC, Huang FJ, Lin YC, Kung FT, Hsieh CH, et al. (2003) The predictive value of using a combined Z-score and day 3 embryo morphology score in the assessment of embryo survival on day 5. *Hum Reprod* 18: 1299–1306.
- Nagy ZP, Dozortsev D, Diamond M, Rienzi L, Ubaldi F, et al. (2003) Pronuclear morphology evaluation with subsequent evaluation of embryo morphology significantly increases implantation rates. *Fertil Steril* 80: 67–74.
- Stone BA, Greene J, Vargyas JM, Ringler GE, Marrs RP (2005) Embryo fragmentation as a determinant of blastocyst development in vitro and pregnancy outcomes following embryo transfer. *Am J Obstet Gynecol* 192: 2014–2019; discussion 2019–2020.
- Volpes A, Sammartano F, Cofarò F, Mistretta V, Scaglione P, et al. (2004) Number of good quality embryos on day 3 is predictive for both pregnancy and implantation rates in in vitro fertilization/intracytoplasmic sperm injection cycles. *Fertil Steril* 82: 1330–1336.
- (2006) Guidelines on number of embryos transferred. *Fertil Steril* 86 (5 Suppl): S51–52.
- Racowsky C (2002) High rates of embryonic loss, yet high incidence of multiple births in human ART: is this paradoxical? *Theriogenology* 57: 87–96.
- Milki AA, Hinckley MD, Gebhardt J, Dasig D, Westphal LM, et al. (2002) Accuracy of day 3 criteria for selecting the best embryos. *Fertil Steril* 77: 1191–1195.
- Neuber E, Mahutte NG, Arici A, Sakkas D (2006) Sequential embryo assessment outperforms investigator-driven morphological assessment at selecting a good quality blastocyst. *Fertil Steril* 85: 794–796.
- Rijnders PM, Jansen CA (1998) The predictive value of day 3 embryo morphology regarding blastocyst formation, pregnancy and implantation rate after day 5 transfer following in-vitro fertilization or intracytoplasmic sperm injection. *Hum Reprod* 13: 2869–2873.
- Masterbroek S, Twisk M, van Echten-Arends J, Sikkema-Raddatz B, Korevaar JC, et al. (2007) In vitro fertilization with preimplantation genetic screening. *N Engl J Med* 357: 9–17.
- Guzick DS, Overstreet JW, Factor-Litvak P, Brazil CK, Nakajima ST, et al. (2001) Sperm morphology, motility, and concentration in fertile and infertile men. *N Engl J Med* 345: 1388–1393.
- Friedman J (1999) Greedy function approximation: A stochastic boosting machine. Technical Report, Department of Statistics, Stanford University.
- Friedman J (1999) Stochastic gradient boosting. Technical Report, Department of Statistics, Stanford University.
- Friedman J (2002) Tutorial: Getting started with MART in R. Department of Statistics, Stanford University.
- Friedman J, Meulmann JJ (2003) Multiple additive regression trees with application in epidemiology. *Statistics in medicine* 22: 1365–1381.
- Gardner DK, Schoolcraft WB, Wagley L, Schlenker T, Stevens J, et al. (1998) A prospective randomized trial of blastocyst culture and transfer in in-vitro fertilization. *Hum Reprod* 13: 3434–3440.
- Racowsky C, Jackson KV, Cekleniak NA, Fox JH, Hornstein MD, et al. (2000) The number of eight-cell embryos is a key determinant for selecting day 3 or day 5 transfer. *Fertil Steril* 73: 558–564.
- Braude P, Bolton V, Moore S (1988) Human gene expression first occurs between the four- and eight-cell stages of preimplantation development. *Nature* 332: 459–461.
- Fisch JD, Milki AA, Behr B (1999) Sibling embryo blastocyst development correlates with the in vitro fertilization day 3 embryo transfer pregnancy rate in patients under age 40. *Fertil Steril* 71: 750–752.
- Gardner DK, Surrey E, Minjarez D, Leitz A, Stevens J, et al. (2004) Single blastocyst transfer: a prospective randomized trial. *Fertil Steril* 81: 551–555.
- Lane M, Gardner DK (1992) Effect of incubation volume and embryo density on the development and viability of mouse embryos in vitro. *Hum Reprod* 7: 558–562.
- Rijnders PM, Jansen CAM (1998) The predictive value of day 3 embryo morphology regarding blastocyst formation, pregnancy, and implantation rate after day 5 transfer following in vitro fertilization or intracytoplasmic sperm injection. *Hum Reprod* 13: 2869–2873.
- Spyropoulou I, Karamalegos C, Bolton VN (1999) A prospective randomized study comparing the outcome of in-vitro fertilization and embryo transfer following culture of human embryos individually or in groups before embryo transfer on day 2. *Hum Reprod* 14: 76–79.
- Paria BC, Dey SK (1990) Preimplantation embryo development in vitro: cooperative interactions among embryos and role of growth factors. *Proc Natl Acad Sci U S A* 87: 4756–4760.

Figure S1 Variables and their relative importance in determining A) number of 8-cell embryos, B) day 3 FSH, and C) the total number of embryos.

Found at: doi:10.1371/journal.pone.0002562.s003 (2.65 MB TIF)

Acknowledgments

We thank the Embryology staff and L. Morcom for their efforts on data entry, L. Wu and C. Mak for valuable discussion, M. Zamah for help with preliminary data collection.

Author Contributions

Conceived and designed the experiments: WW MY SJ BC. Performed the experiments: BC. Analyzed the data: WW BC. Contributed reagents/materials/analysis tools: WW LW BB. Wrote the paper: WW MY SJ BC RR LS. Other: Acquired data: MY SJ LW BB. Interpreted data analysis: SJ MY. Interpreted findings: BB RRP LW.

42. Broekmans FJ, Kwee J, Hendriks DJ, Mol BW, Lambalk CB (2006) A systematic review of tests predicting ovarian reserve and IVF outcome. *Hum Reprod Update* 12: 685–718.
43. Jain T, Soules MR, Collins JA (2004) Comparison of basal follicle-stimulating hormone versus the clomiphene citrate challenge test for ovarian reserve screening. *Fertil Steril* 82: 180–185.
44. Srouji SS, Mark A, Levine Z, Betensky RA, Hornstein MD, et al. (2005) Predicting in vitro fertilization live birth using stimulation day 6 estradiol, age, and follicle-stimulating hormone. *Fertil Steril* 84: 795–797.
45. Watt AH, Legedza AT, Ginsburg ES, Barbieri RL, Clarke RN, et al. (2000) The prognostic value of age and follicle-stimulating hormone levels in women over forty years of age undergoing in vitro fertilization. *J Assist Reprod Genet* 17: 264–268.
46. Scott RT Jr, Elkind-Hirsch KE, Styne-Gross A, Miller KA, Frattarelli JL (2007) The predictive value for in vitro fertility delivery rates is greatly impacted by the method used to select the threshold between normal and elevated basal follicle-stimulating hormone. *Fertil Steril*.