Correcting misperceptions about cryopreserved embryos and stem cell research

To the Editor:

Lawrence Goldstein[1] recently commented that "details matter" but that "regrettably, considerable inaccuracy has found its way into debates about stem cells, on all sides, with consequent fogging of the issues." Here we reevaluate the number of cryopreserved embryos in the United States and discuss the rate at which they are used. An updated accounting of these details serves to improve both public understanding of stem cell science and the informed-consent process for donating cryopreserved embryos to research.

Thirty-four years after the birth of Louise Brown, the first 'test tube baby', an estimated 5 million births aided by *in vitro* fertilization (IVF) have occurred. The International Committee Monitoring Assisted Reproductive Technologies reported this milestone at the 2012 European Society of Human Reproduction and Embryology annual meeting[2]. In 2010, 61,564 infants (>1% of all births in the United States), were conceived using assisted reproductive technology (ART)[3]. Through ongoing refinement of ART procedures, success rates in the United States, as measured by the percentage of embryo transfers resulting in live births, have improved over the last decade[3].

During ART clinical procedures, the number of embryos produced is routinely in excess of the number that can be transferred. Extra embryos are usually frozen and then stored. Ninety-nine percent of centers ($n = 443$) reporting to the 2010 US National Summary of Fertility Centers Report offer cryopreservation to patients[3]. Interviews with families who have stored frozen embryos ($n = 77$) revealed that the median number stored was seven[4] and that the average storage time was five years[4]. One-third of these families choose to store excess embryos indefinitely[4]. National and international data suggest that the absolute number of embryos in storage is increasing[5]. A survey involving 430 US fertility clinics reported nearly 400,000 embryos in storage as of 2003 (ref. 6). Research articles and media reports published in 2012 continue to cite this figure, despite broad recognition that it is inaccurate[7–9].

Misperceptions also exist concerning the utilization rate of embryos in stem cell research. Public discussion of human embryonic stem cell (hESC) research often invokes a need for a continuous supply of embryos. For example, former presidential candidate Mitt Romney echoed a familiar talking point in affirming his opposition to 'embryo farming'[10]. In reality, prospective donors may be surprised to learn that opportunities to donate embryos for the derivation of hESC lines are limited at present. Thus, to clarify this 'fogging' of the issues, reconsideration of the number of existing cryopreserved embryos and the rate at which they are used in hESC research is warranted.

On the basis of discussions with practitioners providing ART services, we devised a conservative formula to support a revision of the estimated number of cryopreserved embryos in the United States. Our formula uses the 400,000 figure from 2003 as a baseline and then predicts the number of additional embryos cryopreserved between 2005 and 2010. (Data from 2004, 2011 and 2012 were not available.) Our primary assumption is that for every successful live ART birth, an additional five embryos will remain in storage. Recognizing that some centers do not provide cryopreservation services, we assume that 95% of excess embryos resulting from IVF will be subsequently stored.

We conclude that through 2012, more than 1 million embryos would have been cryopreserved in the United States (**Table 1**). We consider this a conservative estimate for four reasons. First, this figure is based on successful births. The majority of women undergoing ART procedures do not become pregnant and may discontinue treatment, leaving additional cryopreserved embryos in storage[3]. Second, published reports suggest that the assumption that five embryos remain for each live birth and that 95% of these are stored is probably an underestimate. Third, improvements in ART pregnancy success rates may have resulted in a higher proportion of unused embryos. And fourth, additional embryos have been stored since 2010.

One must also consider the rate at which embryos are taken out of storage to be discarded, provided to another couple for fertility treatment or donated to research. Surveys indicate that >85% of those already frozen are being stored for patient treatment[11]. A review of previous studies suggests that up to half of individuals or couples avoid making a decision about unused embryos by not informing the clinic of a change in address[12]. We routinely interact with researchers and clinicians involved in the operation of fertility centers. They report that their default procedure for lost-to-follow-up (abandoned) embryos is to keep them in storage out of concern that individuals with dispositional authority may contact the center in the future. For these reasons, we believe that the estimate that the number of embryos in storage is increasing in proportion to the number of live ART births is valid.

Interest in embryo donation for research seems to be growing. A recent survey of 1,020 individuals or couples with cryopreserved embryos found that donation for research was the most popular disposition option[13]. In California, about one-third of donors choose human development research, another third choose stem cell research and the remaining third choose both types[14,15]. These results suggest that >60% of future donations will be available for stem cell research.

Notably, there seems to be an imbalance between the potential supply of embryos and the ability to use them in research. Although new hESC lines continue to be crucial for the development of regenerative medicine, the need for new cryopreserved embryos is now limited, as derivation efforts have become targeted and more specialized.

Table 1 Estimated minimum number of cryopreserved embryos

Year	Centers for Disease Control and Prevention (CDC) reported live births	Total number of embryos cryopreserved
2005	38,910	194,550
2006	30,787	153,935
2007	29,556	147,780
2008	31,365	156,825
2009	30,787	153,935
2010	47,090	235,450
2005–2010 total		1,042,475
95% of stored embryos		990,351
2003 baseline		400,000
Total		1,390,351

In the past five years, the number of hESC lines available to researchers has expanded rapidly. As of December 2012, there were 198 hESC lines in the US National Institutes of Health (NIH) Human Embryonic Stem Cell Registry. There are >1,200 hESC lines identified in the International Stem Cell Registry[16]. We have reported that the California Institute for Regenerative Medicine (CIRM) grantees have used 138 unique hESC lines[17]. To date, researchers have tended to use a relatively small number of lines[18]. NIH eligibility status, material transfer agreements, a policy decision to support National Stem Cell Bank distribution, and resulting scientific utility are cited as reasons why a smaller number of lines are being utilized[17].Therefore, at this time, the scientific need for additional lines remains limited.

There continues to be interest in developing 'disease-specific' hESC lines from embryos that have undergone pre-implantation genetic diagnosis (PGD). For example, nine of the 12 hESC lines added to the NIH Human Embryonic Stem Cell Registry in December 2012 were from PGD[19]. These new derivations are crucial for understanding the disease process, but PGD embryos carry genetic mutations. Absent research donation, these embryos would otherwise be discarded, so these derivations do not impact the number cryopreserved.

One laboratory reports focusing on the derivation of clinical grade lines under xeno-free and feeder-free conditions (http://www.stemcells.wisc.edu/node/858). The three non-PGD hESC lines registered with the NIH in December 2012 are clinical grade[19]. Given the expense associated with derivation, estimated to be $100,000–120,000, and with maintenance of clinical grade lines under conditions compliant with Current Good Manufacturing Practice regulations, future efforts are likely to be selective and limited in number[20].

Research repositories in California and other states have successfully implemented embryo donation programs. Pruksananonda *et al.*[21] reported the derivation of an hESC line from an embryo preserved for 18 years, further suggesting that current numbers of stored embryos may be sufficient to fulfill research needs in the foreseeable future.

Although research articles and media reports continue to cite a decade-old figure of nearly 400,000 cryopreserved embryos in storage, on the basis of reported ART utilization and trends for embryo storage, we conservatively estimate that there are now more than 1 million cryopreserved embryos in the United States. The cumulative number of individuals confronting disposition decisions has grown over time, and embryo donation for stem cell research has emerged as a preferred option. In many cases, however, it may not be possible to accommodate this choice. CIRM attracts national interest because certain states have prohibitions on embryo research, and federal restrictions also limit the ability of many repositories to accept donations. This uneven policy environment results in the limiting of research donation by certain legal and geographic considerations[13]. Between 2007 and 2010, CIRM could refer individuals to numerous research programs with active embryo donation and hESC derivation protocols. Programs could accept embryos from a range of donors, provided they were able to complete a comprehensive informed-consent process. Recently, our ability to facilitate referrals has become more limited. After completing CIRM's New Cell Lines Awards program, many CIRM grantees did not renew their rigorous informed-consent protocols required for embryo donation[22,23]. Grantees cite administrative, staffing, storage and data-management costs as reasons for not maintaining ongoing consent protocols. In November 2012, we could identify only one center accepting embryos from donors not enrolled in their clinical program.

Collectively, this evidence suggests that the ability to donate embryos specifically for hESC research may be limited in the future. This finding has implications for the informed-consent process. Embryo donors cite aspirations such as "wanting to help find cures for diseases like Alzheimer's or diabetes" as reasons for choosing stem cell research[13]. We concur with previous authors who advocate an effective education and consent process for infertility treatment that involves embryo cryopreservation[14]. Given the current number of stored embryos consented for research and their effective utilization rate, we support the following revisions to the informed-consent process. First, donors should be informed that specific research options may not be feasible because of capacity considerations. Prospectively, providers may want to consider ranked choice disposition options for research donors. Second, fertility centers should disclose disposition options to all gamete donors, including third-party sperm and oocyte donors, to support consent protocols consistent with the US National Academies' *Guidelines for Human Embryonic Stem Cell Research*[24]. Comprehensive disclosure to gamete donors can improve the efficacy of the informed-consent process. Third, consent protocols should disclose that embryos may not be stored indefinitely and could be discarded without being used in scientific studies. The lack of such a provision can result in the need to store existing embryos indefinitely. Finally, centers should coordinate their efforts nationally to ensure comprehensive disclosure and consent.

We have attempted to address misperceptions and inaccuracies that persist in policy deliberations, media reports and research articles concerning stem cell research. Our evaluation suggests that the use of cryopreserved embryos in stem cell research is limited. Consequently, the number of embryos used for stem cell research relative to the total number of unused cryopreserved embryos available is negligible. We hope that this information will enhance public understanding of science and policy deliberations and will provide a more complete understanding of the scientific context to prospective embryos donors. As Goldstein reminds us, "details are important when novel technologies encounter law, politics, and ethics"[1].

ACKNOWLEDGMENTS

We would like to thank our CIRM grantees, the other state stem cell research programs and the national research institutions who related their experiences with embryo donation and hESC line derivation. In addition, we thank D. Adamson of International Committee Monitoring Assisted Reproductive

Technologies for comments on an early draft of this publication.

COMPETING FINANCIAL INTERESTS
The authors declare no competing financial interests.

Geoffrey P Lomax[1] & Alan O Trounson[1,2]

[1]California Institute for Regenerative Medicine, San Francisco, California, USA. [2]The Ritchie Centre, Monash Institute of Medical Research, Clayton, Australia.
e-mail: glomax@cirm.ca.gov

1. Goldstein, L. *J. Law Med. Ethics* **38**, 204–211 (2010).
2. Nordqvist, C. *Medical News Today* <http://www.medcalnewstoday.com/articles/247343.php> (2012).
3. CDC. 2010 National Summary, Assisted Reproductive Technology National Summary Report (Centers for Disease Control and Prevention, Atlanta, 2013).
4. Nachtigall, R.D. *et al. Fertil. Steril.* **92**, 2094–2096 (2009).
5. Bankowski, B.J., Lyerly, A.D., Faden, R.R. & Wallach, E.E. *Fertil. Steril.* **84**, 823–832 (2005).
6. Hoffman, D.I. *et al. Fertil. Steril.* **79**, 1063–1069 (2003).
7. Ertelt, S. LifeNews.com <http://www.lifenews.com/2012/03/05/obama-end-funding-for-snowflake-embryo-adoption-program/> (2012).
8. Palca, J. *Talk of the Nation* <http://www.npr.org/templates/story/story.php?storyId=96392644&ft=1&f=5> (2008).
9. Brezina, P.R. & Zhao, Y. *Obstet. Gynecol. Int.*, 686253 (2012).
10. Stein, J. *Milwaukee Journal Sentinel* (2012) <http://www.jsonline.com/news/statepolitics/how-thompson-baldwin-differ-on-stem-cell-research-im75rfj-174113021.html>.
11. Hug, K. *Fertil. Steril.* **89**, 263–277 (2008).
12. Hammarberg, K. & Tinney, L. *Fertil. Steril.* **86**, 86–91 (2006).
13. Lyerly, A.D. *et al. Fertil. Steril.* **93**, 499–509 (2008).
14. Kalista, T., Freeman, H.A., Behr, B., Pera, R.R. & Scott, C.T. *Cell Stem Cell* **8**, 360–362 (2011).
15. CIRM Medical and Ethical Standards Working Group July 25, 2008. <http://www.cirm.ca.gov/sites/default/files/files/agenda/transcripts/07-25-08.pdf>
16. International Stem Cell Registry. <http://www.umassmed.edu/iscr/index.aspx> (accessed 21 March 2013).
17. Patel, R. & Lomax, G.P. *J. Stem Cell Res. Ther.* **1**, 107 (2011).
18. Löser, P., Schirm, J., Guhr, A., Wobus, A.M. & Kurtz, A. *Stem Cells* **28**, 240–246 (2010).
19. Advisory, N.I.H. Committee to the Director GENEA Submission #2012-ACD-003. <http://acd.od.nih.gov/06142012_HeSC_002GENEA.pdf>.
20. Jacquet, L. *et al. EMBO Mol. Med.* **5**, 10–17 (2013).
21. Pruksananonda, K. *et al. BioResearch Open Access* **1**, 166–173 (2012).
22. CIRM New Cell Line Awards Request for Application, San Francisco, 2005. <http://www.cirm.ca.gov/our-funding/research-rfas/new-cell-lines> (accessed 21 March 2013).
23. CIRM. Medical and Ethical Standards Regulations. (California Institute for Regenerative Medicine, San Francisco, 2012).
24. National Research Council and Institute of Medicine. *Final Report of the National Academies' Human Embryonic Stem Cell Research Advisory Committee and 2010 Amendments to the National Academies' Guidelines for Human Embryonic Stem Cell Research* (National Academies Press, Washington, DC, 2010).

ProtoNet: charting the expanding universe of protein sequences

To the Editor:

As next-generation sequencing technologies continue to generate staggering amounts of raw protein sequences, it has become very difficult to thoroughly annotate the emerging protein-sequence space. Complete proteomes (that is, the collection of all valid proteins from a sequenced genome) as well as partial sequencing efforts have resulted in the archiving of more than 20 million protein sequences in UniProtKB (release 2012_1, 25 January 2012; http://www.uniprot.org). This repository is compiled from millions of viral sequences, thousands of microbial genomes and sequences from thousands of multicellular organisms. These sequences comprise what may be considered the now-known parts of the protein space. At present, the functional characterization available for the vast majority of this space is based mostly on sequence-similarity approaches. In fact, the characterized part of this space is orders of magnitude smaller than the whole, and only 3.5% of sequences in UniProtKB[1] have any experimental support. From this view, only a robust, unsupervised and automated method can realistically achieve comprehensive and functional annotation of this rapidly expanding protein space.

Protein three-dimensional structures provide the most reliable information on biochemical function. At present, there are 80,000 solved protein structures (http://www.rcsb.org/pdb/home/home.do) that are indirectly associated with a large fraction of the protein space. Through semi-automatic classifications, these three-dimensional solved proteins are organized in an inventory of ~1,500 basic folds[2,3]. However, these folds are consistent with local domains rather than full-length proteins. Complementary sequence–based approaches for protein family assignment rely primarily on the notion of domains as the building blocks of proteins. The general scheme starts with multiple sequence alignment, which is then translated into statistically based models (e.g., Pfam)[4]. The integration of different resources (e.g., InterPro)[5] leads to a substantial increase in domain coverage of the protein space. The curated portion of the protein space is already enormous. Still, one-third (6.7 million) of UniProtKB sequences are marked as putative or hypothetical. For these sequences, current methods for direct inference of function with high confidence have mostly failed. Furthermore, most sequence- and structure-based assignments rely on local information such as structural fold, sequence domain and functional signature). Consequently, functional annotations at the level of the full-length protein are prone to erroneous inference. It is realistic to expect an even faster growth in the number of protein sequences (e.g., from large-scale sequencing of environmental samples). This creates a pressing need for accurate methods of annotation inference.

We offer the ProtoNet 6.1 family tree (http://www.protonet.cs.huji.ac.il), a classification resource created by an unsupervised analysis of protein sequences[6]. The families in the ProtoNet tree are generated through the following steps: (i) precalculation of sequence-similarity values for all possible pairwise relationships (all against all BLAST values), (ii) application of an unsupervised bottom-up clustering algorithm (this algorithm organizes large sets of proteins in a hierarchical tree that yields high-quality protein families) (**Supplementary Table 1**) and (iii) a process of pruning the ProtoNet tree to retain only the most informative clusters. This computational process yields a tree-like skeleton of the entire known protein space. In the next stage, each cluster is assessed through a comprehensive battery of descriptors for domains, three-dimensional structures, enzymes, gene ontology, taxonomy and more (**Supplementary Tables 2** and **3**). In addition, rigorous annotation-based quality tests are carried out to assign a statistically based quality measure for each stable cluster. Each cluster is then assigned the set of descriptors that reflect the most significant annotation(s) of its proteins. In this way, ProtoNet circumvents many of the pitfalls in annotation inference discussed above.

There are several features that allow ProtoNet to cope with the scale of the known protein space. First, it applies a scalable, efficient and accurate algorithm for clustering millions of sequences[7]. Second, family construction is 'model free'; the process of tree construction is continuous and data driven. Third, all sequences are dealt with on an equal basis, irrespective of length, domain organization, taxonomy or prior knowledge. Thus, putative proteins play an integral part in the construction of ProtoNet.

Nearly 19 million full-length protein sequences are included in ProtoNet 6.1