



^aInstitute for Stem Cell Biology and Regenerative Medicine, Stanford University Medical School, Stanford, California, USA; ^bMeta-Research Innovation Center at Stanford (METRICS), Stanford University, Stanford, California, USA; ^cDepartment of Medicine, Stanford University School of Medicine, Stanford, California, USA; ^dDepartment of Health Research and Policy, Stanford University School of Medicine, Stanford, California, USA; ^eDepartment of Biomedical Data Science, Stanford University School of Medicine, Stanford, California, USA; ^fDepartment of Statistics, Stanford University School of Humanities and Sciences, Stanford, California, USA


Correspondence: Rebecca Marton, B.S., Institute for Stem Cell Biology and Regenerative Medicine, Stanford University School of Medicine, 3165 Porter Drive, Stanford, California 94305, USA. Telephone: (650) 725-6659; e-mail: rmarton@stanford.edu; or John P. A. Ioannidis, M.D., D.Sc., Meta-Research Innovation Center at Stanford (METRICS), Disease Prevention and Medicine, Health Research and Policy, Biomedical Data Science, and Statistics, Stanford University, 1265 Welch Road, Medical School Office Building, Room X306, Stanford, California 94305, USA. Telephone: (650) 725-5465; e-mail: jiannid@stanford.edu

Received April 16, 2018; accepted for publication November 6, 2018; first published December 10, 2018.

<http://dx.doi.org/10.1002/sctm.18-0088>

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

A Comprehensive Analysis of Protocols for Deriving Dopaminergic Neurons from Human Pluripotent Stem Cells

REBECCA M. MARTON ^{a,b}, JOHN P. A. IOANNIDIS^{b,c,d,e,f}

Key Words. Embryonic stem cells • Induced pluripotent stem cells • Differentiation • Dopaminergic neurons

ABSTRACT

The potential applications of human embryonic and induced pluripotent stem cells has led to immense interest in developing new protocols to differentiate specific cell types or modifying existing protocols. To investigate to what extent and why new protocols for the same cell types are developed and adopted, we systematically evaluated 158 publications (2004-2017) that differentiated human stem cells into dopaminergic neurons. We categorized each article by degree of novelty and recorded motivations for protocol development. 74 novel or modified protocols were developed. Most (65%) were not used again in subsequent studies. Diverse motivations were recorded and performance of new methods was assessed with substantially different approaches across studies. There was improvement over time in yield of neuron production, but not in yield of dopaminergic neurons or time required for getting neurons. Standardized reporting of performance metrics may help rational choice of the best methods. *STEM CELLS TRANSLATIONAL MEDICINE* 2019;8:366–374

SIGNIFICANCE STATEMENT

The ability to generate neurons from human pluripotent stem cells has made possible the study of previously inaccessible cell types in vitro. However, these cells are only as useful as they are able to recreate key characteristics of their in vivo counterparts. Tremendous effort has been poured into developing differentiation protocols, and new methodologies have been published every year since the advent of stem cell culture. Yet, the majority of these protocols are never reused, and inconsistencies in reporting make it difficult to form an educated decision when selecting a methodology. Understanding the motivations behind and outcomes of newly developed protocols will help the field to consider the effectiveness and utility of these efforts and how they are reported.

INTRODUCTION

The development of methodologies to culture human embryonic stem cells (hESCs) [1] and induced pluripotent stem cells (hiPSCs) [2] allows the use of stem cells to study human disease. hESCs and hiPSCs (referred to here collectively as human pluripotent stem cells (hPSCs)) can be expanded in vitro and differentiated into cell types that were previously inaccessible for study [1, 2]. Human stem cell-derived models provide advantages over animal models as they avoid species-specific differences and, in the case of hiPSCs, contain a patient's own genetic background [3, 4]. Due to these advantages, cells derived from hPSCs are widely useful for biomedical research including applications in screening drug compounds, investigating cellular

disease mechanisms, and producing cellular therapies.

There has been a rapid expansion of the stem cell field. An analysis of all scientific publications from 2008 to 2012 found that stem cell related publications had an annual growth rate that was twice as high as all other topics [5]. hESC and hiPSC related publications were also cited two times and three times more often, respectively, than all other publications in related disciplines [5]. This interest in stem cell research has led to the accelerated development of methods to differentiate specific cell types from hPSCs.

Stem cell-based neuroscience is one such field undergoing a rapid expansion of differentiation methods. Subtypes of neurons can vary widely in morphology and behavior, creating

the need for protocols that specify desired neural fates. Hundreds of papers have been published describing the production of neuronal subtypes from hPSCs. While some of these publications detail entirely new methods, others modify already existing methods. It would be useful to understand the relative merits of different methods and the main driving motivations for developing new methods.

Neurons produced *in vitro* from hPSCs have the potential to be translated into therapeutic treatments. Reproducible, standardized, optimal methods are essential in this regard. The aim of this study is to systematically evaluate methods to differentiate neurons from hPSCs using dopaminergic neuron differentiation as a case study. We assess the metrics used in these articles to measure dopaminergic neuron yield and protocol utility, as well as the extent to which the larger stem cell community adopts each method. Finally, we identify areas for improving reporting to better allow protocol comparisons for the purpose of applying these methods to translational medicine.

METHODS

Search Method

We performed within the PubMed database the following search query: (human[Title/abstract] AND stem cells[Title/abstract] AND neurons[Title/abstract]) AND (pluripotent[Title/abstract] OR embryonic[Title/abstract] or induced[Title/abstract]). The search was last updated on May 8, 2017 and identified 1,909 items.

Selection Criteria

We considered only primary studies that differentiate hESCs or hiPSCs into neurons and screened abstracts according to these criteria. We further refined the papers under consideration by screening the full text and only accepting methods that produce neurons that express neural lineage markers, exit the cell cycle, and adopt the morphology of the subtype of interest. As neurons are often produced along with other cell types, we only considered papers with the expressed goal of generating neurons. Differentiation methods must demonstrate the production of neurons *in vitro*.

Data Collection

We collected the following types of information for each article that meets the above selection criteria: PubMed ID, authors, year of publication, month of publication, title, full citation, and subtype of neuron produced. The approved articles were organized by neuronal subtype, and articles involving the differentiation of dopaminergic neurons were selected for further analysis. Of these articles, we selected for in depth scrutiny only those that involved the directed differentiation of dopaminergic neurons, as we were interested in methods with the specific intent to differentiate dopaminergic neurons. We collected the following additional information for each of the eligible articles on dopaminergic neurons: the location of the first author's primary institution, the type of stem cells used (hiPSCs and/or hESCs), the methodological details of the differentiation, the degree of novelty of the methods, and the citation of previous dopaminergic differentiation papers for the method of differentiation. The degree of novelty was assigned to one of three categories based on the authors' own description: A. the methods are novel; B. the methods involve significant

modification to a previously published method; or C. the methods are largely the same as a previously published method. For articles in Novelty Category B and C, we recorded whether there were authors that were common to both the article under examination and the article from which the methods were cited. Using this information, we determined the number of times the differentiation methods of Novelty Category A and B papers were reused in other studies in our dataset, and whether they shared co-authors with the papers that cited them. Protocols were only considered to be reused if they were specifically cited in reference to the methods used by another article.

If the methods used in the article were designated as novel (Novelty Category A) or an adaptation of previous methods with significant modification (Novelty Category B), we assessed the authors' reported motivations for new method development. Motivations were grouped into the following categories: reduction in time until neural production, increase in efficiency or yield of neurons produced, improvement in similarity to *in vivo* counterparts, reduction in variability between batches/improved reproducibility, creation of defined/good manufacturing practice (GMP) compliant culture conditions, optimization for transplantation optimized for starting stem cell type, other, or none given. When multiple motivations applied, all of them were captured. We also recorded the temporal length of differentiation protocols, quantifications of yield of dopaminergic neurons, the number of control lines used in each Novelty Category A and B article, progenitor origin marker expression, and small molecules and proteins used to induce differentiation. As multiple approaches were used to quantify yield, we recorded both the method of quantification and the corresponding values. For papers that reported dopaminergic neuron yield at multiple time points, we recorded the values for the time point with the maximum yield. For articles that sorted to enrich yield, yield quantifications were recorded for unsorted populations.

To compare hPSC differentiation protocols to human mesenchymal stem cell (hMSC) differentiation protocols, we identified articles in our original PubMed search that met the criteria defined above except that they used hMSCs from bone marrow or the umbilical cord instead of hPSCs as the starting cell type. We identified 16 articles that differentiate hMSCs into dopaminergic neurons and repeated the data collection procedures above to sort articles by novelty category, determine protocol adoption, and identify motivations, dopaminergic neuron yield, and protocol length.

Comparison of Protocols

For Novelty Category A and B papers, we estimated the Pearson correlation coefficient between year and quantifiable performance metrics such as neuron yield, dopaminergic neuron yield, number of control lines, and the length of time of differentiation from hPSC to mature neuron. The top five most frequently cited methods were highlighted in the respective scatter plots for comparison against the other, less-cited methods.

RESULTS

Studies Included in the Analysis of hPSC Differentiation

Our search identified 1,909 items and 750 articles met our predetermined criteria for primary literature that differentiated

hESCs or hiPSCs to neurons following abstract screening (Supporting Information Fig. S1). Three hundred thirty one of the 750 articles did not specify what types of neurons were differentiated in the abstract, and in these cases the full text was reviewed. Despite the heterogeneity of neurons subtypes found in vivo, 137 articles provided no description or validation of neuronal subtype within the abstract or full text. The rest of the articles were characterized by the type of neuron produced, and a summary of the types of neurons produced is provided in Supporting Information Table S1. Dopaminergic neurons were the most common type of neurons differentiated from hPSCs and were produced in 25% of papers in our dataset. Of 186 identified articles that involved the differentiation of hPSCs to dopaminergic neurons, 28 were excluded for being the result of undirected differentiation.

Publication and Adoption of New Dopaminergic Differentiation Protocols

The differentiation methods used in the 158 remaining dopaminergic neuron differentiation papers from hPSCs were classified by Novelty Category. Novelty Category A articles ($n = 34$) involve the development of new methods, Novelty Category B articles ($n = 40$) describe substantially modified methods, and Novelty Category C articles ($n = 84$) reuse previously published methods. The first article that differentiated hPSCs to dopaminergic neurons was published in 2004, and the annual number of articles involving the differentiation of dopaminergic neurons from hPSCs has since increased (Fig. 1A). Articles containing new or significantly modified differentiation methods continued to be published nearly every year since 2004 contributing 74 different published protocols over 13 years; however, 65% of these protocols were not used again by any of the dopaminergic differentiation papers in our dataset (Fig. 1B). Of the 26 (35%) of the methods that were reused, 46% were referenced only by articles in which they shared one or more co-author. Overall, only 19% of protocols were reused by a research group without shared co-authors with the original article.

Among the 26 Novelty Category A and B articles containing protocols that were subsequently reused, a few articles accumulated the most citations. A breakdown of these papers by year of publication and number of differentiation method citations reveals that five of these articles were especially popular among outside research groups (≥ 7 such citations each, ≥ 4 outside research groups), while the remainder were predominantly reused by shared research groups or a small number (≤ 2) of separate research groups (Fig. 1C). Three of the five most frequently cited methods were published by the Studer Lab at Memorial Sloan Kettering, and the other two were published by the Zhang Lab at the University of Wisconsin, Madison and the Yang Lab at Thomas Jefferson University. The two most cited papers were both published in journals with the highest impact factors of all of the papers describing new or significantly modified differentiation methods (Supporting Information Fig. S2).

Motivations for New Protocol Development

To investigate the incentives for new method development to differentiate hPSCs to dopaminergic neurons, we assessed each Novelty Category A and B article for statements of motivation. Seventy out of 74 papers listed some form of motivation, with

most articles listing multiple motivations. Articles most commonly mentioned improving the yield or efficiency of dopaminergic neuron production as a driving motivation for new protocol development ($n = 23$), followed by achieving defined or GMP compliant culture conditions ($n = 16$), reducing variability between batches or improving reproducibility ($n = 10$), and optimizing neurons for transplantation ($n = 20$) (Fig. 2A). Other motivations included improving similarity of hPSC-derived neurons to their in vivo counterparts, optimizing for particular types of hPSCs, and reducing the time until neuron production.

Motivations for protocol development were depicted by publication year to determine whether reasons for method development changed over time (Fig. 2B). Some motivations were listed during specific time periods, while others were mentioned consistently. Optimization for starting cell type was given as a motivation in 2004 and 2005 when methods were being developed for hESCs, in 2010 when methods were being developed for iPSCs, and then in later years as methods were being developed for specific types of hiPSCs. In contrast, efficiency or yield of neuron production was frequently listed as driving motivation for protocol development over time, with peaks of interest in 2009 and 2013. Reducing variability between batches was also a consistently popular reason for developing new methods from 2008 to 2014, while other motivations lacked clear temporal patterns.

Comparison of Popular Versus Other Protocols

To investigate potential reasons why certain protocols were cited more widely, we compared the five most popular protocols against other dopaminergic neuron differentiation protocols in our dataset based on three criteria that we identified as motivations for new protocol development: efficiency or yield of dopaminergic neurons, variability between batches or reproducibility, and time to neuron production. These motivations were chosen due to the availability of quantified measurements allowing comparisons among protocols and over time.

Yield of Dopaminergic Neurons

We extracted dopaminergic neuron quantifications from each of the Novelty Category A and B hPSC differentiation paper and found that 12% of papers did not provide dopaminergic neuron quantifications, and 11% were unclear in what quantifications were performed (Fig. 3A). Within the papers with quantifications, 15 different metrics were used to assess dopaminergic neuron yield, with some papers using combinations of metrics (Fig. 3A, Supporting Information Fig. S3). The exclusion of quantifications, lack of clarity, and use of multiple metrics made the comparison of yields between protocols difficult. All articles used tyrosine hydroxylase (TH), an enzyme essential to dopamine synthesis, as a marker for dopaminergic neurons; however, they differed in the way they presented the percentage of TH+ cells. The two most common quantifications reported were manual cell counting of the percentage of TH+ cells out of total cells and the percentage of TH+ cells out of total neurons as detected with the neuronal marker TUJ1 (Fig. 3A). Using these metrics, we found a statistically significant correlation between year of publication and the yield of TH+ neurons out of total cells ($r = .418$, $p = .030$), although we found no correlation between year of publication and the yield of TH+ neurons out of TUJ1+ cells ($r = -.245$, $p = .312$), suggesting that the development of new protocols apparently

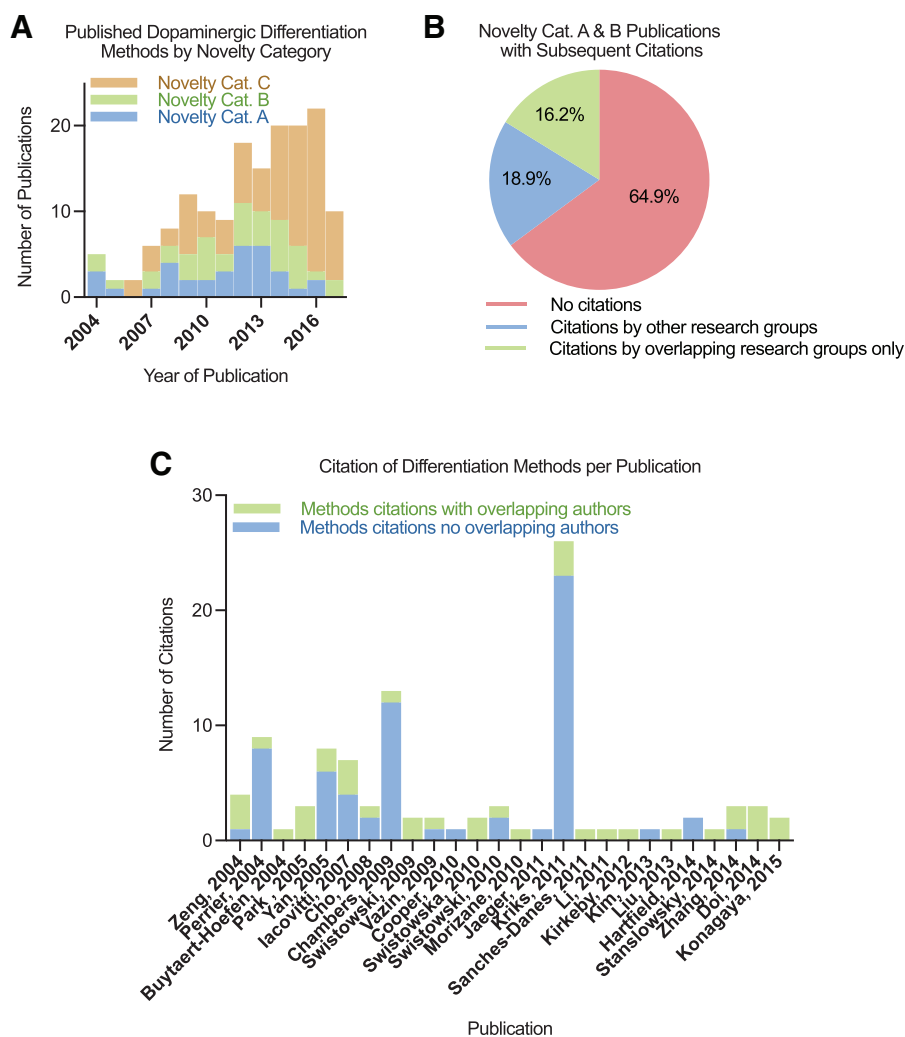


Figure 1. Analysis of hPSC dopaminergic differentiation method development and adoption. **(A):** Quantification of publications per novelty category for the differentiation of hPSCs to dopaminergic neurons. **(B):** The percentage of papers with differentiation methods cited by future publications conducted by separate or shared research groups. Publications by shared research groups contain one or more mutual co-authors. **(C):** The number of subsequent citations of the methods described per publication conducted by separate or shared research groups.

increased the overall yield of TH⁺ cells in hPSC-derived cultures, but not necessarily the percentage of neurons that acquired the dopaminergic fate (Fig. 3B, 3C). To determine if the most widely adopted protocols have higher TH⁺ cell yields, we highlighted the yield of the methods in the most frequently cited Novelty Category A and B articles methods in comparison to the rest of the Novelty Category A and B papers that use the same quantification metrics. All three of these widely adopted protocols had high yields for their year of publication (Fig. 3B, 3C).

While TH reactivity was the most commonly used marker of dopaminergic neurons, TH alone is insufficient to conclude that neurons are dopaminergic or useful for studying or treating diseases such as Parkinson's disease. TH is not specific to dopaminergic neurons as it is expressed in other catecholaminergic neurons. Additionally, the origin of dopaminergic neurons and whether they are generated from midbrain progenitors as those affected by Parkinson's disease are derived *in vivo* or from dorsal forebrain progenitors can impact their phenotype and usefulness in disease modeling and treatment. Twenty nine percent of papers that produced TH⁺ neurons did so from

midbrain progenitors (FOXA2⁺, LMX1A⁺) and 35% differentiated TH⁺ neurons from forebrain progenitors (PAX6⁺, SOX1⁺). Twenty seven percent of papers did not characterize progenitor origin, and 9% observed progenitors expressing both midbrain and forebrain markers. The proportion of papers describing midbrain versus forebrain progenitors increased over time (Fig. 3D); however, only eight papers quantified either the percentage of FOXA2⁺ out of TH⁺ cells, providing too small a sample to determine whether midbrain dopaminergic neuron yield increased over time. Only one of the five most frequently cited papers generated FOXA2⁺/LMX1A⁺ progenitors, and this paper was the most frequently cited paper methodologically.

Variability Between Batches

While many articles used multiple hPSC lines ($n = 46$), only 12 Novelty Category A and B papers provided information regarding the production of dopaminergic neurons by line individually. Therefore, we compared the number of control lines used in each Novelty Category A and B article as a proxy for reproducibility. There was no correlation between the number

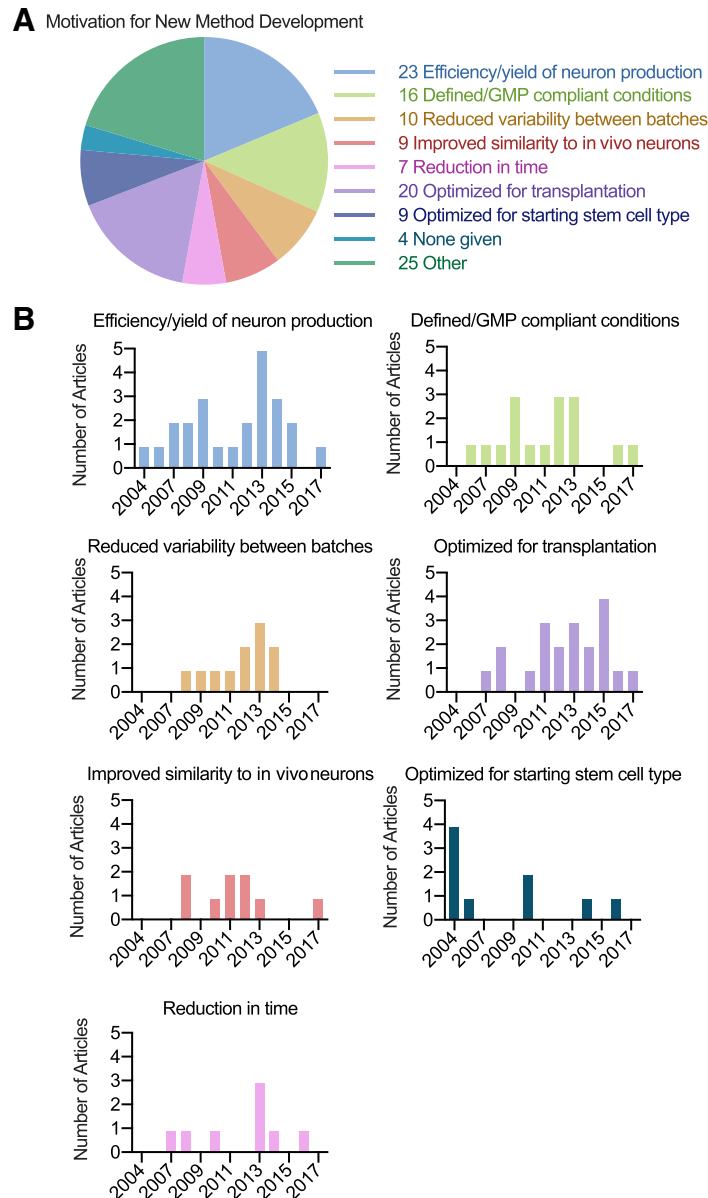


Figure 2. Motivation for the development of new methods to differentiate hPSCs to dopaminergic neurons. **(A):** Motivations were categorized based on the authors' descriptions and publications may include more than one motivation. The numbers listed depict the total number of references to each motivation. **(B):** The number of articles that list each motivation depicted by year of publication. Abbreviation: GMP, good manufacturing practice.

of lines used and the year of publication ($r = .193$, $p = .107$). Four out of the five most popular papers used at least three independent control hPSC lines, which is slightly above the average number of 2.24 lines used per study but the difference was not statistically significant (Fig. 3E).

Time to Neuron Production

The temporal length of protocols to reach neuron production was a reported motivation in seven published methods. Longer protocols delay the availability of neurons for experimentation or transplantation, making a reduction in protocol length valuable to the field. A plot of the time from hPSCs to neurons for each Novelty Category A and B article published per year revealed no correlation ($r = .129$, $p = .333$), suggesting that differentiation length did not decrease with new protocol

development (Fig. 3F). The length of the five most frequently referenced protocols was not significantly different from contemporaneously published papers ($p = .581$), suggesting that protocol length was not a key contributing factor driving the popularity of these protocols.

Small Molecules Used to Induce Differentiation

Across the 37 hPSC studies that provided one or both of the two most common types of dopaminergic neuron yield quantifications, we collected information on the types of small molecules used to induce differentiation as well as their concentration and length of treatment (Supporting Information File S1). A total of 27 different small molecules were reported to be used to induce dopaminergic neuron differentiation from hPSCs (LDN193189, SB431542, dorsomorphin, noggin, smoothed agonist, sonic

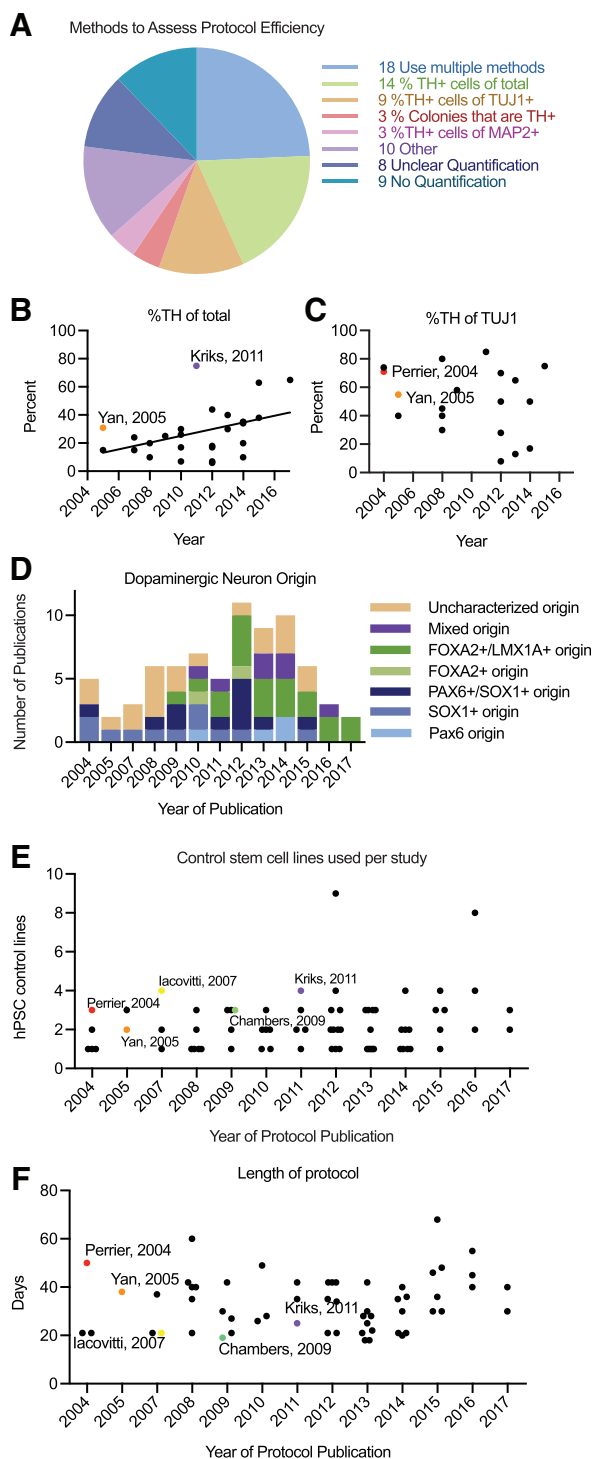


Figure 3. Analysis of commonly listed motivations among papers with new methods or significantly modified methods. **(A):** Number of articles using metrics to quantify yield of dopaminergic neurons. **(B):** Percentage of TH+ cells out of total cells. **(C):** Percentage of TH+ cells out of TUJ1+ cells. **(D):** The number of articles describing dopaminergic neuron differentiation by progenitor origin marker. **(E):** Number of control hPSC lines used per study. **(F):** Temporal length of protocols to differentiate dopaminergic neurons from hPSCs. Frequently cited methods are depicted with colored dots. The five frequently cited methods did not differ significantly from other methods in lines used or length of protocol ($p > .05$ by t test). Abbreviations: hPSC, human pluripotent stem cells; TH, tyrosine hydroxylase.

hedgehog, purmorphamine, CHIR99021, brain-derived neurotrophic factor (BDNF), glial cell-derived neurotrophic factor (GDNF), ascorbic acid, transforming growth factor beta-3, cyclic adenosine monophosphate (cAMP), interleukin 1 β , N-[N-(3,5-difluorophenacetyl)-L-alanyl]-S-phenylglycine t-butyl ester, fibroblast growth factor 8 (FGF8), basic fibroblast growth factor (also known as FGF2), fibroblast growth factor 20, epidermal growth factor, heparin, laminin, WNT family member 1 (WNT1), retinoic acid, leukemia inhibitory factor, stromal cell-derived factor 1a, secreted frizzled related protein 1, and vascular endothelial growth factor D). The median number of those reported was 6 (interquartile range 3–9) across these studies. As shown in Supporting Information file 1, there were no two studies that reported an identical mix of small molecules. Only four compounds were reported in more than 20 studies, but even for those the dose and/or duration varied: BDNF (10–25 ng/ml, 7–37 days), GDNF (2 ng/ml–20 μ g/ml, 4–37 days), cAMP (1 μ M–1 mM, 4–37 days), FGF8 (20 ng/ml–100 mg/ml, 4–31 days).

Comparison to Mesenchymal Stem Cell Differentiation Protocols

hMSCs are an alternative source of multipotent stem cells to hPSCs. Like hPSCs, they are able to differentiate into a number of cell types; however, they do not have the pluripotent capabilities of hPSCs and therefore may pose a smaller risk of producing unwanted cell types following transplantations. To compare protocols for hMSCs differentiation protocols and hPSC differentiation, we identified 16 articles in our original PubMed search that differentiated hMSCs to dopaminergic neurons. Of these articles, nine involved the development of new methods (Category A), five described substantially modified methods (Category B), and two reused previously published methods (Category C, Supporting Information Fig. S4A). Only one of these articles was cited methodologically by a subsequent publication in our data set, and this study was conducted by an overlapping research group. Interestingly, none of the hMSC differentiation articles cited hPSC articles methodologically or vice versa. The motivations given for new protocol development among Category A and B hMSC differentiation articles were conceptually similar to those given in the hPSC differentiation articles and included efficiency or yield of neuron production ($n = 3$), improved similarity to *in vivo* neurons ($n = 1$), optimization for transplantation ($n = 4$), and optimization for starting cell type ($n = 4$, Supporting Information Fig. S4B). Eight of the Category A and B hMSC differentiation papers quantified the percentage of TH+ cells out of all cells and two quantified the percentage of TH+ cells out of TUJ1+ neurons. The yields reported in these articles were similar to those reported in hPSC differentiation articles (Supporting Information Fig. S4C, 4D). However, 10 out of the 11 hMSC differentiation articles that reported protocol length were shorter than the hPSC differentiation protocols published in the same year (Supporting Information Fig. S4E). This difference in protocol length may be due to the relatively mature starting state of hMSCs compared to hESCs and hiPSCs.

DISCUSSION

Extensive effort has been put into developing protocols to differentiate hPSCs into specific cell types for applications in

disease modeling and cell transplantation. However, it is unclear to what extent the development of new protocols contributes to the field, especially as new protocols are published for years after initial methodologies were developed. Using the differentiation of dopaminergic neurons as a case study, we found that 74 new or substantially modified methodologies for dopaminergic neuron differentiation have accrued over 13 years since the first protocols were published. However, the majority of these methods were not used again, and only 19% were used by groups that were independent of the original developers. Similar patterns of fragmentation of effort and translational inefficiency were seen also in 16 differentiation protocols based on hMSCs.

To understand the incentives driving new protocol development, we identified and categorized statements of motivation from papers involving new or significantly modified methods. It is possible that some of these statements do not represent the initial motivations driving method development and are rather benefits discovered post hoc; however, their mention indicates some degree of importance to the field. Almost all studies with new protocol development stated some motivation, which suggests that authors almost ubiquitously make an effort to justify that they have something worthwhile to add to the literature. However, we should caution that mentioning a motivation in the development of a new method does not mean necessarily that the specific method was more successful than other methods in this regard. We therefore set out to investigate whether some key metrics improve over time with new method development and whether methods that are more widely adopted perform favorably in some specific areas.

Inconsistencies in Reporting Preclude Protocol Comparisons

An examination of dopaminergic neuron yield in newly developed protocols quickly revealed that numerous different metrics were being used to quantify yield, making cross-study comparison challenging. Additionally, the metrics chosen to report yield can influence how the results are interpreted. Standardizing how the yield of the desired cell type is quantified and reported would enhance the community's ability to make accurate comparisons between methods. In the current environment where multiple definitions abound and there is no common standard, authors may cherry pick one or a few metrics where the method appears to have best performance.

In addition to yield, another important characteristic when choosing a protocol is the origin of the dopaminergic neurons produced. Many experiments using *in vitro*-derived dopaminergic neurons aim to study or develop treatments for Parkinson's disease, which affects dopaminergic neurons located in the substantia nigra of the midbrain. It is therefore relevant whether dopaminergic neurons are produced from midbrain neural progenitors *in vitro*. Twenty seven percent of the hPSC-based papers did not report any characterization of progenitor origin, making it impossible to know whether forebrain or midbrain markers were not expressed, not assayed, or assayed but not reported. Additionally, of the 29% of the hPSC-based articles that did identify midbrain progenitor markers, it is unclear how many of these also express a mixture of forebrain progenitor markers as observed in 9% of papers that report both forebrain and midbrain marker expression. More

consistent reporting of cell origin will allow better selection of protocols that produce cell types that are most authentic to their *in vivo* counterparts.

Comparing protocols for variability between batches of differentiations also proved to be challenging as the majority of papers did not include data from individual hPSC lines. The publication of only aggregate data, even when measurements of variance are provided, precludes the possibility of distinguishing between factors contributing to the variance, such as the performance of an individual hPSC line or the same lines over repeated differentiations. Using the number of control lines as a proxy for an assessment of variability between lines, we found no increase in the number of control lines used in newly developed methods over time. hiPSC lines can vary greatly in gene expression and differentiative tendencies, creating a barrier for *in vitro* studies and clinical translation [6–8]. Genes that vary in expression among hiPSC lines are enriched for pathways involved in development [6], and it is therefore unsurprising that studies comparing the differentiation of multiple hiPSC lines to cells of the neuroectodermal lineage vary in yield [9]. Currently, there is no consensus on the appropriate number of lines to use in a study involving iPSC differentiation, but the fewer lines used in a study the higher the likelihood of specific traits of an individual line influencing the outcomes. While maintaining and differentiating large numbers of lines can be costly and labor intensive, it is particularly important to invest in using many iPSC lines when developing a new differentiation method since creating a robust protocol that works similarly across lines may improve the reproducibility of subsequent studies. Additional simple, cost-free steps can be taken to address issues stemming from iPSC variability, such as making data comparing the performance of individual lines available upon publication.

Examination of Protocol Length

Examining the length of time of protocols from hPSCs to dopaminergic neurons showed no correlation with year of publication. When each experiment requires a month of preparation, progress in the field is hindered. The temporal length of protocols is also problematic for applications in cellular transplantation therapies. The use of patient-derived iPSCs to create autologous cells for transplantation is one of the promises of stem cell research, but an extensive length of time required for the reprogramming and differentiation of iPSCs prevents treatment for diseases that require rapid intervention. It is possible that with the current strategies, there is a minimum amount of time that dopaminergic neurons cells require to mature *in vitro*, thus explaining the lack of reduction in protocol length over the past 13 years.

Unclear Motivations Driving Protocol Adoption

Given the abundance of protocols to differentiate hPSCs to dopaminergic neurons, it is a natural question as to why some are adopted by other research groups and others are not. It is possible that protocols that were not adopted were tried by other groups and not published due to failure to replicate or to function well in their hands. Other potential reasons as to why one protocol might be more popular include the status of the lab and the journal in which it was published. The two protocols that were most frequently cited methodologically were published in journals with the highest impact factors.

However, we cannot distinguish whether these methods became popular due to their visibility in high profile journals, or whether they were published in high profile journals due to the superiority of their methods. These papers often also demonstrated applications of their differentiated dopaminergic neurons, and publication in such high impact factor journals could reflect other findings within these papers besides the differentiation method itself.

To investigate whether protocols were chosen based on the factors identified in our motivation analysis, we compared the five most popular protocols to contemporaneously published methods on the basis of yield, variability, and temporal length. These five protocols performed well in terms of yield, but were not remarkably favorable overall. This suggests that either the lack of consistent reporting obscures the superiority of these protocols along these metrics, or there is a disconnect between the qualities motivating new protocol development and the qualities that lead to a protocol being more widely adopted.

Establishing Reporting Guidelines to Facilitate Method Development and Comparison

Creating new methods requires a great deal of time and resources. It is therefore important to investigate why new protocols continue to be developed, why some methodologies are adopted over others, and whether the persistent development of new methods benefits the field as a whole. Ultimately, the neurons created using these methods act as the starting point for investigations into cellular mechanisms and disease and can potentially be used for cellular transplantation therapies. A recent study of transplantation efficiencies found that expression of genes that have previously been used to identify dopaminergic progenitors and neurons during *in vitro* differentiation including *FOXA2*, *LMX1A*, *CORIN*, *TH*, *NURR1*, and *DDC* do not correlate significantly with dopaminergic yield following engraftment into rats, suggesting that the current array of markers used in assessing protocol yields are not ideal for optimizing success rate for transplantation [10]. The same study identified genes of the caudal ventral mesencephalon (VM) and midbrain-hindbrain boundary including *PAX5*, *FGF8*, *SPRY1*, *EN1/2*, *SP5*, *ETV4/5*, *CNPY1*, *TLE4*, and *WNT1* as correlated with dopaminergic neuron yield following transplantation *in vivo* [10]. An additional study that performed single cell sequencing on *LMX1A*+ cells from the mouse ventral mesencephalic and diencephalic region found that several genes associated with dopaminergic neuron production including *Lmx1a/b*, *Foxa1/a2*, *Otx2*, *Msx1*, and *Nurr1* are also expressed in progenitors that ultimately produce glutamatergic subthalamic nucleus neurons, and like Kirkeby et al. (2017), they also found that genes expressed in the caudal VM, *En1/2*, *Wnt1*, and *Cnpy1*, were useful for identifying progenitors that produce dopaminergic neurons [11]. Together these studies suggest that transition from the markers previously used in the differentiation protocols assessed in this study

to using markers for a caudal VM may improve the comparison of dopaminergic neuron differentiation protocols for the purpose of transplantation. Additional studies comparing marker expression to desired outcomes in transplantation or disease modeling applications will be helpful in identifying the most useful standards for protocol evaluation.

CONCLUSION

The results of this study indicate a lack of understanding regarding the extent to which new method development benefits the field. Inconsistency in reporting makes it difficult for researchers to draw comparisons and make educated decisions for protocol selection. Manageable steps can be taken to standardize the transparency and reporting of important metrics to make inter-study comparisons and replications more feasible [12, 13]. For example, biases and inconsistencies in reporting have been previously identified as an issue in clinical studies, leading to the creation of reporting guidelines such as CONSORT [14]. These guidelines inspired the development of similar frameworks in other fields and ultimately the establishment of the Enhancing the QUALity and Transparency Of health Research (EQUATOR) network, an international initiative that encourages good reporting practices by serving as a guideline repository and educational initiative for responsible reporting of various types of studies [14–16]. While standards for the ethical conduct and portrayal of stem cell research and publication of stem cell-related clinical trials have been developed, similar guidelines relating basic stem cell research with an outlook for translational, clinical application are not covered by these initiatives [17, 18]. The development of a network to serve as a repository for stem cell differentiation protocols and to establish reporting guidelines would be useful in ensuring optimal, reproducible differentiation practices for disease modeling, and clinical applications. In addition, a discussion of issues such as the number of cell lines that should be used and the reporting of consistent quantifications of performance metrics is essential to best use efforts in new method development and to optimally apply stem cell-derived models to *in vitro* studies and clinical translation.

AUTHOR CONTRIBUTIONS

R.M.M.: conception and design, collection of data, data analysis and interpretation, manuscript writing; J.P.A.I.: conception and design, interpretation, manuscript writing.

DISCLOSURE OF POTENTIAL CONFLICTS OF INTEREST

The authors indicated no potential conflicts of interest.

REFERENCES

- 1 Thomson JA, Itskovitz-Eldor J, Shapiro SS et al. Embryonic stem cell lines derived from human blastocysts. *Science* 1998;282:1145–1147.
- 2 Takahashi K, Tanabe K, Ohnuki M et al. Induction of pluripotent stem cells from adult

human fibroblasts by defined factors. *Cell* 2007;131:861–872.

- 3 Mestas J, Hughes CC. Of mice and not men: Differences between mouse and human immunology. *J Immunol* 2004;172:2731–2738.

- 4 Martignoni M, Groothuis GM, de Kanter R. Species differences between mouse,

rat, dog, monkey and human CYP-mediated drug metabolism, inhibition and induction. *Expert Opin Drug Metab Toxicol* 2006;2:875–894.

- 5 van Servellen A, Oba I. Stem cell research: Trends in and perspectives on the evolving international landscape. *Research Trends* 2014.

6 Carcamo-Orive I, Hoffman GE, Cundiff P et al. Analysis of transcriptional variability in a large human iPSC library reveals genetic and non-genetic determinants of heterogeneity. *Cell Stem Cell* 2017;20:518–532. e9.

7 Liang G, Zhang Y. Genetic and epigenetic variations in iPSCs: Potential causes and implications for application. *Cell Stem Cell* 2013;13:149–159.

8 Tapia N, Scholer HR. Molecular obstacles to clinical translation of iPSCs. *Cell Stem Cell* 2016;19:298–309.

9 Hu BY, Weick JP, Yu J et al. Neural differentiation of human induced pluripotent stem cells follows developmental principles but with variable potency. *Proc Natl Acad Sci USA* 2010;107:4335–4340.

10 Kirkeby A, Nolbrant S, Tiklova K et al. Predictive markers guide differentiation to improve graft outcome in clinical translation of hESC-based therapy for Parkinson's disease. *Cell Stem Cell* 2017;20:135–148.

11 Kee N, Volakakis N, Kirkeby A et al. Single-cell analysis reveals a close relationship between differentiating dopamine and subthalamic nucleus neuronal lineages. *Cell Stem Cell* 2017;20:29–40.

12 Munafò MR, Nosek BA, Bishop DV et al. A manifesto for reproducible science. *Nat Hum Behav* 2017;1:0021.

13 Ioannidis JP. How to make more published research true. *PLoS Med* 2014;11:e1001747.

14 Altman DG, Simera I. A history of the evolution of guidelines for reporting medical

research: The long road to the EQUATOR Network. *J R Soc Med* 2016;109:67–77.

15 Simera I, Moher D, Hirst A et al. Transparent and accurate reporting increases reliability, utility, and impact of your research: Reporting guidelines and the EQUATOR Network. *BMC Med* 2010;8:24.

16 Moher D, Simera I, Schulz KF et al. Helping editors, peer reviewers and authors improve the clarity, completeness and transparency of reporting health research. *BMC Med* 2008;6:13.

17 Caulfield T, Sipp D, Murry CE et al. Confronting stem cell hype. *Science* 2016;352:776–777.

18 Daley GQ, Hyun I, Apperley JF et al. Setting global standards for stem cell research and clinical translation: The 2016 ISSCR guidelines. *Stem Cell Reports* 2016;6:787–797.



See www.StemCellsTM.com for supporting information available online.