

Special Issue: Immunoregulation in regenerative medicine

Mini Review

Induced pluripotent stem cell generation-associated point mutations

Ryoko Araki, Mayumi Sugiura[§], Yuko Hoki, Misato Sunayama, Miki Nakamura, Yasuji Kasama and Masumi Abe^{*}

Transcriptome Research Group, National Institute of Radiological Sciences, Chiba, Japan

The discovery of induced pluripotent stem cells (iPSCs) has the potential to drastically alter the future of medicine. Further, the achievement of cell lineage conversion by gene transduction is expected to make a dramatic contribution to the advancement of basic biology. Currently, various iPSC applications have been the focus of much attention due to their potential in regenerative medicine. Nevertheless, because the molecular mechanisms underlying the creation of these cells have remained elusive, confidence in their safe use in a clinical setting has remained rather shaky. In our present review, we discuss genome integrity during iPSC generation, with a particular focus on point mutations, to further address the issue of whether iPSC generation causes genetic aberrations.

Rec.9/7/2015, Acc.//2015, pp226-232

*Correspondence should be addressed to:

Masumi Abe, Transcriptome Research Group, National Institute of Radiological Sciences, Chiba 263-8555, Japan. Phone: +81-43-206-4692, Fax: +81-4-206-4593, E-mail: abemasum@nirs.go.jp

[§]Present address: Division of Natural Sciences, Research Group of Biological Sciences, Nara Women's University, Nara 630-8506, Japan

Key words iPS cell, Point mutation, reprogramming, transversion, whole genome sequencing

Introduction

The maintenance of genome integrity is of critical importance in the iPSC research field. In this review, we focus on point mutations that arise in the genomes of these stem cells. With the evolution of next generation DNA sequencing methods, genome wide point mutation analysis has become possible, and the presence of single nucleotide variations (SNVs) in the iPSC genome was first demonstrated by exome sequencing of a large number of iPSCs and detemined to be independent of delivery method used to produce these cells¹⁾ (Table 1). The accumulation of point mutations in iPSC genomes has now been confirmed by many groups using exome and whole genome sequencing (WGS)²⁻⁶⁾. Substantial numbers of SNVs (300 ~1,000/entire genome) have been observed in all of the iPSC lines analyzed to date, regardless of the underlying

Table 1. St	udy of iPSC	C SNVs
-------------	-------------	--------

references	species	somatic cell types	reprogramming method	No.of iPSCs	sequencing method	No. of coding SNVs / line (average)	No. of genome SNVs / line (average)
Gore A, <i>et al</i> ., 2011. ¹⁾	human	fibroblasts	viral, mRNA and episomal	22	exome seq.	2-14 (6)	nt
Ji J, <i>et al</i> ., 2011. ²⁾	human	fibroblasts	viral	5	exome seq.	5-16 (6)	nt
Cheng L, <i>et al</i> ., 2012. ³⁾	human	bone marrow cells (CD34+) and mesenchymal stem cells	episomal	3	whole-genome seq.	6-12 (8)	1,058-1,808 (1,325)
Young MA, <i>et al</i> ., 2012. ⁴⁾	mouse	fibroblasts	viral	10	whole-genome seq.	3-22 (11)	190-773 (512)
Ruiz S, <i>et al</i> ., 2013. ⁵⁾	human	keratinocytes, astrocytes, endothelial cells, neural stem cells and mesenchymal stem cells	viral	8	exome seq.	2-12 (5)	nt
Sugiura M, <i>et al.</i> , 2014. ⁶⁾	mouse	fibroblasts	episomal and viral	6 iPSCs and 4 ESCs	whole-genome seq.*	iPSCs: 1-7 (3.5) ESCs: 0-1 (0.5)	iPSCs:215-574 (362) ESCs:13-37 (23)**

nt: not tested

*genome coverage shared with parental cells: 55-61%

**estimated number of SNVs on whole genome region: iPSCs 352-990 (average 640), ESCs 23-67 (average 41)

species or somatic cell type. There is no doubt that the origin of these SNVs is one of the most important issues to be resolved if iPSCs are to have a clinical future. Intriguingly, sample testing focusing on randomly selected SNVs identified from iPSC genomes showed that a substantial body of these mutations already existed in the parent somatic cell¹. In addition, an analysis of sister iPSC clones derived from the same parent somatic cells showed the presence of a large number of commonly shared SNVs⁴. Based on these observations, it has been proposed that most of the SNVs observed in iPSC genomes already existed in the parent cells (hereafter referred to as pre-existing SNVs). On the other hand, it was also suggested that a significant number of SNVs were not pre-existing². Hence, this is still the subject of debate^{2, 7}.

Needless to say, whether the process of iPSC generation intrinsically causes point mutations or not is a crucial issue in terms of the use of these pluripotent stem cells in regenerative medicine, as this has a direct impact on the immunogenicity and tumorigenic potential of these stem cells and their derivatives⁸⁻¹¹. Hence, if iPSC generation causes point mutations to arise, their number, type and location in the genome must be known. A fuller understanding of the mechanisms governing this phenomenon would also enable us to reduce the frequency of point mutations and also provide some important clues to the mechanisms of genome reprogramming.

Exclusive identification of *de novo* point mutations

The ability to exclusively identify de novo point mutations and control for any false background signals is critical for genome-wide point mutation analysis. Following the comprehensive identification of SNVs in iPSC genomes, we used an extremely sensitive sequencing technique known as 'ultra-deep sequencing' to evaluate whether these were pre-existing or *de novo*. Using this sensitive method, only the small regions encompassing SNVs are amplified by PCR from the parent somatic cell genome and sequenced with a standard next-generation methodology. By using this approach, more than 10⁵ reads can be obtained for each SNV region, enabling the detection of the SNV of interest even when it is present in an extremely small number of cells within the parent cell population. Previously reported ultra-deep sequencing results have revealed that significant bodies of such SNVs already existed in the iPSC parent cell genomes, though in a small portion of these parent genomes¹⁾. Furthermore, the possibility cannot be excluded that there are pre-existing but very low frequency SNVs in the parent cell population that fall below the threshold of detection by ultra-deep sequencing. Additionally, common SNVs have been identified among sister iPSC clones⁴. Through such observations, the concept was developed that most SNVs in iPSCs are pre-existing and that the generation of these stem cells does not induce point mutations¹²⁾.





Fig. 1 Comparison of the point mutation profiles of iPSCs versus ESCs

(A)The number and the distribution of point mutations at each chromosome. The dots indicate the positions of the point mutations. The 2A-4F-118, -119, and -136 iPS cell lines were established using episomal vectors. The numbers of identified SNVs are shown in parentheses.

(B)Mutation numbers. The Y-axis indicates the number of mutations per 10⁹ bp genome.

(C)Ratio of transitions to transversions.

Direct comparison was as follows,

1)between the head of single embryos and iPSCs; because the amount of MEF from single embryo is very limited, we used the head of each single embryo.

2)between the livers of both parents and ESCs.

The question that arises from the evidence to date is whether most SNVs identified in iPSC genomes are indeed pre-existing and arose in the parent cell population spontaneously. The concern however is that such preexisting SNVs can be explained by SNPs. This is because significant numbers of unknown and intra-strain SNPs exist even in inbred mice, and also as the parent mouse embryonic fibroblasts (MEFs) used for iPSC generation have usually been prepared from multiple embryos. To overcome this situation in our laboratory, we employed single embryo-derived MEFs and directly compared our iPSC lines with their parent embryos (direct comparison). Furthermore, we conducted analyses with new approaches focusing on base substitution types and on homogeneity within iPSC colonies⁶⁾. The use of single embryo-derived MEFs allowed us to accurately and effectively identify de novo point mutations and determine whether there were pre-existing or common SNVs among sister iPSCs established from the same parent MEF.

Supporting evidence for iPSC generationassociated point mutations 1)ESCs vs iPSCs

A comparison between iPSCs and the gold standard of pluripotent stem cells, embryonic stem (ES) cells, has provided critical information about various aspects of iPSCs, including genetic aberrations. However, there has been no success to date with genome-wide point mutation analysis of ES cells because it has not been possible to evaluate their parent cell genomes. A comprehensive profile of the point mutations in ES-cell genomes would provide a vital clue to addressing whether genome reprogramming causes such mutations. In this regard, we have in our laboratory newly established ES cell lines from C57BL/6 mice and also obtained both parent genomes from the livers of these animals. Further, we have passaged ESCs and iPSCs a similar number of times (~3 passages) to minimize any differences that could occur due to spontaneous point mutations. The pluripotency of our ES cell lines was verified



using a chimera mouse formation test.

Point mutation analysis of these pluripotent stem cell lines via whole genome sequencing revealed an approximately 10-fold higher frequency of point mutations in iPSC genomes compared with ESC genomes and also clear differences in the base substitution pattern, (transition-predominant for ES cells but transversion-predominant for iPSCs) (Fig. 1)⁶. This first genome-wide point mutation analysis of ESC genomes strongly suggested that the point mutations observed in iPSC genomes arose during the genome reprogramming process.

2)Substantial frequency of de novo SNVs

Even when using single embryo-derived MEFs as the parental cells in our laboratory, we still identified a large number of point mutations within iPSC genomes. We conducted a conventional investigation through the ultradeep sequencing of material MEF genomes for randomly selected SNVs identified in iPSC genomes to determine how many pre-existing SNVs were detected by our system. In the results, no sequence reads was observed in the parent MEF genomes for any of the SNVs assessed. In addition, very few common SNVs were observed among the sister clones established from the same MEF fraction. Taken together, these data indicate that our system of employing a direct comparison between iPSC genomes and their parent embryo genomes, and the use of single embryo-derived MEFs, detects de novo point mutations exclusively.

3)A unique base-substitution profile

The accumulation of substantial numbers of *de novo* SNVs allowed us to conduct a statistically significant investigation of the base substitution types in each case. The results of this analysis were of great interest: transversion-predominant base substitutions were detected in the SNVs in iPSC genomes whereas transition-predominant mutations were found in the ESC genomes (Fig. 1C)⁶. Point mutations can be categorized into transition and transversion types. Transitions are interchanges between purines ('A' \Leftrightarrow 'G') or pyrimidines ('C' \Leftrightarrow 'T'), resulting in a similar DNA configuration. In contrast, transversions are interchanges between purines and pyrimidines.

Moreover, additional analysis that incorporated the timeline of iPSC generation clearly indicated that the transversion-predominant SNV pattern is reprogramming process-specific (Fig. 2A)⁶⁾. In contrast, SNVs showed a



Fig. 2 Base substitution profile during each step of iPSC clone establishment

(A)Base substitution profile during the "reprogramming" process showing a predominantly transversion pattern. ts, transition; tv, transversion. 'Pre-existing', SNVs identified in parental somatic cells by ultra-deep sequencing or shared between sister clones; 'reprogramming', SNVs not identified in parental somatic cells by ultradeep sequencing or not shared among sister clones; 'passaging'; SNVs newly identified in extended cultures. The two studies, by Gore et al. and Ji et al., used ultra-deep sequencing with a focus on randomly selected point mutation candidates, but the report by Young et al. discriminated between each SNV based on whole genome sequencing data from sister clones derived from the same somatic cell fraction: shared SNVs were categorized as 'pre-existing' but each sister clone-specific (private) SNV was defined as not pre-existing, i.e. was the result of 'reprogramming' or 'passaging'^{1,2,4)}. Note that the former two studies employed exome sequencing followed by sampling using amplicon sequencing.

(B)Base substitution profiles of SNPs. Approximately 15 million SNPs from the dbSNP database were analysed.

Special Issue (Mini Review) Heterogeneity of point mutations within an iPSC colony Inflammation and Regeneration Vol.35 No.5 November 2015



Fig. 3 Heterogeneity within an iPSC colony

(A)Establishment of sublines from single cells within an iPSC colony.

(B)Timing of the occurrence of point mutations. We identified a large number of *de novo* SNVs by whole iPSC genome sequencing and examined these variants in each single cell-derived subline to investigate clonal heterogeneity. This schematic representation indicates the timing of the occurrence of point mutations at the conversion from fibroblasts to iPSCs.

transition-predominance before and after the generation steps. Namely, both the pre-existing SNVs which had been identified by ultra-deep sequencing and the SNVs which arose during the prolonged culture process to establish iPSC clones exhibited a transition-predominance. Hence, our investigation concluded that iPSC generation-associated point mutations have a unique base substitution profile that is transversion-predominant.

The reason for the transversion-predominance of the base substitutions in iPSC genomes became an issue of great interest and was also a promising clue to the mechanism behind the onset of these mutations. Because transition-predominant base substitutions arise in ESC genomes and are responsible for most SNPs (Fig. 2B)⁶, the occurrence of the SNVs identified in iPSC genomes cannot be explained by usual spontaneous point mutations.

4)Heterogeneity within an iPSC colony

Although the large number of *de novo* point mutations in iPSCs, the unique base substitution profile within iPSC genomes, and the clear differences between ES cells and iPS cells in this regard, was a strong indication of the presence of iPSC generation-associated point mutations, direct evidence for this was still lacking. To obtain such evidence, we next focused on the allele frequency of each SNV within an iPSC clone.

Almost all SNVs that have been reported in iPSC genomes to date exhibit only a 50% allele frequency, and not less than a 50% frequency, indicating that they arose in only one allele and are present in every cell in an iPSC colony⁴⁾. This simple fact seems to be crucially important because if these were not pre-existing SNVs, they would have to have arisen in almost all instances via a single step in an ancestral fibroblast which had subsequently developed into a pluripotent stem cell colony. If this was the case, a huge number of the point mutations identified in iPSCs would have to arise simultaneously during the initial stages of their generation. This challenging concept warrants a consideration that these SNVs already existed in the parent cells. We conducted an investigation based on the hypothesis that iPSC generation-associated point mutations arise in a short period around the initial stages of the generation of these stem cells. If this hypothesis was indeed correct, a heterogeneous point mutation pattern would be expected within a single iPSC colony, and this could be detected even with whole genome sequencing.

Furthermore, to verify the heterogeneity suggested by WGS, we independently established sublines from individual cells in iPSC clone and investigated the point mutation pattern in each case. To establish a single cell-



derived cell line, we picked randomly selected single cells with a micro-capillary under a microscope, transferred them into a 96-well culture dish, and propagated them under standard culture conditions (Fig. 3A). Because our deep whole genome sequencing of the iPSC genome successfully detected a ~25% variant allele frequency peak in addition to a ~50% frequency, we focused on the SNVs at a ~25% allele frequency and investigated these variants in each individual cell-derived subline. The results clearly demonstrated a heterogeneity in the mutation profiles in each iPSC clone and the history of the emergence of each mutation was also revealed (Fig. 3B)⁶⁾. Thus, cells with different point mutation profiles were found to be present within a single iPSC clone, and such mutations arose within only a few cell divisions of the lineage conversion from the parent fibroblast. Crucially, this was direct evidence of point mutations associated with iPSC generation.

Although variant allele frequencies in the iPSC clones were evaluated with WGS followed by amplicon sequencing in a previous study, the presence of subclones showing lower allele frequencies (such as a ~25% frequency) was confirmed in some of the 10 iPSC clones examined⁴⁾. The allele frequencies of 9/10 of these clones averaged 50%, suggesting that most variants were present in nearly all of the cells in each clone. With regard to the lower allele frequent SNVs observed in a single iPSC clone, this was considered to be the consequence of the growth expansion of a cell in which certain point mutations providing a selective advantage occurred during iPSC generation or subsequent prolonged culture. A similar mechanism of serial point mutations during passaging has been suggested previously for iPSC generation as well as tumorigenesis¹³⁾. In contrast, our current analysis focusing on individual cells within a colony demonstrated that lower frequency SNVs are basically not the result of the expansion of certain cells through any growth advantage⁶⁾.

Conclusions and future directions

The new approaches we have employed to further investigate the biology and mutation profile of iPSCs have produced findings that a number of unique point mutations emerge during the iPSC generation process which is contrary to the widely accepted view. However, it must be still noted that the possible existence of iPSC generationassociated point mutations does not necessarily mean that the generation process for these stem cells intrinsically causes point mutations. The possibility that iPSCs can be established with only spontaneous point mutations has not been totally excluded by the current data. Although a number of point mutations have actually been observed in all of the iPSC genomes analyzed thus far without exception, a significant difference in their number has been suggested among iPSCs established using different methods. Further clarification of the molecular mechanisms behind this phenomenon is crucial to address this issue and also to exploit the possibility of reducing the point mutation frequency in iPSCs. Needless to say, a better understanding of these molecular mechanisms would also yield important clues regarding the genome reprogramming mechanism itself. Finally, it is noteworthy that the onset of transversionpredominant point mutations is the first molecular signature to be identified for the initial steps of iPSC generation.

Acknowledgments and Sources of funding

We thank Akemi Ishibashi, Hana Yoshida and Yukiko Shindo for technical assistance. This work was partially funded by Precursory Research for Embryonic Science and Technology and Grant-in-Aid for Scientific Research, 25290068, JSPS.

Conflicts of interest

None

References

- Gore A, Li Z, Fung HL, Young JE, Agarwal S, Antosiewicz-Bourget J, Canto I, Giorgetti A, Israel MA, Kiskinis E, Lee JH, Loh YH, Manos PD, Montserrat N, Panopoulos AD, Ruiz S, Wilbert ML, Yu J, Kirkness EF, Izpisua Belmonte JC, Rossi DJ, Thomson JA, Eggan K, Daley GQ, Goldstein LS, Zhang K: Somatic coding mutations in human induced pluripotent stem cells. Nature. 2011; 471: 63-67.
- 2) Ji J, Ng SH, Sharma V, Neculai D, Hussein S, Sam M, Trinh Q, Church GM, McPherson JD, Nagy A, Batada NN: Elevated coding mutation rate during the reprogramming of human somatic cells into induced pluripotent stem cells. Stem Cells. 2012; 30: 435-440.
- 3) Cheng L, Hansen NF, Zhao L, Du Y, Zou C, Donovan FX, Chou BK, Zhou G, Li S, Dowey SN, Ye Z, Chandrasekharappa SC, Yang H, Mullikin JC, Liu PP: Low incidence of DNA sequence variation in human induced pluripotent stem cells generated by nonintegrating plasmid expression. Cell Stem Cell. 2012; 10: 337-344.
- 4) Young MA, Larson DE, Sun CW, George DR, Ding L,



Miller CA, Lin L, Pawlik KM, Chen K, Fan X, Schmidt H, Kalicki-Veizer J, Cook LL, Swift GW, Demeter RT, Wendl MC, Sands MS, Mardis ER, Wilson RK, Townes TM, Ley TJ: Background mutations in parental cells account for most of the genetic heterogeneity of induced pluripotent stem cells. Cell Stem Cell. 2012; 10: 570-582.

- 5) Ruiz S, Gore A, Li Z, Panopoulos AD, Montserrat N, Fung HL, Giorgetti A, Bilic J, Batchelder EM, Zaehres H, Scholer HR, Zhang K, Izpisua Belmonte JC: Analysis of protein-coding mutations in hiPSCs and their possible role during somatic cell reprogramming. Nature communications. 2013; 4: 1382.
- 6) Sugiura M, Kasama Y, Araki R, Hoki Y, Sunayama M, Uda M, Nakamura M, Ando S, Abe M: Induced pluripotent stem cell generation-associated point mutations arise during the initial stages of the conversion of these cells. Stem Cell Reports. 2014; 2: 52-63.
- 7) Hussein SM, Elbaz J, Nagy AA: Genome damage in induced pluripotent stem cells: assessing the mechanisms and their consequences. Bioessays. 2013; 35: 152-162.
- 8)Zhao T, Zhang ZN, Rong Z, Xu Y: Immunogenicity of induced pluripotent stem cells. Nature. 2011; 474: 212-215.

- 9) Araki R, Uda M, Hoki Y, Sunayama M, Nakamura M, Ando S, Sugiura M, Ideno H, Shimada A, Nifuji A, Abe M: Negligible immunogenicity of terminally differentiated cells derived from induced pluripotent or embryonic stem cells. Nature. 2013; 494: 100-104.
- 10)Guha P, Morgan JW, Mostoslavsky G, Rodrigues NP, Boyd AS: Lack of immune response to differentiated cells derived from syngeneic induced pluripotent stem cells. Cell Stem Cell. 2013; 12: 407-412.
- Kaneko S, Yamanaka S: To Be Immunogenic, or Not to Be: That's the iPSC Question. Cell Stem Cell. 2013; 12: 385-386.
- 12) Liang G, Zhang Y: Genetic and epigenetic variations in iPSCs: potential causes and implications for application. Cell Stem Cell. 2013; 13: 149-159.
- 13) Laurent LC, Ulitsky I, Slavin I, Tran H, Schork A, Morey R, Lynch C, Harness JV, Lee S, Barrero MJ, Ku S, Martynova M, Semechkin R, Galat V, Gottesfeld J, Izpisua Belmonte JC, Murry C, Keirstead HS, Park HS, Schmidt U, Laslett AL, Muller FJ, Nievergelt CM, Shamir R, Loring JF: Dynamic changes in the copy number of pluripotency and cell proliferation genes in human ESCs and iPSCs during reprogramming and time in culture. Cell Stem Cell. 2011; 8: 106-118.