Messenger RNA- Versus Retrovirus-Based Induced Pluripotent Stem Cell Reprogramming Strategies: Analysis of Genomic Integrity

CLARA STEICHEN,a,b,c ELÉANOR LUCE,a,b,c JÉRÔME MALUENDA,d,* LUCIE TOSCA,e,* INMACULADA MORENO-GIMENO,f CHRISTOPHE DESTERKE,a,b,c NOUSHN DIANAT,a,b,c SYLVIE GOULINET-MAINOT,a,b,c SARAH AWAN-TOOR,a,b,c DEBORAH BURKS,f JOËLLE MARIE,g ANNE WEBER,a,b,c GÉRARD TACHDJIAN,e JUDITH MELKU,d ANNE DUBART-KUPPERSCHMITT,a,b,c

Key Words. iPSCs • mRNA reprogramming • Genome integrity • SNP/CNV analysis • Hepatic differentiation

ABSTRACT

The use of synthetic messenger RNAs to generate human induced pluripotent stem cells (iPSCs) is particularly appealing for potential regenerative medicine applications, because it overcomes the common drawbacks of DNA-based or virus-based reprogramming strategies, including transgene integration in particular. We compared the genomic integrity of mRNA-derived iPSCs with that of retrovirus-derived iPSCs generated in strictly comparable conditions, by single-nucleotide polymorphism (SNP) and copy number variation (CNV) analyses. We showed that mRNA-derived iPSCs do not differ significantly from the parental fibroblasts in SNP analysis, whereas retrovirus-derived iPSCs do. We found that the number of CNVs seemed independent of the reprogramming method, instead appearing to be clone-dependent. Furthermore, differentiation studies indicated that mRNA-derived iPSCs differentiated efficiently into hepatoblasts and that these cells did not load additional CNVs during differentiation. The integration-free hepatoblasts that were generated constitute a new tool for the study of diseased hepatocytes derived from patients’ iPSCs and their use in the context of stem cell-derived hepatocyte transplantation. Our findings also highlight the need to conduct careful studies on genome integrity for the selection of iPSC lines before using them for further applications.

INTRODUCTION

Induced pluripotent stem cells (iPSCs) have considerable potential for use in regenerative medicine, including for liver diseases [1, 2]. However, the genomic stability of these cells remains a matter of concern. Copy number variations (CNVs) [3–5] and single-nucleotide variations (SNVs) [6, 7] have been reported in recent years. These mutations may already be present in the fibroblast population [8] or may be acquired during reprogramming [9]. Efforts to overcome this genomic instability generally focus on generating human iPSCs devoid of integrated reprogramming transgenes and on defining a robust reprogramming protocol with a minimal impact on genome integrity.

Many reprogramming strategies are used, but two methods excluding the use of DNA are particularly appealing: Sendai virus-based vectors [10] and synthetic mRNAs [11]. The synthetic mRNA approach has an advantage over the Sendai vector approach in that no clean-up phase is required to eliminate viral proteins. However, this method is technically challenging, repeated transfections are stressful for the cells, and little is known about the genomic integrity of the iPSCs generated.

We report in this study the generation of human iPSCs by transfection with mRNAs produced in the laboratory. We investigated the genomic integrity of these cell lines by single-nucleotide polymorphism (SNP) and CNV analyses on whole-genome microarrays, comparing the results with those obtained for retrovirus-derived iPSC clones generated in parallel from the same fibroblasts and cultured in identical conditions. Moreover, their genomic integrity was also evaluated after differentiation into hepatoblasts.

MATERIALS AND METHODS

We generated iPSC lines by the daily transfection of fibroblasts, for 16 days, with mRNAs produced in the laboratory or by a single transduction with a polycistronic retrorival vector. All the iPSC lines were generated by the same experimenter and were cultured in identical conditions.
Two independent reprogramming assays with the reprogramming mRNA cocktail OCT4, KLF4, SOX2, cMYC with or without LIN28 gave rise to embryonic stem cell-like colonies (approximately 20 colonies per 10^4 parental fibroblasts, efficiency of approximately 0.2%). The removal of LIN28 from the reprogramming cocktail in the second experiment (Fig. 1) had no effect on reprogramming efficiency. The characterization of these mRNA-derived iPSCs, and of the retrovirus-derived iPSCs obtained in parallel (referred to hereafter as mRNA-iPSCs and retrovirus-iPSCs, respectively), showed that all these clones had the typical features of human pluripotent stem cells. The expression of stemness markers was analyzed by immunostaining, flow cytometry, and RT-PCR (Fig. 2A; supplemental online Table 3). Pluripotency was initially demonstrated in vitro by the generation of embryoid bodies expressing markers from the three germ layers (Fig. 2F). Secondly, the in vivo injection of these cells into immunodeficient mice resulted in teratomas (Fig. 2G). Finally, all the selected iPSC lines had a normal karyotype (46,XY) (supplemental online Table 2). Based on published protocols [12–14], we demonstrated that mRNA-iPSCs differentiated efficiently into definitive endoderm, one of the critical barriers to the generation of iPSC-derived hepatocytes in vitro [15]. These cells differentiated further, into a homogeneous population of cells expressing a specific combination of markers of hepatoblasts, including α-fetoprotein, α-antitrypsin, cytokeratin 19, and hepatic nuclear factor 4α (Fig. 2H, 2I). Our results indicate that the use of modified mRNAs is an efficient and flexible strategy for generating nonintegrated human iPSCs that can be induced to differentiate efficiently into cells of the hepatic lineage.

We analyzed the 763,953 SNPs that gave a result for all samples and had a confidence value ≥99.95%. We determined the number of SNVs between each iPSC line and each of the two parental fibroblast samples separately. SNV numbers were between 494 and 2,067. The number of SNVs between a given iPSC line and each of the two fibroblast samples was very similar, demonstrating the robustness of the analysis (Fig. 3A). The two fibroblast samples corresponded to separate genomic DNA extractions, both at passage 6, in two independent experiments, and were thus genetically identical. The presence of 504 SNVs between these two samples corresponds to the background of our microarray analysis. This value is consistent with the threshold of ≥99.95% selected for the confidence value, implying that we would expect approximately 500 false SNP calls for every million SNPs analyzed. We therefore removed the 504 SNPs identified as SNVs between the two fibroblast samples, resulting in a new list of 763,449 SNPs. In this list, the number of SNVs between a given iPSC line and the fibroblasts was between 340 and 1,799 (Fig. 3B). A mean of 1,575 SNVs was found for group P, versus 352 and 416 for groups A1 and A2, respectively, corresponding to a fourfold difference (p value = 0.0006). Thus, the SNP profile of retrovirus-iPSC lines was more affected by the reprogramming process than that of mRNA-iPSC lines. Principal component analysis and hierarchical clustering were performed on our quantitative data (Fig. 3C) and revealed the existence of two clusters, corresponding to the reprogramming methods. Groups A1 and A2 clustered together, closer to the fibroblast population than group P, consistent with higher levels of genomic integrity in these mRNA-iPSC lines. Our analysis was based on SNV at sites throughout the genome that were known to be polymorphic. Our results are not directly comparable to those of studies based on exome capture, which have reported a mean of six protein-coding mutations per iPSC genome [7]. Overall, these results show that mRNA-iPSCs do not differ significantly from the parental fibroblasts in SNP analyses, whereas retrovirus-iPSCs do. This could be explained by the limited period of time needed for mRNA reprogramming (colonies appeared from day 12), which avoids the overexpression of reprogramming factors as a result of the persistence of retroviral vectors.

The Affymetrix 6.0 microarray contains approximately 946,000 CNV probes. We looked for large areas of CNV, with a minimal size of 100 kb, and found three deletions and four duplications in the fibroblast samples (versus the Hg19 reference sequence). All these CNVs are polymorphic (reported in the Database of Genomic Variants), and their presence was confirmed in all the iPSC lines. We then focused on additional CNVs detected in at least one iPSC line (Fig. 4A). The number of deletions for each iPSC line, with respect to the parental fibroblasts, was between 0 and 7, and the number of duplications was between 0 and 4. Results were highly heterogeneous, depending on the iPSC line considered: some iPSC lines seemed less affected (P1, A22) than
others (P2, A17). However, by contrast to our observations for SNV analyses, no clear relationship was found between the number of CNVs and the reprogramming method, suggesting that the reprogramming method used had no direct impact on the CNV number. This should be confirmed by analyzing a larger number of iPSC lines. The visualization of these CNVs on an ideogram (Fig. 4B) showed that they had no preferential localization. Fourteen of the 16 variants contained the coding region of at least one gene (Fig. 4B). Interestingly, two CNVs contained tumor suppressor genes, such as NEDD4L [16] or LRP1B, which is...
Figure 3. SNV analysis. (A): Absolute number of SNV detected for each iPSC line with respect to fibroblast samples 1 and 2, and percentage (in brackets) of the total number of SNPs analyzed. (B): Histograms representing the mean number ± SD of SNV per reprogramming group. The background (% confidence × number of SNPs analyzed) of the SNP analysis is indicated by the hatched area. The difference between group P and group A (four mRNA-iPSC lines) was highly significant (p = .0006), with the presence of four times as many SNVs in group P (*, p ≤ .05, ***, p < .001, Student’s t test). (C): Principal component analysis and hierarchical clustering based on SNV values for each iPSC line with respect to the two fibroblast samples. Abbreviations: iPSC, induced pluripotent stem cell; SNP, single-nucleotide polymorphism; SNV, single-nucleotide variation.

Figure 4. Copy number variation (CNV) analysis. (A): Number of deletions and duplications for each induced pluripotent stem cell line with respect to the parental fibroblasts. (B): Ideogram showing the localization of each CNV and the genes included in the mutated region.
located in a common fragile site [17, 18] already shown to be affected in iPSC lines [3]. Some genes affected by CNVs were involved in cell division [WAPAL] [19]. Two of the four genes affected by duplications, RAB40C and ELK3 [20, 21], have been reported to be associated with tumorigenicity. We completed this analysis by oligonucleotide-based array CGH 135K, which made it possible to detect smaller CNVs, but with a lower resolution than for Affymetrix 6.0 microarray. These two techniques did not cover the same regions of the genome and were used in this study as complementary techniques, although one deletion was confirmed by both methods (supplemental online Table 1). Our results confirmed that some clones were more affected than others, whatever the reprogramming method used. Five of these CNVs affected genes associated with cancer or cell proliferation (supplemental online Table 1), including duplications of oncogenes that were confirmed by fluorescence in situ hybridization analysis (supplemental online Fig. 1). Eighteen of the 54 CNVs detected with both techniques were common to at least two iPSC lines, and 11 were found in iPSCs reprogrammed by different methods. This probably reflects the presence of these CNVs in a subpopulation of fibroblasts before reprogramming, as previously described [3]. All our iPSC lines were analyzed at comparable early passages, to prevent the long-term passaging effect, which tends to select against highly damaged iPSCs [3]. Finally, these data are consistent with the findings of another study based on the comparison of iPSC lines generated with retrovirus or piggyBac transposon systems [3].

Finally, we extended our study with Affymetrix SNP 6.0 microarrays by comparing genomic analysis of iPSCs before and after differentiation into hepatoblasts in three independent differentiation experiments (supplemental online Fig. 2A). Interestingly, using the same criteria as previously described, no additional CNV was acquired by the cells during differentiation. Furthermore, the percentage of SNVs between undifferentiated iPSCs and hepatoblasts was not different from those measured between either the same undifferentiated iPSCs at two successive passages or two different iPSC lines derived with the two reprogramming methods (supplemental online Fig. 2B).

**CONCLUSION**

Our study demonstrates that the mRNA reprogramming strategy yields iPSCs that do not differ significantly from the parental fibroblasts in SNP analysis by contrast to retrovirus-iPSC lines. However, like other reprogramming techniques, this strategy can lead to the occurrence of CNV that are either selected or acquired during reprogramming or early passages whatever the reprogramming method used. This highlights the need for careful studies of genomic integrity to select iPSC lines, even though this nonintegrative technology appears safer for therapeutic applications.

**ACKNOWLEDGMENTS**

We thank Vincent Saravaki, Dominique Pineau, and Mélanie Point for excellent technical assistance. C.S. and N.D. were supported by fellowships from Association Française Contre les Myopathies and Région Ile de France (DIM Stem Pôle), respectively. E.L. was supported by Grant 2011-Rare-006 “HEMO-iPS,” and S.A.-T. by FP7-HEALTH.2011.1.4-2-278152 “InnovaLiv.” This work was supported by ANR-2010-RFCS-004 “Liv-iPS” and FP7-HEALTH.2011.1.4-2-278152 “InnovaLiv.”

**AUTHOR CONTRIBUTIONS**

C.S. and E.L.: conception and design, collection and assembly of data, data analysis and interpretation, manuscript writing; J. Maluenda, L.T., and I.M.-G.: collection and assembly of data, data analysis and interpretation; C.D.: data analysis and interpretation; N.D., S.G.-M., and S.A.-T.: collection and assembly of data; D.B., G.T., and J. Marie: data analysis and interpretation; J. Melki and A.W.: conception and design, data analysis and interpretation, manuscript writing.

**DISCLOSURE OF POTENTIAL CONFLICTS OF INTEREST**

The authors indicate no potential conflicts of interest.

**REFERENCES**

1. Dianat N, Steichen C, Vallier L et al. Human pluripotent stem cells for modelling human liver diseases and cell therapy. Curr Gene Ther 2013;13:120–132.
2. Lanzeni G, Oikawa T, Wang Y et al. Concise review: Clinical programs of stem cell therapies for liver and pancreas. STEM CELLS 2013;31:2047–2060.
3. Hussein SM, Batada NN, Vuoristo S et al. Copy number variation and selection during reprogramming to pluripotency. Nature 2011;471:58–62.
4. Laurent LC, Ulltisky I, Slavin I et al. Dynamic changes in the copy number of pluripotency and cell proliferation genes in human ESCs and iPSCs during reprogramming and in culture. Cell Stem Cell 2011;8:106–118.
5. Mayshar Y, Ben-David U, Lavon N et al. Identification and classification of chromosomal aberrations in human induced pluripotent stem cells. Cell Stem Cell 2010;7:521–531.

6. Cheng L, Hansen NF, Zhao L et al; NISC Comparative Sequencing Program. Low incidence of DNA sequence variation in human induced pluripotent stem cells generated by nonintegrating plasmid expression. Cell Stem Cell 2012;10:337–344.
7. Gore A, Li Z, Fung HL et al. Somatic coding mutations in human induced pluripotent stem cells. Nature 2011;471:63–67.
8. Young MA, Larson DE, Sun CW et al. Background mutations in parental cells account for most of the genetic heterogeneity of induced pluripotent stem cells. Cell Stem Cell 2012;10:570–582.
9. Ji J, Ng SH, Sharma V et al. Elevated coding mutation rate during the reprogramming of human somatic cells into induced pluripotent stem cells. STEM CELLS 2012;30:435–440.
10. Fusaki N, Ban H, Nishiyama A et al. Efficient induction of transgene-free human pluripotent stem cells using a vector based on Sendai virus, an RNA virus that does not integrate into the host genome. Proc Jpn Acad, Ser B, Phys Biol Sci 2009;85:348–362.
11. Warren L, Manos PD, Ahfeldt T et al. Highly efficient reprogramming to pluripotency and directed differentiation of human cells with synthetic modified mRNA. Cell Stem Cell 2010;7:618–630.
12. Ma X, Duan Y, Tschudy-Seney B et al. Highly efficient differentiation of functional hepatocytes from human induced pluripotent stem cells. STEM CELLS TRANSLATIONAL MEDICINE 2013;2:409–419.
13. Medine GN, Lucendo-Villarin B, Zhou W et al. Robust generation of hepatocyte-like cells from human embryonic stem cell populations. J Vis Exp 2011;(56):e2969.
14. Touboul T, Hannan NR, Corbineau S et al. Generation of functional hepatocytes from human embryonic stem cells under chemically defined conditions that recapitulate liver development. Hepatology 2010;51:1754–1765.
15. D’Amour KA, Agulnick AD, Eliazer S et al. Efficient differentiation of human embryonic
stem cells to definitive endoderm. Nat Biotechnol 2005;23:1534–1541.

16 Gao C, Pang L, Ren C et al. Decreased expression of Nedd4L correlates with poor prognosis in gastric cancer patient. Med Oncol 2012;29:1733–1738.

17 Ding D, Lou X, Hua D et al. Recurrent targeted genes of hepatitis B virus in the liver cancer genomes identified by a next-generation sequencing-based approach. PLoS Genet 2012;8:e1003065.

18 Ferber MJ, Thorland EC, Brink AA et al. Preferential integration of human papillomavirus type 18 near the c-myc locus in cervical carcinoma. Oncogene 2003;22:7233–7242.

19 Peters JM, Nishiyama T. Sister chromatid cohesion. Cold Spring Harb Perspect Biol 2012;4:pii:a011130.

20 Wasylyk C, Zheng H, Castell C et al. Inhibition of the Ras-Net (Elk-3) pathway by a novel pyrazole that affects microtubules. Cancer Res 2008;68:1275–1283.

21 Yang Q, Jie Z, Cao H et al. Low-level expression of let-7a in gastric cancer and its involvement in tumorigenesis by targeting RAB40C. Carcinogenesis 2011;32:713–722.