

## Fibroblast-Derived Induced Pluripotent Stem Cells Show No Common Retroviral Vector Insertions

FLORENCIO VARAS,<sup>a</sup> MATTHIAS STADTFELD,<sup>b</sup> LUISA DE ANDRES-AGUAYO,<sup>a</sup> NIMET MAHERALI,<sup>b</sup> ALESSANDRO DI TULLIO,<sup>a</sup> LORENA PANTANO,<sup>c</sup> CEDRIC NOTREDAME,<sup>c</sup> KONRAD HOCHEDLINGER,<sup>b</sup> THOMAS GRAF<sup>a,d</sup>

<sup>a</sup>Differentiation and Cancer and <sup>c</sup>Bioinformatics Program, Center for Genomic Regulation and Pompeu Fabra University, Barcelona, Spain; <sup>b</sup>Cancer Center and Center for Regenerative Medicine, Massachusetts General Hospital, Harvard Stem Cell Institute, Boston, Massachusetts, USA; <sup>d</sup>Institució Catalana de Recerca i Estudis Avançats

**Key Words.** Induced pluripotent stem cells • Retroviral insertions • Cellular reprogramming • Retroviral tagged mouse genes • Insertional mutagenesis

### ABSTRACT

Several laboratories have reported the reprogramming of mouse and human fibroblasts into pluripotent cells, using retroviruses carrying the *Oct4*, *Sox2*, *Klf4*, and *c-Myc* transcription factor genes. In these experiments the frequency of reprogramming was lower than 0.1% of the infected cells, raising the possibility that additional events are required to induce reprogramming, such as activation of genes triggered by retroviral insertions. We have therefore determined by ligation-mediated polymerase chain reaction (LM-PCR) the retroviral insertion sites in six induced pluripotent stem (iPS) cell clones derived from mouse fibroblasts. Seventy-nine insertion sites were as-

signed to a single mouse genome location. Thirty-five of these mapped to gene transcription units, whereas 29 insertions landed within 10 kilobases of transcription start sites. No common insertion site was detected among the iPS clones studied. Moreover, bioinformatics analyses revealed no enrichment of a specific gene function, network, or pathway among genes targeted by retroviral insertions. We conclude that *Oct4*, *Sox2*, *Klf4*, and *c-Myc* are sufficient to promote fibroblast-to-iPS cell reprogramming and propose that the observed low reprogramming frequencies may have alternative explanations. STEM CELLS 2009;27:300–306

Disclosure of potential conflicts of interest is found at the end of this article.

### INTRODUCTION

A major goal of current stem cell research is the use of specialized cells obtained from patient-derived embryonic stem (ES) for therapeutic purposes. A giant leap closer to this goal was made with the discovery that the expression of the transcription factors *Oct4*, *Sox2*, *Klf4*, and *c-Myc* induced the reprogramming of mouse skin-derived fibroblasts into induced pluripotent stem (iPS) cells capable of differentiating into cells of all three germ layers in teratomas and viable chimeric mice [1]. Similar results were also reported from other laboratories [2, 3] and for human cells [4]. Common to these studies is the low frequency of cell reprogramming, estimated to be approximately 0.1% or less in all cases. The reason for the observed low frequencies is not clear, but it is possible that additional genes have to be activated by insertion of retroviral vectors [5].

Integration sites of gammaretroviruses such as Moloney leukemia virus are known to be biased toward transcription start sites of actively transcribed genes [6, 7]. The retroviral long terminal repeats (LTR) act as promoter/enhancer elements that can modulate the expression of adjacent genes, frequently leading to their upregulation [8]. On the basis of repeated detections of specific genes targeted by integration in independently derived tumors from retrovirus infected mice, viral insertion strategies have been used to identify proto-oncogenes [9]. More recently this strategy was also successful in identifying genes that help to expand the hematopoietic stem cell pool [10, 11].

A recent report concluded that liver- and stomach-derived iPS cells exhibit no common retroviral integrations, suggesting that expression of *Oct4*, *Sox2*, *Klf4*, and *c-Myc* is sufficient for reprogramming [12]. However, since these cell types exhibited only 4–6 integration sites per clone, it is still possible that fibroblasts, in which a larger number of inte-

Author contributions: F.V.: conception and design, collection and assembly of data, data analysis and interpretation, manuscript writing; M.S.: conception and design, provision of study materials, collection and assembly of data, data analysis and interpretation; L.D.A.-A.: collection and assembly of data; N.M.: provision of study materials; A.D.T.: assembly of data; L.P. and C.N.: data analysis and interpretation; K.H. and T.G.: conception and design, data analysis and interpretation, manuscript writing, final approval of manuscript.

Correspondence: Thomas Graf, Ph.D., Centre for Genomic Regulation (CRG), C/Dr. Aiguader, 88, 08003 Barcelona, Spain. Telephone: 34-93-316-01-27; Fax: 34-93-316-00-99; e-mail: thomas.graf@crg.es; Konrad Hochedlinger, Ph.D., Center for Regenerative Medicine and Cancer Center, Massachusetts General Hospital, 185 Cambridge Street, CPZN 4242, Boston, Massachusetts 02114, USA. Telephone: 617-643-2075; Fax: 617-643-3170; e-mail: khochedlinger@helix.mgh.harvard.edu Received July 18, 2008; accepted for publication October 4, 2008; first published online in STEM CELLS EXPRESS November 13, 2008; available online without subscription through the open access option. ©AlphaMed Press 1066-5099/2009/\$30.00/0 doi: 10.1634/stemcells.2008-0696

gration sites are typically observed, differ in this respect. Therefore, to explore whether fibroblast reprogramming requires the activation of additional host genes by retroviral insertion, we exhaustively determined the integration sites in six iPS clones obtained from mouse embryo- and tail tip-derived fibroblasts after infection with retroviral vectors encoding the four Yamanaka transcription factors. We identified and sequenced 93 retroviral insertion sites and mapped 79 insertions to a single location in the mouse genome. No evidence was obtained for an insertion site common to several or all clones, nor was any gene function, gene network, or canonical pathway preferentially associated with the targeted genes. Our data therefore indicate that Oct4, Sox2, Klf4, and Myc are sufficient to induce iPS cell reprogramming in fibroblasts, extending the conclusion reached by Ai et al. with liver- and stomach-derived cells [12].

## MATERIALS AND METHODS

### Generation of iPS Cell Clones

Stable iPS cell lines were established as previously described [3]. Briefly, fibroblast cultures were established from postnatal tail-tip biopsies (for iPS clones A, B, C, and F) or from E14.5 mouse embryos (iPS clones D and E). cDNAs for murine Oct4, Sox2, and Klf4, as well as human c-Myc (the constitutively active T58A mutant), were cloned into the retroviral pMX vector and transfected into PlatE packaging cells cultured in standard ES medium (Dulbecco's modified Eagle's medium supplemented with 15% fetal bovine serum, nonessential amino acids, L-glutamine, penicillin-streptomycin,  $\beta$ -mercaptoethanol, and 1,000 U/ml leukemia inhibitory factor) on 15-cm plates using Fugene reagent (Roche Applied Science, Basel, Switzerland, <https://www.roche-applied-science.com>). Viral supernatant was harvested at 24, 48, 72, and 96h after transfection, filtered using a 0.45- $\mu$ m filter, supplemented with 5  $\mu$ g/ml polybrene, and added to the respective fibroblasts cultured on 10-cm plates. Colonies with iPS-like morphology were picked 3–5 weeks later and expanded on irradiated feeder cells in ES medium.

### Retroviral Insertion Sites into iPS Cell Clones Quantified by Southern Blotting

Ten micrograms of genomic DNA of each of the six iPS lines, as well as of V6.5 ES cells, was digested with BglIII (for c-Myc), BamHI (for Klf4 and Oct4), and HindIII (for Sox2) and separated on a 0.8% agarose gel. DNA was blotted onto HybondXL membrane (Amersham Biosciences, Piscataway, NJ, <http://www.amersham.com>) and hybridized with the respective cDNA probes labeled with [<sup>32</sup>P] $\alpha$ -dCTP as described [3].

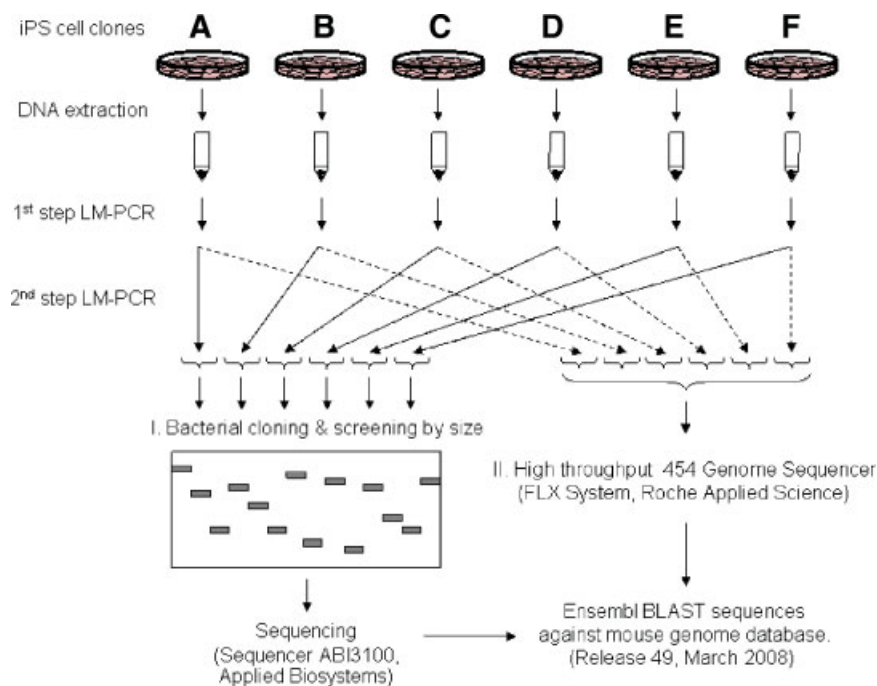
### Mapping of Retroviral Insertion Sites into iPS Clones

Retroviral insertion sites into iPS were identified by sequencing the junction fragments between the proviral 5'LTR and the host mouse genome, amplified by LM-PCR essentially as previously reported [13]. Briefly, DNA extracted from each iPS clone was digested overnight at 65°C with the frequent cutter restriction enzyme Tsp509I (Fermentas Life Sciences, Burlington, ON, Canada, <http://www.fermentas.com>), which recognizes the tetranucleotide AATT as its target sequence. The biotin-labeled and LTR-hybridizing primer LTR-Irev (AGCTGTTCCATCTGTTCTTGGCCCT) was annealed to the digested DNA and extended with the proofreading Pfu DNA polymerase (Stratagene, La Jolla, CA, <http://www.stratagene.com>). The resulting biotin-tagged double-stranded DNA fragments were recovered with a magnet after incubating with streptavidin-coated magnetic beads (Dynabeads M280; Invitrogen, Carlsbad, CA, <http://www.invitrogen.com>). An asymmetric linker cassette (LC) resulting from annealing LC1 (GACCCGGGAGATCTGAATTCAGTGGCACAG) and LC2 (CTGTGCCACTG) oligonucleotides was attached with T4 DNA ligase (Fermentas) to the

biotin-free edge, the resulting DNA was denatured with 0.1 N NaOH, and single-stranded biotin-labeled DNA was recovered. Afterward, a first polymerase chain reaction (PCR) (first-step LM-PCR) was performed with primers LTR-Irev (GACCTTGATCT-GAACTTCTC) and LCPCR1for (GACCCGGGAGATCTGAATTC), followed by a nested PCR (second-step LM-PCR) with LTR-IIIrev (TCCATGCCTTGCAAAATGGC) and LC-PCR2for (GATCTGAATTCAGTGGCACAG). An LM-PCR internal control was provided by a 278-base pair (bp) amplicon resulting from the 3'LTR located LTR-IIIrev primer up to a retroviral vector backbone carried Tsp509I target sequence. The positioning for all LM-PCR-related primers on the resulting amplicons is shown in supporting information Figure 2. Amplicons for each iPS DNA sample were shortly electrophoresed in a 2% agarose gel and excised in three regions, corresponding to high, medium, and low molecular weight (supporting information Fig. 3). Double-stranded DNA from these three regions was separately extracted with MinElute columns (Qiagen, Germantown, MD, <http://www1.qiagen.com>) and directly cloned for sequencing into pGEM-T Easy Vector (Promega, Madison, WI, <http://www.promega.com>). A preliminary PCR-based screen on individually grown bacterial colonies was performed using T7 and SP6 primers that anneal to the cloning plasmid. Amplicons were electrophoresed in 1% agarose gels, and their sizes were determined with the Quantity One software (Bio-Rad, Hercules, CA, <http://www.bio-rad.com>) using a Molecular Imager Gel Doc XR System (Bio-Rad). Amplicons differing by more than 5 bp were considered potentially different and sequenced.

To increase coverage of retroviral insertions, amplicons were directly sequenced by 454 Genome Sequencer (FLX System; Roche Applied Science). With that goal, the nested primers for the second-step LM-PCR described above were replaced by the primers A-LCPCR2for and B-code-LTR-IVrev, respectively. Primer A-LCPCR2for (GCCTCCCTCGCGCCATCAGGATCTGAATTC-AGTGGCACAG) consisted of Roche's primer A (19-mer, shown in italics) as required for the high-throughput sequencing, fused to LCPCR2for. On the other hand, the primer B-code-LTR-IVrev (GCCTTGCCAGCCCGCTCAGNNNNNGCTTGCCAAACCTACAGGTG) results from Roche's primer B (19-mer, shown in italics) fused to a 4-mer oligonucleotide tag (shown as underlined n) plus a final LTR-IVrev primer annealing up to 10 nucleotides apart from the 5'LTR edge. The tags correspond to the sequences GATC, CGAT, TCGA, ATCG, AACG, and CGAA for the A, B, C, D, E, and F iPS clones, respectively. These tags act as code bars that enable unequivocal tracking of the iPS clone contributing each single amplicon and so allow mixing of the second-step LM-PCR products from the six iPS clones. This amplicon mix was analyzed by high-throughput sequencing (Lifesequencing S.L., Paterna, Spain, <http://www.lifesequencing.com>) from primer B-code-LTR-IVrev. The sequences obtained were then processed to determine the different sequences and their relative abundances.

Amplicons sequenced either after cloning in bacteria or by high throughput were analyzed for homologies using a publicly available mouse genome sequence data-base (Ensembl BLAST [<http://www.ensembl.org/Multi/blastview>]; Ensembl release 49, March 2008). Insertion sites obtained by bacterial cloning consisted of amplicons containing mouse genome sequences flanked by the viral sequence from the LTR-IIIrev primer up to a CA dinucleotide and the linker cassette primer LCPCR2for joined by the tetranucleotide AATT (supporting information Fig. 2). High-throughput sequencing of amplicons was performed only from B-code-LTR-IVrev primer. In addition, short amplicons enabled to read even the complementary sequence to the primer A-LCPCR2for. Eighty-one percent of the sequences contained the expected AATT linker, with 13% replacing AATT with AATC, and 1%–3% carried AACT, CACT, or CCCA sequences. We defined genuine integration sites as those showing at least 95% sequence homology with the annotated mouse genome and matching one single genomic locus [6]. The precise location of the retroviral insertions was defined as the first nucleotide at the junction between the mouse genome sequence and the proviral 5'LTR.



**Figure 1.** Experimental strategy for mapping retroviral insertion sites in iPS clones. DNA was extracted from individual iPS cell clones, and the fragments between the retroviral 5' long terminal repeats and mouse genomic DNA were amplified by LM-PCR. The resulting amplicons were then sequenced either after cloning in bacterial plasmids or directly by high-throughput sequencing. Retroviral insertions were then determined by performing Ensembl BLAST searches against a mouse genome database. Abbreviations: iPS, induced pluripotent stem; LM-PCR, ligation-mediated polymerase chain reaction.

## Gene Network Analysis for Retroviral Insertion Tagged Genes

The retroviral insertions were considered to hit gene transcription units when they landed between the transcription start and stop sites. Transcription units were considered only when included in the Refseq database (<http://www.ncbi.nlm.nih.gov/RefSeq>) or when they were Ensembl-known transcripts (<http://www.ensembl.org>). The retroviral targeted genes included those in which insertions landed in the transcription unit, as well as those with a transcription start site closest to the retroviral insertions. The tagged genes were studied using the Ingenuity Pathway Analysis software (Ingenuity Systems, Redwood City, CA, <http://www.ingenuity.com>) to uncover enrichments in shared gene functions, pathways, or networks. Details on the symbols used for the gene network and the connections as shown in supporting information Figure 6 can be found at <https://analysis.ingenuity.com/pa/info/help/help.htm#legend.htm>. To evaluate the relevance of our findings, the same software was also applied to 50 randomly selected groups of mouse genes listed in the Refseq database. To that end, all genes were arbitrarily assigned numbers and groups of random numbers selected using the Random Integer Generator software (Random.org, Dublin, Ireland, <http://www.random.org/integers>).

## RESULTS

### Mapping of Retroviral Insertion Sites in iPS Clones Reveals No Common Targeted Genes

To test whether there are recurrent integration sites in individually derived iPS clones generated with retroviral vectors that might point to as yet unknown collaborating reprogramming factors, we determined the integration sites of six different iPS clones derived from murine fibroblasts. Four of these clones (A, B, C, and D) have previously been determined to be pluripotent by the formation of teratomas and to generate chimeras after injection into blastocysts [3]. The other two clones (E and F) also form teratomas and differentiate into distinct mature cell types *in vitro* but have not been tested by blastocyst injection (M. Stadtfeld and K. Hochedlinger, unpublished data). We first performed Southern blot analyses with probes against the transcription factors Oct4, Sox2, Klf4, and c-Myc on genomic DNA

extracted from the different iPS clones. The resulting hybridization bands (supporting information Fig. 1) enabled us to estimate 6–18 integration sites per iPS clone.

To identify and map the retroviral insertion sites we used two strategies (Fig. 1). In both approaches we first amplified the junction fragments between proviral and mouse genomic sequences by LM-PCR. The resulting amplicons were then either cloned into the bacterial plasmid pGEM-T Easy vector followed by sequencing or directly subjected to high-throughput amplicon sequencing with the 454 Genome Sequencer (FLX System; Roche Applied Science). Bacterial inserts were run on gels and screened by size (supporting information Fig. 3). Screening more than 600 bacterial colonies allowed the sequencing of 8, 7, 11, 9, 9, and 14 different junction fragments for iPS clones A–F, respectively. The coverage of insertion sites relative to those estimated by Southern blotting was below 80%, probably because of the predominant amplification of internal control bands and the heterogeneous PCR amplification efficiency of different insertions.

The high-throughput sequencing yielded 19,103 sequences that clustered into 93 amplicons. Every insertion identified by bacterial cloning was also contained in these amplicons. The retroviral insertions so identified raised the numbers of insertions per clone to 16, 12, 14, 13, 11, and 27 sites for clones A–F, respectively. Seventy-nine of those 93 amplicons/insertion sites could be mapped to specific mouse chromosomes (Table 1). Unmapped insertions correspond to sequences that aligned with more than one genomic locus, were too short to be mapped, or fell into incompletely sequenced genomic stretches.

The retroviral insertion sites were widely distributed throughout the mouse genome (Fig. 2). Eighteen of the autosomal chromosomes contained at least one insertion, whereas chromosome 9 and both sex chromosomes had no insertions. No relationship was observed between the chromosomal length and the number of retroviral insertions per chromosome. However, the insertions frequently hit the same chromosomes more than once. This is exemplified by iPS clone F, where 17 of the 23 insertions mapped were restricted to only four chromosomes (9, 3, 3, and 2 insertions in chromosomes 2, 4, 8, and 16, respectively). Most insertions (Table 1) landed in gene-rich areas, with

**Table 1. Retroviral insertion sites in fibroblast derived iPS clones**

| Insertion       | Chromosome | Nucleotide position | Tagged gene (retrovirus orientation) | Distance to transcription start site |
|-----------------|------------|---------------------|--------------------------------------|--------------------------------------|
| A1              | 2B         | 34483975            | <i>Gapvd1</i> (F)                    | 126777                               |
| A2 <sup>a</sup> | 2B         | 35116715            | <i>Gsn</i> (F)                       | 4765 (intron 1)                      |
| A3              | 2H2        | 156800645           | <i>Ndr3</i> (R)                      | 17168 (intron 1)                     |
| A4              | 5B1        | 34869389            | <i>Sh3bp2</i> (F)                    | 956 (intron 1)                       |
| A5              | 11E2       | 118226086           | <i>Timp2</i> (R)                     | - 9447                               |
| A6              | 12F1       | 103463226           | <i>Slc24a4</i> (F)                   | 95597 (intron 7)                     |
| A7              | 13A5       | 35914762            | <i>Cdyl</i> (R)                      | 163013 (intron 5)                    |
| A8              | 13C3       | 82707525            | <i>Cetm3</i> (R)                     | 785273                               |
| A9              | 13D1       | 94774501            | <i>Lhfp12</i> (R)                    | 53250                                |
| A10             | 18E1       | 68218379            | D18Erd653e (R)                       | 125468 (intron 1)                    |
| A11             | 18E2       | 69917746            | <i>Ccdc68</i> (R)                    | 167467                               |
| B1              | 4B3        | 55541032            | <i>Klf4</i> (F)                      | 4315 (exon 5)                        |
| B2              | 4C7        | 108060530           | OTTMUSG0000008243 (F)                | 4576                                 |
| B3              | 6A1        | 4368969             | <i>Col1a2</i> (R)                    | 86728                                |
| B4              | 6A1        | 4453028             | <i>Col1a2</i> (R)                    | 2669                                 |
| B5              | 6C1        | 67216395            | <i>Serbp1</i> (R)                    | 610                                  |
| B6              | 7B2        | 39014778            | <i>Plekhf1</i> (R)                   | -1768                                |
| B7              | 10B3       | 59289089            | <i>Cbara1</i> (R)                    | 123742 (intron 10)                   |
| B8              | 11B4       | 76139823            | <i>Nxn</i> (R)                       | 72820 (intron 1)                     |
| B9              | 13C3       | 91061729            | <i>Rps23</i> (F)                     | -1043                                |
| B10             | 16B3       | 35495343            | <i>Pdia5</i> (F)                     | -4384                                |
| C1              | 1H6        | 191532238           | <i>Ptpn14</i> (F)                    | 19909                                |
| C2              | 2C3        | 73209524            | Sp9 (R)                              | 14976 (intron 4)                     |
| C3              | 5A1        | 5709239             | <i>Steap2</i> (R)                    | 14671                                |
| C4              | 8E1        | 122211976           | <i>Wfdc1</i> (F)                     | 21553 (exon 7)                       |
| C5              | 10C1       | 78186016            | <i>Casp14</i> (F)                    | 5032                                 |
| C6              | 10D3       | 126748434           | <i>Mars</i> (F)                      | 261 (intron 1)                       |
| C7              | 13A3.2     | 24892273            | BC005537 (R)                         | 2879                                 |
| C8              | 13D1       | 105318214           | <i>Adams6</i> (F)                    | 240299                               |
| C9              | 15A2       | 11453709            | <i>Tars</i> (R)                      | 124296                               |
| C10             | 15D2       | 47534486            | <i>Sh3pxd2a</i> (R)                  | 4405 (intron 1)                      |
| C11             | 19D3       | 58585304            | 1700011F14Rik (F)                    | 1188                                 |
| D1              | 1H1        | 163968862           | <i>Dnm3</i> (R)                      | 439299 (intron 17)                   |
| D2              | 2G2        | 141059046           | OTTMUSG00000015674 (F)               | -940                                 |
| D3              | 3F2        | 95030931            | <i>Mll11</i> (R)                     | 1294 (intron 1)                      |
| D4              | 5F         | 118953105           | <i>Med13l</i> (F)                    | 57647                                |
| D5              | 5F         | 112228069           | C130026L21Rik (R)                    | 217488                               |
| D6              | 6D2        | 87823102            | <i>Copg</i> (F)                      | 14820                                |
| D7              | 10B3       | 59415137            | <i>Ddit4</i> (F)                     | 619                                  |
| D8              | 11C        | 88857493            | <i>Trim25</i> (R)                    | -3197                                |
| D9              | 13D1       | 96319422            | <i>F2rl1</i> (R)                     | 24240                                |
| D10             | 14E4       | 120697059           | <i>Mbnl2</i> (R)                     | 22130 (intron 1)                     |
| D11             | 16B1       | 20457862            | <i>Abcc5</i> (R)                     | -31395                               |
| D12             | 18E2       | 75374255            | <i>Dym</i> (R)                       | 195829 (intron 1)                    |
| D13             | 19A        | 4233573             | <i>Pold4</i> (F)                     | 1595 (exon 4)                        |
| E1              | 3F1        | 88318454            | <i>Lmna</i> (F)                      | 11225                                |
| E2              | 3F3        | 102331352           | <i>Ngfb</i> (F)                      | 57501                                |
| E3              | 4D1        | 119753070           | <i>Edn2</i> (F)                      | 80736                                |
| E4              | 11E2       | 118291101           | <i>C1qtnf1</i> (R)                   | 1288 (intron 1)                      |
| E5              | 13A4       | 33179504            | <i>Serp1b1b</i> (R)                  | 3466 (intron 3)                      |
| E6              | 16B2       | 30596895            | BC022623 (R)                         | 2914                                 |
| E7              | 17C        | 47012823            | A330017A19Rik (R)                    | 14331                                |
| E8              | 17E4       | 84368563            | <i>Haa0</i> (R)                      | 122433                               |
| E9              | 18B3       | 38009039            | <i>Diap1</i> (R)                     | 86091 (intron 28)                    |
| E10             | 18E3       | 77736014            | <i>Rnf165</i> (R)                    | 28996                                |
| E11             | 19B        | 8787296             | <i>Slc3a2</i> (F)                    | 1070 (exon 2)                        |
| F1              | 2B         | 33120311            | <i>Ralgps1</i> (F)                   | 106695 (intron 7)                    |
| F2              | 2B         | 39204166            | <i>Ppp6c</i> (F)                     | 122347                               |
| F3              | 2C1        | 52178122            | <i>Neb</i> (F)                       | 12532 (intron 7)                     |
| F4              | 2C3        | 62411646            | <i>Fap</i> (F)                       | 432 (intron 1)                       |
| F5              | 2C3        | 68693947            | <i>Lass6</i> (F)                     | 5667                                 |
| F6              | 2F2        | 125226449           | <i>Fbn1</i> (F)                      | 107280 (intron 7)                    |
| F7              | 2G3        | 145685702           | <i>Rin2</i> (F)                      | 73986 (intron 8)                     |
| F8              | 2H1        | 152752982           | <i>Fkhl18</i> (R)                    | 5962                                 |
| F9              | 2H2        | 156219035           | 2900097C17Rik (R)                    | 470                                  |
| F10             | 3E2        | 68387400            | <i>Schip1</i> (R)                    | 89270 (intron 2)                     |
| F11             | 4A2        | 11918546            | ENSMUSG00000073995 (R)               | 11097                                |

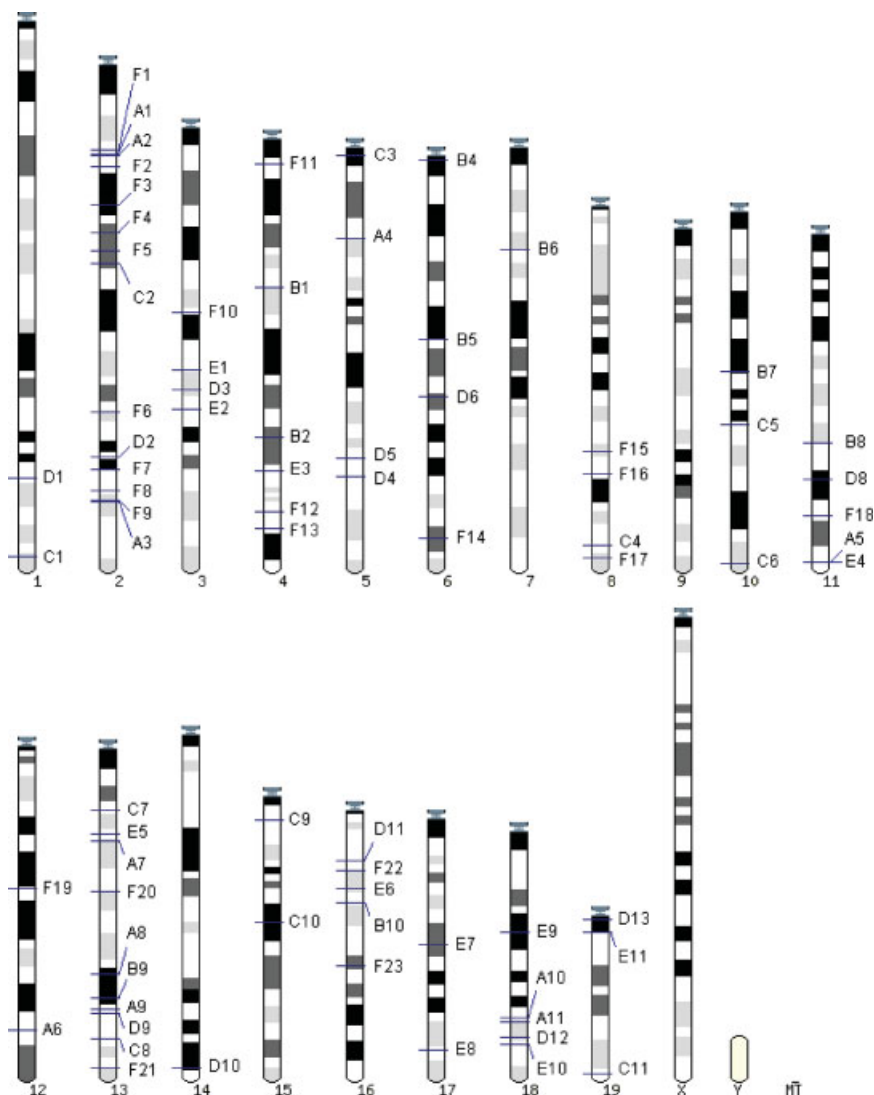
(continued)



**Table 1.** (Continued)

| Insertion | Chromosome | Nucleotide position | Tagged gene (retrovirus orientation) | Distance to transcription start site |
|-----------|------------|---------------------|--------------------------------------|--------------------------------------|
| F12       | 4D3        | 134374117           | <i>Tmem57</i> (F)                    | 35143 (intron 6)                     |
| F13       | 4D3        | 140187666           | <i>Arhgef10l</i> (R)                 | 34261 (intron 1)                     |
| F14       | 6G1        | 137507993           | <i>Eps8</i> (F)                      | 89813 (intron 2)                     |
| F15       | 8C4        | 88970931            | <i>Abcc12</i> (F)                    | 133654                               |
| F16       | 8C5        | 96786499            | <i>Nup93</i> (R)                     | 48019 (intron 3)                     |
| F17       | 8E2        | 126808432           | <i>Galnt2</i> (F)                    | 53138 (intron 1)                     |
| F18       | 11D        | 102017387           | <i>Tmem101</i> (F)                   | 288 (intron 1)                       |
| F19       | 12C2       | 53593700            | <i>Arhgap5</i> (R)                   | 11259                                |
| F20       | 13B2       | 53640111            | <i>Msx2</i> (F)                      | 71962                                |
| F21       | 13D2.2     | 115375507           | 4930544M13Rik (F)                    | 21938                                |
| F22       | 16B1       | 24237514            | <i>Lpp</i> (R)                       | 156062                               |
| F23       | 16C1.1     | 57420543            | Q6P6L0-2 (R)                         | 67153 (intron 1)                     |

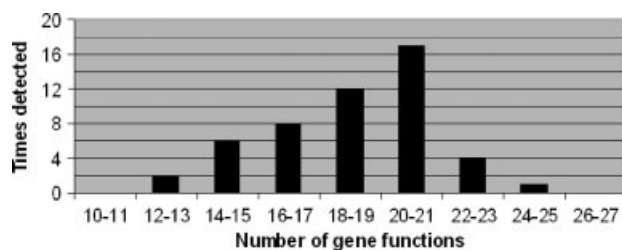
The table lists retroviral insertions found in the six iPS clones (A to F) examined. Underlined insertions were found both by bacterial sequencing and by high throughput sequencing. Genes shown in italics were included in the Ingenuity Pathway Analysis. The orientation of retroviral vector transcription was designated as forward (F) when it proceeded in the same sense than the tagged gene and reverse (R) when it pointed in the opposite direction. The distance between the retroviral insertion site and the transcription start site of the closest gene is shown in the last column. Insertions that occurred within transcription units and their locations are shown in brackets.



**Figure 2.** Distribution of retroviral insertion sites among mouse chromosomes. The horizontal lines across the chromosomes indicate sites of insertions.

78% of the insertions located within 100 kilobases (kb) from the closest transcription start site. Four insertions mapped to exons

and 31 to introns, with approximately half of the latter being located in the first two introns. Another 13 and 16 insertions



**Figure 3.** Distribution of the number of gene functions included in the best gene networks of 50 groups of 65 randomly selected genes plus *Oct4*, *Sox2*, *Klf4*, and *c-Myc*.

were within 10 kb downstream and upstream of the transcription start site, respectively.

Thirty of the 79 mapped retroviral insertions in iPS clones are included in the mouse Retrovirus Tagged Cancer Gene Database (<http://rtcgd.abcc.ncicrf.gov>). Importantly, our data showed that not a single gene was shared by more than one iPS clone. The two closest insertions identified (B3 and B4) were located 84 kb apart, upstream of the transcription start site for gene *Colla2* on chromosome 6. However, both were found in the same clone. We therefore conclude that no gene was targeted more than once by retroviral insertions in the different iPS clones studied, suggesting that *Oct4*, *Sox2*, *Klf4*, and *c-Myc* are sufficient to trigger fibroblast reprogramming into iPS cells.

### No Evidence That Genes Tagged by Retroviral Insertions Have Shared Gene Functions or Belong to Common Canonical Pathways or Gene Networks

Although no common tagged gene was uncovered by mapping the insertion sites in iPS cells, it is still possible that some of the insertions preferentially affect specific gene functions, gene networks or signaling pathways. To determine whether retrovirally tagged genes fall into functional gene categories, the genes listed in Table 1 were grouped together with *Oct4*, *Sox2*, *Klf4*, and *c-Myc* and subjected to bioinformatics analyses using Ingenuity Pathway Analysis (Ingenuity Systems). The total number of genes so analyzed was reduced from 79 to 69 since insertions B1 and B4 were left out (they correspond to the endogenous *Klf4* gene and to *Colla2*, which was detected twice), and another 12 genes were subtracted because not enough biological information was available. The remaining 69 genes showed no evident enrichment for any specific gene function (supporting information Fig. 4). Similarly, no bias toward any canonical gene pathway was observed (supporting information Fig. 5). Finally, the network analysis revealed 18 genes as part of a best gene network that was enriched in components of nervous system development and function, cell cycle and cellular development (supporting information Fig. 6). To assess whether this network might be biologically relevant or a product of chance, the Ingenuity Pathway Analysis was repeated 50 times with different groups of 65 randomly selected genes plus the four Yamanaka transcription factors. This analysis revealed that the number of genes participating in best networks ranged between 13 and 28 (Fig. 3). The value obtained with the targeted genes from Table 1, 18 gene functions, therefore falls well within the range of values obtained with the randomized approach. We conclude that the retroviral insertions in the six iPS clones studied have no shared gene functions, nor do they belong to a common canonical pathway or gene network.

www.StemCells.com

**Table 2.** Retroviral insertion numbers determined by different methods

| iPS clone | Southern blotting | Bacterial sequencing | High-throughput sequencing |
|-----------|-------------------|----------------------|----------------------------|
| A         | 15                | 8                    | 16 (6–185) <sup>b</sup>    |
| B         | 12–14             | 7                    | 12 (15–169)                |
| C         | 4–5 <sup>a</sup>  | 11                   | 15 (20–317)                |
| D         | 9                 | 9                    | 13 (26–228)                |
| E         | 10                | 9                    | 11 (3–216)                 |
| F         |                   | 14                   | 26 (6–121)                 |

<sup>a</sup>Underestimation.

<sup>b</sup>Frequency of detection.

Abbreviation: iPS, induced pluripotent stem.

## DISCUSSION

Our results show that no genes tagged by retroviral insertions were repeatedly detected in the six iPS clones analyzed. In addition, bioinformatics analyses showed no enrichment of tagged genes that are functionally related or are part of the same pathways or networks. On the basis of the estimated numbers of integrations by Southern blotting, bacterial cloning and sequencing, and high-throughput sequencing (summarized in Table 2), the insertion site coverage for the six iPS clones analyzed appears to be complete. This is reinforced by the fact that the minimum frequency with which each insertion was sequenced by high throughput was 3–26 times per amplicon per clone. However, it is possible that the number of integrations detected using the *Tsp509I* restriction enzyme to generate the samples used for high-throughput sequencing is not exhaustive. Thus, it has been reported that the restriction enzyme used for genomic DNA digestion preceding PCR-based protocols influences the coverage of retroviral insertion sites detected [14].

Twelve of the 28 retrovirally tagged genes previously reported in liver and hepatic-derived iPS [12] are included in the Retrovirus Tagged Cancer Gene Database [15]. Similarly, we found that an elevated percentage (38%) of the tagged genes in fibroblast-derived iPS is contained in this database, as three sets of 69 random control genes yielded 20%. This might represent a bias of integrations into actively transcribed genes, although a biological relevance of the enriched genes cannot be excluded. Importantly, bioinformatic analyses of our data showed no enrichment of tagged genes that are functionally related or are part of the same pathway or network. In addition, we found no overlap between the genes targeted in our study and those reported by another group [12]. Our results therefore indicate that *Oct4*, *Sox2*, *Klf4*, and *c-Myc* are sufficient to induce the reprogramming of fibroblasts into iPS cells and extend the observations of Aoi et al. [12] to one of the most commonly used and easily available cell types used in reprogramming studies.

There are several alternative explanations for the observed low frequencies of reprogramming. One possibility is that the cell type of origin dictates the frequency and that fibroblasts are heterogeneous, being composed of a few highly susceptible and many resistant cell types. This is unlikely, since it has been shown that epithelial cells derived from stomach, liver, and pancreatic  $\beta$  cells can be reprogrammed by the four Yamanaka factors, at frequencies similar to or above those seen with fibroblasts [12, 16]. However, reprogramming of mature B cells required an additional transcription factor [17]. Another possibility is that the stoichiometry of the Yamanaka factors ex-

pressed after infection is not adequate in most cells, leading to no or incomplete reprogramming, growth disadvantage, or death.

The most likely scenario, therefore, is that only a small proportion of somatic cells have a chromatin configuration amenable to transcription factor-induced changes resulting in iPS cell formation. Two lines of evidence support this notion. First, daughter cells derived from a partially reprogrammed iPS cell clone have been shown to reactivate Oct4 at different times, consistent with the involvement of stochastic epigenetic events leading to the reacquisition of pluripotency [18]. Second, the treatment of partially reprogrammed iPS cells or of somatic cells undergoing reprogramming with compounds targeting epigenetic modifications, including DNA and histone methylation as well as histone acetylation, improves the efficiency and fidelity of reprogramming significantly [19–21].

## CONCLUSION

The analysis of six fibroblast-derived iPS cell clones induced after infection with retroviruses expressing Oct4, Sox2, Klf4, and c-Myc showed no evidence of common retroviral integration sites. Taking these data together with supporting bioinformatics analyses, we suggest that the activation of additional endogenous genes is not required for the reprogramming induced by the four factors. It will now be interesting to see whether the four Yamanaka factors are sufficient to generate viral integration-free iPS cells for research and regenerative medicine.

## REFERENCES

- Okita K, Ichisaka T, Yamanaka S. Generation of germline-competent induced pluripotent stem cells. *Nature* 2007;448:313–317.
- Wernig M, Meissner A, Foreman R et al. In vitro reprogramming of fibroblasts into a pluripotent ES-cell-like state. *Nature* 2007;448:318–324.
- Maherali N, Sridharan R, Xie W et al. Directly reprogrammed fibroblasts show global epigenetic remodeling and widespread tissue contribution. *Cell Stem Cell* 2007;1:55–70.
- 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.
- Yamanaka S. Strategies and new developments in the generation of patient-specific pluripotent stem cells. *Cell Stem Cell* 2007;1:39–49.
- Wu X, Li Y, Crise B et al. Transcription start regions in the human genome are favored targets for MLV integration. *Science* 2003;300:1749–1751.
- Mitchell RS, Beitzel BF, Schroder AR et al. Retroviral DNA integration: ASLV, HIV, and MLV show distinct target site preferences. *PLoS Biol* 2004;2:E234.
- Mikkers H, Berns A. Retroviral insertional mutagenesis: Tagging cancer pathways. *Adv Cancer Res* 2003;88:53–99.
- Akagi K, Suzuki T, Stephens RM et al. RTCGD: Retroviral tagged cancer gene database. *Nucleic Acids Res* 2004;32:D523–D527.
- Kustikova O, Fehse B, Modlich U et al. Clonal dominance of hematopoietic stem cells triggered by retroviral gene marking. *Science* 2005;308:1171–1174.
- Kustikova OS, Geiger H, Li Z et al. Retroviral vector insertion sites

## NOTE ADDED IN PROOF

Two communications have now shown that the formation of iPS cells can be generated without using retroviruses as vectors and in the absence of detectable integrations:

Okita K, Nakagawa M, Hyenjong H, Ichisaka T, Yamanaka S. Generation of mouse induced pluripotent stem cells without viral vectors. *Science*. 2008;322:949–953.

Stadtfeld M, Nagaya M, Utikal J, Weir G, Hochedlinger K. Induced pluripotent stem cells generated without viral integration. *Science*. 2008;322:945–949.

## ACKNOWLEDGMENTS

This work was funded by Center for Genomic Regulation and ICREA (to T.G.), as well as by a grant from Fondo de Investigación Sanitaria-ISCIII (to F.V.). Work was also supported by an NIH Director's Innovator Award and the Harvard Stem Cell Institute (to K.H.). M.S. was supported by a Schering postdoctoral fellowship.

## DISCLOSURE OF POTENTIAL CONFLICTS OF INTEREST

The authors indicate no potential conflicts of interest.

associated with dominant hematopoietic clones mark “stemness” pathways. *Blood* 2007;109:1897–1907.

- Aoi T, Yae K, Nakagawa M et al. Generation of pluripotent stem cells from adult mouse liver and stomach cells. *Science* 2008;321:699–702.
- Schmidt M, Hoffmann G, Wissler M et al. Detection and direct genomic sequencing of multiple rare unknown flanking DNA in highly complex samples. *Hum Gene Ther* 2001;12:743–749.
- Wang GP, Garrigue A, Ciuffi A et al. DNA bar coding and pyrosequencing to analyze adverse events in therapeutic gene transfer. *Nucleic Acids Res* 2008;36:e49.
- Hawley RG. Does retroviral insertional mutagenesis play a role in the generation of induced pluripotent stem cells? *Mol Ther* 2008;16:1354–1355.
- Stadtfeld M, Brennand K, Hochedlinger K. Reprogramming of pancreatic Beta cells into induced pluripotent stem cells. *Curr Biol* 2008;18:890–894.
- Hanna J, Markoulaki S, Schorderet P et al. Direct reprogramming of terminally differentiated mature B lymphocytes to pluripotency. *Cell* 2008;133:250–264.
- Meissner A, Wernig M, Jaenisch R. Direct reprogramming of genetically unmodified fibroblasts into pluripotent stem cells. *Nat Biotechnol* 2007;25:1177–1181.
- Shi Y, Do JT, Despons C et al. A combined chemical and genetic approach for the generation of induced pluripotent stem cells. *Cell Stem Cell* 2008;2:525–528.
- Huangfu D, Maehr R, Guo W et al. Induction of pluripotent stem cells by defined factors is greatly improved by small-molecule compounds. *Nat Biotechnol* 2008;26:795–797.
- Mikkelsen TS, Hanna J, Zhang X et al. Dissecting direct reprogramming through integrative genomic analysis. *Nature* 2008;454:49–55.



See [www.StemCells.com](http://www.StemCells.com) for supporting information available online.