Few Single Nucleotide Variations in Exomes of Human Cord Blood Induced Pluripotent Stem Cells

Rui-Jun Su1,3,4, Yadong Yang2,*, Amanda Neises1, Kimberly J. Payne3,4, Jasmin Wang1, Kasthuribai Viswanathan5, Edward K. Wakeland5, Xiangdong Fang2,*, Xiao-Bing Zhang1,*

1 Department of Medicine, Loma Linda University, Loma Linda, California, United States of America, 2 CAS Key Laboratory of Genome Sciences, Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing, China, 3 Division of Anatomy, Loma Linda University, Loma Linda, California, United States of America, 4 Center for Health Disparities and Molecular Medicine, Loma Linda University, Loma Linda, California, United States of America, 5 Department of Immunology and Genomics Core Facility, the University of Texas Southwestern Medical Center, Dallas, Texas, United States of America

Abstract

The effect of the cellular reprogramming process per se on mutation load remains unclear. To address this issue, we performed whole exome sequencing analysis of induced pluripotent stem cells (iPSCs) reprogrammed from human cord blood (CB) CD34+ cells. Cells from a single donor and improved lentiviral vectors for high-efficiency (2–14%) reprogramming were used to examine the effects of three different combinations of reprogramming factors: OCT4 and SOX2 (OS), OS and ZSCAN4 (OSZ), OS and MYC and KLF4 (OSMK). Five clones from each group were subject to whole exome sequencing analysis. We identified 14, 11, and 9 single nucleotide variations (SNVs), in exomes, including untranslated regions (UTR), in the five clones of OSMK, OS, and OSZ iPSC lines. Only 8, 7, and 4 of these, respectively, were protein-coding mutations. An average of 1.3 coding mutations per CB iPSC line is remarkably lower than previous studies using fibroblasts and low-efficiency reprogramming approaches. These data demonstrate that point nucleotide mutations during cord blood reprogramming are negligible and that the inclusion of genome stabilizers like ZSCAN4 during reprogramming may further decrease reprogramming-associated mutations. Our findings provide evidence that CB is a superior source of cells for iPSC banking.

Citation: Su R-J, Yang Y, Neises A, Payne KJ, Wang J, et al. (2013) Few Single Nucleotide Variations in Exomes of Human Cord Blood Induced Pluripotent Stem Cells. PLoS ONE 8(4): e59908. doi:10.1371/journal.pone.0059908

Editor: Zoran Ivanovic, French Blood Institute, France

Received January 20, 2013; Accepted February 20, 2013; Published April 1, 2013

Copyright: © 2013 Su et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This research was supported by Loma Linda University Department of Medicine (XBX), and DOD Concept Award W81XWH-11-1-0667 (XBX), the Department of Pathology and Human Anatomy and the Center for Health Disparities and Molecular Medicine at Loma Linda University (KJP and RJS). This work was also supported by the “863 Projects” of Ministry of Science and Technology of China (2011AA020118 to XF) and the “Strategic Priority Research Program” of the Chinese Academy of Sciences, Stem Cell and Regenerative Medicine Research (XDA01040405 to XF). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: xzhang@llu.edu (XZ); fangxd@big.ac.cn (XF)

These authors contributed equally to this work.

Introduction

The discovery of a simple approach for reprogramming human somatic cells into induced pluripotent stem cells (iPSCs) has revolutionized regenerative medicine [1,2,3]. Technological breakthroughs have made it possible to generate integration-free iPSCs with modified mRNAs [4,5], non-integrating Sendai virus [6,7,8,9,10] or oriP/EBNA1-based episomal vectors [11,12,13,14,15,16,17,18,19] and other methods, which brings breakthroughs have made it possible to generate integration-free iPSCs with modified mRNAs [4,5], non-integrating Sendai virus [6,7,8,9,10] or oriP/EBNA1-based episomal vectors [11,12,13,14,15,16,17,18,19] and other methods, which brings revolutionized regenerative medicine [1,2,3]. Technological breakthroughs have made it possible to generate integration-free iPSCs with modified mRNAs [4,5], non-integrating Sendai virus [6,7,8,9,10] or oriP/EBNA1-based episomal vectors [11,12,13,14,15,16,17,18,19] and other methods, which brings revolutionized regenerative medicine [1,2,3].

The effect of the cellular reprogramming process per se on mutation load remains unclear. To address this issue, we performed whole exome sequencing analysis of induced pluripotent stem cells (iPSCs) reprogrammed from human cord blood (CB) CD34+ cells. Cells from a single donor and improved lentiviral vectors for high-efficiency (2–14%) reprogramming were used to examine the effects of three different combinations of reprogramming factors: OCT4 and SOX2 (OS), OS and ZSCAN4 (OSZ), OS and MYC and KLF4 (OSMK). Five clones from each group were subject to whole exome sequencing analysis. We identified 14, 11, and 9 single nucleotide variations (SNVs), in exomes, including untranslated regions (UTR), in the five clones of OSMK, OS, and OSZ iPSC lines. Only 8, 7, and 4 of these, respectively, were protein-coding mutations. An average of 1.3 coding mutations per CB iPSC line is remarkably lower than previous studies using fibroblasts and low-efficiency reprogramming approaches. These data demonstrate that point nucleotide mutations during cord blood reprogramming are negligible and that the inclusion of genome stabilizers like ZSCAN4 during reprogramming may further decrease reprogramming-associated mutations. Our findings provide evidence that CB is a superior source of cells for iPSC banking.

mutations in the parental cell pool [27]. These seemingly conflicting reports warrant further investigation into whether the process of iPSC generation is mutagenic and if so, the extent of such mutations.

Three mechanisms have been proposed to account for the up to 10-fold higher rate of genetic alterations in iPSCs as compared to anticipated background mutations. First, the fixation of rare mutations in the parent cell population has been implicated. Early studies suggest that ~50% of SNVs are pre-existing in parent cell cultures [20]. A recent report demonstrates that 30% of skin fibroblasts have somatic CNVs in their genomes [28]. Second, the selection of clones harboring mutations that improve reprogramming efficiency and/or promote cell survival/proliferation has been suggested as a contributing factor. This idea is supported by enrichment analysis that found that the observed genetic variations are strongly associated with cancer [20]. A third proposed mechanism is proliferative stress induced by reprogramming factor overexpression. In support of this hypothesis, some reprogramming factors such as MYC are strongly oncogenic [29]. Furthermore, the downregulation of genome guardians like...
p53 substantially increases reprogramming efficiency [15,19,30,31].

A careful examination of reported data suggests that several factors might affect the number of SNVs identified in the coding regions of each iPSC clone. First, reprogramming efficiency is a potential factor. Extremely low reprogramming efficiency (10⁻⁶) is associated with very high levels of SNVs (more than 10 per iPSC) [32,33]. Thus, low reprogramming efficiency might also contribute to the outgrowth of clones with mutations in genes that promote cell growth and exert causative effects in cancer [20]. Second, long-term culture may lead to the accumulation of rare SNVs, since longer durations of in vitro culture after harvest of the primary cells is associated with increased numbers of SNVs [27,32]. Third, source cells from reprogramming may also play a role; hematopoietic CD34⁺ cell-derived iPSCs harbor less than half the mutations detected in iPSC clones from MSC or fibroblasts [32].

Given the potential contribution of the above factors, we propose that an accurate estimate of reprogramming-induced SNVs requires the use of a high-efficiency approach (>1%) for the reprogramming of homogenous primary cells from a single donor with minimal in vitro manipulation. The majority of CD34⁺ hematopoietic stem/progenitor cells in adults reside in the bone marrow niche and are protected from environmental insults, thus are presumably more homogeneous than fibroblasts from skin biopsy [28,32]. Umbilical cord blood (CB) is a source of CD34⁺ hematopoietic cells that is superior to and more homogeneous than adult blood or marrow cells. This is because CB is a source from earlier in life and the pool of CD34⁺ cells in the baby has been less extensively expanded than adult blood or marrow [34,35]. Thus, CB CD34⁺ cells are less likely to harbor unique rare mutations than cells from other sources. In addition, we recently reported that CB CD34⁺ cells can be very efficiently reprogrammed to iPSCs (2%) using improved lentiviral vectors [18], thus providing us with the unique opportunity to address an important and largely unanswered question: What is the contribution of reprogramming per se to genetic alterations in iPSC?

Materials and Methods

Cord Blood

The use of CB was approved by the Institutional Review Board of Loma Linda University and written informed consent was obtained from all participants. After treating CB with red blood lysis buffer, CD34⁺ cells were purified from nucleated cells by MACS (Miltenyi Biotec, Auburn, CA). All the iPSC clones for exome sequencing analysis were derived from a single CB.

Constructs and Lentiviral Vector Packaging

In conducting work involving the use of recombinant DNA, we adhered to the current version of the National Institutes of Health (NIH) Guidelines for Research Involving Recombinant DNA Molecules. The lentiviral vector constructs have been detailed previously [36]. In brief, a strong promoter SFFV was used to drive the expression of OS (OCT4 and SOX2) or MK (MYC and KLF4), which are linked with a 2A self-cleavage peptide sequence [18,37]. Vector containing the SC4MV gene was obtained from Applied Biological Materials Inc. (ABM; Richmond, BC, Canada). Detailed methods for lentiviral vector packaging and titering have been published [36]. After a 100-fold concentration by ultracentrifugation, biological titer of 5–10×10⁷ were achieved.

iPSC Generation

CB CD34⁺ cells were cultured in hematopoietic cell culture conditions: Iscove’s modified Dulbecco’s medium (IMDM)/10% FBS supplemented with cytokines TPO, SCF, FL, and G-CSF each at 100 ng/ml, and IL-3 at 10 ng/ml [38,39]. After 2 days of pre-stimulation, 1×10⁶ cells per well were seeded into a CH-296 (Takara Bio, Inc., Shiga, Japan)-treated non-TC 24-well plate. Lentiviral vectors were added at an MOI of 4 and co-cultured for 16 hours. Protamine sulphate at a final concentration of 8 μg/ml was added to increase the transduction efficiency. After transduction, cells were harvested and transferred to 6-well plates pre-seeded with inactivated rat embryonic fibroblast (REF) feeder cells (ABM). Cells were maintained in the hematopoietic cell culture condition for 2 more days before being gradually replaced with iPSC media. The iPSC medium used in our study is composed of Knockout DMEM/F12 medium (Invitrogen; Carlsbad, CA) supplemented with 20% Knockout Serum Replacement (KSR) (Invitrogen), 1 mM GlutaMAX (Invitrogen), 2 mM nonessential amino acids (ABM), 1×penicillin/streptomycin (ABM), 0.1 mM β-mercaptoethanol (Sigma-Aldrich Corp, St. Louis, MO), 20 ng/ml FGF2 (ABM), and 50 μg/ml ascorbic acid [40,41]. To increase reprogramming efficiency, sodium butyrate [42,43] was added at 0.25 mM from day 2 to 10, and cells were cultured under hypoxia [44,45].

Flow Cytometry

iPSCs were harvested with Accutase (Innovative Cell Technologies, Inc., San Diego, CA) and fixed for 10 min at room temperature in fixation buffer (eBioscience, Inc., San Diego, CA). For staining with TRA-1-60-PE (eBioscience), cells were incubated with the antibody for 30 min at room temperature. Flow cytometry analysis was performed using a FACS Aria II (BD Biosciences, San Jose, CA) with a 488-nm laser. 30,000 events were collected for each sample. For flow cytometry analysis, gates were set based on isotype controls.

Confocal Imaging

For immunostaining of iPSC colonies, iPSCs were cultured in 2-well chamber culture slides for 4–5 days. Cells were treated with fixation buffer supplemented with permeabilization buffer (eBioscience) for 10 min before being stained overnight with PE or FITC conjugated antibodies OCT4 (eBioscience), NANOG (BD), or SSEA4 (eBioscience). The samples were washed twice with permeabilization buffer, and then coverslipped. Imaging was performed using the Zeiss LSM 710 NLO laser scanning confocal microscope with a 20× objective at the Loma Linda University Advanced Imaging and Microscopy Core. High resolution monochrome image was captured using a Zeiss HRm CCD camera.

Teratoma Assay

The use of NOD/SCID/IL2RG⁻/⁻ (NSG) immunodeficient mice for the teratoma formation assay was approved by the Institutional Animal Care and Use Committee at Loma Linda University (LLU). NSG mice were purchased from the Jackson Laboratory and maintained at the LLU animal facility. iPSCs were harvested by Dispase (Invitrogen) digestion, and approximately 1×10⁶ iPSCs were re-suspended in 200 μl DMEM/F12 diluted (1:1) Matrigel solution (BD) before subcutaneous injection into NSG mice. At 2 months after injection of iPSCs, teratomas were dissected and fixed in 10% formalin. After paraffin embedding and microsectioning, samples were stained with hematoxylin and eosin (H & E), following standard protocol.
Pictures of differentiated tissues were captured with a Nikon microscope using a 20× objective.

**Exome Sequencing**

To deplete feeder cells, iPSCs were cultured in TeSR medium (StemCell Technologies) for 1 passage before cell harvest. Genomic DNA from passage 5 iPSCs was extracted using the Gentra Puregene Cell Kit (Qiagen). Libraries were prepared using the Illumina TruSeq DNA Sample Prep Kit. In brief, DNA was fragmented (~200–350 bp) and ligated to the Illumina sequencing adaptor oligonucleotides. The adaptors-ligated fragments were amplified by PCR and then hybridized to the Illumina TruSeq Exome Enrichment Kit, which covers 1.22% of human genomic regions corresponding to the CDS (coding sequence) exons. The hybridized fragments were captured by streptavidin-coated magnetic beads, followed by sequencing on a HiSeq2000 sequencer using 100-bp paired-end reads. The image analysis and base calling were performed using the Illumina pipeline (v1.8) with default settings.

**Bioinformatic Analysis**

All reads were aligned to human reference sequence (release hg19) from University of California - Santa Cruz (UCSC) with the Burrows-Wheeler Aligner (BWA) version 0.6.1-r104 [46]. Picard version 1.57 was used to convert, sort, and index the aligned data files and remove PCR duplicates. For discovery of variations, we implemented a pipeline based on the Genome Analysis Toolkit (GATK) version 1.6–9 [47]. First, sequence reads were locally realigned and base-quality scores recalibrated. Second, variants were identified by the Unified Genotyper program in GATK. Third, low-quality variants were filtered using the Variant Filtration Walker tools in GATK and in-house developed codes. A minimum read depth of five and consensus quality of 50 was required at every examined location. Variants flanking homopolymer longer than 5 were removed. Any three or more variants located in a 50-bp window were discarded. Variants that had a record in the dbSNP database (version 135) were removed from consideration to reduce the false-positive rate [20]. For the heterozygous sites, both normal and variant depth should be more than five. For the homozygous sites, normal depth should be less than 1 and variant depth should be greater than 5. Variants that occurred in all the iPSC lines were removed from consideration. The filtered variants were annotated with ANNOVAR [48] and the effect of each variant was predicted with SIFT [49].

**Verification of SNVs**

To validate SNVs identified by bioinformatic analysis, we used a real-time PCR approach. We designed 3 primers for each point mutation. Two forward primers, with SNV site at the 3’ end, were manually designed and had melting temperatures of 50–55°C. One forward primer matches the wildtype allele, while another matches the SNV allele. The reverse primer was designed using Primer3Plus (http://frodo.wi.mit.edu/) with a melting temperature of 60°C. Equal amount of DNA (100 ng) was used for the sample that harbors a particular SNV and 4 controls that do not.

**Ingenuity Pathways Analysis**

To examine significantly over-represented networks and pathways, we analyze all the identified SNVs pooled from all the iPSC clones by Ingenuity Pathways Analysis (IPA, Ingenuity® Systems, www.ingenuity.com). Ingenuity knowledge base is the largest manually curated database for pathway analysis [50].

**COSMIC Analysis**

To test if the genes harboring variants occurred during the reprogramming process are enriched in gene set bearing cancer-associated mutations, we queried Catalogue of Somatic Mutations in Cancer (COSMIC) v62 (http://www.sanger.ac.uk/genetics/CGP/cosmic/).

**Statistics**

Data are presented as mean ± standard deviation (SD). Two-tailed Student t test was performed. P values of <0.05 were considered statistically significant.

**Results**

**Generation of iPSCs from CB CD34+ Cells with Three Different Combinations of Reprogramming Factors**

We are interested in reprogramming CB CD34+ cells, because CB has been proposed as a cell source in iPSC banking for allogeneic cell replacement therapy and CB may possess fewer genetic mutations than skin fibroblasts and PB [35]. To minimize the likelihood of clonal selection in low-efficiency reprogramming, we used lentiviral vectors to reprogram CB. Using an lentiviral vector optimized to achieve high-level transgene expression in hematopoietic cells [18], we have been able to reprogram 2% CB CD34+ cells into iPSCs with OCT4 and SOX2 (OS) alone, an efficiency that is ~1000-fold higher than previously reported [34]. This ability allowed us to compare the effects of different combinations of reprogramming factors on SNV loads in iPSC clones without the confounding effect of low reprogramming efficiency. For this purpose, we generated iPSCs using OS alone, using OS and ZSCAN4 (abbreviated as OSZ or Z for simplicity), or using OS and MK (abbreviated as OSMK or MK for simplicity). ZSCAN4 was used in combination with OS because it has been shown to enhance telomere lengthening, regulate genomic stability, and improve the quality of iPSCs [51,52,53]. The Yamanaka combination, OSMK, served as a control in our experiments since it has been employed in the majority of previous iPSC exome sequencing studies [20,22,54].

To minimize the accumulation of random mutations during long-term in vitro culture, we cultured CB CD34+ cells for only 2 days before lentiviral transduction. Consistent with our previous report, we found that 2% of CB CD34+ cells can be reprogrammed into iPSCs with OS (Figure 1A). However, in contrast to early studies [52,53], inclusion of ZSCAN4 appeared to decrease the reprogramming efficiency, albeit not reaching statistical significance (n = 3, P = 0.2; Figure 1A). This result is reminiscent of our early finding that KLF4, alone, does not increase OS-mediated reprogramming, likely because OS-mediated reprogramming is highly efficient [18]. However, the inclusion of both MYC and KLF4 (MK), expressed in a single vector, substantially increased reprogramming efficiency to 14% (n = 3, P < 0.05 compared to OS or OSZ; Figure 1A).
To accurately compare SNVs in each iPSC clone, we picked iPSC colonies generated from a single cord blood. Most iPSC clones were able to be passaged long-term and maintained typical iPSC morphology. We randomly selected 5 clones from each group for further analysis. No obvious differences were observed in the expression of pluripotency markers like OCT4, NANOG, SSEA4 and TRA-1-60 after passage 10 (Figure 1B and Figure S1). However, we did observe that higher portions of cells express...
TRA-1-60 at passages 3 in OSZ compared to OS iPSC clones (62 ± 9% vs. 42 ± 6%; n = 5, P < 0.01), suggesting that ZSCAN4 can increase the quality of OS-mediated reprogramming. This result is consistent with reports showing that inclusion of ZSCAN4 improves the quality of mouse iPSCs [52,53]. To further characterize the iPSC clones, we performed teratoma assays. Histological analysis showed that teratomas generated from all of the 15 iPSC lines consisted of tissues from three germ layers such as cartilage, gut-like structures, neurotubules, and pigmented epithelial cells (Figure S2ABC). Taken together, these data demonstrate that the 15 iPSC clones are bona fide pluripotent stem cells.

**Exome Sequencing**

To evaluate SNVs, we focused our analysis on mutations accumulated during reprogramming only, thus iPSCs at passage 5 were used. To minimize contamination of feeder cells, iPSCs were cultured in TeSR medium for 1 passage before harvest. To prevent unintended bias during cell culture, sample processing, exome capture and sequencing, all the 15 iPSC clones were cultured and processed in tandem. We enriched for protein coding genes using Illumina TruSeq Exome Enrichment Kit and sequenced the captured DNA from 15 samples using Illumia Genome Analyzer IIX with one sample per lane. After aligning the reads to the reference human genome (release hg19), we obtained 37–80 million uniquely aligned reads per sample (Table 1).

We searched for single base changes, small insertions/deletions and alternative splicing variants and identified more than 20,000 known and novel variants that had a minimum read depth of five and consensus quality of 30 for the majority of iPSC lines (Table 1). An iPSC variant is defined as a mutation if it is present only in one clone and absent in other iPSC lines. We reason that if there is a rare preexisting SNV fixed in 1 out of 15 iPSC clones, this SNV is unlikely to be detectable in the parent CB sample, because we set the algorithms to call an SNV positive only if it is present in more than 10% reads. Given this, we did not sequence the parental sample. We identified 548 heterozygous novel SNVs shared by all of the samples, indicating that they were pre-existing variants in the parent CB sample. In contrast to earlier report that some samples share the same SNVs, we found that none of SNVs in our study was shared by 2 or more out of 15 iPSC clones, suggesting that CB CD34+ cells are very homogenous and that our identified SNVs are unlikely to arise from rare pre-existing variants. We also identified 34 SNVs that were unique to specific clones.

**Verification of SNVs**

To verify the 34 SNVs identified by bioinformatics (Table 2), we developed a real-time PCR approach, which compared the differences in amplification efficiency when using a matched and a one-nucleotide mismatched primer at the 3’ end. This approach is demonstrated in Figure S3. The presence of a particular SNP led to more efficient PCR amplification when the relevant primer was used. When the difference amplification cycle or ΔΔCt was more than 1, the SNP was validated. We analyzed all the identified SNVs and 74% were verified by real-time PCR (Table S1). Due to technical limitations, SNVs that are present in 10% or less cells or located in repeat regions of the genome may not be validated. Some of the unvalidated SNVs may be false positives. However, to prevent underestimation of SNVs in CB iPSC lines, we pooled all the SNVs identified in any of the 5 iPSC clones generated with a particular factor combination for the following analyses.

**Few SNVs in Exomes of CB iPSCs and OSZ Appears to be a Better Combination for Generating iPSCs Harboring Fewer SNVs**

As shown in Table 2, we identified 14, 11, and 9 SNVs on exomes including untranslated regions (UTR) in five clones of OSMK, OS, and OSZ iPSC lines. Among them, there are only 8, 7, and 4 protein-coding mutations in the 3 groups of iPSC lines. There is a trend that OS iPSC lines appear to harbor fewer SNVs than OSMK iPSC lines, and inclusion of Z to OS further decreases SNV loads during reprogramming, but the differences

---

**Table 1. Summary of the exome sequencing data and the identified single nucleotide variants.**

| iPSC lines | # Total reads (M) | # Unique reads (M) | # Heterozygous variants | % in dbSNP | Reads of variants |
|------------|------------------|------------------|------------------------|-----------|-----------------|
| MK4        | 130.9            | 54.5             | 6150                   | 97        | 23              |
| MK5        | 127.1            | 51.8             | 24943                  | 97        | 50              |
| MK7        | 125.0            | 51.3             | 24907                  | 97        | 50              |
| MK8        | 143.4            | 58.7             | 25935                  | 97        | 54              |
| MK9        | 194.6            | 79.8             | 28150                  | 97        | 72              |
| OS1        | 89.6             | 36.8             | 19400                  | 97        | 38              |
| OS3        | 142.5            | 58.4             | 26142                  | 97        | 56              |
| OS5        | 141.0            | 58.0             | 25753                  | 97        | 54              |
| OS6        | 143.8            | 59.2             | 26247                  | 97        | 55              |
| OS7        | 124.8            | 51.2             | 24425                  | 97        | 49              |
| Z1         | 148.7            | 61.2             | 26187                  | 97        | 59              |
| Z2         | 126.7            | 52.2             | 23747                  | 97        | 52              |
| Z5         | 153.8            | 63.1             | 26277                  | 97        | 61              |
| Z6         | 154.3            | 62.8             | 26422                  | 97        | 62              |
| Z7         | 145.3            | 60.1             | 26046                  | 97        | 59              |

The numbers of heterozygous variants are those that have a minimum of 5× coverage. The dbSNP percentage represents the portion of identified variants present in the Single Nucleotide Polymorphism Database.

doi:10.1371/journal.pone.0059908.t001
are not statistically significant. In each clone, 1.3 (range: 0–3) coding mutations was identified, which is remarkably lower than 5–10 SNVs identified in previous studies using fibroblasts and low-efficiency reprogramming approaches. Of note, 2 out of 5 OS or OSZ iPSC lines did not acquire any coding mutations during reprogramming (Figure 2A).

Synonymous SNVs do not alter amino acid sequence and thus may not be harmful to the cells. Accordingly, we also analyzed nonsynonymous SNVs. Of significant interest, only 1 nonsynonymous SNV was observed in OSZ iPSC lines compared to 7 and 5 for MK and OS iPSC clones (Figure 2B). OSZ iPSCs harbor significantly fewer nonsynonymous SNVs than OSMK iPSCs (0.2 vs. 1.4; \( P < 0.05 \)). This result suggests that the combination of OSZ may be used to generate “safer” iPSCs with fewer potentially risky SNVs than the commonly used OSMK factors.

Pathway Analysis

Due to limited numbers of SNVs in each group, we combined all the 34 SNVs from 15 iPSC lines for analysis. Ingenuity Pathway analysis showed that the top network is cell development, cell growth and proliferation, hair and skin development and function (Figure S4). This result suggests that some SNVs might have improved iPSC proliferation.

Table 2. Genes found to be mutated in exomes of 15 CB iPSC lines.

| iPSC line | Chrom | Position | Ref | Alt | Codon | Substitution | Region | SNP Type | Gene Name | COSMIC gene |
|-----------|-------|----------|-----|-----|-------|--------------|--------|----------|-----------|-------------|
| MK4       | chr2  | 160086962| G   | A   | ATG-ATa| M1569I       | CDS    | NS       | TANC1     | No          |
| MK4       | chr3  | 24971717 | G   | –   | –     | –            | UTR    | S        | ELA1L2    | No          |
| MK4       | chr12 | 147435563| A   | G   | –     | –            | UTR    | S        | ZNF132    | No          |
| MK4       | chr17 | 70230071 | C   | G   | –     | –            | UTR    | S        | ZNF630    | No          |

The full details of each SNV including reads of SNV and wildtype alleles are in Table S1. MK: iPSC lines generated with OCT4, SOX2, MYC and KLF4; OS: iPSC lines generated with OCT4 and SOX2; Z: iPSC lines generated with OCT4, SOX2 and ZSCAN4. CDS: coding sequence; UTR: untranslated region; Downstream: SNV is at downstream of 5' UTR; S: synonymous coding mutation; NS: nonsynonymous coding mutation. doi:10.1371/journal.pone.0059908.t002

Few SNVs in Exomes of CB iPSCs
cancer. Only one out of 34 SNVs was found in this database, which is remarkably lower than the 50 out of 124 SNVs identified in the early report [20].

Discussion

Here we report that CB iPSCs harbor an average of 1.3 coding mutations per line. The SNV load appears to be dependent on factors used during reprogramming: Each OSMK iPSC line showed 1.6 protein-coding mutations, while OSZ iPSCs only acquired 0.8 such variations per line. In comparison, previous studies reported an average of 3–10 coding SNVs per iPSC line, a mutation rate that is estimated to be ~10-fold higher than background mutation during in vitro culture [20,22,27,32,33]. For the first time, we observed SNVs acquired during iPSC generation that is similar to or only slightly higher than that expected by random mutation. In addition, our novel finding that genome stabilizers like ZSCAN4 can significantly decrease genetic mutation rates during reprogramming should have important implications for the clinical application of cellular reprogramming.

Several factors might have contributed to exceeding low SNV loads that we report in our 15 CB iPSC lines. First, SNVs in previous studies may have been overestimated. Several studies have concluded that 50% or even the majority of identified SNVs are pre-existing in fibroblasts. However, the number of pre-existing SNVs may still be an underestimate because the rare mutations (such as those occurred at the rate of 10^-6) in skin fibroblasts and acquired during in vitro culture are unlikely to be detectable by current technologies [20,27,28]. Second, it is possible that SNVs in our study were underestimated. This is unlikely, because we intentionally decreased the required reads for SNVs from 10 in many studies to 5 to prevent false negative results. Accordingly, we identified >20,000 unique SNVs in the majority of our 15 iPSCs clones, which are substantially higher than many other studies [20,22,54]. A third possibility is that SNVs were reduced by our high-efficiency reprogramming approach. We converted 2–14% of transduced CB CD34^+ cells into iPSCs, an efficiency that is 100–10,000 fold higher than those reported in similar studies. This high reprogramming efficiency would reduce the possibility that iPSCs were generated from cells selected because they harbored SNVs favorable for reprogramming. In addition, the duration from gene transduction to reprogramming initiation or the first cell division was 3–7 days in our study, compared to ~2 weeks for fibroblast reprogramming. This may also have decreased chances for CB cells to accumulate more SNVs during seemingly quiescent stage after transduction of reprogramming factors.

We designed our experiments with aims to address two questions: 1) whether omitting MYC and KLF4 in the reprogramming combination can decrease SNV loads, and 2) whether genome stabilizer ZSCAN4 can decrease SNV loads. We did observe lower numbers of total SNVs and coding SNVs in OSZ iPSCs relative to OSMK iPSCs. But the differences are far from significant, suggesting that transient expression of MYC and other factors during reprogramming does not significantly increase mutation rates. With regard to ZSCAN4, we did observe a trend toward decreased coding SNVs from 1.6 in OSMK iPSCs to 0.8 in OSZ iPSCs. However, this difference did not reach statistical significance (P = 0.11), largely because the sample size was still small. However, nonsynonymous SNVs are significantly lower in OSZ iPSC lines compared to OSMK iPSC lines, suggesting that ZSCAN4 does play a positive role in stabilizing the genome and decreasing mutations during reprogramming. Taken together, these data suggest that the optimization of iPSC derivation conditions, through combinations of reprogramming factors and culture conditions, promotes genetic stability of pluripotent stem cells.

Because we identified few coding mutations in each clone of iPSCs, it was not our intent to determine how many of these SNVs were pre-existing in the cord blood sample. We cannot completely exclude the possibility that one or more of the identified SNVs are pre-existing rare mutations in the parent CB CD34^+ cell population. However, given that our reprogramming efficiency is 100–10,000 fold higher than that in the earlier studies and that CB CD34^+ cells are much more homogenous than skin fibroblasts, we believe that almost all the identified SNVs are de novo mutations that occurred during reprogramming.

Our data suggest that reprogramming of CB CD34^+ cells into iPSCs is not mutagenic, particularly when a genome stabilizer is included during reprogramming. However, this conclusion does not necessarily suggest that reprogramming of other types of cells like fibroblasts is not mutagenic. Cells that are difficult to reprogram such as fibroblasts are likely to result in increased mutations as compared to reprogramming of CB CD34^+ cells, because clones that harbor mutations favorable for reprogramming are selected for and extended period of culture required for
reprogramming increases chances for the accumulation of random mutations. Our data, together with these considerations, suggest that cord blood would be the best choice of cells for iPSC banking [35,35,56].

Taken together, our data demonstrate that it is possible to achieve reprogramming to full pluripotency with a very low level of SNV load that is close to the rate of random background mutation. Our finding that the genome stabilizer ZSCAN4 decreases coding mutation rates deserves further investigation on a large scale through whole genome sequencing.

Supporting Information

Figure S1 Flow cytometry analysis of CB iPSC lines. FACS diagrams show the expression of the pluripotency factor TRA-1-60 on bulk populations of 15 CB iPSC lines cultured with feeder support. OS, iPSCs generated with OS alone; Z, iPSCs generated with OSZ; MK, iPSCs generated with OSMK.

Figure S2 Teratoma formation from CB iPSC cells. iPSC cells were subcutaneously injected into NSG mice. After ~2 months, the teratomas were analyzed by haematoxylin and eosin staining. A) iPSCs generated with OS formed teratomas consisted of cartilage (mesoderm), gut-like structures (endoderm), and neurotubules and pigmented epithelium (ectoderm). B) iPSCs generated with OSZ (Z) formed teratomas consisted of cartilage (mesoderm), gut-like structures (endoderm), and neurotubules and pigmented epithelium (ectoderm). C) iPSCs generated with OSMK (MK) lines formed teratomas consisted of cartilage (mesoderm), gut-like structures (endoderm), and neurotubules and pigmented epithelium (ectoderm).

References

1. Takahashi K, Tanabe K, Ohnuki M, Narita M, Ichisaka T, et al. (2007) Induction of pluripotent stem cells from adult human fibroblasts by defined factors. Cell 131: 861-872.
2. Yu J, Vodyanik MA, Smuga-Otto K, Antosiewicz-Bourget J, Frane JL, et al. (2007) Induced pluripotent stem cell lines derived from human somatic cells. Science 318: 1917-1920.
3. Park IH, Zhao R, West JA, Yabuuchi A, Huo H, et al. (2008) Reprogramming of human somatic cells to pluripotency with defined factors. Nature 451: 141-146.
4. Warren L, Manos PD, Alheid T, Loh YH, Li H, et al. (2010) Highly efficient reprogramming to pluripotency and directed differentiation of human cells with synthetic modified mRNA. Cell Stem Cell 7: 618-630.
5. Warren L, Ni Y, Wang J, Guo X (2012) Feeder-free derivation of human induced pluripotent stem cells with messenger RNA. Sci Rep 2: 657.
6. Sori T, Yuasa S, Oda M, Egashira T, Yae K, et al. (2010) Generation of induced pluripotent stem cells from human terminally differentiated circulating T cells. Cell Stem Cell 7: 11-14.
7. Ban H, Nishishita N, Fusaki N, Tabata T, Saki K, et al. (2011) Efficient generation of transgene-free human induced pluripotent stem cells (iPSCs) by temperature-sensitive Sendai virus vectors. Proc Natl Acad Sci U S A 108: 14234-14239.
8. Nishimura K, Sano M, Ohtake M, Furuta B, Uemura Y, et al. (2011) Efficient generation of defect-free and persistent Sendai virus vector: a unique gene delivery/expression system ideal for cell reprogramming. J Biol Chem 286: 4760-4771.
9. Jin ZB, Okamoto S, Xiang P, Takahashi M (2012) Integration-free induced pluripotent stem cell derived from retinitis pigmentosa patient for disease modeling. Stem Cells Transl Med 1: 503-509.
10. Ono M, Hamaeda Y, Horinouchi Y, Matsuo-Takahashi M, Imoto Y, et al. (2012) Generation of induced pluripotent stem cells from human nasal epithelial cells using a Sendai virus vector. PLoS One 7: e24033.
11. Yu J, Hu K, Smuga-Otto K, Tian S, Stewart R, et al. (2009) Human induced pluripotent stem cells free of vector and transgene sequences. Science 324: 797-801.
12. Chou BK, Mah P, Huang X, Ye Z, Dowey SN, et al. (2011) Efficient human iPSC cell derivation by a non-integrating plasmid from blood cells with unique epigenetic and gene expression signatures. Cell Rex 21: 518-529.
13. Yu J, Chau KF, Vodyanik MA, Jiang J, Jiang Y (2011) Efficient feeder-free episomal reprogramming with small molecules. PLoS One 6: e17557.
14. Hu K, Yu J, Sukunthika K, Tian S, Montgomery K, et al. (2011) Efficient generation of transgene-free induced pluripotent stem cells from normal and neoplastic bone marrow and cord blood mononuclear cells. Blood 117: e109-119.
15. Okita K, Matsumura Y, Sato Y, Okada A, Morizane A, et al. (2011) A more efficient method to generate integration-free human iPSCs. Nat Methods 8: 409-412.
16. Okita K, Yamakawa T, Matsumura Y, Sato Y, Amano N, et al. (2012) An Efficient Non-viral Method to Generate Integration-Free Human iPSC Cells from Cord Blood and Peripheral Blood Cells. Stem Cells.
17. Mack AA, Kroboth S, Rajesh D, Wang WB (2011) Generation of induced pluripotent stem cells from CD34+ cells across blood drawn from multiple donors with non-integrating episomal vectors. PLoS One 6: e27956.
18. Meng X, Neises A, Su R-J, Payne KJ, Ritter L, et al. (2012) Efficient Reprogramming of Human Cord Blood CD34+ Cells Into Induced Pluripotent Stem Cells With OCT3/4 and SOX2 Alone. Mol Ther 20: 408-416.
19. Dowey SN, Huang X, Chou BK, Ye Z, Chong L (2012) Generation of integration-free human induced pluripotent stem cells from postnatal blood mononuclear cells by plasmid vector expression. Nat Protoc 7: 2013-2021.
20. Gower A, Li Z, Fung HL, Young JE, Agarwal S, et al. (2011) Somatic coding mutations in human induced pluripotent stem cells. Nature 471: 63-67.
21. Hussein SM, Batada NN, Vuoristo S, Ching RW, Auto R, et al. (2011) Copy number variation and selection during reprogramming to pluripotency. Nature 471: 50-62.
22. Ji J, Ng SH, Sharma V, Ncelai D, Hussein S, et al. (2012) Elevated coding mutation rate during the reprogramming of human somatic cells into induced pluripotent stem cells. Stem Cells 30: 435-440.
23. Laurent LC, Ulikin I, Slavin I, Tran H, Schork A, et al. (2011) 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 8: 106-118.
24. Martinez-Taylor K, Niiler BS, Taapken SM, Compton T, Crandall L, et al. (2011) Recurrent copy number variations in human induced pluripotent stem cells. Nat Biotechnol 29: 489-491.

Figure S3 Validation of SNV by real-time PCR. A) No obvious difference was observed when a pair of wildtype primers was used to amplify DNA from 5 iPSC lines. B) When the primer harboring the SNV at the 3′ end was used, the sample DNA containing the particular SNV amplified more efficiently, leading to lower cycles. C) In samples that ΔCt is substantially lower than the control, the SNV is validated (Left). However, in samples that ΔCt is not significantly lower than control (ΔΔCt < 1), the SNV is not validated (Right).

Figure S4 Ingenuity pathway analysis of all the 34 SNVs identified in 15 iPSC lines.

Table S1 Detailed information of genes found to be mutated in exomes of 15 CB iPSC lines.

Acknowledgments

Imaging was performed in the LLUMC Advanced Imaging and Microscopy Core that is supported by NSF Grant No. MRR-DBI 0923559 (Sean M Wilson) and the Loma Linda University School of Medicine. The authors thank Monica Rubalcava for technical support in confocal imaging.

Author Contributions

Conceived and designed the experiments: XBZ XF. Performed the experiments: RJS AN JW XBZ. Analyzed the data: YY KV EKW XF KJP. Contributed reagents/materials/analysis tools: KJP. Wrote the paper: XBZ.
25. Mayshar Y, Ben-David U, Lavon N, Biancotti JC, Yakir B, et al. (2010) Identification and characterization of chromosomal aberrations in human induced pluripotent stem cells. Cell Stem Cell 7: 521–531.

26. Martins-Taylor K, Xu RH (2012) Concise review: Genomic stability of human induced pluripotent stem cells. Stem Cells 30: 22–27.

27. Young MA, Larson DE, Sun CW, George DR, Ding L, et al. (2012) Background mutations in parental cells account for most of the genetic heterogeneity of induced pluripotent stem cells. Cell Stem Cell 10: 570–582.

28. Abyzov A, Mariani J, Palejev D, Zhang Y, Haney MS, et al. (2012) Somatic copy number mosaicism in human skin revealed by induced pluripotent stem cells. Nature 492: 438–442.

29. Nakagawa M, Takizawa N, Narita M, Ichisaka T, Yamanaka S (2010) Promotion of direct reprogramming by transformation-deficient Myc. Proc Natl Acad Sci U S A 107: 14152–14157.

30. Marion RM, Strati K, Li H, Murga M, Blanco R, et al. (2009) A p53-mediated DNA damage response limits reprogramming to ensure iPS cell genomic integrity. Nature 460: 1149–1153.

31. Kawamura T, Suzuki J, Wang YV, Menendez S, Morera LB, et al. (2009) Linking the p53 tumour suppressor pathway to somatic cell reprogramming. Nature 460: 1140–1144.

32. Cheng L, Hansen NF, Zhao L, Du Y, Zou C, et al. (2012) Low incidence of DNA sequence variation in human induced pluripotent stem cells generated by nonintegrating plasmid expression. Cell Stem Cell 10: 337–344.

33. Howden SE, Gore A, Li Z, Fung HL, Nider BS, et al. (2011) Genetic correction and analysis of induced pluripotent stem cells from a patient with gyrate atrophy. Proc Natl Acad Sci U S A 108: 6537–6542.

34. Giorgetti A, Montserrat N, Aasen T, Gonzalez F, Rodriguez-Piza I, et al. (2009) Oncogenic specification of human mesenchymal stem cells by hTERT immortalization. Cell Gene Therapy and Leaks to Enhanced Endosteal Bone Formation. PLoS One 4: e37569.

35. Hayden EC (2011) California ponders cell-banking venture. Nature 472: 403.

36. Ruiz S, Gore A, Li Z, Panopoulos AD, Montserrat N, et al. (2013) Analysis of DNA sequence variation in human induced pluripotent stem cells generated by nonintegrating plasmid expression. Cell Stem Cell 10: 337–344.

37. Jiang J, Lv W, Ye X, Zhang M, et al. (2013) Zscan4 promotes genomic stability during reprogramming and dramatically improves the quality of iPS cells. Nature 492: 438–442.

38. Wang K, Li M, Hakonarson H (2010) ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res 38: e164.

39. Kumar P, Henikoff S, Ng PC (2009) Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nat Protoc 4: 1073–1081.

40. Foja S, Jung M, Harwardt B, Riemann D, Pelz-Ackermann O, et al. (2013) Hypoxia Supports Reprogramming of Mesenchymal Stromal Cells Via Induction of Embryonic Stem Cell-Specific microRNA-302 Cluster and Pluripotency-Associated Genes. Cell Reprogram 15: 68–79.

41. Zalzman M, Falco G, Sharova LV, Nishiyama A, Thomas M, et al. (2010) Zscan4 regulates telomere elongation and genomic stability in ES cells. Nature 464: 858–863.

42. Zalzman M, Falco G, Sharova LV, Nishiyama A, Thomas M, et al. (2010) Zscan4 regulates telomere elongation and genomic stability in ES cells. Nature 464: 858–863.

43. Carey BW, Markoulaki S, Hanna J, Saha K, Gao Q, et al. (2009) Reprogramming of murine and human somatic cells using a single polycistronic vector. Proc Natl Acad Sci U S A 106: 157–162.

44. Zhang XB, Schwartz JL, Humphries RK, Kim HP (2007) Effects of HOXB4 overexpression on ex vivo expansion and immortalization of hematopoietic cells from different species. Stem Cells 25: 2074–2081.

45. Mali P, Chou BK, Yen J, Ye Z, Zou J, et al. (2010) Butyrate greatly enhances derivation of human induced pluripotent stem cells by promoting epigenetic remodeling and the expression of pluripotency-associated genes. Stem Cells 28: 713–720.

46. Bhadra A, Takahashi K, Okita K, Ichisaka T, Yamanaka S (2009) Hypoxia Enhances the Generation of Induced Pluripotent Stem Cells. Cell Stem Cell 5: 237–241.

47. Abyzov A, Mariani J, Palejev D, Zhang Y, Haney MS, et al. (2012) Somatic copy number mosaicism in human skin revealed by induced pluripotent stem cells. Nature 492: 438–442.

48. Nakagawa M, Takizawa N, Narita M, Ichisaka T, Yamanaka S (2010) Promotion of direct reprogramming by transformation-deficient Myc. Proc Natl Acad Sci U S A 107: 14152–14157.

49. Marion RM, Strati K, Li H, Murga M, Blanco R, et al. (2009) A p53-mediated DNA damage response limits reprogramming to ensure iPS cell genomic integrity. Nature 460: 1149–1153.

50. Kawamura T, Suzuki J, Wang YV, Menendez S, Morera LB, et al. (2009) Linking the p53 tumour suppressor pathway to somatic cell reprogramming. Nature 460: 1140–1144.

51. Cheng L, Hansen NF, Zhao L, Du Y, Zou C, et al. (2012) Low incidence of DNA sequence variation in human induced pluripotent stem cells generated by nonintegrating plasmid expression. Cell Stem Cell 10: 337–344.

52. Howden SE, Gore A, Li Z, Fung HL, Nider BS, et al. (2011) Genetic correction and analysis of induced pluripotent stem cells from a patient with gyrate atrophy. Proc Natl Acad Sci U S A 108: 6537–6542.

53. Giorgetti A, Montserrat N, Aasen T, Gonzalez F, Rodriguez-Piza I, et al. (2009) Generation of induced pluripotent stem cells from human cord blood using OCT4 and SOX2. Cell Stem Cell 3: 335–357.

54. Broxmeyer HE, (2010) Will iPSCs cells enhance therapeutic applicability of cord blood cells and banking? Cell Stem Cell 6: 21–24.

55. Yang J, Wu Y, Wang L, Zhang M, et al. (2012) Zscan4 regulates telomere elongation and genomic stability in ES cells. Nature 464: 858–863.