

ARTICLE

Received 30 May 2014 | Accepted 16 Jan 2015 | Published 18 Feb 2015

DOI: 10.1038/ncomms7318

Unique features of mutations revealed by sequentially reprogrammed induced pluripotent stem cells

Shuai Gao^{1,2,*}, Caihong Zheng^{3,*}, Gang Chang^{2,*}, Wenqiang Liu⁴, Xiaochen Kou⁴, Kun Tan¹, Li Tao¹, Kai Xu⁴, Hong Wang⁴, Jun Cai³, Jianhui Tian¹ & Shaorong Gao^{2,4}

Although viable mice can be generated from induced pluripotent stem cells (iPSCs), the impact of accumulated mutations on the developmental potential of the resulting iPSCs remains to be determined. Here, we demonstrate that all-iPSC mice generated through tetraploid blastocysts complementation can tolerate the accumulation of somatic mutations for up to six generations using a Tet-on inducible reprogramming system. But, the viability of the all-iPS mice decreased with increasing generations. A whole-genome sequencing survey revealed that thousands of single-nucleotide variations (SNVs), including 44 non-synonymous ones, accumulated throughout the sequential reprogramming process. Subsequent analysis provides evidence that these accumulated SNVs account for the gradual reduction in viability of the resultant all-iPSC mice. Unexpectedly, our present reprogramming system revealed that pluripotent stem cells are heterogeneous in terms of possessing a set of copy-number alterations (CNAs). These CNAs are unique for pluripotent cells and subsequently disappear in the differentiating progenies.

¹Ministry of Agriculture Key Laboratory of Animal Genetics, Breeding and Reproduction; National Engineering Laboratory for Animal Breeding; College of Animal Sciences and Technology, China Agricultural University, Beijing 100193, China. ²National Institute of Biological Sciences, NIBS, Beijing 102206, China. ³Key Laboratory of Genomic and Precision Medicine, Beijing Institute of Genomics, CAS Beijing 100101, China. ⁴Clinical and Translational Research Center of Shanghai First Maternity & Infant Hospital, School of Life Sciences and Technology, Tongji University, Shanghai 200092, China. * These authors contributed equally to this work. Correspondence and requests for materials should be addressed to J.C. (email: juncai@big.ac.cn) or to J.T. (email: tianjh@cau.edu.cn) or to S.G. (email: gaoshaorong@tongji.edu.cn).

The successful reprogramming of differentiated somatic cells into pluripotent stem cells has revolutionized our understanding of cell plasticity^{1,2}. The generation of patient-specific pluripotent stem cells directly from somatic cells has raised the possibility of utilizing these personalized pluripotent stem cells in regenerative medicine, pathological studies and drug screening^{3–6}.

Somatic cell nuclear transfer (SCNT) and the induction of pluripotency by defined transcription factors are two major approaches that can reprogram differentiated mammalian somatic cells to a pluripotent and even a totipotent state by SCNT^{1,2,7–11}. The results of a very recent study have demonstrated that mice can be successfully re-cloned to more than 25 generations through serial nuclear transfer, which indicates that a very strict selection strategy to eliminate the defective cloned embryos during development might be utilized in SCNT-mediated reprogramming¹². Similarly, we and others have previously shown that induced pluripotent stem cells (iPSCs) reprogrammed via a Tet-on inducible reprogramming system can also produce all-iPSC clonal mice through tetraploid blastocyst complementation^{13–16}. However, the number of generations of all-iPSC mice that can be serially produced using this inducible iPS system has not been determined.

It has been reported that defined transcription factors that mediate reprogramming can potentially induce or enlarge mutations in the resultant iPSCs, but the impact of the mutations on the developmental potential of resultant iPSCs remains to be evaluated^{17–21}. The Tet-on inducible iPS system provides an invaluable framework that enables the evaluation of the developmental potential of sequentially reprogrammed iPSCs and the determination of the effects of genetic alterations. Here, we demonstrate that the accumulation of somatic mutations was tolerated by all-iPSC mice for up to six generations, but directly impacted the survival of all-iPSC mice in later generations.

Results

Generation of adult all-iPSC mice to three generations. Using a sequential reprogramming approach based on the doxycycline (Dox)-inducible iPS system, we first aimed to determine whether viable all-iPSC mice with normal fertility could be sequentially produced through tetraploid blastocyst complementation in a similar manner to that used to produce serial cloned mice through somatic cell nuclear transfer (SCNT)²² (Fig. 1a). The 1⁰-all-iPSC mice were generated from one Tet-on inducible iPSC line that was initially derived from mouse embryonic fibroblasts (MEFs) with the 129/Sv × M2rtTA genetic background²³. Subsequently, adipocyte progenitor cells (APCs) were retrieved from the adipose tissue of the 1⁰-all-iPSC mice (Supplementary Fig. 1a,b). After the addition of Dox to the induction medium, 2⁰-iPSC colonies emerged that exhibited the typical ESC-like morphology and positive staining for alkaline phosphatase (AP) (Supplementary Fig. 1c,d). After propagation, pluripotency markers, including Oct4 (also known as Pou5f1), Nanog, Sox2 and the ESC-specific surface marker SSEA-1, were positively expressed in the 2⁰-iPSCs (Supplementary Fig. 1e). Bisulfite genomic sequencing analysis of the DNA methylation status of the *Pou5f1* and *Nanog* promoters confirmed that DNA demethylation had occurred during secondary reprogramming (Supplementary Fig. 1f,g). The silencing of the exogenous factors was further confirmed in the iPSC lines that were tested (Supplementary Fig. 1h). Furthermore, after the injection of iPSCs into severe combined immunodeficient mice, the formation of teratomas with three germ layers was observed (Supplementary Fig. 1i). Moreover, chimeric mice (2N) with germline transmission ability could be produced from these 2⁰-iPSC lines

(Supplementary Fig. 1j). Most importantly, viable fertile all-iPSC mice could be generated from the APC-derived 2⁰-iPSC lines through tetraploid complementation (Fig. 1b). Subsequently, we produced a third generation of viable all-iPSC mice with normal fertility from the 3⁰-iPSC lines that were established from APCs retrieved from the 2⁰-all-iPSC mice (Fig. 1b). Simple sequence length polymorphism (SSLP) analyses were performed, which confirmed that the all-iPSC mice were indeed produced from iPSCs (Fig. 1c). We concluded that adult all-iPSC mice with normal fertility could be produced using this Tet-on inducible iPS system.

Reduced viability of all-iPSC mice with increasing generations.

To further ascertain whether viable adult all-iPSC mice with normal fertility could be generated to many generations, we established 4⁰-, 5⁰- and 6⁰-iPSC lines from the somatic cells retrieved from the previous generations of all-iPSC mice, and 4⁰-, 5⁰- and 6⁰-all-iPSC mice were subsequently produced through tetraploid complementation (Fig. 1d–g). In summary, the average induction efficiency of sequential reprogramming based on AP staining was ~0.7% (Supplementary Fig. 1k), and the efficiency of all-iPSC mouse formation (average 1.4%) was comparable to that described in recent reports¹⁴ (Supplementary Table 1). However, the viability of the all-iPSC mice produced was greatly reduced with increasing generations. For example, the 1⁰–3⁰-all-iPSC mice could grow into fertile adults, while the 4⁰-all-iPSC and 5⁰-all-iPSC mice only survived up to 4 weeks and 2 days, respectively. Most strikingly, all of the 27 6⁰-all-iPSC mice died immediately after caesarean section (Fig. 1h and Supplementary Table 1).

We first performed 5-methylcytosine (5-mC) MeDIP-seq and RNA-seq analysis of the sequentially reprogrammed iPSC lines. The results revealed that only a few differential methylation regions (DMRs) located in gene promoter regions displayed an accumulating pattern. However, the expression traces in sequentially reprogrammed iPSC of downstream genes demonstrated that no genes were cis-regulated by an accumulating pattern of differential methylation in the promoter region (Supplementary Fig. 2). Moreover, in a recent study, we have demonstrated that the core histone modifications (H3K4me2, H3K4me3, H3K27me3) are comparable among iPSC lines that can produce all-iPSC mice²⁴. Taken together, we conclude that although all-iPSC mice can be produced up to six generations, their viability decreases with increasing generations, and epigenetic effects might not be a major cause of this reduction in viability.

Single-nucleotide variations accumulated during sequential reprogramming.

To further investigate whether sequential reprogramming altered the genome integrity of the resultant iPSCs and subsequently affected the viability of all-iPSC mice, a whole-genome sequencing approach was applied to investigate the effects of genetic alterations on the developmental potential of the sequentially reprogrammed iPSCs. DNA was extracted from a total of eight sequentially reprogrammed iPSC lines and five parental somatic cells. Paired-end reads of up to 150 Giga and 50 × for the mouse genome were generated for each sample using a HiSeq 2000 sequencer (Supplementary Fig. 3a). The paired-end reads and gene expression analyses confirmed that there were nine stable lentiviral integration sites in the exogenous *Oct4*, *Sox2*, *Klf4* and *c-Myc* vectors in the genomes throughout the sequential reprogramming process, and these sites did not disrupt any endogenous genes (Supplementary Fig. 3b and Supplementary Data 1).

Thousands of single-nucleotide variations (SNVs) that occurred throughout the sequential iPS process were surveyed

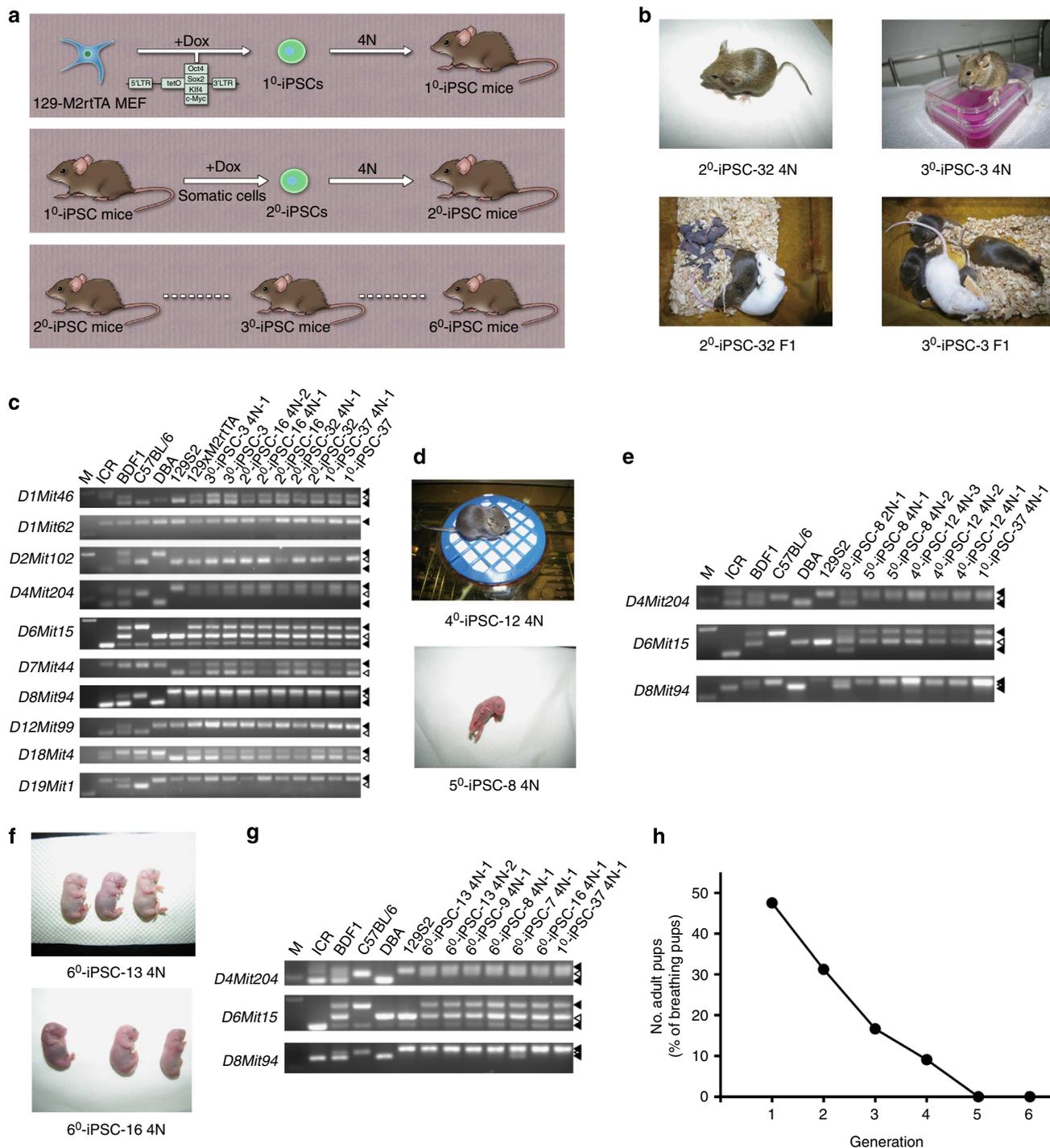


Figure 1 | Sequentially reprogrammed iPSCs support the production of all-iPSC mice for up to six generations. (a) A schematic diagram for the production of all-iPSC mice for up to six generations using sequentially reprogrammed iPSCs. (b) The viable all-iPSC mice that were derived from sequentially reprogrammed iPSC lines (2⁰-iPSCs and 3⁰-iPSCs) through tetraploid complementation (top) and the offspring (F1) of the all-iPSC mice (bottom). (c) SSLP analysis for the genetic identification of the all-iPSC mice. (d,e) The all-iPSC mice derived from 4⁰-iPSC, 5⁰-iPSC (d) and SSLP analysis for the genetic identification of all-iPSC mice (e). (f,g) The all-iPSC mice derived from 6⁰-iPSC (f) and SSLP analysis for the genetic identification of all-iPSC mice (g). (h) The percentage of live-born all-iPSC pups that reached adulthood in each generation.

by pair-wise comparisons among the samples. The SNVs were fractionated according to the various accumulating patterns (Fig. 2a and Supplementary Fig. 3c). A total of 189 SNVs, including all of the coding SNVs, were verified via Sequenom genotyping in all 13 samples in addition to the feeder cell line (Supplementary Fig. 3d,e and Supplementary Data 2). The genotyping results also suggested that pollution of the feeder

cells in the DNA isolation from the iPSC lines caused the recurrence of some SNVs in all of the iPSC lines (Supplementary Data 2).

The reduction in viability of the sequential all-iPSC mice demonstrated the impact of one type of SNV on the developmental potential of the resultant sequentially reprogrammed iPSCs. This SNV accumulated in each generation and then was

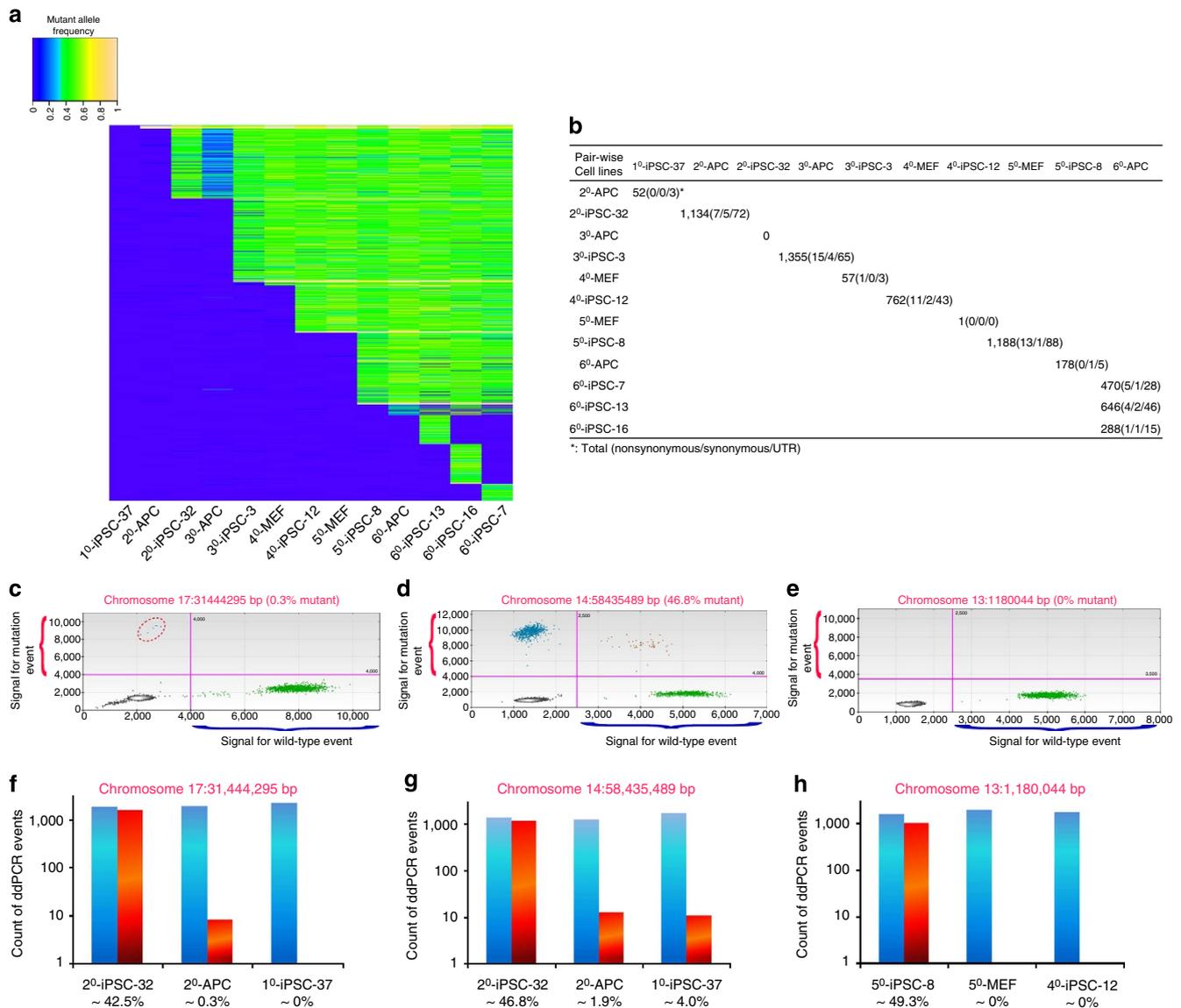


Figure 2 | The accumulated SNVs and their frequencies during sequential reprogramming. (a) A heat map of the dynamic frequencies representative of the type of accumulated SNVs on the trace of the sequential iPSC procedures. Note: The DNA of the 3⁰-APCs was extracted from three all-iPSC mice that were derived from the same iPSC line. (b) The numbers of accumulated SNVs emerged in each stage throughout the sequential reprogramming. (c) Detection of low frequencies of SNVs in the 2⁰-APC by droplet digital PCR. The signal intensities associated with the PCR amplification across the mutant allele (y axis, red) and the reference allele (x axis, blue). The scatter dot surrounded by a red dashed circle represents the mutant allele. There were significantly fewer dots for PCR events in the SNV than in the reference allele. (d,e) A scatter plot showing the fixed accumulated SNVs confirmed in 6⁰-iPSC-13 (d); the SNVs were not detected in 5⁰-MEF somatic cells of the parental generation (e). (f) Droplet digital PCR analysis of an SNV showing that the accumulated SNV pre-existed in the somatic cells of the parental generation at low frequencies and reached high frequencies in subsequent iPSCs. Counts of ddPCR events for SNVs (blue bar) and the reference (red bar) resulted in an estimated cell frequency in 2⁰-iPSC-32, 2⁰-APC, and 1⁰-iPSC-37 of 42.5%, 0.3%, 0%, respectively. (g) The accumulated SNV pre-existed in somatic cells of the parental generation and previous iPSC cell lines at a low frequency as detected by ddPCR analysis. (h) Droplet digital PCR analysis showing that the accumulated SNV was not detectable in the somatic cells of the parental generation and was detected in subsequent iPSCs with a relatively high frequency.

inherited and fixed in subsequent offspring due to the induction of a genetic bottleneck. The heat map of the frequencies illustrated a dynamic trace of the accumulation of these SNVs, which were confirmed to have a high validation rate (Fig. 2a). The SNVs were subdivided according to the time of emergence across the sequential reprogramming, and then their numbers were calculated (Fig. 2b).

Fifty-nine validated coding SNVs, including 44 non-synonymous SNVs, accumulated throughout the sequential process (Table 1). Among them, 16 were annotated in the genes with lethal homozygous phenotypes in the Mouse Genome Informatics

database (Supplementary Table 2)²⁵. In addition, on the basis of a survey of the literature, seven that accumulated during a later period of the sequential reprogramming were involved in genes with some heterozygous phenotypes (Supplementary Table 2)^{26–46}. In particular, *Ep300* in 3⁰-iPSCs and subsequently *Ryr2* in 5⁰-iPSCs shared similar one-allele mutant and haploinsufficiency phenotypes that were associated with cardiac developmental defect. In addition, mice with haploinsufficiency for *Cd2ap* exhibited a phenotype similar to human focal segmental glomerulosclerosis (Supplementary Table 2). These phenotypes were supported by our histopathological examination of

Table 1 | Comprehensive information about the SNVs located in the coding sequence (CDS) regions.

Chromosome	Position	Reference allele	Mutant allele	Gene name	Amino acid changes	Step	Sequencing allele frequency		Frequency in iPSCs (Sequenom)		Frequency in parental somatic cells (ddPCR)
							Upstream sample	Downstream sample	Upstream sample	Downstream sample	
1	75,129,287	C	T	Slc23a3	R→H	5 ⁰ -MEF⇒5 ⁰ -iPSC-8	0	0.39	0.01	0.45	—
1	75,553,365	T	G	Slc4a3	F→V	4 ⁰ -MEF⇒4 ⁰ -iPSC-12	0	0.45	0	0.46	0.8%
1	88,283,154	A	C	Nmur1	F→V	5 ⁰ -MEF⇒5 ⁰ -iPSC-8	0	0.43	0	0.60	F
1	89,849,004	G	A	Usp40	L⇒L	6 ⁰ -APC⇒6 ⁰ -iPSC-16	0	0.45	0.02	0.51	0.9%
1	91,685,765	C	T	Agap1	S→F	5 ⁰ -MEF⇒5 ⁰ -iPSC-8	0	0.41	0	0.64	—
1	159,088,959	C	G	Rasal2	A→P	3 ⁰ -APC⇒3 ⁰ -iPSC-3	0	0.52	0	0.31	0.3%
2	28,695,619	A	C	Gtf3c4	F→L	5 ⁰ -MEF⇒5 ⁰ -iPSC-8	0.02	0.40	0	0.27	0
2	34,751,348	G	A	Phf19	A→V	5 ⁰ -MEF⇒5 ⁰ -iPSC-8	0	0.47	—	—	—
2	67,352,953	A	C	Xirp2	T→P	2 ⁰ -APC⇒2 ⁰ -iPSC-32	0	0.52	0	0.42	0
2	70,425,725	A	C	Gad1	I→L	4 ⁰ -MEF⇒4 ⁰ -iPSC-12	0	0.41	0.05	0.44	0
2	88586936	G	A	Olfr1198	L⇒L	5 ⁰ -iPSC-8⇒6 ⁰ -APC	0	0.08	0	0.18	—
2	111,323,964	G	A	Olfr1289	V→I	3 ⁰ -APC⇒3 ⁰ -iPSC-3	0	0.35	0	0.42	—
3	106,624,742	G	T	Olfr266	S→Y	6 ⁰ -APC⇒6 ⁰ -iPSC-13	0.02	0.41	0.01	0.64	0
4	86,061,429	G	C	Adamts1	R→T	3 ⁰ -APC⇒3 ⁰ -iPSC-3	0	0.39	0	0.29	—
4	89,891,735	T	C	Zfp352	Y→H	3 ⁰ -APC⇒3 ⁰ -iPSC-3	0	0.40	0.07	0.43	—
4	141,259,094	G	T	Ddi2	G⇒G	6 ⁰ -APC⇒6 ⁰ -iPSC-16	0.02	0.47	0	0.44	—
4	143,749,551	C	T	Oog3	E→K	6 ⁰ -APC⇒3 ⁰ -iPSC-3	0	0.47	0	0.45	0.2%
5	108,030,814	G	A	Rpap2	V→I	3 ⁰ -APC⇒3 ⁰ -iPSC-3	0	0.41	0	0.42	F
5	121,810,124	C	A	Gm15800	F→L	3 ⁰ -APC⇒3 ⁰ -iPSC-3	0	0.45	0	0.37	—
5	139,048,292	G	T	Zfp68	F→L	5 ⁰ -MEF⇒5 ⁰ -iPSC-8	0	0.54	0.03	0.65	—
6	83,596,856	G	A	Clec4f	P→S	4 ⁰ -MEF⇒4 ⁰ -iPSC-12	0.08	0.43	0	0.54	1.0%
6	118,078,765	C	T	Csgalnact2	E⇒E	2 ⁰ -APC⇒2 ⁰ -iPSC-32	0	0.33	0	0.40	—
6	144,065,075	C	T	Sox5	V→M	4 ⁰ -MEF⇒4 ⁰ -iPSC-12	0	0.47	0.16	0.54	0
7	13,764,009	C	T	Vmn1r88	L⇒L	2 ⁰ -APC⇒2 ⁰ -iPSC-32	0.01	0.51	0.07	0.61	—
7	16,573,743	A	C	Gltscr1	T⇒T	3 ⁰ -APC⇒3 ⁰ -iPSC-3	0.02	0.44	0.03	0.53	—
7	23,575,675	A	G	Vmn1r158	L⇒L	2 ⁰ -APC⇒2 ⁰ -iPSC-32	0	0.48	—	—	—
7	31,200,656	C	T	Tyrobp	R→W	6 ⁰ -APC⇒6 ⁰ -iPSC-16	0	0.30	0.18	0.39	0.4%
7	80,594,099	G	C	Chd2	P→A	6 ⁰ -APC⇒6 ⁰ -iPSC-16	0	0.43	0.01	0.41	F
8	29,835,629	T	C	Unc5d	S→G	6 ⁰ -APC⇒6 ⁰ -iPSC-7	0	0.49	0.17	0.62	—
8	78,193,344	T	C	1700007B14Rik	H⇒H	6 ⁰ -APC⇒6 ⁰ -iPSC-7	0	0.40	0.02	0.37	F
8	128,194,446	A	T	4933403G14Rik	E→V	3 ⁰ -APC⇒3 ⁰ -iPSC-3	0	0.47	—	—	—
9	35,928,654	A	T	Gm5916	V→D	3 ⁰ -APC⇒3 ⁰ -iPSC-3	0	0.52	—	—	—
9	38,256,777	T	G	Olfr902	F→V	4 ⁰ -MEF⇒4 ⁰ -iPSC-12	0	0.55	—	—	—
9	38,868,101	T	C	Olfr937	K→R	4 ⁰ -MEF⇒4 ⁰ -iPSC-12	0	0.35	—	—	—
9	98,408,306	C	A	Rbp2	Y→*	3 ⁰ -APC⇒3 ⁰ -iPSC-3	0	0.48	0	0.32	—
9	105,651,236	T	C	Col6a6	D→G	4 ⁰ -MEF⇒4 ⁰ -iPSC-12	0.02	0.37	—	—	—
10	29,037,515	G	T	Echdc1	M→I	5 ⁰ -MEF⇒5 ⁰ -iPSC-8	0	0.59	0.04	0.59	0.2%
10	60,689,870	T	A	Adamts14	H→L	6 ⁰ -APC⇒6 ⁰ -iPSC-16	0	0.39	0.03	0.30	1.8%
10	129,855,780	T	A	Vmn2r85	K→M	5 ⁰ -MEF⇒5 ⁰ -iPSC-8	0	0.47	—	—	—
10	21,900,990	G	A	Raet1e	A→T	4 ⁰ -MEF>4 ⁰ -iPSC-12	0	0.33	—	—	—
11	6,174,740	G	T	Tmed4	H→N	2 ⁰ -APC⇒2 ⁰ -iPSC-32	0	0.58	0	0.66	0.2%
11	69,297,838	C	T	Dnahc2	V→I	6 ⁰ -APC⇒6 ⁰ -iPSC-13	0	0.49	0.02	0.48	0.09%
11	69,481,221	G	T	Eif4a1	T→N	6 ⁰ -APC⇒6 ⁰ -iPSC-13	0.02	0.45	0.06	0.55	F
11	103,248,606	T	C	Plekhl1	N→S	3 ⁰ -APC⇒3 ⁰ -iPSC-3	0	0.55	0.03	0.45	—
12	29,279,600	C	T	Colec11	V⇒V	5 ⁰ -MEF⇒5 ⁰ -iPSC-8	0	0.52	0.02	0.57	—
12	72,000,398	A	G	Frmd6	T→A	4 ⁰ -MEF⇒4 ⁰ -iPSC-12	0.02	0.50	0	0.48	—
13	11,800,441	C	A	Ryr2	L→F	5 ⁰ -MEF⇒5 ⁰ -iPSC-8	0	0.52	0.09	0.47	0
13	93,154,737	A	G	Dhr	E⇒E	4 ⁰ -MEF⇒4 ⁰ -iPSC-12	0	0.42	—	—	—
14	21,267,806	G	A	Ttc18	H→Y	6 ⁰ -APC⇒6 ⁰ -iPSC-13	0	0.36	—	—	—
14	33,239,472	A	G	Chat	S→P	6 ⁰ -APC⇒6 ⁰ -iPSC-16	0	0.41	0.01	0.64	0
14	55,683,018	T	C	Zfhx2	K⇒K	2 ⁰ -APC⇒2 ⁰ -iPSC-32	0	0.44	0.34	0.54	—
14	56,513,904	C	C	Khny1	R→R	4 ⁰ -MEF⇒4 ⁰ -iPSC-12	0	0.51	0	0.66	—
14	58,435,489	T	A	F630043A04Rik	E→D	2 ⁰ -APC⇒2 ⁰ -iPSC-32	0	0.42	0	0.62	1.9%
14	66,349,684	G	A	Scara5	V→M	3 ⁰ -APC⇒3 ⁰ -iPSC-3	0	0.29	—	—	—
15	78,629,558	C	G	Card10	A→P	3 ⁰ -APC⇒3 ⁰ -iPSC-3	0	0.52	0.01	0.63	F
15	80,042,465	T	G	Mgat3	G⇒G	2 ⁰ -APC⇒2 ⁰ -iPSC-32	0.02	0.44	0	0.44	—
15	81,481,092	C	G	Ep300	P→A	3 ⁰ -APC⇒3 ⁰ -iPSC-3	0	0.59	0	0.50	—
15	98,915,211	A	T	C1ql4	S→T	5 ⁰ -MEF⇒5 ⁰ -iPSC-8	0	0.40	0	0.51	—
15	101,380,713	C	A	Krt82	D→Y	3 ⁰ -APC⇒3 ⁰ -iPSC-3	0	0.58	0.12	0.53	1.2%
16	4,850,255	A	T	4930562C15Rik	G⇒G	3 ⁰ -APC⇒3 ⁰ -iPSC-3	0	0.38	0	0.53	—
16	19,769,277	A	G	B3gnt5	N→S	2 ⁰ -APC⇒2 ⁰ -iPSC-32	0	0.62	0.06	0.50	0.5%
17	3,144,4295	A	G	Slc37a1	S→G	2 ⁰ -APC⇒2 ⁰ -iPSC-32	0	0.52	0	0.39	0.3%
17	35,195,283	C	G	Clic1	C→W	3 ⁰ -APC⇒3 ⁰ -iPSC-3	0	0.57	0	0.38	—
17	42,942,180	A	T	Cd2ap	N→K	5 ⁰ -MEF⇒5 ⁰ -iPSC-8	0	0.40	0	0.56	3.4%
17	46,893,992	C	A	Ptcr1	S→I	6 ⁰ -APC⇒6 ⁰ -iPSC-13	0	0.53	0	0.56	F
18	10,458,346	T	G	Greb11	L→V	4 ⁰ -MEF⇒4 ⁰ -iPSC-12	0	0.54	0	0.42	F
18	37462679	C	A	Pcdhb3	V⇒V	3 ⁰ -APC⇒3 ⁰ -iPSC-3	0	0.39	0.01	0.43	—
18	59,128,425	T	G	Adamts19	L→V	4 ⁰ -MEF⇒4 ⁰ -iPSC-12	0	0.35	0	0.34	—
18	65,450,956	C	T	Alpk2	R→K	3 ⁰ -APC⇒3 ⁰ -iPSC-3	0	0.40	0.18	0.47	—
18	80,326,096	G	T	Adnp2	I⇒I	6 ⁰ -APC⇒6 ⁰ -iPSC-13	0	0.41	0.03	0.51	0
18	89,278,694	T	G	Rtt1	L→R	3 ⁰ -iPSC-3⇒4 ⁰ -MEF	0.08	0.49	—	—	—
19	11,015,047	T	C	Gpr44	V→A	5 ⁰ -MEF⇒5 ⁰ -iPSC-8	0	0.52	—	—	—
19	39,479,161	T	A	Cyp2c38	P⇒P	3 ⁰ -APC⇒3 ⁰ -iPSC-3	0	0.45	—	—	—
X	136,098,234	G	A	Rnf128	G→D	5 ⁰ -MEF⇒5 ⁰ -iPSC-8	0	0.75	—	—	—

In the column of 'Frequency in iPSCs (Sequenom)', '—' indicates these sites failed in the primer design step. In the column of 'Frequency in parental somatic cells (ddPCR)', '—' denotes not attempted. An 'F' indicates a failed reaction due to probe design fail or insufficient ddPCR events for analysis.

6⁰-all-iPSC mouse tissues, which revealed an abnormal morphogenesis of the heart and kidney (Supplementary Fig. 4). The combined effect of missense mutations in *Ep300*, *Ryr2* and *Cd2ap* might gradually disrupt the tolerable genetic load and account for the developmental failure of all-iPSC mice in late generations.

Tracking causes of the accumulated single-nucleotide variations.

To distinguish the origins of these accumulated SNVs, we subsequently applied droplet digital PCR (ddPCR) to monitor the frequencies of these SNVs in parental somatic cells and iPSCs of previous generations (Fig. 2c–e and Supplementary Table 3)^{47,48}. Three primary frequency patterns emerged of SNVs in iPSCs and

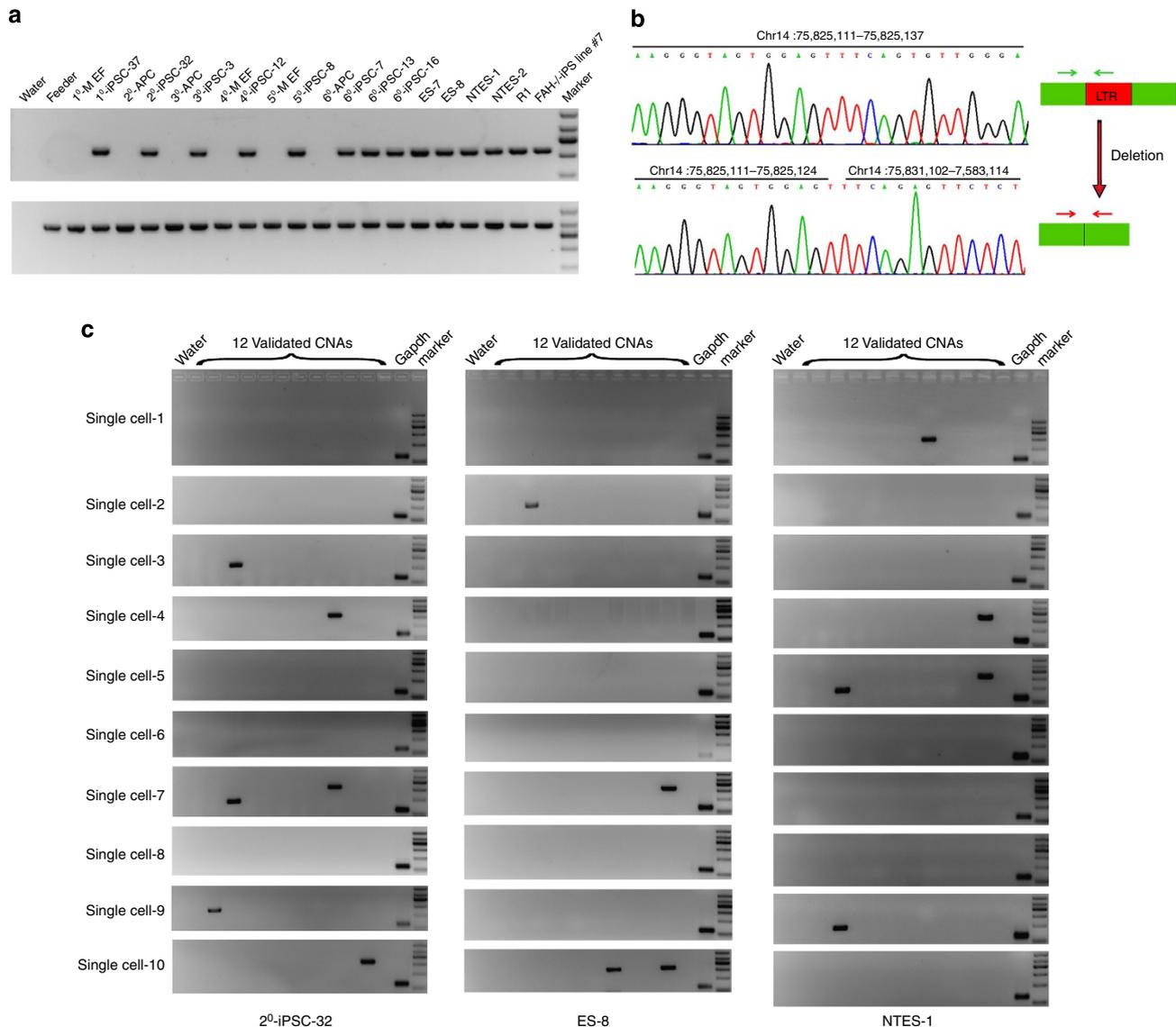


Figure 3 | Repeated copy-number alterations (CNAs) in pluripotent stem cells. (a) PCR results for the iPSC-specific CNA at chr14:75,825,104–75,831,195 (upper panel) and the control intact allele that retained the CNA (lower panel). A unique amplified breakpoint-spanning PCR product was present in all of the pluripotent stem cell lines but not in all of the somatic cell lines. (b) Sanger sequencing of the amplified band and schematic of the PCR validation assay design. The exact deletion breakpoints were confirmed, and the deleted regions were in accordance with the annotation of the retrotransposon (LINE). (c) The single-cell PCR results for all of the 12 validated CNAs (from lane 2 to lane 13 according to the chromosome position of the CNAs) in 2⁰-iPSC-32, ES-8 and NTES-1.

are summarized in Fig. 2f–h. Two types of SNVs were pre-existing but distinct with regard to their time of origin. The first type of SNV with a fixed frequency in iPSCs had a low frequency only in the parental somatic cells (Fig. 2f), which occurred in cell divisions during the development of the all-iPSC mice. The second type exhibited a low frequency in both the parental somatic cells and the iPSCs of the early generations (Fig. 2g). Each SNV of the second type was initiated in an individual cell during cell culture of the early generations of the iPSC line but was retained in a mosaic all-iPSC mouse that was derived from the tetraploid complementation of the 10–15 iPSCs injection. The SNVs in the last category, which were considered *de novo* SNVs, had no detectable frequencies via ddPCR in parental somatic cells and iPSCs of the previous generation but were detected in subsequent iPSCs with a high frequency (Fig. 2h). In summary, at least two-thirds of the SNVs originated before iPSC induction and pre-existed in the all-iPSC mouse tissues (Supplementary Table 4).

Repeated CNAs observed in pluripotent stem cells. Unlike SNVs, no apparent copy-number alterations (CNAs) accumulated during the sequential reprogramming. Instead, the genomic data for the sequential reprogramming revealed a pattern in which several uniform CNAs were found in each sequential iPSC lines but were not detected in the somatic cells derived from the all-iPSC mice. We excluded the pseudo-CNAs caused by replication timing, which could result in a gain or a loss in the read-depth profiles of the iPSC lines (Supplementary Fig. 5a–c)^{49–51}. The breakpoints of the CNAs surveyed from the paired-end reads and the subsequent PCR results for multiple breakpoints supported the repeated occurrences of the true CNAs in the iPSC lines but not in the somatic cells and feeder cells during the sequential reprogramming (Fig. 3a, Supplementary Fig. 2b and Supplementary Figs 6a–12a, Supplementary Data 3). The extended analysis further revealed that these repeated CNAs were obtained not only in the sequential iPSC lines but also in

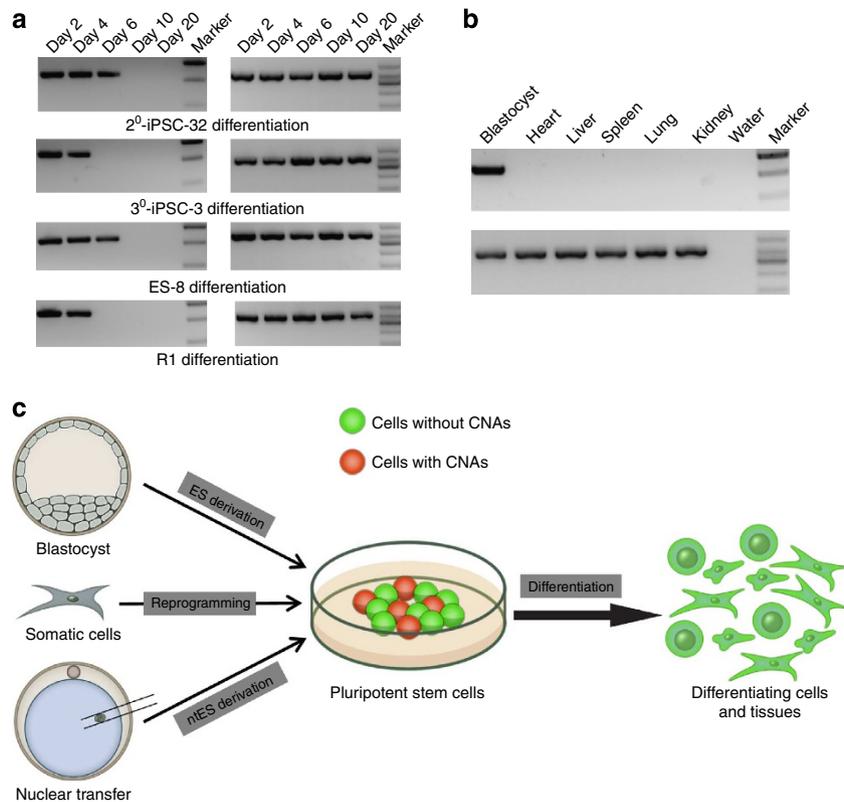


Figure 4 | Disappearance of CNAs during the *in vitro* differentiation of pluripotent stem cells and tissues. (a) PCR analysis of the gradually differentiated cells demonstrating that the validated CNA at chr14:75,825,104–75,831,195 became undetectable during the *in vitro* differentiation of iPSCs and ESCs (left image). The corresponding PCR results for the control are shown in the right image. (b) The CNA validated above could not be detected in all of the tested tissues but was detected in pre-implantation blastocysts by PCR (upper panel). A unique band retaining the LINE region was also produced by PCR (lower panel). (c) Summary model illustrating that the pluripotent stem cells were heterogeneous in terms of the repeated CNAs and that the CNAs were unique to pluripotent cells and subsequently disappeared in differentiating cells.

other SCNT-derived ES (ntES) cell lines, normal fertilization-derived ESC lines and an iPSC line generated in another laboratory (Fig. 3a, Supplementary Figs 6a–12a and Supplementary Data 3). Sanger sequencing of PCR products of breakpoints revealed that each deleted fragment caused by the verified CNA included an annotation of a retrotransposon element, such as long interspersed nucleotide elements (LINEs) and long terminal repeats (LTRs) (Fig. 3b and Supplementary Figs 6b–12b and Supplementary Data 3). PCR and Sanger sequencing results also revealed the presence of another intact allele that retained the LINE/LTR elements (Fig. 3a,b and Supplementary Figs 6a,b–12a,b). We further demonstrated the presence of these CNAs in a portion, but not all, of the pluripotent stem cells, including ESCs, ntESCs and iPSCs, by single-cell PCR analysis (Fig. 3c)^{52,53}. We conclude that the pluripotent cells are heterogeneous in terms of these repeated CNAs. Subsequently, eight single iPS cells that were randomly isolated from the 2⁰-iPSC line were cultured separately into subclonal iPS cell lines. In terms of the proportion, approximately half, of cells with deletion allele in the 2⁰-iPSC line (Fig. 3c), the probability function of binomial distribution determined that the eight single iPS cells, at a probability of 99%, should include at least one cell with deletions and one cell without any deletion. Surprisingly, single-cell PCR analysis also demonstrated a mixed composite of cells with and without the repeated deletions in each subclonal iPS cell line (Supplementary Fig. 13). This additional experiment supports that genetic heterogeneity is a fundamental characteristic of pluripotent cells, regardless of the identity of the initial single cell. The additional differentiation experiments

confirmed that the CNAs preferentially occurred in the pluripotent stem cell lines and then disappeared during the process of differentiation (Fig. 4a and Supplementary Figs 6c–9c, 11c, 12c, 14). This pattern had no relationship with the cell culture passages because the CNAs were retained in distinct passages of iPSC lines but did not emerge due to extended cell culture passages in other differentiated cell lines (Supplementary Fig. 15). Furthermore, we demonstrated that these CNAs could be detected in blastocysts but disappeared in the fresh tissues of adult mice with the same genetic background as iPSCs (Fig. 4b and Supplementary Figs 6d, 7d,e, 8d, 9d–f, 10c, 11d,e, 12d). Taken together, these data provide evidence that these CNAs are unique to pluripotent stem cells (Fig. 4c).

Discussion

In summary, our study demonstrates for the first time the impact of induced or enlarged mutations on the developmental potential of the resultant iPSCs via a sequential reprogramming system. Forty-four non-synonymous accumulated SNVs may account for the success of the sequential all-iPSC mice for up to six generations through tetraploid complementation, but the viability of the mice decreased with increasing generations. Notably, the pivotal genes (*EP300*, *Ryr2*, *Chd2*) involved in heart and kidney development were found to be disrupted in later-generation iPSCs. And the corresponding histopathological examination provides good evidence that the accumulation of SNVs directly impacted the survival of later-generation all-iPSC mice. However, the exact functional effects of SNVs associated with the gradually

reduced viability of all-iPSC mice must be experimentally characterized in future studies. Moreover, compared with the serial SCNT experiments, the repeated generation of pluripotent cell lines might have experienced additional stress from the *in vitro* cell culture, potentially resulting in an increased rate of mutations, as observed in the present study. Therefore, a joint effect of the direct reprogramming strategy and the addition of an HDAC inhibitor may eliminate cloned embryos with abnormalities and facilitate serial re-cloning of mice up to 25 generations by SCNT. Importantly, the results of our present study provide information to better understand the association between gene mutations and developmental effect, which is essential for screening pre-clinical bio-safety iPSCs.

Unexpectedly, the comparison of multiple sequential profiles highlights a unique pattern of CNAs in pluripotent stem cells that were ignored in previous studies on the basis of a paired comparison^{18,20,54–56}. Moreover, not only PCR and Sanger sequencing data but also the genomic sequencing reads do confirm the presence of the repeated CNAs in pluripotent stem cells and the disappearance of the CNAs in differentiating cells and somatic cells. These subtle genome alterations occur primarily in the retrotransposons of pluripotent stem cells, including iPSCs, ntESCs, ESCs and pre-implantation blastocysts from which ESCs are derived, suggesting a common characteristic of self-renewing pluripotent stem cells⁵⁷. The single-cell PCR and subclonal iPSC lines analysis further indicate that the allele deletion of LINR/LTR may contribute to the heterogeneity in pluripotent stem cells. Although this experimental phenomenon is interesting, the exact explanation for why such a phenomenon occurred requires more subsequent experimentation. Previous study provided a recurrent deletion mechanism of LTR transposable elements in *Drosophila* due to double-strand breaks beside the retrotransposon sequence with an opening of the chromatin⁵⁸. However, the underlying molecular mechanism of repeated LINE/LTR deletion occurring in pluripotent stem cells necessitates a more thorough investigation in the future.

Methods

Animal use and care. All of the animal procedures were performed according to the ethical guidelines of the National Institute of Biological Sciences, NIBS.

Cell culture. Mouse embryonic fibroblasts (MEFs) were derived from 13.5 days post coitum (dpc) embryos collected from female 129/Sv mice that were mated with Rosa26-M2rtTA transgenic mice. Adipocyte progenitor cells (APCs) were generated and cultured as previously described⁵⁹. Briefly, APCs were derived from the stoma-vascular fraction from the inguinal fat deposits of 10-week-old or E19.5 all-iPSC mice. ES cells and iPSC cells were cultured on mitomycin C-treated MEFs in ES medium, which contained DMEM (Gibco Invitrogen, Carlsbad, CA) supplemented with 15% FBS, 1 mM L-glutamine, 0.1 mM mercaptoethanol, 1% non-essential amino acid stock and 1,000 U ml⁻¹ LIF (all from Chemicon, Temecula, CA).

Generation of iPSCs. The lentivirus-based plasmids and the procedure employed for iPSC cell derivation have been previously reported¹⁵. The viral supernatants containing the TetO-FUW-Oct4, Sox2, Klf4 and c-Myc plasmids and the packaging plasmids ps-PAX-2 and pMD2G were harvested, and the MEFs were infected with supernatants containing viruses at a density of $\sim 5 \times 10^5$ cells per 6-cm dish. The infection medium was replaced with ES medium supplemented with 1 μ g ml⁻¹ Dox 12 h after the infection. The ES-like colonies appeared after ~ 12 days, and, 4 days after the withdrawal of Dox, smooth colonies were isolated and passaged for an additional 3 days to derive the 1⁰-iPSC lines. APCs were retrieved from the all-iPSC mice generated from the 1⁰-iPSCs through tetraploid complementation. Under Dox induction, 2⁰-iPSCs were subsequently established. The 3⁰-iPSCs, 4⁰-iPSCs, 5⁰-iPSCs and 6⁰-iPSCs were all derived from the preceding 'all-iPSC' mice using the same induction method.

AP and immunofluorescence staining. AP staining was performed using the alkaline phosphatase detection kit (Millipore) according to the manufacturer's instructions. Immunofluorescence staining was performed as previously described¹⁵. iPSC cells growing on the gelatin-coated cover slides were fixed in

4% paraformaldehyde overnight at 4 °C. After permeabilization with 0.1% Triton X-100 and blocking with 2.5% bovine serum albumin for 1 h at room temperature, iPSC cells were incubated with the primary antibodies to anti-Oct4 (1:500; Santa Cruz), Sox2 (1:1,000, Abcam), Nanog (1:1,000; Cosmo BioCo) and SSEA-1 (1:50; Abcam) for 2 h at room temperature separately. After three times of washing, the samples were incubated with Alexa-Fluor goat anti-rabbit or goat anti-mouse IgG secondary antibodies (1:1000; Invitrogen) for 1 h at room temperature. DNA was labelled by DAPI. Stained cells mounted on slides were observed on a LSM 510 META microscope (Zeiss) using Plan Neofluar $\times 63/1.4$ Oil DIC objective.

***In vitro* differentiation of iPSCs and adipogenesis.** The iPSCs were trypsinized and cultured in 20- μ l hanging drops (1,000 cells per drop) supplemented with DMEM, 10% FBS, 1 mM L-glutamine, 0.1 mM mercaptoethanol and 1% non-essential amino acid stock without LIF. Embryoid bodies from the hanging drops were collected at 2 days and transferred to ultra-low cluster plates (Costar) for 3 days. Next, the embryonic bodies were harvested and plated onto gelatin-coated tissue culture dishes for another 15 days. Spontaneous differentiation was examined by quantitative PCR and PCR with reverse transcription for representative lineage-specific marker genes at various time points (day 2, day 4, day 6, day 10, day 20). Differentiation of the APCs into adipocytes was performed using the human mesenchymal stem cell kit (Lonza) according to the manufacturer's instructions.

Bisulfite sequencing. Genomic DNA was isolated with phenol:chloroform:isoamyl alcohol (25:24:1) using a standard protocol and treated with the EpiTeck Bisulfite Kits (Qiagen). Two rounds of nested PCR was performed to amplify the promoter region of Oct4 and Nanog. The PCR products were cloned into the vector using the pEASY-T5 Zero cloning kit (TransGen Biotech), and 12 randomly selected clones were sequenced.

Pluripotency validation of iPSCs. iPSCs ($2-5 \times 10^6$) were subcutaneously injected into the forelimb of severe combined immunodeficient mice. Four weeks after the injection, the tumours were dissected and processed for haematoxylin-eosin staining. To produce chimeric mice, 10–15 iPSCs were microinjected into eight-cell ICR embryos using a piezo-actuated microinjection pipette. After culturing for 1 day, the embryos were transplanted into the uteruses of pseudo-pregnant mice. To perform tetraploid embryo complementation, two cell-stage ICR embryos were electroporated to produce tetraploid embryos, and 10–15 iPSCs were subsequently injected into the reconstructed tetraploid blastocysts. Caesarean sections were performed at E19.5, and the pups were fostered by lactating ICR mothers. The primer sequences used for SSLP analysis were obtained from the Mouse Genome Informatics website (<http://www.informatics.jax.org>).

Sample preparation and genome sequencing. All of the pluripotent stem cells (sequentially iPSCs, ntESCs, ESCs) were cultured under the same standardized conditions and collected at approximately passage 10. However, R1 ESCs were cultured and collected at passage 30. The somatic cells from all-iPSC mice were cultured and collected at passage 2. After removing the feeder cells, genomic DNA was extracted from the cell pellets using the DNeasy Mini Kit (Qiagen), and the DNA was quantified using a Qubit 2.0 Fluorometer (Life Technologies). The library construction and sequencing, including five somatic cell lines and eight iPSC lines, were performed according to the standard manufacturer's instructions for the HiSeq 2000 system. Whole-genome sequencing was applied to generate paired-end reads with a length of 101 bp. Raw reads were aligned to the mouse genome NCBI37/mm9 assembly by the Burrows–Wheeler aligner with default parameters⁶⁰. Only uniquely mapped reads with a quality score > 15 were retained for downstream analysis. For each sample, the data provided ~ 50 -fold coverage of the mouse genome (Supplementary Fig. 2a).

Analysis of MeDIP sequencing and RNA sequencing. MeDIP-seq and RNA-seq of all eight sequential iPSC lines were conducted as previously described²⁴. In brief, the genomic DNA was sonicated to 100–500 bp and adaptor was then ligated to the end of the DNA fragments according to the paired-end DNA sample prep kit (Illumina). Then the DNA was immunoprecipitated with the 5 mC antibody using the magnetic methylated DNA Immunoprecipitation kit (Diagenod). After the immunoprecipitated DNA was amplified for approximately 12–15 cycles, fragments of the proper size (200–300 bp) were gel-purified using the Gel Extraction Kit (Qiagen). The paired-end sequencing was performed at the Beijing Genomics Institute (ShenZhen, China) using the HiSeq 2000 system developed by Illumina. The MeDIP sequencing reads were aligned to mouse genome NCBI37/mm9 via using the Burrows–Wheeler aligner⁶⁰. The uniquely aligned reads were retained and fed into MACS2 (ref. 61). The 'macs2 callpeak' command with an adjusted cut-off *P* value of $1 - e^5$ was used to evaluate the DNA methylation peaks. The peak profiles of all-iPSC lines were utilized to define the DMRs across the sequential reprogramming. We further obtained the accumulating pattern of DMRs in promoter regions, (-1.5 kbp, 500 bp) around the transcription start sites (annotated in UCSC genome browser, <http://genome.ucsc.edu>).

Total RNA was isolated from cell pellets using TRIzol reagent (Life Technologies). The mRNA was enriched using oligo (dT) magnetic beads and sheared to create short fragments of ~200 bp. Subsequently, cDNA was synthesized using random hexamer primers and purified. Finally, the sequencing primers linked to the cDNA fragments were isolated by gel electrophoresis and enriched by PCR amplification to construct the library. Single-end sequencing was applied to the RNA sequencing. The RNA sequencing reads were aligned by TopHat with the annotation file fetched from the UCSC genome browser (<http://genome.ucsc.edu>). The FPKM for each gene was then calculated using cufflinks based on the coverage information generated from TopHat^{62,63}. Differentially expressed genes were identified on the basis of a comprehensive consideration of the *P* value and fold-change via the 'Cuffdiff' command.

Tracing SNVs during sequential reprogramming. The 'mpileup' command in SAMtools was used to generate the basic information for identifying variants⁶⁴. The following criteria were applied: (1) the minimum coverage of the variant sites was 20; (2) the Phred-scaled base quality was > 15; (3) the mutant allele frequency was > 0.3, however, there could be no reads for the mutant allele in the control sample; and (4) the mutant allele was supported by both the forward and reverse strand reads. We collected each SNV through pair-wise comparisons and calculated the mutant allele frequency in all 13 sequential samples. The SNVs were divided into four categories on the basis of the characteristics of the allele frequencies among the 13 samples.

SNV validation. The single-nucleotide variants were validated via multiplexed genotyping reactions with a Sequenom MassARRAY Analyzer⁶⁵. The primers were designed using the online genotyping design tools available in the Assay Design Suite (<https://www.mysequenom.com/Tools>). All of the experimental processes, including the PCR reactions, SAP reactions and extend reactions, were manipulated according to the standard procedures supplied by Sequenom in triplicate. Among the 20,000 mutant sites detected in the sequencing data, 189 sites were selected that were randomly distributed in the four SNV categories for validation. The mutant allele frequencies in all 13 sequential samples were verified, and the frequency in the feeder cells was also verified to exclude the effects of feeder cell contamination.

Digital PCR to estimate the SNV cell frequency in somatic cells and iPSCs.

Digital PCR was performed as previously described using the Bio-Rad QX100 ddPCR system^{47,48}. In our experience, for samples of this nature, the optimal concentration of genomic DNA is 24 ng per 20 µl, although a series of concentrations (3 ng per 20 µl, 12 ng per 20 µl, 24 ng per 20 µl, 48 ng per 20 µl) was tested to reduce the solution viscosity and ensure a sufficient number of droplets. For each SNV, we performed Custom TaqMan SNP Genotyping Assays (Life Technologies), which contained sequence-specific forward and reverse primers to amplify the polymorphic sequence of the SNV and two TaqMan MGB probes. One probe was labelled with VIC dye and used to detect the reference allele sequence, and another probe was labelled with FAM dye and used to detect the mutant allele sequence. TaqMan assays were conducted at a final concentration of 900 nM of each primer and 250 nM of each probe in all ddPCR reactions. According to the manufacturer's instructions, DG8 cartridges were loaded with 20 µl PCR reaction mixtures consisting of ddPCR Supermix for the TaqMan assay, and 70 µl of droplet generator oil was used for each sample. The cartridges were placed into a droplet generator for the emulsification of 15,000–20,000 water-in-oil droplets, and the emulsified samples were transferred onto a 96-well plate. PCR was then performed using the following programme: 95 °C for 10 min, 40 cycles of 94 °C for 30 s, 60 °C for 1 min, and 98 °C for 15 min. Each chemically homogenous droplet supported PCR amplification in a thermal cycler (C1000 Touch, Bio-Rad Laboratories).

Finally, the PCR products were analysed using a droplet reader (QuantaLife) to count the number of positive and negative droplets based on the fluorescence. It should be noted that the FAM-positive droplets indicated the occurrence of SNVs, and the VIC-positive droplets represented the reference base. The application of Poisson statistics allowed us to estimate the absolute number of copies of both genotypes in the cell lines. After comparing the concentrations between the mutant and wild type, the mutant allele frequency of each SNV was evaluated. Moreover, no-template controls containing TE buffer (10 mM Tris, 0.1 mM EDTA, pH 8.0) in place of DNA in the TaqMan assays were analysed to eliminate results representative of low-level template contamination.

Detection of CNAs via read depth and paired-end analysis. By adopting a mean-shift approach after GC correction, CNVnator was used to detect abnormal segments on read-depth ratio profiles⁶⁶. To obtain ultra-deep coverage in our experiment, 100-bp bins were used for segmentation. The parameter '-relax' was adjusted from default frequencies of 0.25 to 0.125, which made the CNVnator more sensitive for the detection of 'gain' or 'loss' segments with lower cell frequencies in pluripotent stem cells. Breakdancer, a method that is used to determine the orientation of reads and the abnormal span range of paired reads in contrast to the library insert size in library construction, was used to detect the specifically occurring CNA pattern in iPSCs⁶⁷. A CNA that was found to be repeated in at least five iPSC lines but was not detected in any of the somatic cells was regarded as a potential CNA for further study pending validation. We manually checked the

repeated CNA candidates in 13 sequential samples, allowing both ends of a CNA in multiple samples to deviate up to 50 bp due to variability in the CNA determination.

PCR and Sanger sequencing for CNA validation. To validate the repeated CNAs in pluripotent stem cells, we used Primer-BLAST (www.ncbi.nlm.nih.gov/tools/primer-blast/) to design specific primers to amplify the breakpoint-spanning or other complex region with structural variation such as a duplication or insertion. To demonstrate the presence of an intact allele that retained the deleted region, primers were designed to target the deleted region and its adjacent areas. The optimal primers were defined to be 20–25 bp in length and had a *T*_m of 58–62 °C and a GC content of 45–55% to ensure the amplification of only specific PCR products in the presence of deletions or other structural variations. PCR was conducted with 20-µl reactions consisting of 10 × Ex Taq Buffer (TaKaRa), 0.1 µl TaKaRa Ex Taq HS (5 units µl⁻¹), 1.6 µl dNTP mixture (2.5 mM), 0.5 µl each of forward and reverse primers at a concentration of 10 µM and ~100 ng of genomic DNA in sterilized distilled water. The reactions were performed using a thermal cycler (Bio-Rad) with the following conditions: 95 °C for 3 min, 35 cycles of 95 °C for 30 s, 58–62 °C (according to the primers), 72 °C for 30 s and a final extension of 72 °C for 5 min⁴⁷. The CNA candidates were screened using all of the somatic cell lines and pluripotent stem cell lines. If a single amplified product was present in each pluripotent stem cell line but no somatic cell lines, the repeated CNAs were validated. These validated CNAs were subsequently analysed in gradually differentiated cells. All of the specifically amplified PCR products were purified using the QIAquick PCR Purification Kit. Sanger sequencing of the extracted DNA was performed using both forward and reverse primers. The resulting sequences were aligned to the reference genome using the online sequence alignment tool BLAT.

To validate the CNAs in single cell of the 2⁰-iPSC-32, ES-8 and NTES-1 cell lines, we used the single-cell whole genome amplification kit (Yikon Genomics) to efficiently amplify genomic DNA from single cells according to the manufacturer's instructions^{52,53}. The products were then used as a template for PCR validation of the CNAs.

References

- Takahashi, K. & Yamanaka, S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126**, 663–676 (2006).
- Takahashi, K. *et al.* Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* **131**, 861–872 (2007).
- Park, I. H. *et al.* Disease-specific induced pluripotent stem cells. *Cell* **134**, 877–886 (2008).
- Soldner, F. *et al.* Parkinson's Disease patient-derived induced pluripotent stem cells free of viral reprogramming factors. *Cell* **136**, 964–977 (2009).
- Takebe, T. *et al.* Vascularized and functional human liver from an iPSC-derived organ bud transplant. *Nature* **499**, 481–484 (2013).
- Yamanaka, S. A fresh look at iPSC cells. *Cell* **137**, 13–17 (2009).
- Hou, P. *et al.* Pluripotent stem cells induced from mouse somatic cells by small-molecule compounds. *Science* **341**, 651–654 (2013).
- Kato, Y. *et al.* Eight calves cloned from somatic cells of a single adult. *Science* **282**, 2095–2098 (1998).
- Wakayama, T., Perry, A. C., Zuccotti, M., Johnson, K. R. & Yanagimachi, R. Full-term development of mice from enucleated oocytes injected with cumulus cell nuclei. *Nature* **394**, 369–374 (1998).
- Wilmot, L., Schnieke, A. E., McWhir, J., Kind, A. J. & Campbell, K. H. Viable offspring derived from fetal and adult mammalian cells. *Nature* **385**, 810–813 (1997).
- Yu, J. *et al.* Induced pluripotent stem cell lines derived from human somatic cells. *Science* **318**, 1917–1920 (2007).
- Wakayama, S. *et al.* Successful serial recloning in the mouse over multiple generations. *Cell Stem Cell* **12**, 293–297 (2013).
- Boland, M. J. *et al.* Adult mice generated from induced pluripotent stem cells. *Nature* **461**, 91–94 (2009).
- Carey, B. W. *et al.* Reprogramming factor stoichiometry influences the epigenetic state and biological properties of induced pluripotent stem cells. *Cell Stem Cell* **9**, 588–598 (2011).
- Kang, L., Wang, J. L., Zhang, Y., Kou, Z. H. & Gao, S. R. iPSC cells can support full-term development of tetraploid blastocyst-complemented embryos. *Cell Stem Cell* **5**, 135–138 (2009).
- Stadtfeld, M. *et al.* Aberrant silencing of imprinted genes on chromosome 12qF1 in mouse induced pluripotent stem cells. *Nature* **465**, 175–181 (2010).
- Young, M. A. *et al.* Background mutations in parental cells account for most of the genetic heterogeneity of induced pluripotent stem cells. *Cell Stem Cell* **10**, 570–582 (2012).
- Laurent, L. C. *et al.* 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 (2011).
- Gore, A. *et al.* Somatic coding mutations in human induced pluripotent stem cells. *Nature* **471**, 63–76 (2011).

20. Hussein, S. M. *et al.* Copy number variation and selection during reprogramming to pluripotency. *Nature* **471**, 58–67 (2011).
21. Ji, J. F. *et al.* Elevated coding mutation rate during the reprogramming of human somatic cells into induced pluripotent stem cells. *Stem Cells* **30**, 435–440 (2012).
22. Wakayama, T. *et al.* Ageing: cloning of mice to six generations. *Nature* **407**, 318–319 (2000).
23. Kou, Z. H. *et al.* Mice cloned from induced pluripotent stem cells (iPSCs). *Biol. Reprod.* **83**, 238–243 (2010).
24. Chang, G. *et al.* High-throughput sequencing reveals the disruption of methylation of imprinted gene in induced pluripotent stem cells. *Cell Res.* **24**, 293–306 (2014).
25. Eppig, J. T. *et al.* The Mouse Genome Database (MGD): comprehensive resource for genetics and genomics of the laboratory mouse. *Nucleic Acids Res.* **40**, 881–886 (2012).
26. Roelfsema, J. H. *et al.* Genetic heterogeneity in Rubinstein-Taybi syndrome: mutations in both the CBP and EP300 genes cause disease. *Am. J. Hum. Genet.* **76**, 572–580 (2005).
27. Le Gallo, M. *et al.* Exome sequencing of serous endometrial tumors identifies recurrent somatic mutations in chromatin-remodeling and ubiquitin ligase complex genes. *Nat. Genet.* **44**, 1310–1315 (2012).
28. Peifer, M. *et al.* Integrative genome analyses identify key somatic driver mutations of small-cell lung cancer. *Nat. Genet.* **44**, 1104–1110 (2012).
29. Zimmermann, N., Acosta, A. M., Kohlhaase, J. & Bartsch, O. Confirmation of EP300 gene mutations as a rare cause of Rubinstein-Taybi syndrome. *Eur. J. Hum. Genet.* **15**, 837–842 (2007).
30. Foley, P., Bunyan, D., Stratton, J., Dillon, M. & Lynch, S. A. Further case of Rubinstein-Taybi syndrome due to a deletion in EP300. *Am. J. Med. Genet. A* **149A**, 997–1000 (2009).
31. Bartsch, O. *et al.* Two patients with EP300 mutations and facial dysmorphism different from the classic Rubinstein-Taybi syndrome. *Am. J. Med. Genet. A* **152A**, 181–184 (2010).
32. Yao, T. P. *et al.* Gene dosage-dependent embryonic development and proliferation defects in mice lacking the transcriptional integrator p300. *Cell* **93**, 361–372 (1998).
33. Straub, R. E. *et al.* Allelic variation in GAD1 (GAD(67)) is associated with schizophrenia and influences cortical function and gene expression. *Mol. Psychiatry* **12**, 854–869 (2007).
34. Laitinen, P. J. *et al.* Mutations of the cardiac ryanodine receptor (RyR2) gene in familial polymorphic ventricular tachycardia. *Circulation* **103**, 485–490 (2001).
35. Priori, S. G. *et al.* Mutations in the cardiac ryanodine receptor gene (hRyR2) underlie catecholaminergic polymorphic ventricular tachycardia. *Circulation* **103**, 196–200 (2001).
36. Tiso, N. *et al.* Identification of mutations in the cardiac ryanodine receptor gene in families affected with arrhythmogenic right ventricular cardiomyopathy type 2 (ARVD2). *Hum. Mol. Genet.* **10**, 189–194 (2001).
37. Kannankeril, P. J. *et al.* Mice with the R176Q cardiac ryanodine receptor mutation exhibit catecholamine-induced ventricular tachycardia and cardiomyopathy. *Proc. Natl Acad. Sci. USA* **103**, 12179–12184 (2006).
38. Lehnart, S. E. *et al.* Leaky Ca(2+) release channel/ryanodine receptor 2 causes seizures and sudden cardiac death in mice. *J. Clin. Invest.* **118**, 2230–2245 (2008).
39. Bhuiyan, Z. A. *et al.* Expanding spectrum of human RYR2-related disease: new electrocardiographic, structural, and genetic features. *Circulation* **116**, 1569–1576 (2007).
40. Zou, Y. Z. *et al.* Ryanodine receptor type 2 is required for the development of pressure overload-induced cardiac hypertrophy. *Hypertension* **58**, 1099–1110 (2011).
41. Gigante, M. *et al.* CD2AP mutations are associated with sporadic nephrotic syndrome and focal segmental glomerulosclerosis (FSGS). *Nephrol. Dial. Transplant.* **24**, 1858–1864 (2009).
42. Benoit, G. *et al.* Analysis of recessive CD2AP and ACTN4 mutations in steroid-resistant nephrotic syndrome. *Pediatr. Nephrol.* **25**, 445–451 (2010).
43. Kim, J. M. *et al.* CD2-associated protein haploinsufficiency is linked to glomerular disease susceptibility. *Science* **300**, 1298–1300 (2003).
44. Kuroda, R. *et al.* A novel compound heterozygous mutation in the DAP12 gene in a patient with Nasu-Hakola disease. *J. Neurol. Sci.* **252**, 88–91 (2007).
45. Carvill, G. L. *et al.* Targeted resequencing in epileptic encephalopathies identifies de novo mutations in CHD2 and SYNGAP1. *Nat. Genet.* **45**, 825–830 (2013).
46. Chenier, S. *et al.* CHD2 haploinsufficiency is associated with developmental delay, intellectual disability, epilepsy and neurobehavioural problems. *J. Neurodev. Disord.* **6**, 9 (2014).
47. Abyzov, A. *et al.* Somatic copy number mosaicism in human skin revealed by induced pluripotent stem cells. *Nature* **492**, 438–442 (2012).
48. Pinheiro, L. B. *et al.* Evaluation of a droplet digital polymerase chain reaction format for DNA copy number quantification. *Anal. Chem.* **84**, 1003–1011 (2012).
49. Farkash-Amar, S. *et al.* Global organization of replication time zones of the mouse genome. *Genome Res.* **18**, 1562–1570 (2008).
50. Hiratani, I. *et al.* Global reorganization of replication domains during embryonic stem cell differentiation. *PLoS Biol.* **6**, 2220–2236 (2008).
51. Lu, J. *et al.* The distribution of genomic variations in human iPSCs is related to replication-timing reorganization during reprogramming. *Cell Rep.* **7**, 70–78 (2014).
52. Lu, S. J. *et al.* Probing meiotic recombination and aneuploidy of single sperm cells by whole-genome sequencing. *Science* **338**, 1627–1630 (2012).
53. Zong, C. H., Lu, S. J., Chapman, A. R. & Xie, X. S. Genome-wide detection of single-nucleotide and copy-number variations of a single human cell. *Science* **338**, 1622–1626 (2012).
54. Cheng, L. Z. *et al.* Low incidence of DNA sequence variation in human induced pluripotent stem cells generated by nonintegrating plasmid expression. *Cell Stem Cell* **10**, 337–344 (2012).
55. Quinlan, A. R. *et al.* Genome sequencing of mouse induced pluripotent stem cells reveals retroelement stability and infrequent DNA rearrangement during reprogramming. *Cell Stem Cell* **9**, 366–373 (2011).
56. Martins-Taylor, K. *et al.* Recurrent copy number variations in human induced pluripotent stem cells. *Nat. Biotechnol.* **29**, 488–491 (2011).
57. Cahan, P. & Daley, G. Q. Origins and implications of pluripotent stem cell variability and heterogeneity. *Nat. Rev. Mol. Cell Biol.* **14**, 357–368 (2013).
58. Schuster, A. T., Sarvepalli, K., Murphy, E. A. & Longworth, M. S. Condensin II subunit dCAP-D3 restricts retrotransposon mobilization in *Drosophila* somatic cells. *PLoS Genet.* **9**, e1003879 (2013).
59. Zheng, B., Cao, B. H., Li, G. H. & Huard, J. Mouse adipose-derived stem cells undergo multilineage differentiation *in vitro* but primarily osteogenic and chondrogenic differentiation *in vivo*. *Tiss. Eng.* **12**, 1891–1901 (2006).
60. Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **26**, 589–595 (2010).
61. Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).
62. Trapnell, C., Pachter, L. & Salzberg, S. L. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25**, 1105–1111 (2009).
63. Roberts, A., Pimentel, H., Trapnell, C. & Pachter, L. Identification of novel transcripts in annotated genomes using RNA-Seq. *Bioinformatics* **27**, 2325–2329 (2011).
64. Li, H. *et al.* The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
65. Tang, K. *et al.* Chip-based genotyping by mass spectrometry. *Proc. Natl Acad. Sci. USA* **96**, 10016–10020 (1999).
66. Abyzov, A., Urban, A. E., Snyder, M. & Gerstein, M. CNVnator: an approach to discover, genotype, and characterize typical and atypical CNVs from family and population genome sequencing. *Genome Res.* **21**, 974–984 (2011).
67. Chen, K. *et al.* BreakDancer: an algorithm for high-resolution mapping of genomic structural variation. *Nat. Methods* **6**, 677–681 (2009).

Acknowledgements

We thank Dr Dangsheng Li for his critical comments on the manuscript. We thank Professor Rudolf Jaenisch from the Whitehead Institute of MIT for generously supplying the lentivirus vectors. We thank Professor Hans Scholer and Dr Guangming Wu from the Max Planck Institute for providing one high-quality iPSC line. We are also grateful to our laboratory colleagues for their assistance with the experiments and in the preparation of the manuscript. This work was supported by the Ministry of Science and Technology (grants 2011CB812700 and 2011CB964800 to S.G., 2013AA102506, 2011BAD19B01 and 2011BAD19B04 to J.T., 2012CB316505 to J.C.), the Natural Science Foundation of China (31325019 and 91319306 to S.G., 31472092 to J.T., 31000656 to G.C., 31171265 to J.C.) and the Natural Science Foundation of SZU (201407 to G.C.)

Author contributions

S.G., G.C., J.T. and S.R.G. designed the experiments. S.G., G.C., W.L., X.K., K.T., L.T. K.X. and H.W. performed the experiments. C.Z. and J.C. performed the bioinformatics analysis. S.G., C.Z., G.C., J.C., J.T. and S.R.G. analysed and interpreted the data. S.G., C.Z., G.C., J.C., J.T. and S.R.G. wrote the manuscript.

Additional information

Accession codes: The raw whole-genome sequencing data reported in this study have been submitted to the Sequence Read Archive (SRA) at NCBI (accession number SRP029308).

Supplementary Information accompanies this paper at <http://www.nature.com/naturecommunications>

Competing financial interests: The authors declare no competing financial interests.

Reprints and permission information is available online at <http://npg.nature.com/reprintsandpermissions/>

How to cite this article: Gao, S. *et al.* Unique features of mutations revealed by sequentially reprogrammed induced pluripotent stem cells. *Nat. Commun.* **6**:6318 doi: 10.1038/ncomms7318 (2015).