ISSN 1948-0210 (online)

World Journal of *Stem Cells*

World J Stem Cells 2017 May 26; 9(5): 77-88

Published by Baishideng Publishing Group Inc

WJSC

World Journal of Stem Cells

Contents

Monthly Volume 9 Number 5 May 26, 2017

World Journal of Stem Cells
Volume 9 Number 5 May 26, 2017

Contents

NAME OF JOURNAL
World Journal of Stem Cells

ISSN
ISSN 1948-0210 (online)

LAUNCH DATE
December 31, 2009

FREQUENCY
Monthly

EDITORIAL OFFICE
Xiu-Xia Song, Director
World Journal of Stem Cells
Baishideng Publishing Group Inc
7901 Stoneridge Drive, Suite 501, Pleasanton, CA 94588, USA
Telephone: +1-925-2238242
Fax: +1-925-2238243
E-mail: editorialoffice@wjgnet.com
Help Desk: http://www.f6publishing.com/helpdesk
http://www.wjgnet.com

PUBLISHER
Baishideng Publishing Group Inc
7901 Stoneridge Drive, Suite 501,
Pleasanton, CA 94588, USA
Telephone: +1-925-2238242
Fax: +1-925-2238243
E-mail: bpgoffice@wjgnet.com
Help Desk: http://www.f6publishing.com/helpdesk
http://www.wjgnet.com

PUBLICATION DATE
May 26, 2017

World Journal of Stem Cells

Submit a Manuscript: http://www.f6publishing.com

World J Stem Cells 2017 May 26; 9(5): 77-88

DOI: 10.4252/wjsc.v9.i5.77

ISSN 1948-0210 (online)

ORIGINAL ARTICLE

Basic Study

Next-generation sequencing traces human induced pluripotent stem cell lines clonally generated from heterogeneous cancer tissue

Tetsuya Ishikawa

Tetsuya Ishikawa, Cell Biology, Core Facilities for Research and Innovative Medicine, National Cancer Center Research Institute, Tokyo 104-0045, Japan

Tetsuya Ishikawa, Central Animal Division, Fundamental Innovative Oncology Core Center, National Cancer Center Research Institute, Tokyo 104-0045, Japan

Author contributions: Ishikawa T substantially contributed to the conception and design of the study as well as the acquisition, analysis and interpretation of the data, drafted the manuscript, made critical revisions related to the intellectual content of the manuscript, and approved the final version to be published.

Supported by the JSPS KAKENHI, No. 16K07135.

Institutional review board statement: This study was conducted with the approval of the Institutional Review Boards of the National Cancer Center of Japan and the Japanese Collection of Research Bioresources, National Institutes of Biomedical Innovation, Health and Nutrition. Written informed consent was obtained from donors for the use of their tissue in this study.

Institutional animal care and use committee statement: N/A.

Conflict-of-interest statement: The Life-Science Intellectual Property Platform Fund (LSIP) supported this work. The author confirms that the LSIP had no influence over the study design, content of the article, or selection of this journal. The author discloses patents pending (WO2013081188 A1) relevant to the work presented here.

Data sharing statement: The next-generation sequencing data in this study will be available to the public through the DDBJ Sequence Read Archive (DRA).

Manuscript source: Invited manuscript

Correspondence to: Tetsuya Ishikawa, PhD (DMSc), Laboratory Head, Central Animal Division, Fundamental Innovative Oncology Core Center, National Cancer Center Research Institute, 1-1, Tsukiji 5-chome, Chuo-ku, Tokyo 104-0045, Japan. humanipscells@gmail.com
Telephone: +81-3-35475201
Fax: +81-3-35422548

Received: January 31, 2017
Peer-review started: February 7, 2017
First decision: March 7, 2017
Revised: April 3, 2017
Accepted: May 3, 2017
Article in press: May 5, 2017
Published online: May 26, 2017

Abstract

AIM

To investigate genotype variation among induced pluripotent stem cell (iPSC) lines that were clonally generated from heterogeneous colon cancer tissues using next-generation sequencing.

METHODS

Human iPSC lines were clonally established by selecting independent single colonies expanded from heterogeneous primary cells of S-shaped colon cancer tissues by retroviral gene transfer (*OCT3/4*, *SOX2*, and *KLF4*). The ten iPSC lines, their starting cancer tissues, and the matched adjacent non-cancerous tissues were analyzed using next-generation sequencing and bioinformatics analysis using the human reference genome hg19. Non-synonymous single-nucleotide variants (SNVs) (missense, nonsense,

and read-through) were identified within the target region of 612 genes related to cancer and the human kinome. All SNVs were annotated using dbSNP135, CCDS, RefSeq, GENCODE, and 1000 Genomes. The SNVs of the iPSC lines were compared with the genotypes of the cancerous and non-cancerous tissues. The putative genotypes were validated using allelic depth and genotype quality. For final confirmation, mutated genotypes were manually curated using the Integrative Genomics Viewer.

RESULTS

In eight of the ten iPSC lines, one or two non-synonymous SNVs in *EIF2AK2*, *TTN*, *ULK4*, *TSSK1B*, *FLT4*, *STK19*, *STK31*, *TRRAP*, *WNK1*, *PLK1* or *PIK3R5* were identified as novel SNVs and were not identical to the genotypes found in the cancer and non-cancerous tissues. This result suggests that the SNVs were *de novo* or pre-existing mutations that originated from minor populations, such as multifocal pre-cancer (stem) cells or pre-metastatic cancer cells from multiple, different clonal evolutions, present within the heterogeneous cancer tissue. The genotypes of all ten iPSC lines were different from the mutated *ERBB2* and *MKNK2* genotypes of the cancer tissues and were identical to those of the non-cancerous tissues and that found in the human reference genome hg19. Furthermore, two of the ten iPSC lines did not have any confirmed mutated genotypes, despite being derived from cancerous tissue. These results suggest that the traceability and preference of the starting single cells being derived from pre-cancer (stem) cells, stroma cells such as cancer-associated fibroblasts, and immune cells that co-existed in the tissues along with the mature cancer cells.

CONCLUSION

The genotypes of iPSC lines derived from heterogeneous cancer tissues can provide information on the type of starting cell that the iPSC line was generated from.

Key words: Colon cancer; Next-generation sequencing; Single-nucleotide variant; Genotype; Heterogeneous cancer tissue; Cancer associated fibroblast; Pre-cancer cell; Induced pluripotent stem cell; Single cell; Clonal evolution

Core tip: Ten induced pluripotent stem cell (iPSC) lines were clonally generated from heterogeneous colon cancer tissues and analyzed with next-generation sequencing. Non-synonymous single-nucleotide variants (SNVs) of the iPSC lines were not identical to the genotypes of the cancer tissues. The SNVs were *de novo* or pre-existing mutations that originated from a minor population within the cancer tissue. Meanwhile, the genotypes of the iPSC lines were not mutated genotypes of the cancer tissues, suggesting that the starting cells for the iPSC lines were not mature cancer cells. Thus, the genotypes of iPSC lines can be used to trace the genomic origins of single cells within heterogeneous cancer tissue.

Ishikawa T. Next-generation sequencing traces human induced pluripotent stem cell lines clonally generated from heterogeneous cancer tissue. *World J Stem Cells* 2017; 9(5): 77-88 Available from: URL: http://www.wjgnet.com/1948-0210/full/v9/i5/77.htm DOI: http://dx.doi.org/10.4252/wjsc.v9.i5.77

INTRODUCTION

Gene transfer of *OCT3/4, SOX2*, *KLF4,* and *c-Myc* to somatic cells generates human induced pluripotent stem cells (iPSCs)[1-3] although *c-MYC* is not required for iPSC generation[4]. Human iPSCs are indistinguishable from human embryonic stem cells (ESCs) in terms of their long-term self-renewal ability and their *in vivo* pluripotency[3,5]. The starting cells for iPSC generation should be appropriately chosen to generate normal or aberrant iPSC lines for the purpose of regenerative medicine or cancer research/therapy. Human iPSC lines for regenerative medicine would be ideally generated from normal neonatal tissues[3] that are typically free of postnatal aberrant mutations and epigenetic changes. Human iPSCs (or iPSC-like cells) have also been generated from cancer cell lines[6,7], the somatic cells from familial cancer patients[8,9], and pancreatic ductal adenocarcinomas[10]. For cancer research/therapy, it is of great interest to generate iPSCs from heterogeneous cancer tissues. In our recent study[11], human iPSC lines were clonally generated from a heterogeneous mixture of primary cells derived from gastric tissues or colon cancer tissues and were subjected to microarray gene expression analysis. The resultant iPSC lines expressed all ESC-enriched genes including *POU5F1, SOX2* and *NANOG* that are essential for self-renewal ability and pluripotency[5,12] at a level equivalent to those of the typical human iPSC line (201B7)[1]. Genome-wide gene expression patterns were used to categorize the reference iPSC line 201B7 and the iPSC lines derived from distinct cancer tissues into three different groups. The gene expression profiles of these iPSC lines demonstrated differences derived from their distinct starting tissues and similarity and heterogeneity derived from their common starting heterogeneous tissues. More recently, it was reported that reference component analysis (RCA), an algorithm that substantially improves clustering accuracy, was developed to robustly cluster single-cell transcriptomes[13]. The RCA of single-cell transcriptomes elucidated cellular heterogeneity in human colorectal cancer[13].

In this study, iPSC technology and next-generation sequencing were used to resolve genotype variation among single cells within a heterogeneous cancer tissue. The genomic DNA of ten iPSC lines that were clonally generated from human colon cancer tissue was analyzed and compared with the genomic DNA from their cancer tissue of origin and matched adjacent non-cancerous

tissue.

MATERIALS AND METHODS

Tissues derived from a single colon cancer patient

This study was conducted with the approval of the Institutional Review Boards of the National Cancer Center of Japan and the Japanese Collection of Research Bioresources (JCRB), National Institutes of Biomedical Innovation, Health and Nutrition. Written informed consent from a single donor was obtained for the use of the tissues for research. The anonymous remnant non-cancerous and cancerous tissues were provided by the JCRB Tissue Bank. The tissues were derived from the surgical waste material from an operation performed on a 55-year-old Japanese male S-shaped colon cancer patient.

Primary cell culture from cancer tissues

Heterogeneous primary cell culture from the colon cancer tissues was prepared as previously described[11]. Briefly, the tissues were washed with Hank's balanced salt solution (HBSS) and minced into pieces with scissors. The pieces were further washed with HBSS. DMEM with collagenase was added to the tissue precipitates and mixed at 37 ℃ for 1 h on a shaker. After washing with DMEM, cells were seeded on collagen-coated dishes and cultured in DMEM supplemented with 10% FBS.

Generation of human iPSC lines

The study was approved by the Institutional Recombinant DNA Advisory Committee. Heterogeneous primary cells from the cancer tissue were cultured for 24 h at approximately 5%-10% confluency and then incubated with a pantropic retrovirus vector solution (*OCT3/4*, *KLF4*, and *SOX2*) at 37 ℃ for an additional 24 h. The vector solution was prepared as previously described[14]. Mitomycin C-treated mouse embryonic fibroblasts (MEFs, ReproCell) were seeded and co-cultured with the primary cells following the retroviral infection. The culture medium was replaced with MEF-conditioned ESC medium every 3 d until the cell layer was fully confluent and then further refreshed with mTeSR1 medium (STEMCELL Technologies) every day. Each independent colony was isolated from the culture using forceps under a microscope. The independent iPSC lines were sub-cultured with MEF in gelatin-coated 24-well plates.

Expansion and passage culture of human iPSC lines

Human iPSC lines were cultured with the MEFs in primate ESC, ReproStem (ReproCell) or mTeSR1 medium in gelatin-coated dishes[11]. The expanded iPSC lines were treated with a dissociation solution (ReproCell) or 0.25% trypsin-EDTA (Gibco) and passaged in media supplemented with 10-20 μmol/L Y-27632 to avoid cell death[3]. Independent iPSC lines were passaged from the 24-well plates into 6-well plates, further expanded into 100-mm dishes, and minimally passaged in 100-mm dishes under similar culture conditions. Each genomic DNA sample was prepared from independent iPSC lines.

Real-time RT-PCR analysis

Total RNA was prepared using the miRNeasy Mini Kit (Qiagen). Reverse transcription of the RNA was carried out using an iScript™ Advanced cDNA Synthesis Kit for RT-qPCR (Bio-Rad). Quantitative PCR was carried out with an SsoAdvanced Universal SYBR® Green Supermix using the CFX96 Real-Time PCR Detection System (Bio-Rad). PCR primer sets for *OCT3/4*, *SOX2*, *NANOG*, *ZFP42,* and *GAPDH* are listed in Supplemental Table 1. PCR data were analyzed using CFX Manager Software (Bio-Rad). PCR data from the iPSC 201B7[1] RNA were used as a positive control, and PCR data from cancer tissue-derived iPSC lines are presented as quantification cycle (Cq) values.

Target selection and sequencing

Target sequencing was conducted for twelve DNA samples from the cancer tissues, the non-cancerous tissues, and the ten iPSC lines. Genomic DNA was extracted from each of twelve samples using the DNeasy Blood AND Tissue Kit (Qiagen), sheared into approximately 150-bp fragments, and used to make a library for multiplexed paired-end sequencing with the SureSelect^{XT} Reagent Kit (Agilent Technologies). The constructed library was hybridized to biotinylated cRNA oligonucleotide baits from the SureSelect^{XT} Human Kinome Kit (Agilent Technologies) for target enrichment. Targeted sequences were purified by magnetic beads, amplified, and sequenced on an Illumina HiSeq2000 platform in a paired-end 101 bp configuration.

Mapping and single-nucleotide variant calling

Adapter sequences were removed by cutadapt (v1.2.1). After quality control, reads were mapped to the human reference genome hg19 using BWA (ver.0.6.2). Mapping results were corrected using Picard (ver.1.73) for removing duplicates and GATK (ver.1.5-32) for local alignment and base quality score recalibration. Single-nucleotide variant (SNV) calls were performed with multi-sample calling using GATK (UnifiedGenotyper) and filtered to coordinates with a variant call quality score $\geqslant$ 30 and a depth $\geqslant$ 8. SNVs were further classified based on their predicted functions of missense, nonsense or read-through. For final confirmation, SNVs were manually curated using the Integrative Genomics Viewer. Annotations of SNVs were based on dbSNP135, CCDS (NCBI release 20111122), RefSeq (UCSC Genome Browser, dumped 20111122), GENCODE (UCSC Genome Browser, ver. 7), and 1000 Genomes (release 20111011) sequences.

RESULTS

Human iPSC lines derived from colon cancer tissues

The human iPSC lines CC1-1, CC1-2, CC1-7, CC1-8, CC1-9, CC1-11, CC1-12, CC1-17, CC1-18, and CC1-25

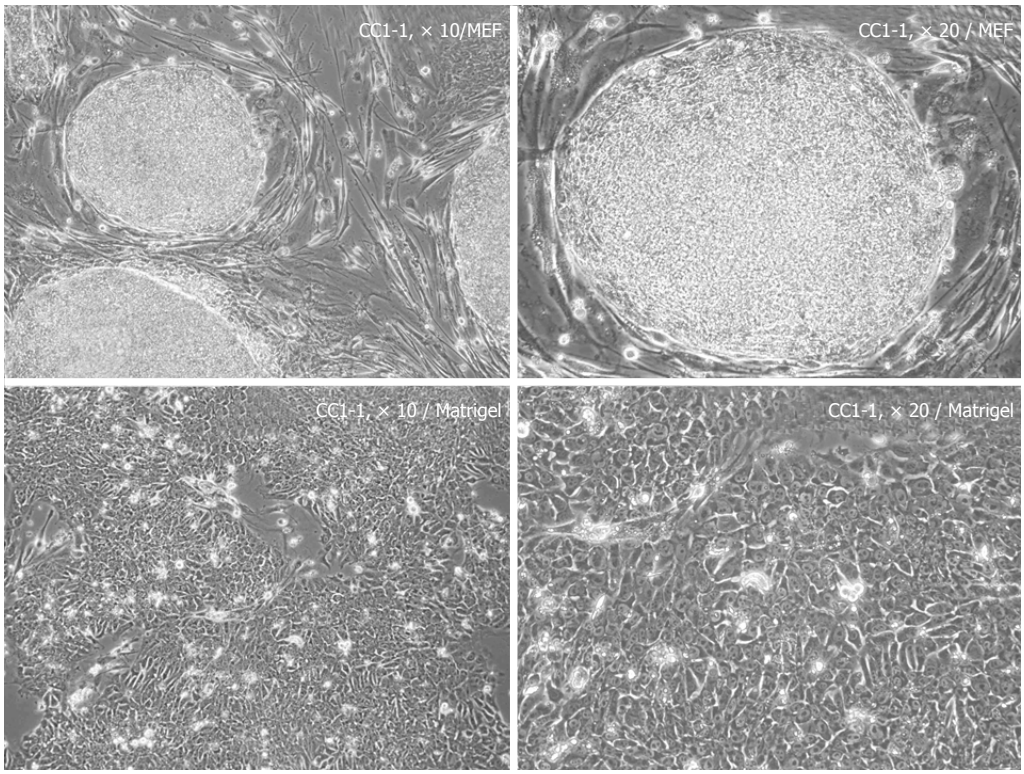

Figure 1 Phase contrast micrographs of colon cancer tissue-derived induced pluripotent stem cells. The human iPSC line CC1-1 was expanded with mitomycin C-treated mouse embryonic fibroblasts in gelatin-coated dishes (upper left panel: × 10, upper right panel: × 20) and cultured with feeder-free mTeSR1 medium in BD Matrigel™-coated dishes (lower left panel: × 10, lower right panel: × 20). iPSC: Induced pluripotent stem cell; MEF: Mouse embryonic fibroblast.

were clonally generated from heterogeneous primary cells cultured from colon cancer tissue. The iPSC lines were expanded serially with MEFs in gelatin-coated dishes. The cancer tissue-derived iPSCs were indistinguishable in morphology from typical (fibroblast-derived) human iPSCs under conventional culture with MEFs (upper panels in Figure 1 and Supplemental Figure 1). The human iPSCs formed colonies consisting of very small cells and were efficiently passaged at a high recovery ratio with the addition of 10-20 μmol/L Y-27632 to the cell culture medium. Human iPSCs were also cultured with feeder-free mTeSR1 medium in BD Matrigel™-coated 100-mm dishes and showed a high nucleus-to-cytoplasm ratio (lower panels in Figure 1 and Supplemental Figure 1).

Expression of human ESC-essential genes

ESC-essential gene expression of the cancer tissue-derived iPSC lines was quantitatively analyzed by real-time RT-PCR. All ten iPSC lines expressed *POU5F1, SOX2,* and *NANOG*, which are essential for self-renewal and pluripotency, at a level equivalent to those of the reference iPSC line (Supplemental Table 2). The study results support previously published microarray data showing that cancer tissue-derived iPSCs equally express ESC-enriched genes[11].

Next-generation sequencing

The target region (SureSelect Human Kinome) in genomic DNA samples from the ten iPSC lines, their starting cancer tissues, and the matched adjacent non-cancerous tissues was analyzed using next-generation sequencing. The target region of approximately 3.2 Mb covers the genome of the coding region of all known protein kinase genes and selected oncogenes and tumor suppressor genes, for a total of 612 genes (Supplemental Table 3). The original reads (2.6-4.0 Gb of sequence) were obtained from each genomic DNA sample by sequencing (Table 1). The modified reads were generated from the original reads (Table 2). The results of the mapped reads, the sequencing depth, and target capture are summarized in Tables 3-5. The average depth on the target region ranged from 317 to 496. More than 99.76% of the target region was covered with at least 8 × depth for high-quality genotype calls (variant call quality score ⩾ 30).

Non-synonymous SNVs compared with hg19

After sequencing, the reads underwent bioinformatics analysis (Figure 2). Through comparison with the human reference genome hg19, the non-synonymous SNVs (missense, nonsense or read-through) were called on the target region (on and near DNA target enrichment baits) of twelve samples (the ten iPSC lines, their starting cancer tissues, and the matched adjacent non-cancerous tissues). Of the resulting 378 non-synonymous SNVs (Supplemental Table 4), 50 were novel SNVs (not reported in dbSNP135 or 1000 Genomes, Supplemental Table 5).

Confirmed genotypes of the twelve samples

Of the 378 non-synonymous SNVs from the twelve

Table 1 Read number (original)

Sample	Original		
	No. of reads	Read length (b)	No. of bases (Gb)
NCC1	18260718	101	3.7
	18260718	101	
CC1	16706190	101	3.4
	16706190	101	
CC1-1	13045740	101	2.6
	13045740	101	
CC1-2	17725772	101	3.6
	17725772	101	
CC1-7	14780507	101	3.0
	14780507	101	
CC1-8	17311972	101	3.5
	17311972	101	
CC1-9	16664067	101	3.4
	16664067	101	
CC1-11	15455638	101	3.1
	15455638	101	
CC1-12	15391361	101	3.1
	15391361	101	
CC1-17	19009957	101	3.8
	19009957	101	
CC1-18	19746313	101	4.0
	19746313	101	
CC1-25	15492560	101	3.1
	15492560	101	

NCC1: The matched adjacent non-cancerous tissue; CC1: The starting cancer tissue of the induced pluripotent stem cell lines; CC1-1 to CC1-25: Each induced pluripotent stem cell line.

Table 2 Read number (modified)

Sample	Modified[1]		
	No. of reads	Read length (b)	Ratio (%) (Mod/Ori)
NCC1	18146940	101[2]	99.38
	18146940	101[2]	99.38
CC1	16597436	101[2]	99.35
	16597436	101[2]	99.35
CC1-1	12942132	101[2]	99.21
	12942132	101[2]	99.21
CC1-2	17614866	101[2]	99.37
	17614866	101[2]	99.37
CC1-7	14687008	101[2]	99.37
	14687008	101[2]	99.37
CC1-8	17180329	101[2]	99.24
	17180329	101[2]	99.24
CC1-9	16545785	101[2]	99.29
	16545785	101[2]	99.29
CC1-11	15346749	101[2]	99.30
	15346749	101[2]	99.30
CC1-12	15281269	101[2]	99.28
	15281269	101[2]	99.28
CC1-17	18880292	101[2]	99.32
	18880292	101[2]	99.32
CC1-18	19618808	101[2]	99.35
	19618808	101[2]	99.35
CC1-25	15378555	101[2]	99.26
	15378555	101[2]	99.26

[1]Modified read file is a data set from the original read file with the adapter sequences and low-quality bases removed; [2]Therefore, there were reads shorter than the number indicated by the read length (b) in a portion of the modified read file. NCC1: The matched adjacent non-cancerous tissue; CC1: The starting cancer tissue of the induced pluripotent stem cell lines; CC1-1 to CC1-25: Each induced pluripotent stem cell line.

samples, 40 were distinct heteroallelic genotypes and included known SNVs in the 612 sequenced gene target region. Supplemental Table 6 lists the forty SNVs that were distinct among the human iPSC lines, their starting cancer tissue, and the matched adjacent non-cancerous tissues. The allelic depth and genotype quality of thirteen of the forty SNVs were validated and manually curated using the Integrative Genomics Viewer (Figure 2).

Mutated genotypes of cancer tissue-derived iPSC lines

The chromosome number, genome position, novelty, gene symbol, and mutation type of the thirteen confirmed SNVs are shown in Table 6; the allelic depth is shown in Table 7; and the genotype is shown in Table 8. The respective SNVs of the ten iPSC lines were compared to those of their starting cancer tissue and the matched non-cancerous tissue. The genotypes, which showed nonsense or missense mutations in *EIF2AK2, TTN, ULK4, TSSK1B, FLT4, STK19, STK31, TRRAP, WNK1, PLK1,* or *PIK3*R5 (Table 6), of the iPSC lines were different from that of the non-cancerous tissue sample (Table 8). Nevertheless, the genotypes of the iPSC samples were also different from that of the starting cancer tissue sample. The heteroallelic read sequences of *ULK4, TRRAP,* and *WNK1* of the starting cancer tissue sample consisted of 247|2 of A|C, 240|1 of G|C, and 246|2 of C|T, respectively (Table 7). Although the major read sequences indicated the genotypes of the non-cancerous tissues, the minor reads indicated missense mutations. The potential heteroallelic genotypes were identical to the mutated genotypes of the CC1-25, CC1-12, and CC1-8 iPSC lines. Meanwhile, the genotypes of all ten of the iPSC lines were different from the mutated genotypes in *ERBB2* and *MKNK2* of the cancer tissues and were identical to those of the non-cancerous tissues and human reference genome hg19 (Table 8). Thus, all analyzed iPSC lines were preferentially generated from starting cells without mutations in *ERBB2* and *MKNK2*, except for those generated from mature cancer cells. Furthermore, the iPSC lines CC1-7 and CC1-17 did not have any confirmed mutated genotypes despite originating from the cancer tissue.

DISCUSSION

The ten iPSC lines were clonally generated from a heterogeneous mixture of primary cells derived from the colon cancer tissue of a single patient. The genomes of the ten iPSC lines were analyzed using next-generation sequencing. The genomes of the starting cancer tissue and matched adjacent non-cancerous tissue from the same donor were also analyzed. The target region for analysis was the human kinome and cancer-related genes that are typically mutated in human tumors. A total of 378 non-synonymous SNVs identified from samples of the ten iPSC lines and the cancerous and non-cancerous tissues were identified by comparing the sequence reads

Table 3 Mapped reads

	NCC1	CC1	CC1-1	CC1-2	CC1-7	CC1-8	CC1-9	CC1-11	CC1-12	CC1-17	CC1-18	CC1-25
No. of total reads ①	36293880	33194872	25884264	35229732	29374016	34360658	33091570	30693498	30562538	37760584	39237616	30757110
No. of mapped reads ② (③+ ④ + ⑤)	36210841	33066194	25545096	34845596	28911555	33842665	32976780	29898986	30464308	36670526	38078508	30173511
No. of mapped reads with Paired-End ③	26935180	26333748	21884450	25981140	21498822	26807330	26704280	25386830	25103580	29513694	30869186	21622036
No. of mapped reads with Single-End ④	15741	24050	14051	16508	12336	17984	14651	27429	14354	27336	25456	19687
No. of discarded reads[1] ⑤	9259920	6708396	3646595	8847948	7400397	7017351	6257849	4484727	5346374	7129496	7183866	8531788
No. of unmapped reads (① - ②)	83039	128678	339168	384136	462461	517993	114790	794512	98230	1090058	1159108	583599
No. of effective reads (③ + ④)	26950921	26357798	21898501	25997648	21511158	26825314	26718931	25414259	25117934	29541030	30894642	21641723

[1]Discarded reads were as follows: Reads mapped to chromosomes other than the target; Reads where each paired-end is mapped to a different chromosome; Reads not used for single-nucleotide variant/InDel detection such as PCR duplicates. Each number of ② consists of each total number of "③ plus ④ plus ⑤"; "① - ②" means "each number of ① minus each number of ②"; "③ + ④" means "each number of ③ plus each number of ④". NCC1: The matched adjacent non-cancerous tissue; CC1: The starting cancer tissue of the iPSC lines; CC1-1 to CC1-25: Each induced pluripotent stem cell line.

Table 4 Sequence depth

Sample	Theoretical depth[1]		Observed depth[2]	
	Total bases (Mb)	Depth[3]	Effective bases on target (Mb)	Average depth on target
NCC1	3689	1173.31	1399	445.06
CC1	3375	1073.43	1084	344.89
CC1-1	2635	838.23	999	317.89
CC1-2	3581	1138.94	1279	406.72
CC1-7	2986	949.69	1195	380.13
CC1-8	3497	1112.35	1352	430.01
CC1-9	3366	1070.72	1355	431.06
CC1-11	3122	993.07	1077	342.50
CC1-12	3109	988.94	1359	432.12
CC1-17	3840	1221.45	1407	447.66
CC1-18	3989	1268.76	1559	495.99
CC1-25	3129	995.45	1083	344.37

[1]Theoretical depth calculated from the total number of bases obtained by DNA sequencing; [2]Observed depth used for single-nucleotide variant/InDel identification; [3]Theoretical depth [Total Bases (Mb)]/[Target region (Mb)]. Target region: SureSelect Human Kinome Kit (approximately 3.1 Mb). NCC1: The matched adjacent non-cancerous tissue; CC1: The starting cancer tissue of the induced pluripotent stem cell lines; CC1-1 to CC1-25: Each induced pluripotent stem cell line.

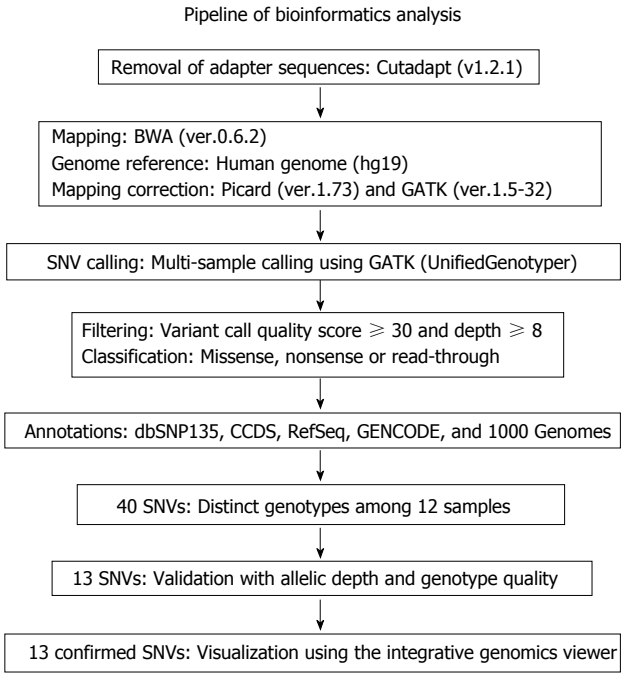

Figure 2 Pipeline of bioinformatics analysis following next-generation sequencing. The thirteen confirmed SNVs are shown in Tables 6-8. SNVs: Single-nucleotide variants.

to the human reference genome hg19. Most of the non-synonymous SNVs showed the genotype of the non-cancerous tissue, suggesting their germline origin. The SNVs of the ten iPSC lines were compared with those of the cancerous and non-cancerous tissues. Forty of the SNVs were distinct genotypes among all twelve samples. Thirteen of the forty SNVs were confirmed using allelic depth, genotype quality, and the Integrative Genomics Viewer.

In eight of the ten iPSC lines, one or two novel, non-synonymous SNVs (heteroallelic missense or nonsense mutation) in *EIF2AK2, TTN, ULK4, TSSK1B, FLT4, STK19, STK31, TRRAP, WNK1, PLK1* or *PIK3R5* were identified as genotypes different from those of the non-cancerous tissue. Unexpectedly, all the SNVs were not identical to the genotypes found in the cancer tissues. Because of minor read sequences, the potential genotype of *ULK4, TRRAP* or *WNK1* in the cancer tissues was implied. The sequences indicated a missense mutation in *ULK4, TRRAP* or *WNK1* identical to that found in the iPSC lines CC1-25, CC1-12, and CC1-8. Accordingly, there is a possibility that each iPSC line was generated from a starting cell from a minor cell population with a mutation in *ULK4, TRRAP,* or *WNK1* that was present within the cancer tissue. The minor read sequences could be confirmed by ultra-deep sequencing to support the potential heteroallelic genotypes. Interestingly, two iPSC lines CC1-7 and CC1-17 did not have any confirmed mutated genotypes despite originating from the cancer

Table 5 Target capture

	NCC1	CC1	CC1-1	CC1-2	CC1-7	CC1-8	CC1-9	CC1-11	CC1-12	CC1-17	CC1-18	CC1-25
Initial bases on target ①	3143812	3143812	3143812	3143812	3143812	3143812	3143812	3143812	3143812	3143812	3143812	3143812
Initial bases near target ②	3790645	3790645	3790645	3790645	3790645	3790645	3790645	3790645	3790645	3790645	3790645	3790645
Initial bases on or near target ③	6934457	6934457	6934457	6934457	6934457	6934457	6934457	6934457	6934457	6934457	6934457	6934457
Total effective reads ④	26950921	26357798	21898501	25997648	21511158	26825314	26718931	25414259	25117934	29541030	30894642	21641723
Total effective bases (Mb) ⑤	2663	2602	2157	2573	2125	2643	2637	2503	2480	2914	3047	2133
Read length mean (b)	98.91	98.83	98.58	99.02	98.84	98.61	98.77	98.58	98.79	98.74	98.69	98.65
Read length median (b)	101	101	101	101	101	101	101	101	101	101	101	101
Effective bases on target (Mb) ⑥	1399	1084	999	1279	1195	1352	1355	1077	1359	1407	1559	1083
Effective bases near target (Mb) ⑦	440	342	355	405	380	438	448	327	445	430	450	342
Effective bases on or near target (Mb) ⑧	1839	1426	1354	1683	1575	1790	1804	1403	1804	1837	2009	1425
Fraction of effective bases on target (%) (⑥/⑤)	52.54	41.67	46.32	49.70	56.25	51.14	51.38	43.01	54.78	48.30	51.18	50.75
Fraction of effective bases near target (%) (⑦/⑤)	16.52	13.14	16.45	15.73	17.88	16.56	17.00	13.05	17.96	14.75	14.77	16.05
Fraction of effective bases on or near target (%) (⑧/⑤)	69.06	54.82	62.78	65.43	74.12	67.70	68.38	56.06	72.74	63.05	65.95	66.80
Average sequencing depth on target (⑥/①)	445.06	344.89	317.89	406.72	380.13	430.01	431.06	342.50	432.12	447.66	495.99	344.37
Average sequencing depth near target (⑦/②)	116.05	90.22	93.63	106.76	100.21	115.48	118.31	86.16	117.51	113.43	118.68	90.33
Average sequencing depth on or near target (⑧/③)	265.21	205.68	195.30	242.75	227.11	258.07	260.10	202.38	260.14	264.96	289.74	205.50
Base covered on target ⑨	3143221	3143152	3142784	3143540	3143280	3143263	3143277	3142887	3143338	3143035	3143296	3142818
Coverage of target region (%) (⑨/①)	99.98	99.98	99.97	99.99	99.98	99.98	99.98	99.97	99.98	99.98	99.98	99.97
Base covered near target ⑩	3775671	3773869	3774076	3771915	3768892	3773823	3776942	3760218	3776060	3766215	3762031	3762823
Coverage of near target region (%) (⑩/②)	99.60	99.56	99.56	99.51	99.43	99.56	99.64	99.20	99.62	99.36	99.25	99.27
Fraction of target covered with at least 15 × (%)	99.72	99.62	99.59	99.69	99.65	99.68	99.70	99.55	99.69	99.60	99.68	99.58
Fraction of target covered with at least 8 × (%)	99.86	99.78	99.78	99.83	99.81	99.82	99.83	99.76	99.83	99.78	99.83	99.77
Fraction of target covered with at least 4 × (%)	99.93	99.89	99.89	99.90	99.90	99.90	99.91	99.87	99.92	99.89	99.91	99.89
Fraction of flanking region covered with at least 15 × (%)	86.88	84.24	87.52	85.84	84.62	87.92	89.85	79.85	90.17	84.50	80.72	82.36
Fraction of flanking region covered with at least 8 × (%)	94.32	93.13	94.66	93.66	92.95	94.70	95.77	89.93	95.87	92.55	89.96	91.39
Fraction of flanking region covered with at least 4 × (%)	97.85	97.33	97.91	97.50	97.15	97.87	98.27	95.76	98.32	96.86	95.71	96.37

The target region, as covered by the SureSelect Human Kinome Kit, was approx. 3.1 Mb. Near target region was 200 bases forward and backward of the target region. "⑥/⑤" means "each number of ⑥ devided by each number of ⑤"; "⑦/⑤" means "each number of ⑦ devided by each number of ⑤"; "⑧/⑤" means "each number of ⑧ devided by each number of ⑤"; "⑥/①" means "each number of ⑥ devided by each number of ①"; "⑦/②" means "each number of ⑦ devided by each number of ②"; "⑧/③" means "each number of ⑧ devided by each number of ③"; "⑨/①" means "each number of ⑨ devided by each number of ①"; "⑩/②" means "each number of ⑩ devided by each number of ②". NCC1: The matched adjacent non-cancerous tissue; CC1: The starting cancer tissue of the induced pluripotent stem cell lines; CC1-1 to CC1-25: Each induced pluripotent stem cell line.

tissues. Therefore, the two iPSC lines might be generated from non-cancerous cells such as pre-cancer (stem) cells and cancer-associated fibroblasts[15,16].

The SNVs of the ten iPSC lines could be *de novo* or pre-existing mutations that originated from minor cell populations, such as multifocal cancer (stem) cells and pre-metastatic cancer cells, present within the heterogeneous cancer tissue. Primary cancer tissues include multifocal pre-, mature and pre-metastatic cancer cells, so it makes sense that their genomes would be heterogeneous. The genotypes of pre-cancer (stem) cells would not be identical to those of germline or mature cancer cells, as colon cancer develops from an adenoma to carcinoma through the accumulation of a number

Table 6 Chromosome number, genome position, reference *vs* single-nucleotide variant, novelty *vs* dbSNP135, gene symbol, and mutation types of single-nucleotide variants

SNV No.	Chromosome No.	Genome position	Ref.|SNV	Novel/known	Gene symbol	Mutation types
1	chr2	37336419	C|T	Novel	*EIF2AK2*	Missense
2	chr2	179408086	A|G	Novel	*TTN*	Missense
3	chr3	41705179	A|C	Novel	*ULK4*	Missense
4	chr5	112769527	C|T	Novel	*TSSK1B*	Missense
5	chr5	180048626	C|T	Novel	*FLT4*	Missense
6	chr6	31947203	T|C	Novel	*STK19*	Missense
7	chr7	23808650	G|T	Novel	*STK31*	Missense
8	chr7	98490141	G|C	Novel	*TRRAP*	Missense
9	chr12	1009680	C|T	Novel	*WNK1*	Missense
10	chr16	23690401	C|T	Novel	*PLK1*	Missense
11	chr17	8789811	G|A	Novel	*PIK3R5*	Nonsense
12	chr17	37881392	A|G	Novel	*ERBB2*	Missense
13	chr19	2046399	G|A	Novel	*MKNK2*	Missense

Ref.: The allele of the human reference genome hg19; SNV: Single-nucleotide variant.

of genetic mutations and epigenetic aberration[17]. It is likely that the genotypes of pre-metastatic cancer cells in multiple clonal evolutions would be different from those of non-metastatic cancer cells. Meanwhile, genotypes of major mature cancer cells would be identical to those of cancer tissues; therefore, it was expected that genotypes of cancer tissue-derived iPSC lines would be identical to those of their starting cancer tissues. It was reported that *ERBB2* mutations were persistent in 3.6% of patients with colorectal cancer[18]. Indeed, a mutated genotype in *ERBB2* of the colon cancer tissues was also identified in this study.

Nevertheless, the genotypes of the ten iPSC lines were different from the mutated *ERBB2* and *MKNK2* genotypes in the cancer tissues and were identical to those of the non-cancerous tissues and the human reference genome hg19. This result suggests that the starting cells for the iPSC lines did not carry the mutations in *ERBB2* and *MKNK2* present in the cancer tissues. It is conceivable that the non-mutated genotypes of each iPSC line were identical to those of non-cancerous cells such as pre-cancer (stem) cells, stroma cells and immune cells that existed within the tissue. Each iPSC line was clonally established by selecting an independent single colony expanded from a putative single starting cell originating from heterogeneous cancer tissue. The genome sequence of each iPSC line was derived from its starting single cell. As a result, each iPSC line conserved the non-mutated *ERBB2* and *MKNK2* genotypes that originated from their respective starting single cells. Interestingly, all ten iPSC lines were not generated from cell populations containing either a mutated *ERBB2* and/or a mutated *MKNK2*. Thus, the genotypes of each iPSC line provide information on the genomic origin of the starting single cell derived from the heterogeneous cancer tissue.

Although the cause of the preference for the genomic origin of their starting cells was not clarified in this study, it seems that chemicals[19], gene sets[1,4], gene transfer[20,21], or inventive pre-culture[22,23], in which the starting cells might be preferentially specified, can affect iPSC generation. Accordingly, materials and methods can be optimized to generate normal or aberrant iPSC lines for the purposes of regenerative medicine or cancer research/therapy. Cancer tissues comprise (pre-) cancer (stem) cells, pre-metastatic cancer cells, stromal cells (such as mesenchymal stem cells, cancer-associated fibroblasts[15,16,24] and tumor endothelial cells) and immune cells (such as tumor-associated macrophages[25], dendritic cells[26] and tumor-infiltrating T cells[23]). Therefore, such a cell-derived iPSC line might be useful for immune-cell therapy[27] with cellular vaccines[28], dendritic cells[29-32] or tumor antigen-specific cytotoxic T cells[23], in addition to the development of models of carcinogenesis[33-35] and drug discovery tools[36,37]. For the purposes of regenerative medicine, human iPSCs are ideally generated from normal neonatal tissues[3,38-40] that are typically inexperienced of postnatal aberrant mutations or epigenetic changes. By contrast, aging and sun-exposed skin carries thousands of evolving clonal cells carrying cancer-causing mutations[41,42]. Indeed, genetic mutations accumulate gradually over a lifetime, even in human somatic stem cells[43]. For this reason, cell sources for iPSC generation should be selected based on the given field of research. Furthermore, iPSC lines with few or no mutations need to be established by the modification of existing methodology[39,44,45], as cell lines with *de novo* mutations not originating from the starting cells are not desired[46-50].

Nevertheless, cancer tissue-derived iPSCs might give rise to such *de novo* mutations, as their starting cells might have already suffered from an aberration (epigenetics or gene expression) associated with *de novo* mutations or cancer. Indeed, colon cancer tissue-derived iPSC lines exhibited unique gene expression profiles, with particular upregulation of *FAM19A5* and *SLC39A7*[11], in comparison with those of the typical iPSC line 201B7[1]. *FAM19A5* and *SLC39A7* were found to be expressed at lower levels in many human iPSC and ESC lines based on a free online expression atlas (Amazonia!, http://amazonia.transcriptome.eu/search.php)[51]. *FAM19A5* was reported as a novel cholangiocarcinoma biomarker[52], while *SLC39A7* is an intracellular zinc transporter and a hub for tyrosine kinase activation related to diseases such as cancer[53]. The analysis of iPSC genomes might expose rare single cells, such as an authentic cancer stem cells present within cancer tissues. Thus, next-generation sequencing of heterogeneous cancer tissue-derived iPSC lines might reveal potential aberrations or changes originating from the cancer tissue.

In conclusion, the genotypes of iPSC lines can be used to trace the genotype of the original single cells derived from heterogeneous cancer tissues.

ACKNOWLEDGMENTS

The author would like to thank Dr. Takanori Washio, Mr. O Kobayashi and Mr. Wataru Kurihara of Riken Genesis

Table 7 Allelic depth of single-nucleotide variants among the matched adjacent non-cancerous tissue, the starting cancer tissue, and the cancer tissue-derived induced pluripotent stem cell lines

	Allelic depth of SNVs											
SNV No.	NCC1	CC1	CC1-1	CC1-2	CC1-7	CC1-8	CC1-9	CC1-11	CC1-12	CC1-17	CC1-18	CC1-25
1	250 | 0	246 | 0	232 | 0	250 | 0	250 | 0	250 | 0	250 | 0	248 | 0	250 | 0	250 | 0	129 | 121	250 | 0
2	249 | 0	240 | 0	240 | 0	248 | 0	248 | 1	250 | 0	129 | 121	248 | 0	242 | 0	250 | 0	250 | 0	244 | 0
3	246 | 0	247 | 2	249 | 0	238 | 1	246 | 0	248 | 0	233 | 0	241 | 0	238 | 1	241 | 0	245 | 0	132 | 106
4	250 | 0	239 | 0	243 | 0	248 | 0	245 | 0	120 | 129	250 | 0	236 | 0	250 | 0	250 | 0	250 | 0	249 | 0
5	216 | 0	150 | 0	75 | 79	189 | 0	184 | 0	180 | 0	200 | 1	131 | 0	176 | 0	221 | 0	207 | 0	179 | 0
6	249 | 0	238 | 0	250 | 0	132 | 114	250 | 0	250 | 0	242 | 0	248 | 0	248 | 0	250 | 0	250 | 0	249 | 0
7	250 | 0	248 | 0	250 | 0	250 | 0	245 | 0	246 | 0	245 | 0	135 | 111	249 | 0	250 | 0	250 | 0	246 | 1
8	233 | 0	240 | 1	243 | 0	250 | 0	245 | 0	242 | 0	247 | 0	248 | 0	132 | 113	240 | 1	247 | 0	241 | 0
9	249 | 0	246 | 2	250 | 0	250 | 0	249 | 0	220 | 30	244 | 0	249 | 0	250 | 0	249 | 1	250 | 0	249 | 0
10	247 | 0	177 | 0	188 | 0	119 | 121	198 | 0	244 | 0	241 | 0	176 | 0	221 | 0	224 | 0	249 | 0	174 | 0
11	246 | 1	172 | 0	181 | 0	208 | 0	209 | 0	198 | 0	189 | 0	175 | 0	244 | 0	182 | 0	233 | 0	95 | 87
12	249 | 1	195 | 54	241 | 0	249 | 0	249 | 0	249 | 1	249 | 0	250 | 0	249 | 0	250 | 0	249 | 1	250 | 0
13	137 | 0	91 | 10	79 | 0	131 | 0	102 | 0	103 | 0	103 | 0	83 | 0	106 | 0	111 | 0	142 | 0	90 | 0

NCC1: The matched adjacent non-cancerous tissue; CC1: The starting cancer tissue of the induced pluripotent stem cell lines; CC1-1 to CC1-25: Each induced pluripotent stem cell line; SNV: Single-nucleotide variant.

Table 8 Genotypes of single-nucleotide variants among the matched adjacent non-cancerous tissue, the starting cancer tissue, and the cancer tissue-derived induced pluripotent stem cell lines

	Genotypes of SNVs											
SNV No.	NCC1	CC1	CC1-1	CC1-2	CC1-7	CC1-8	CC1-9	CC1-11	CC1-12	CC1-17	CC1-18	CC1-25
1	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/T	C/C
2	A/A	A/A	A/A	A/A	A/A	A/A	A/G	A/A	A/A	A/A	A/A	A/A
3	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/C
4	C/C	C/C	C/C	C/C	C/C	C/T	C/C	C/C	C/C	C/C	C/C	C/C
5	C/C	C/C	C/T	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
6	T/T	T/T	T/T	T/C	T/T	T/T	T/T	T/T	T/T	T/T	T/T	T/T
7	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/T	G/G	G/G	G/G	G/G
8	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/C	G/G	G/G	G/G
9	C/C	C/C	C/C	C/C	C/C	C/T	C/C	C/C	C/C	C/C	C/C	C/C
10	C/C	C/C	C/C	C/T	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
11	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/A
12	A/A	A/G	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A	A/A
13	G/G	G/A	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G

NCC1: The matched adjacent non-cancerous tissue; CC1: The starting cancer tissue of the induced pluripotent stem cell lines; CC1-1 to CC1-25: Each induced pluripotent stem cell line; SNV: Single-nucleotide variant.

for supporting the informatics analysis of the sequencing data, Dr. Momoko Kobayashi and Ms. Natsumi Suda for supporting the iPSC culture and genomic DNA preparation, and the members of the Fundamental Innovative Oncology Core Center, the Division of Molecular and Cellular Medicine, the Division of Genetics, and the Division of Carcinogenesis and Prevention for the helpful discussions. The author would also like to acknowledge the professional language editing service of Springer Nature Author Services. The author also acknowledges Dr. Toshio Kitamura, the JCRB Tissue Bank, and the RIKEN BioResource Center for providing the Plat-GP packaging cells, remnant tissues from cancer patients, and the iPSC line (201B7), respectively.

COMMENTS

Background

Starting cells for induced pluripotent stem cell (iPSC) generation should be appropriately adopted to generate normal or aberrant iPSC lines for use in regenerative medicine or cancer research/therapy. Human iPSC lines for regenerative medicine would be ideally generated from normal neonatal tissues, as they are typically free of postnatal aberrant mutations and epigenetic changes. For cancer research/therapy, it is of great interest to generate iPSCs that originate from heterogeneous cancer tissues.

Research frontiers

Microarray experiments have profiled the gene expression of human iPSC lines clonally generated from a heterogeneous mixture of primary cells derived from gastric tissue or colon cancer tissue. The gene expression profiles of such iPSC lines demonstrate differences derived from their distinct starting tissues and

similarity and heterogeneity derived from their common starting heterogeneous tissue.

Innovations and breakthroughs

This is the first study to analyze human iPSC lines clonally generated from a heterogeneous mixture of primary cells derived from cancer tissues using next-generation sequencing. Eight of the ten iPSC lines had single-nucleotide variants with *de novo* or pre-existing mutations originating from a minor population within the cancer tissues. Meanwhile, all other genotypes of the iPSC lines were not mutated as in the original cancer tissues. Two of the ten iPSC lines did not possess any confirmed mutated genotypes despite having been derived from cancer tissue. These results suggest that the majority of iPSC lines originated from starting cells other than major cancer cells. Thus, the genotypes of iPSC lines can be used to trace the genotypes of the starting single cells.

Applications

It is conceivable that cancer tissues are made up of not only pre-cancer (stem) cells and pre-metastatic cancer cells but also stroma cells (such as mesenchymal stem cells, cancer-associated fibroblasts and tumor endothelial cells) and immune cells (such as tumor-associated macrophages, dendritic cells and tumor-infiltrating T cells). These other cell types might serve as targets for drug discovery and immune-cell therapy against cancer. Therefore, such a cell-derived iPSC line might be useful for immune-cell therapies such as cancer vaccines, dendritic cells and tumor antigen-specific cytotoxic T cells, in addition to the development of models of carcinogenesis and drug discovery tools.

Terminology

Most single-nucleotide variants are heteroallelic genotypes that are validated with allelic depth and genotype quality and manually curated using the Integrative Genomics Viewer. Deeper allelic depth of next-generation sequencing further resolves genotype variations among the starting single cells present within heterogeneous cancer tissues. In this way, the genotypes of the iPSC lines may be used to trace the genomic identity of their starting single cells derived from a heterogeneous cancer tissue.

Peer-review

The manuscript is well written and easy to follow.

REFERENCES

1 **Takahashi K**, Tanabe K, Ohnuki M, Narita M, Ichisaka T, Tomoda K, Yamanaka S. Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* 2007; **131**: 861-872 [PMID: 18035408 DOI: 10.1016/j.cell.2007.11.019]

2 **Lowry WE**, Richter L, Yachechko R, Pyle AD, Tchieu J, Sridharan R, Clark AT, Plath K. Generation of human induced pluripotent stem cells from dermal fibroblasts. *Proc Natl Acad Sci USA* 2008; **105**: 2883-2888 [PMID: 18287077 DOI: 10.1073/pnas.0711983105]

3 **Masaki H**, Ishikawa T, Takahashi S, Okumura M, Sakai N, Haga M, Kominami K, Migita H, McDonald F, Shimada F, Sakurada K. Heterogeneity of pluripotent marker gene expression in colonies generated in human iPS cell induction culture. *Stem Cell Res* 2007; **1**: 105-115 [PMID: 19383391 DOI: 10.1016/j.scr.2008.01.001]

4 **Nakagawa M**, Koyanagi M, Tanabe K, Takahashi K, Ichisaka T, Aoi T, Okita K, Mochiduki Y, Takizawa N, Yamanaka S. Generation of induced pluripotent stem cells without Myc from mouse and human fibroblasts. *Nat Biotechnol* 2008; **26**: 101-106 [PMID: 18059259 DOI: 10.1038/nbt1374]

5 **Thomson JA**, Itskovitz-Eldor J, Shapiro SS, Waknitz MA, Swiergiel JJ, Marshall VS, Jones JM. Embryonic stem cell lines derived from human blastocysts. *Science* 1998; **282**: 1145-1147 [PMID: 9804556]

6 **Miyoshi N**, Ishii H, Nagai K, Hoshino H, Mimori K, Tanaka F, Nagano H, Sekimoto M, Doki Y, Mori M. Defined factors induce reprogramming of gastrointestinal cancer cells. *Proc Natl Acad Sci USA* 2010; **107**: 40-45 [PMID: 20018687 DOI: 10.1073/pnas.0912407107]

7 **Carette JE**, Pruszak J, Varadarajan M, Blomen VA, Gokhale S, Camargo FD, Wernig M, Jaenisch R, Brummelkamp TR. Generation of iPSCs from cultured human malignant cells. *Blood* 2010; **115**: 4039-4042 [PMID: 20233975 DOI: 10.1182/blood-2009-07-231845]

8 **Soyombo AA**, Wu Y, Kolski L, Rios JJ, Rakheja D, Chen A, Kehler J, Hampel H, Coughran A, Ross TS. Analysis of induced pluripotent stem cells from a BRCA1 mutant family. *Stem Cell Reports* 2013; **1**: 336-349 [PMID: 24319668 DOI: 10.1016/j.stemcr.2013.08.004]

9 **Lee DF**, Su J, Kim HS, Chang B, Papatsenko D, Zhao R, Yuan Y, Gingold J, Xia W, Darr H, Mirzayans R, Hung MC, Schaniel C, Lemischka IR. Modeling familial cancer with induced pluripotent stem cells. *Cell* 2015; **161**: 240-254 [PMID: 25860607 DOI: 10.1016/j.cell.2015.02.045]

10 **Kim J**, Hoffman JP, Alpaugh RK, Rhim AD, Reichert M, Stanger BZ, Furth EE, Sepulveda AR, Yuan CX, Won KJ, Donahue G, Sands J, Gumbs AA, Zaret KS. An iPSC line from human pancreatic ductal adenocarcinoma undergoes early to invasive stages of pancreatic cancer progression. *Cell Rep* 2013; **3**: 2088-2099 [PMID: 23791528 DOI: 10.1016/j.celrep.2013.05.036]

11 **Seno A**, Kasai T, Ikeda M, Vaidyanath A, Masuda J, Mizutani A, Murakami H, Ishikawa T, Seno M. Characterization of Gene Expression Patterns among Artificially Developed Cancer Stem Cells Using Spherical Self-Organizing Map. *Cancer Inform* 2016; **15**: 163-178 [PMID: 27559294 DOI: 10.4137/CIN.S39839]

12 **Boyer LA**, Lee TI, Cole MF, Johnstone SE, Levine SS, Zucker JP, Guenther MG, Kumar RM, Murray HL, Jenner RG, Gifford DK, Melton DA, Jaenisch R, Young RA. Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell* 2005; **122**: 947-956 [PMID: 16153702 DOI: 10.1016/j.cell.2005.08.020]

13 **Li H**, Courtois ET, Sengupta D, Tan Y, Chen KH, Goh JJ, Kong SL, Chua C, Hon LK, Tan WS, Wong M, Choi PJ, Wee LJ, Hillmer AM, Tan IB, Robson P, Prabhakar S. Reference component analysis of single-cell transcriptomes elucidates cellular heterogeneity in human colorectal tumors. *Nat Genet* 2017; **49**: 708-718 [PMID: 28319088 DOI: 10.1038/ng.3818]

14 **Ishikawa T**, Kobayashi M, Yanagi S, Kato C, Takashima R, Kobayashi E, Hagiwara K, Ochiya T. Human induced hepatic lineage-oriented stem cells: autonomous specification of human iPS cells toward hepatocyte-like cells without any exogenous differentiation factors. *PLoS One* 2015; **10**: e0123193 [PMID: 25875613 DOI: 10.1371/journal.pone.0123193]

15 **Huang L**, Xu AM, Liu S, Liu W, Li TJ. Cancer-associated fibroblasts in digestive tumors. *World J Gastroenterol* 2014; **20**: 17804-17818 [PMID: 25548479 DOI: 10.3748/wjg.v20.i47.17804]

16 **Mukaida N**, Sasaki S. Fibroblasts, an inconspicuous but essential player in colon cancer development and progression. *World J Gastroenterol* 2016; **22**: 5301-5316 [PMID: 27340347 DOI: 10.3748/wjg.v22.i23.5301]

17 **Fearon ER**, Vogelstein B. A genetic model for colorectal tumorigenesis. *Cell* 1990; **61**: 759-767 [PMID: 2188735 DOI: 10.1016/0092-8674(90)90186-I]

18 **Jauhri M**, Bhatnagar A, Gupta S, Shokeen Y, Minhas S, Aggarwal S. Targeted molecular profiling of rare genetic alterations in colorectal cancer using next-generation sequencing. *Med Oncol* 2016; **33**: 106 [PMID: 27568332 DOI: 10.1007/s12032-016-0820-2]

19 **Ma X**, Kong L, Zhu S. Reprogramming cell fates by small molecules. *Protein Cell* 2017; **8**: 328-248 [PMID: 28213718 DOI: 10.1007/s13238-016-0362-6]

20 **Fusaki N**, Ban H, Nishiyama A, Saeki K, Hasegawa M. Efficient induction of transgene-free human pluripotent stem cells using a vector based on Sendai virus, an RNA virus that does not integrate into the host genome. *Proc Jpn Acad Ser B Phys Biol Sci* 2009; **85**: 348-362 [PMID: 19838014]

21 **Cheng L**, Hansen NF, Zhao L, Du Y, Zou C, Donovan FX, Chou BK, Zhou G, Li S, Dowey SN, Ye Z, Chandrasekharappa SC, Yang H, Mullikin JC, Liu PP. Low incidence of DNA sequence variation in human induced pluripotent stem cells generated by nonintegrating plasmid expression. *Cell Stem Cell* 2012; **10**: 337-344 [PMID:

22385660 DOI: 10.1016/j.stem.2012.01.005]
22 **Sato T**, Stange DE, Ferrante M, Vries RG, Van Es JH, Van den Brink S, Van Houdt WJ, Pronk A, Van Gorp J, Siersema PD, Clevers H. Long-term expansion of epithelial organoids from human colon, adenoma, adenocarcinoma, and Barrett's epithelium. *Gastroenterology* 2011; **141**: 1762-1772 [PMID: 21889923 DOI: 10.1053/j.gastro.2011.07.050]
23 **Vizcardo R**, Masuda K, Yamada D, Ikawa T, Shimizu K, Fujii S, Koseki H, Kawamoto H. Regeneration of human tumor antigen-specific T cells from iPSCs derived from mature CD8(+) T cells. *Cell Stem Cell* 2013; **12**: 31-36 [PMID: 23290135 DOI: 10.1016/j.stem.2012.12.006]
24 **Koliaraki V**, Pallangyo CK, Greten FR, Kollias G. Mesenchymal Cells in Colon Cancer. *Gastroenterology* 2017; **152**: 964-979 [PMID: 28111227 DOI: 10.1053/j.gastro.2016.11.049]
25 **Herrera M**, Herrera A, Domínguez G, Silva J, García V, García JM, Gómez I, Soldevilla B, Muñoz C, Provencio M, Campos-Martin Y, García de Herreros A, Casal I, Bonilla F, Peña C. Cancer-associated fibroblast and M2 macrophage markers together predict outcome in colorectal cancer patients. *Cancer Sci* 2013; **104**: 437-444 [PMID: 23298232 DOI: 10.1111/cas.12096]
26 **Legitimo A**, Consolini R, Failli A, Orsini G, Spisni R. Dendritic cell defects in the colorectal cancer. *Hum Vaccin Immunother* 2014; **10**: 3224-3235 [PMID: 25483675 DOI: 10.4161/hv.29857]
27 **Rami F**, Mollainezhad H, Salehi M. Induced Pluripotent Stem Cell as a New Source for Cancer Immunotherapy. *Genet Res Int* 2016; **2016**: 3451807 [PMID: 27019752 DOI: 10.1155/2016/3451807]
28 **Koido S**, Ito M, Sagawa Y, Okamoto M, Hayashi K, Nagasaki E, Kan S, Komita H, Kamata Y, Homma S. Vaccination with vascular progenitor cells derived from induced pluripotent stem cells elicits antitumor immunity targeting vascular and tumor cells. *Cancer Immunol Immunother* 2014; **63**: 459-468 [PMID: 24627093 DOI: 10.1007/s00262-014-1531-1]
29 **Li Y**, Liu M, Yang ST. Dendritic cells derived from pluripotent stem cells: Potential of large scale production. *World J Stem Cells* 2014; **6**: 1-10 [PMID: 24567783 DOI: 10.4252/wjsc.v6.i1.1]
30 **Iwamoto H**, Ojima T, Hayata K, Katsuda M, Miyazawa M, Iida T, Nakamura M, Nakamori M, Iwahashi M, Yamaue H. Antitumor immune response of dendritic cells (DCs) expressing tumor-associated antigens derived from induced pluripotent stem cells: in comparison to bone marrow-derived DCs. *Int J Cancer* 2014; **134**: 332-341 [PMID: 23824921 DOI: 10.1002/ijc.28367]
31 **Zeng J**, Wu C, Wang S. Antigenically Modified Human Pluripotent Stem Cells Generate Antigen-Presenting Dendritic Cells. *Sci Rep* 2015; **5**: 15262 [PMID: 26471005 DOI: 10.1038/srep15262]
32 **Fairchild PJ**, Leishman A, Sachamitr P, Telfer C, Hackett S, Davies TJ. Dendritic cells and pluripotency: unlikely allies in the pursuit of immunotherapy. *Regen Med* 2015; **10**: 275-286 [PMID: 25933237 DOI: 10.2217/rme.15.6]
33 **Curry EL**, Moad M, Robson CN, Heer R. Using induced pluripotent stem cells as a tool for modelling carcinogenesis. *World J Stem Cells* 2015; **7**: 461-469 [PMID: 25815129 DOI: 10.4252/wjsc.v7.i2.461]
34 **Kim J**, Zaret KS. Reprogramming of human cancer cells to pluripotency for models of cancer progression. *EMBO J* 2015; **34**: 739-747 [PMID: 25712212 DOI: 10.15252/embj.201490736]
35 **Laplane L**, Beke A, Vainchenker W, Solary E. Concise Review: Induced Pluripotent Stem Cells as New Model Systems in Oncology. *Stem Cells* 2015; **33**: 2887-2892 [PMID: 26179060 DOI: 10.1002/stem.2099]
36 **Kim JJ**. Applications of iPSCs in Cancer Research. *Biomark Insights* 2015; **10**: 125-131 [PMID: 26279620 DOI: 10.4137/BMI.S20065]
37 **Papapetrou EP**. Patient-derived induced pluripotent stem cells in cancer research and precision oncology. *Nat Med* 2016; **22**: 1392-1401 [PMID: 27923030 DOI: 10.1038/nm.4238]
38 **Cai J**, Li W, Su H, Qin D, Yang J, Zhu F, Xu J, He W, Guo X, Labuda K, Peterbauer A, Wolbank S, Zhong M, Li Z, Wu W, So KF, Redl H, Zeng L, Esteban MA, Pei D. Generation of human induced pluripotent stem cells from umbilical cord matrix and amniotic membrane mesenchymal cells. *J Biol Chem* 2010; **285**: 11227-11234 [PMID: 20139068 DOI: 10.1074/jbc.M109.086389]
39 **Su RJ**, Yang Y, Neises A, Payne KJ, Wang J, Viswanathan K, Wakeland EK, Fang X, Zhang XB. Few single nucleotide variations in exomes of human cord blood induced pluripotent stem cells. *PLoS One* 2013; **8**: e59908 [PMID: 23573220 DOI: 10.1371/journal.pone.0059908]
40 **Lo Sardo V**, Ferguson W, Erikson GA, Topol EJ, Baldwin KK, Torkamani A. Influence of donor age on induced pluripotent stem cells. *Nat Biotechnol* 2017; **35**: 69-74 [PMID: 27941802 DOI: 10.1038/nbt.3749]
41 **Martincorena I**, Roshan A, Gerstung M, Ellis P, Van Loo P, McLaren S, Wedge DC, Fullam A, Alexandrov LB, Tubio JM, Stebbings L, Menzies A, Widaa S, Stratton MR, Jones PH, Campbell PJ. Tumor evolution. High burden and pervasive positive selection of somatic mutations in normal human skin. *Science* 2015; **348**: 880-886 [PMID: 25999502 DOI: 10.1126/science.aaa6806]
42 **Saini N**, Roberts SA, Klimczak LJ, Chan K, Grimm SA, Dai S, Fargo DC, Boyer JC, Kaufmann WK, Taylor JA, Lee E, Cortes-Ciriano I, Park PJ, Schurman SH, Malc EP, Mieczkowski PA, Gordenin DA. The Impact of Environmental and Endogenous Damage on Somatic Mutation Load in Human Skin Fibroblasts. *PLoS Genet* 2016; **12**: e1006385 [PMID: 27788131 DOI: 10.1371/journal.pgen.1006385]
43 **Blokzijl F**, de Ligt J, Jager M, Sasselli V, Roerink S, Sasaki N, Huch M, Boymans S, Kuijk E, Prins P, Nijman IJ, Martincorena I, Mokry M, Wiegerinck CL, Middendorp S, Sato T, Schwank G, Nieuwenhuis EE, Verstegen MM, van der Laan LJ, de Jonge J, IJzermans JN, Vries RG, van de Wetering M, Stratton MR, Clevers H, Cuppen E, van Boxtel R. Tissue-specific mutation accumulation in human adult stem cells during life. *Nature* 2016; **538**: 260-264 [PMID: 27698416 DOI: 10.1038/nature19768]
44 **Bhutani K**, Nazor KL, Williams R, Tran H, Dai H, Džakula Ž, Cho EH, Pang AW, Rao M, Cao H, Schork NJ, Loring JF. Whole-genome mutational burden analysis of three pluripotency induction methods. *Nat Commun* 2016; **7**: 10536 [PMID: 26892726 DOI: 10.1038/ncomms10536]
45 **Yoshihara M**, Hayashizaki Y, Murakawa Y. Genomic Instability of iPSCs: Challenges Towards Their Clinical Applications. *Stem Cell Rev* 2017; **13**: 7-16 [PMID: 27592701 DOI: 10.1007/s12015-016-9680-6]
46 **Gore A**, Li Z, Fung HL, Young JE, Agarwal S, Antosiewicz-Bourget J, Canto I, Giorgetti A, Israel MA, Kiskinis E, Lee JH, Loh YH, Manos PD, Montserrat N, Panopoulos AD, Ruiz S, Wilbert ML, Yu J, Kirkness EF, Izpisua Belmonte JC, Rossi DJ, Thomson JA, Eggan K, Daley GQ, Goldstein LS, Zhang K. Somatic coding mutations in human induced pluripotent stem cells. *Nature* 2011; **471**: 63-67 [PMID: 21368825 DOI: 10.1038/nature09805]
47 **Ji J**, Ng SH, Sharma V, Neculai D, Hussein S, Sam M, Trinh Q, Church GM, McPherson JD, Nagy A, Batada NN. Elevated coding mutation rate during the reprogramming of human somatic cells into induced pluripotent stem cells. *Stem Cells* 2012; **30**: 435-440 [PMID: 22162363 DOI: 10.1002/stem.1011]
48 **Ruiz S**, Gore A, Li Z, Panopoulos AD, Montserrat N, Fung HL, Giorgetti A, Bilic J, Batchelder EM, Zaehres H, Schöler HR, Zhang K, Izpisua Belmonte JC. Analysis of protein-coding mutations in hiPSCs and their possible role during somatic cell reprogramming. *Nat Commun* 2013; **4**: 1382 [PMID: 23340422 DOI: 10.1038/ncomms2381]
49 **Rouhani FJ**, Nik-Zainal S, Wuster A, Li Y, Conte N, Koike-Yusa H, Kumasaka N, Vallier L, Yusa K, Bradley A. Mutational History of a Human Cell Lineage from Somatic to Induced Pluripotent Stem Cells. *PLoS Genet* 2016; **12**: e1005932 [PMID: 27054363 DOI: 10.1371/journal.pgen.1005932]
50 **Kwon EM**, Connelly JP, Hansen NF, Donovan FX, Winkler T, Davis BW, Alkadi H, Chandrasekharappa SC, Dunbar CE, Mullikin JC, Liu P. iPSCs and fibroblast subclones from the same fibroblast population contain comparable levels of sequence variations. *Proc Natl Acad Sci USA* 2017; **114**: 1964-1969 [PMID: 28167771 DOI: 10.1073/pnas.1616035114]
51 **Ramirez JM**, Bai Q, Dijon-Grinand M, Assou S, Gerbal-Chaloin

S, Hamamah S, De Vos J. Human pluripotent stem cells: from biology to cell therapy. *World J Stem Cells* 2010; **2**: 24-33 [PMID: 21607113 DOI: 10.4252/wjsc.v2.i2.24]

52 **Janvilisri T**, Leelawat K, Roytrakul S, Paemanee A, Tohtong R. Novel Serum Biomarkers to Differentiate Cholangiocarcinoma from Benign Biliary Tract Diseases Using a Proteomic Approach. *Dis Markers* 2015; **2015**: 105358 [PMID: 26060332 DOI: 10.1155/2015/105358]

53 **Hogstrand C**, Kille P, Nicholson RI, Taylor KM. Zinc transporters and cancer: a potential role for ZIP7 as a hub for tyrosine kinase activation. *Trends Mol Med* 2009; **15**: 101-111 [PMID: 19246244 DOI: 10.1016/j.molmed.2009.01.004]

P- Reviewer: Li SC, Mozdarani H **S- Editor**: Ji FF **L- Editor**: A
E- Editor: Li D

Published by **Baishideng Publishing Group Inc**
7901 Stoneridge Drive, Suite 501, Pleasanton, CA 94588, USA
Telephone: +1-925-223-8242
Fax: +1-925-223-8243
E-mail: bpgoffice@wjgnet.com
Help Desk: http://www.f6publishing.com/helpdesk
http://www.wjgnet.com