Chapter 4:
Strategies for rapidly mapping proviral integration sites and assessing cardiogenic potential of nascent human induced pluripotent stem cell clones

Cheryl Dambrot, Henk P.J. Buermans, Eszter Varga, Georgios Kosmidis, Karin Langenberg, Simona Casini, David A. Elliott, Andras Dinnyes, Douwe E. Atsma, Christine L. Mummery, Stefan R. Braam, Richard P. Davis

Published: Experimental Cell Research 2014 Oct 1;327(2):297-306

Abstract

Recent methodological advances have improved the ease and efficiency of generating human induced pluripotent stem cells (hiPSCs), but this now typically results in a greater number of hiPSC clones being derived than can be wholly characterized. Additionally, individual clones display biological variability including their capacity to differentiate, with genetic heterogeneity thought to be a contributing factor. It is therefore imperative that methods are developed which facilitate rapid selection of the hiPSC clones that are most suited to the downstream research aims. Here we describe a useful combination of procedures that enables the simultaneous screening of multiple iPSC clones to determine their genomic integrity as well as their cardiac differentiation potential within two weeks of the putative reprogrammed colonies initially appearing in culture. By coupling a splinkerette-PCR method with Ion Torrent sequencing, we could ascertain the number of and map the proviral integration sites in lentiviral-reprogrammed hiPSCs, enabling identification of clones containing single copy integrations in genomic regions devoid of known genes or transcriptional control elements. In parallel, we developed an effective cardiac differentiation protocol that generated functional cardiomyocytes within 10 days without requiring line-specific optimization for any of 6 independent human pluripotent stem cell lines tested. Finally, to demonstrate the scalable potential of these procedures, we picked 20 nascent iPSC clones and performed these independent assays concurrently. Before the clones required passaging, we were able to identify clones with a single integrated copy of the reprogramming vector and robust cardiac differentiation potential for further analysis.

Introduction

Since the initial reports that human fibroblasts could be reprogrammed to a self-renewing, pluripotent state [1,2], interest in induced pluripotent stem cells (iPSC) has grown exponentially as their many applications in biomedical research and regenerative medicine have become evident. However, a remaining challenge for the field is to address the technical issues that contribute to individual cell line differences and thereby confound our ability to distinguish disease-associated variability from background noise. These include methods for rapidly assessing the quality and genomic integrity of the reprogrammed cell lines, and robust protocols for differentiation that are completely defined and readily transferable across multiple iPSC lines.

Most reprogramming methods generate a large number of putative iPSC clones that appear morphologically similar to embryonic stem cells (ESCs); however only a subset of these have comparable molecular and functional features [3]. To distinguish bona fide iPSCs from those that are only partially reprogrammed, detailed characterization of the clones is required, which is often lengthy and expensive. This typically limits the number of clones that are examined to fewer than six, with the selection of clones for further analysis predominantly based on subjective criteria like morphology and culture characteristics [4,5]. While this is generally sufficient to identify at least one genuine iPSC clone that can indefinitely self-renew and is pluripotent, it does not assess the clonal variability in differentiation efficiency that is often detected when pluripotent stem cells (PSCs) undergo further guided differentiation to functional cell subtypes [6-9]. One potential source of this clonal variation if integrative reprogramming methods have been used might be insertional mutagenesis caused by the reprogramming factors [10,11]. However, the genomic integrity of iPSCs is frequently not assessed, despite studies demonstrating that the integration of viral vectors can result in the dysregulation of adjacent genes [12,13]. This altered gene expression could lead to perturbed signalling pathways that may alter the pluripotency or differentiation potential of the iPSCs, similar to that observed in some virally-transduced hematopoietic stem cell lines [14].

To address these issues, we developed a novel combination of methods to aid in selecting hiPSC clones for studying cardiogenesis. Furthermore, both procedures are scalable, allowing many potential iPSC clones to be screened simultaneously. To demonstrate this, the number and location of the proviral integrations, as well as the cardiac differentiation potential of 20 putative iPSC clones was determined within 2 weeks following their emergence during reprogramming. These methods require minimal amounts of starting material and can assist in identifying the optimal iPSC clones for further characterization. Furthermore, these procedures potentially could be applied to aid in the selection of transgenic human PSCs in which, for example, overexpression constructs, therapeutic genes or ectopic reporters have been randomly integrated.

Materials and Methods

Ethics Statement

Human skin biopsies were obtained from patients after individual written permission using standard informed consent procedures, and following approval for use in this study by the Leiden University Medical Center's medical ethics committee. Control skin samples and milk teeth were obtained as waste tissue from donors in accordance with the Dutch Federation of Biomedical Scientific Societies "Use of human tissue for scientific research" and "Code of good use" directives. All samples were collected and anonymized by the treating physician.

Human iPSC derivation

Human dermal fibroblasts or dental pulp cells were reprogrammed to iPSCs using the pRRL.PPT. SF.hOKSMidTomato-preFRT polycistronic lentiviral vector as previously published [15,16]. Based on morphology, human (h)ESC-like colonies were manually picked 30 days after transduction. All human samples were obtained after informed consent.

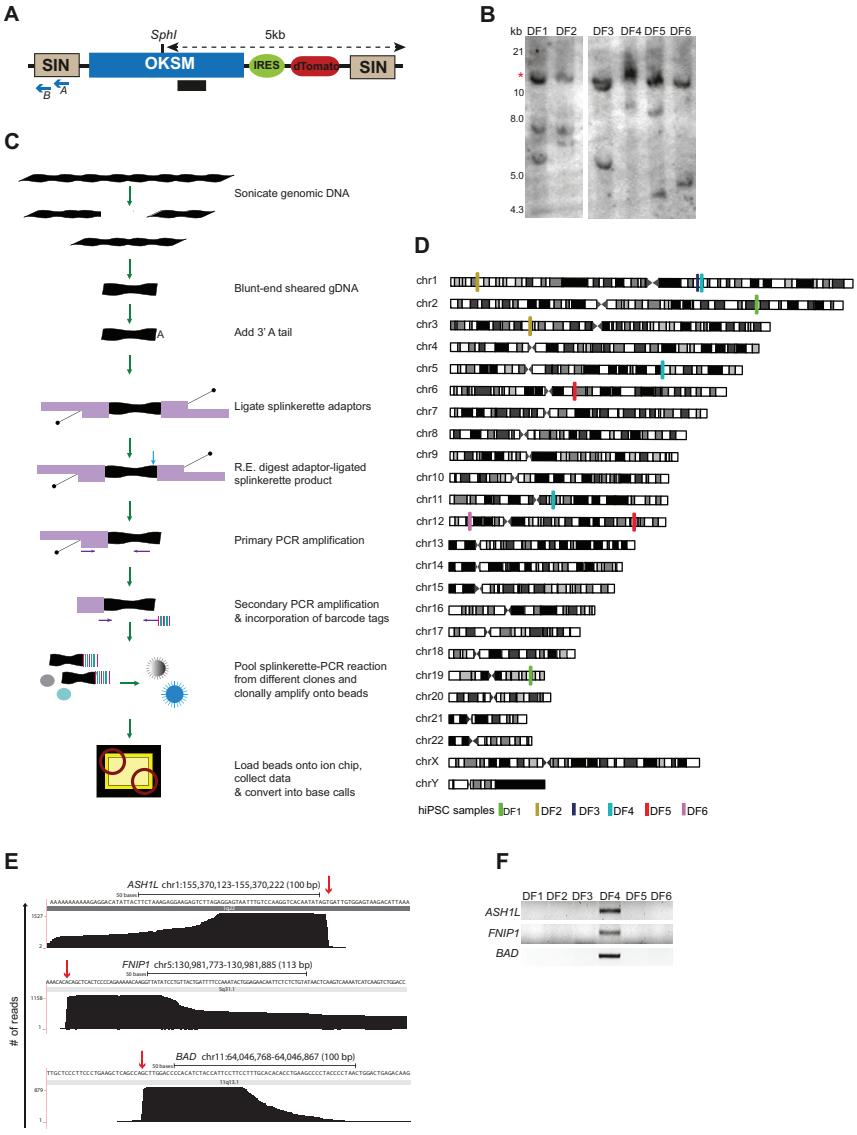

Figure 1. Genomic mapping of proviral integration sites in lentiviral-generated hiPSCs. A) Schematic representation of the polycistronic lentiviral vector used to generate the hiPSCs following integration into the host genome. The four reprogramming factors (OKSM) are linked by 2A peptide sequences, while a red fluorescent protein (dTomato) is downstream of an IRES sequence. The position and distance of the SphI site from the 3' end of the integrated provirus, as well as the probe (thick black line) used for Southern blot analysis are shown. Also depicted are the primers (A, B) within the viral LTR used in the amplification of the adaptor-ligated splinkerette products (SIN, self-inactivating element; OKSM, OCT4/KLF4/SOX2/c-MYC; IRES, internal ribosome entry site). B) Southern blot analysis of 6 control hiPSC lines to assess the number of viral integrations. Genomic DNAs were digested with SphI and hybridized with a probe to c-myc. The * indicates the location of the band corresponding to the endogenous c-MYC locus. C) Schematic representation of the shear-splinkerette procedure coupled to the fragment preparation for multiplexed PCR amplification and Ion Torrent sequencing. D) Idiogram depicting the genomic integration sites of the lentivirus for the 6 control hiPSC lines as determined by shear-splink PCR combined with Ion Torrent sequencing. E) Visualization of the mapped reads obtained for hiPSC line, DF4 following shear-splink PCR and Ion Torrent sequencing. Reads mapped to three genomic loci. The red arrows indicate the position in the genome where the provirus integrated. F) Validation of the lentivirus integration sites identified in E by genomic PCR analysis using primer pairs specific for each integration locus. Insertion of the lentivirus at these three genomic loci was only detected in the hiPSC line, DF4.

PSC culture and differentiation

PSCs were cultured on Matrigel (BD Biosciences)-coated tissue culture dishes in mTeSR1 according to the manufacturer's protocol (Stem Cell Technologies). The control hESC and dermal fibroblast (DF)-derived hiPSC lines used in this study are listed in Table S1. To initiate differentiation to cardiomyocytes, the PSCs were dissociated into small clusters of cells and seeded onto a Matrigel-coated cell culture dish in mTeSR1. Three days later (differentiation day(d) 0), the medium was replaced with low-insulin (1 mg/L), (LI)-BPEL medium [17] and supplemented with BMP4, Activin A, CHIR99021 and XAV939 as indicated. If the cultures were to be maintained for more than 16 days, the cells were overlaid with additional Matrigel (1:100) on differentiation d3 to prevent detachment from the dishes. From differentiation d6, the LI-BPEL medium was refreshed twice a week.

Southern blot

Genomic DNA was digested with SphI (Promega) overnight, resolved by gel electrophoresis and transferred to Hybond-N+ Nylon membrane (GE Healthcare). Hybridization and detection of the probe, encompassing the coding sequences of c-Myc, was performed as previously described [18].

Amplification and sequencing of vector integration sites

To determine the lentiviral integration sites in either the DF hiPSC lines or the putative iPSC clones, genomic DNA was isolated with the DNAeasy Blood and Tissue Kit (Qiagen) and splinkerette PCR performed [19-21]. Oligonucleotides and PCR amplification conditions used are listed in Tables S2 and S3. To fragment the DNA, either 2 µg was sheared for a target peak of 200bp using a Covaris S2 sonicator according to the manufacturer's protocol, or 5µg was digested with the restriction enzymes CviQI, AseI and BclI (all from New England Biolabs (NEB)). The resulting DNA fragments were concentrated, blunt-ended using the NEBNext End Repair module (NEB) and modified at the 3' ends with the addition of an adenosine (NEBNext dA-Tailing Module; NEB). Splinkerette adaptors were ligated to the DNA fragments using Quick T4 DNA Ligase (NEB). To prevent amplification of internal lentiviral sequences, samples were digested overnight with BglII. Primary amplification of the adaptor-ligated splinkerette products was performed using Platinum Taq HiFi (Invitrogen) with the primer pairs, PCR_A1 and PCR_A2, followed by secondary amplification with the primer pair, PCR_B1 and PCR_B2.

To de-multiplex the splinkerette-PCR products, the amplified products were either cloned into the vector pCR®2.1-TOPO® (Invitrogen), transformed into bacteria and sequenced using Sanger sequencing, or barcoded and sequenced using ion semi-conductor sequencing [22]. For Ion Torrent sequencing (Life Technologies) of pooled splinkerette-PCR products from multiple iPSC lines, "barcode" and adaptor sequences were incorporated into the primers PCR_B1 and PCR_B2. Adaptor and barcode sequences are listed in Table S4.

Mapping and Processing of Sequence Reads

The Ion Torrent sequence data was processed in several steps to identify the provirus integration sites in the hiPSC genome. First, the custom 3' splinkerette adaptor sequence was identified and removed from the sequence reads by the Ion Torrent Suite software (v3.4.2). The dataset was demultiplexed into the individual sample sets and the provirus sequences removed from the reads. Reads with a minimal length of 30 bases were aligned to the human genome (Hg19) using the TMAP sequence aligner with standard settings. The Bed-tools package [23] was used to retrieve regions with a peak height of at least 10% reads relative to the total number of aligned reads and a peak base width of at least 40 nucleotides. The selected candidate integration sites were annotated to the nearest gene or transcriptional control element. Lentiviral insertion loci in the DF-iPSCs were confirmed by genomic PCR using lentiviral-specific and genomic-specific primers (Table S5).

Flow cytometry
Differentiation d13 cells were dissociated using TrypLE Select (Invitrogen) and filtered through a 35 µm cell strainer (BD Bioscience). Intracellular flow cytometry with mouse anti-bovine cardiac troponin T (cTnT; Hybridomabank) was performed using the FIX & PERM kit (Invitrogen) according to manufacturer's protocol, with the primary antibody detected with phycoerythrin-conjugated goat anti-mouse IgG2a (Jackson ImmunoResearch). Cells were gated based on forward and side scatter, and flow cytometric gates were set using control cells not expressing GFP, or differentiated cells labeled with the isotype control. Samples were measured using a LSRII flow cytometer (BD Bioscience), and results analyzed using FlowJo 7.6.5 (TreeStar).

Immunofluorescence analysis
Differentiation d20 NKX2-5eGFP/w hESCs were dissociated with TrypLE Select, seeded on Matrigel-coated coverslips and cultured for a minimum of 72 h before fixing with 2% PFA. Coverslips were stained with antibodies against α-Actinin (mouse-α-rabbit), sarcomeric Tropomyosin (mouse-α-chick; both from Sigma Aldrich), Troponin I (rabbit-α-human), Nkx2.5 (rabbit-α-human; both from Santa Cruz), β-MHC (mouse-α-human; Millipore), MLC2a (rabbit-α-mouse; provided by S.W. Kubalak), as previously published [24]. Primary antibodies were detected with IgG-conjugated Cy3 (donkey-α-rabbit, Jackson Immunoresearch) or Alexa 488 (donkey-α-mouse, Invitrogen) antibodies. Images were captured on a Leica SP5 confocal laser scanning microscope (Leica Microsystems).

Table 1: The number of confirmed proviral insertions in the lentiviral-generated DF-hiPSC lines as determined using different mapping approaches.

Clone	Number of inserts as determined by:		
	Southern blot	RE-splinkerette; TOPO Subcloning	Shear-splink; Ion Torrent sequencing
DF1	2	2	2
DF2	2	2	2
DF3	1	0	1
DF4	1	2	3
DF5	2	1	2
DF6	1	1	1

RE: restriction enzyme

Electrophysiology and Ca2+ imaging
Action potential measurements were performed on single cardiomyocytes 7-15 days after cell dissociation and analyzed as described previously [25]. For Ca2+ imaging, differentiation d20 NKX2-5eGFP/w hESCs were dissociated with TrypLE Select, seeded (40,000 cells/well) into a Matrigel-coated 96 well imaging plate (BD Bioscience), and cultured for a further 7 days. Cells were loaded with 5µM Fluo-4 AM (Invitrogen) and imaged in Tyrode's solution using a confocal microscope (Leica AF6000) as previously published [26]. Spontaneous Ca2+ transients were recorded at 37° C for 20 s.

Results

High throughput mapping of lentiviral integration sites in iPSCs
To validate and establish a procedure for high-throughput sequencing and mapping of the lentiviral insertion sites, we initially quantified and mapped the genomic locations of the reprogramming cassettes in six control hiPSC lines (DF1-6) by Southern blot analysis, and by

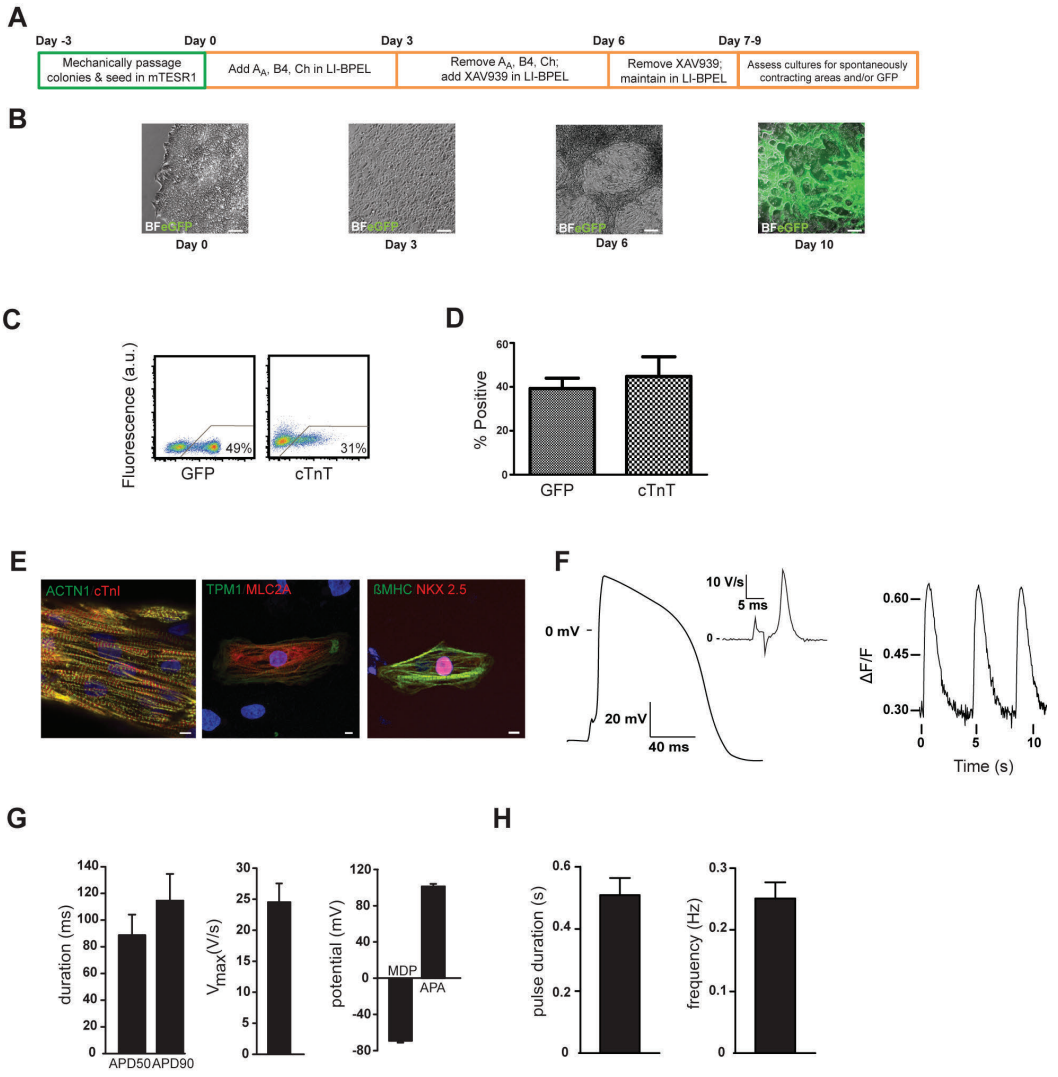

Figure 2. Efficient cardiogenesis of NKX2-5eGFP/w hESCs maintained as colonies in mTeSR1. A) Schematic representation of the optimized differentiation protocol for generating cardiomyocytes from human PSCs. B) Overlaid bright field (BF) and green fluorescence (eGFP) images showing the typical morphology and eGFP expression of NKX2-5eGFP/w hESCs at various time points during the differentiation. Scale bars, 100 µm. C) Representative flow cytometry profiles showing the correlation in the percentage of cells expressing eGFP (NKX2-5) and cardiac troponin T (cTnT) at differentiation d14. D) Histogram showing the mean percentage of differentiation d14 cells expressing eGFP (NKX2-5) or cTnT. Data represents the mean ± SEM from 5 independent experiments. E) Immunofluorescence images of the cardiac markers α-actinin (ACTN1), troponin I (cTnI), sarcomeric tropomyosin (TPM1), myosin regulatory light chain 2, atrial isoform (MLC2A), beta-myosin heavy chain (βMHC), and NKX2-5 in cardiomyocytes differentiated using the protocol depicted in A. All cells were co-stained with the nuclear dye DAPI (blue). Scale bars represent 5 µm. F) Representative electrophysiological recordings from cardiomyocytes derived using the protocol depicted in A. The left panel shows a typical example of an action potential and an upstroke velocity (inset) from a cardiomyocyte triggered at 1Hz. The right panel shows the calcium transient of a spontaneously contracting cardiomyocyte. G) Average data at 1Hz for action potential duration at 50% and 90% repolarization (APD50, APD90), maximal upstroke velocity (Vmax), maximal diastolic potential (MDP), and action potential amplitude (APA) in cardiomyocytes derived from NKX2-5eGFP/w hESCs (n=7, mean ±SEM). H) Summary of spontaneous whole cell Ca2+ pulse duration and frequency in differentiated NKX2-5eGFP/w hESC cardiomyocytes (n=34, mean ±SEM).

combining splinkerette-PCR with TOPO-shotgun cloning and amplicon sequencing. These hiPSC lines had been generated following transfection of fibroblasts with a polycistronic lentivirus at MOIs ranging from 0.25 to 5 (Figure 1A). We had previously ascertained that MOIs above 0.4 would frequently generate hiPSCs containing multiple proviral integrations [16], enabling us to assess the sensitivity of the above procedures.

Southern blot analysis confirmed that each of the lines contained at least 1-3 proviral integrations (Figure 1B). However, a limitation of this technique is that it only provides information on the number of integrations and not their genomic location. Such data is critical to fully assess the risk of insertional mutagenesis in virally-transduced cell lines. Additionally, if two or more integrations result in a similar sized DNA fragment when digested, this can lead to an inaccurate estimation of the total number of integrations in that clone. To resolve these issues, splinkerette PCR was performed on genomic DNA that was fragmented using restriction endonucleases (REs), either individually or in combination [19,20] (Figure S1). Splinkerette PCR is a variant of ligation-mediated PCR and is commonly used in forward genetic screens to map the genomic sites in which insertional mutagens such as viruses or transposons have integrated [19,21]. More recently the technique has been employed to identify the integration sites in iPSCs reprogrammed using transposons, enabling the excision of the reprogramming cassette following re-expression of the transposase to be monitored [18,27].However neither of these splinkerette procedures identified the minimum number of insertion sites as determined by Southern blot analysis for all six clones (Table 1), most likely due to amplification biases resulting from the uneven genomic distribution of RE recognition sites [21].

To limit these amplification and sequencing biases, we applied a modified splinkerette method, termed “shear-splink”, that employed random fragmentation of the genomic DNA resulting in controllable size distribution of DNA fragments for all insertions [21]. By incorporating ‘barcode’ sequences into the primers, the resulting splinkerette PCR products from the different clones could be pooled together and sequenced using ion sequencing technology (Figure 1C). All insertion sites previously identified using RE-splinkerette procedures were also detected using this high-throughput method. Furthermore, the shear-splink method mapped additional insertions not detected either by RE-splinkerette or Southern blot (Figure 1D & E, Table 1, and Table S6). For example, with DF4, Southern blot analysis identified one insert while shear-splink analysis identified three. All insertion events were confirmed by locus-specific genomic PCR (Figure 1F and Figure S2). These results indicated that combining shear-splink with ion semi-conductor sequencing can efficiently uncover proviral insertion sites and is amenable to high-throughput adaptation.

Robust and Scalable Protocol to assess the Cardiac Differentiation Efficiency of human PSCs

To test the cardiac differentiation potential of hiPSC clones as they emerge from the reprogramming process, a rapid and reproducible method requiring no optimization between the clones is needed. Building on previous findings [28-30], we developed a monolayer differentiation protocol that could be carried out directly on mechanically passaged PSC colonies grown in mTeSR1 culture medium (Figure 2A). The method was initially optimized using a hESC line in which GFP reports the expression of the cardiac associated transcription factor NKX2-5 (NKX2-5eGFP/w hESCs) [28]. We then confirmed the procedure also effectively induced cardiac differentiation in a further 5 human PSC lines (Figure S3).

Three days after mechanical passaging, the medium on the colonies was replaced with BPEL medium supplemented with 20 ng/ml Activin A, 20 ng/ml BMP4 and 1.5 μM Chir99021 for 72 h. Although this growth factor combination was sufficient to generate cardiomyocytes in some PSC lines, the yield was very low and variable between individual differentiation experiments. Because it has previously been demonstrated the tankyrase inhibitor XAV939 could promote cardiogenesis when added after mesoderm formation [30,31], we supplemented the medium

with 5μM XAV at day 3 for 72 h. Phase contrast photomicrographs of cultures at various time points during the protocol demonstrate the typical appearance of the cells during the differentiation process (Figure 2B). By differentiation d6, GFP- web-like colonies were visible that subsequently expanded, becoming GFP+ and forming a sheet of contracting cells by differentiation d10. Using this optimized differentiation protocol, on average ~40% of the cells expressed NKX2-5(eGFP) and cTnT by differentiation d14 (Figure 2C & D). Immunofluorescence analysis confirmed these cells expressed sarcomeric markers (Figure 2E). Typical examples of action potential, upstroke velocity and Ca2+ transients in the resulting cardiomyocytes are shown in Figure 2F. The action potential properties of these cells, as well as the spontaneous whole cell Ca2+ pulse duration and frequency (Fig 2G & H), were comparable to those previously reported for human PSC-derived cardiomyocytes [32-34].

Mapping proviral integration sites and assessing the cardiac differentiation potential of 20 newly reprogrammed hiPSC clones

To demonstrate the applicability of these methods to identify the lentiviral insertion sites and assess the cardiac differentiation potential of prospective iPSC clones rapidly, 20 putative iPSC clones were selected from reprogrammed dental pulp cells based on morphology. Colony pieces from these clones were either plated for the cardiac differentiation assay or used to isolate genomic DNA (Figure 3A). The shear-splink procedure in combination with ion semi-conductor sequencing identified that 1 of the 18 clones with mapped reads had 2 proviral insertions, while the remainder had only 1 insertion (Figure 3B). The predominance of single copy integrations was due to a MOI of 0.25 being used for reprogramming. Based on the genomic locations to which the lentivirus was mapped, 11 of these clones were unique with the remaining 6 being subclones. Of these 11 clones, 2 (clones 5 & 17) had the provirus integrated outside any known genes or non-coding RNA sequences. A complete overview of insert locations is summarized in Table S7.

As early as day 10 after initial plating, contracting clusters were observed in the differentiation assay with some of the clones. Analysis of cTnT expression at day 13 by flow cytometry indicated the cardiac differentiation efficiencies of the putative iPSC clones ranged from ~0-22% (Figure 3C & D). Based on these results, several iPSC clones (9, 11, 12, 17, and 19) were identified as being the most suitable for further characterization, while several other clones (1, 4, 6, 8, 13, 14 and 20) were designated as least suitable based on their poor cardiac differentiation efficiency. Combining these findings with the Ion Torrent sequencing data, the lentivirus had inserted within intragenic regions for clones 9 and 19 (NPLOC4 and AP1G1 respectively), while for clone 11 the lentivirus was in exon 2 of ZDBF2. Clone 12 had two proviral integrations within introns of PRPSAP1 and PEX14. Only clone 17 contained a viral insert outside any annotated genes or transcriptional control elements, and also had relatively high cardiac differentiation efficiency.

Discussion

Current methods to identify fully reprogrammed iPSCs following viral infection of somatic cells rely on expensive and often lengthy protocols, highlighting the need for methods that will streamline the selection process. During reprogramming, hundreds of clonal iPSCs are typically generated, often with variable differentiation capacities possibly due to genetic and epigenetic differences [6,9,35]. However representative iPSC clones are generally selected based on morphology, at a time when little or no information is available on the number and location of the proviral integrations or on the differentiation efficiency to particular lineages.

Using shear-splink PCR coupled with high-throughput ion semi-conductor sequencing [21,22], we were able to quickly and reliably determine proviral insert number and genomic location for multiple hiPSC clones simultaneously. Indeed this combined procedure identified integrations that were undetected when the clones were analyzed by Southern blot or RE-based splinkerette procedures.

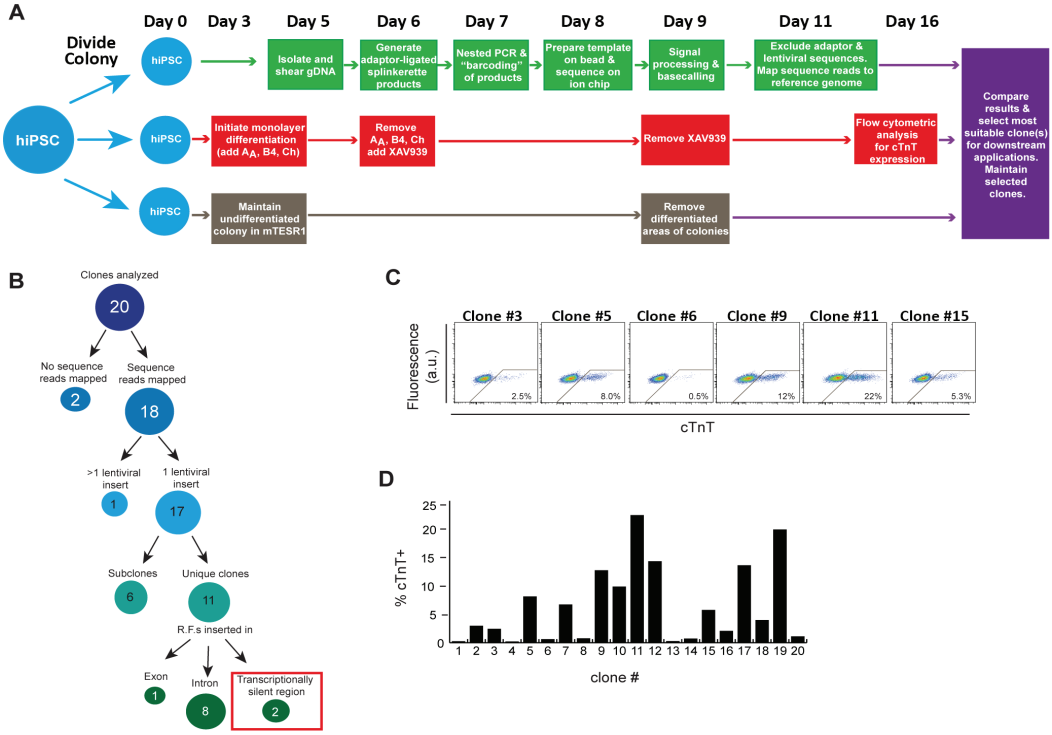

Figure 3. Simultaneous mapping of proviral integration sites and assessment of cardiac differentiation potential for multiple hiPSC clones. A) Schematic representation of the pipeline to assess newly derived hiPSCs. B) Flow chart summarizing the proviral insert number, genomic location and locality of 20 hiPSC as determined by shear-splink and Ion Torrent sequencing. The red box indicates the clones most suitable for further evaluation (RFs, reprogramming factors) C) Representative flow cytometry plots showing the percentage of cardiac troponin T-positive (cTnT+) cells after 13 days of differentiation for a selection of the 20 hiPSC clones analyzed. D) Histogram showing the mean percentage of cells expressing cTnT at day 16 for the 20 hiPSC clones assessed for cardiac differentiation potential.

Significant inter-clonal variability in the differentiation efficiency of hiPSCs, including in the generation of cardiomyocytes has been well documented [6,9,36]. Inadvertently selecting representative patient-specific iPSC clones with poor cardiac differentiation efficiency could adversely influence the ability to distinguish disease-associated phenotypes. Here we developed a cardiac differentiation protocol that was fast and required no optimization between human PSC lines. We favored a monolayer differentiation-based technique due to its simplicity and scalability [28,37,38]. Previously published protocols often required prior adaptation of the cells to enzymatic passaging to enable the seeding of single or small clusters of cells, necessitating an additional 1-4 weeks to establish [39,40]This makes these protocols unsuitable for rapidly determining the cardiac differentiation potential of emerging iPSC colonies, since at this stage the colonies are still manually passaged. Aside from the time required, maintaining >20 colonies while assessing the cardiac potential is not practical and prohibitively expensive for high throughput development.

By using LI-BPEL, a serum-free differentiation medium, we limited the variability due to undefined components present in serum, as well as enhanced the effectiveness of the recombinant growth factors and small molecules [17,41]. While the combination of BMP4, Activin A and either bFGF or WNT3A can induce cardiogenesis in monolayer cultures of PSCs [28,29,37], optimization of the concentrations of BMP4 and activin A for individual cell lines as well as the time of addition

and removal of the growth factors has proved to be important [42]. Although small molecule modulation of the regulatory elements of the Wnt/β-catenin signaling pathway has been shown also to result in more robust cardiac differentiation in multiple human PSC lines [38,43] than the method described here, these protocols as mentioned earlier, have all required prior adaptation of the cells to enzymatic passaging. We found combining both approaches (staged exposure to BMP4, activin A and the GSK3β inhibitor, CHIR99021, followed by the tankyrase inhibitor, XAV939) resulted in robust cardiac differentiation from 6 human PSC lines without the need to individually optimize timing or concentrations for each line. This procedure could successfully induce cardiomyocyte differentiation from hiPSC colonies maintained either with MEFs in KOSR-containing medium or feeder-free with mTeSR1 medium, as well as from PSC lines passaged enzymatically, without any adaptations to the protocol. It is likely that the efficiency of cardiac development could be further optimized for each cell line by determining the optimal concentration and time of application of the growth factors and small molecule [42]. However such an approach is not feasible when screening multiple clones and is more suitable once an iPSC line is established. Similarly, reducing the level of insulin in the differentiation media further might lead to additional improvements in cardiac differentiation efficiency.

To demonstrate that both procedures could be incorporated into the current workflow for identifying fully reprogrammed iPSCs, we analyzed 20 nascent iPSC colonies that were selected for further characterization based on morphology. Within 16 days and prior to the clones requiring passaging, the cardiac potential along with mapping the proviral integration sites had been ascertained. It should be noted that the results from performing both of these procedures does not conclusively demonstrate that the poor cardiac differentiation capabilities of a clone is due to the integration site of the provirus. Impaired differentiation could also be due to the clone not being fully reprogrammed, or other genetic or epigenetic abnormalities [9]. The purpose of performing such a screen is to enable the researcher to make a more informed decision on which iPSC clones to select for further characterization based on the downstream applications of the cell line. From the iPSC colonies that were screened, clone 11 was the most efficient at generating cardiomyocytes. However with this clone, the lentivirus had inserted into exon 2 of the zinc finger gene, ZDBF2. Little is known about the function of ZDBF2, although there is evidence that it is an imprinted gene predominantly expressed in the brain and muscle [44]. It has also been identified as a transcript upregulated in embryonic cardiac fibroblasts from microarray studies [45]. Therefore while disruption of ZDBF2 does not appear to be detrimental to cardiomyocyte differentiation, it could influence the phenotype if the cell line was to be used in developing cardiac models for example in “organ-on-a-chip” formats. Alternatively, clone 17 also had relatively high cardiac differentiation efficiency along with a proviral integration site outside any known coding regions, potentially making this clone more suitable for such applications.

Furthermore, both of these techniques are not limited to evaluating the genomic integrity and differentiation potential of reprogrammed iPSCs. With little modification, these procedures could also be readily incorporated into a screening strategy of human PSCs that have been genetically modified by either viral or transposon-based transfection. Here there is also a recognized need to generate and analyze several transgenic lines, preferably in a short timespan, since positional effects resulting from genomic integration could impact on their differentiation potential.

Conclusion

It is widely acknowledged that hiPSCs hold great promise for disease modeling, safety pharmacology and drug discovery, particularly with respect to the cardiovascular system [46-48]. However, methods to identify the most suitable iPSC clones for further study are often lengthy and laborious. Here, we have described a rapid and scalable pipeline consisting of two independent assays that can simultaneously be applied directly to hiPSC clones as they appear following reprogramming, enabling clones to be selected based on the number and genomic

location of the integrated reprogramming cassette, as well as the cardiac differentiation potential. Solutions to a number of the remaining technical issues, such as those covered here, could contribute to more rapid realization of the potential of iPSCs by reducing the time required between obtaining primary cell cultures for reprogramming and selecting the best hiPSC lines for the question to be addressed.

Acknowledgements

We thank C. Baum and A. Schambach (Department of Experimental Hematology, Hannover Medical School, Hannover, Germany) for the lentiviral vector used for reprogramming, and C. Freund and S. van de Pas (iPSC core facility, LUMC) for generation of dental pulp hiPSC. We also thank A Riem (TandInZicht, Harmelen) for collection of milk teeth and M. Rabelink (Molecular Cell Biology, LUMC) for production of lentivirus.

References

1 Yu, J., Vodyanik, M.A., Smuga-Otto, K.et al (2007) Induced pluripotent stem cell lines derived from human somatic cells Science 318, 1917-1920
2 Takahashi, K., Tanabe, K., Ohnuki, M.et al (2007) Induction of pluripotent stem cells from adult human fibroblasts by defined factors Cell 131, 861-872
3 Cahan, P. and Daley, G.Q. (2013) Origins and implications of pluripotent stem cell variability and heterogeneity Nat Rev Mol Cell Biol 14, 357-368
4 Papapetrou, E.P. and Sadelain, M. (2011) Generation of transgene-free human induced pluripotent stem cells with an excisable single polycistronic vector Nat.Protoc. 6, 1251-1273
5 Park, I.H. and Daley, G.Q. (2009) Human iPS cell derivation/reprogramming Curr.Protoc.Stem Cell Biol. Chapter 4,
6 Boulting, G.L., Kiskinis, E., Croft, G.F.et al (2011) A functionally characterized test set of human induced pluripotent stem cells Nat Biotech 29, 279-286
7 Devine, M.J., Ryten, M., Vodicka, P.et al (2011) Parkinson's disease induced pluripotent stem cells with triplication of the alpha-synuclein locus Nat.Commun. 2, 440
8 Hu, B.Y., Weick, J.P., Yu, J.et al (2010) Neural differentiation of human induced pluripotent stem cells follows developmental principles but with variable potency Proc.Natl.Acad.Sci.U.S.A 107, 4335-4340
9 Thatava, T., Kudva, Y.C., Edukulla, R.et al (2013) Intrapatient variations in type 1 diabetes-specific iPS cell differentiation into insulin-producing cells Mol.Ther. 21, 228-239
10 Carlos Polanco J., Ho, M.S., Wang, B.et al (2013) Identification of Unsafe Human Induced Pluripotent Stem Cell Lines Using a Robust Surrogate Assay for Pluripotency Stem Cells
11 Hawley, R.G. (2008) Does retroviral insertional mutagenesis play a role in the generation of induced pluripotent stem cells? Mol.Ther. 16, 1354-1355
12 Winkler, T., Cantilena, A., Metais, J.Y.et al (2010) No evidence for clonal selection due to lentiviral integration sites in human induced pluripotent stem cells Stem Cells 28, 687-694
13 Zheng, W., Wang, Y., Chang, T.et al (2013) Significant differences in genotoxicity induced by retrovirus integration in human T cells and induced pluripotent stem cells Gene 519, 142-149
14 Wu, C. and Dunbar, C.E. (2011) Stem cell gene therapy: the risks of insertional mutagenesis and approaches to minimize genotoxicity Front Med. 5, 356-371
15 Warlich, E., Kuehle, J., Cantz, T.et al (2011) Lentiviral vector design and imaging approaches to visualize the early stages of cellular reprogramming Mol.Ther. 19, 782-789
16 Dambrot, C., van de Pas, S., van, Z.L.et al (2013) Polycistronic lentivirus induced pluripotent stem cells from skin biopsies after long term storage, blood outgrowth endothelial cells and cells from milk teeth Differentiation 85, 101-109
17 Ng, E.S., Davis, R., Stanley, E.G.et al (2008) A protocol describing the use of a recombinant protein-based, animal product-free medium (APEL) for human embryonic stem cell differentiation as spin embryoid bodies Nat.Protoc. 3, 768-776
18 Davis, R.P., Nemes, C., Varga, E.et al (2013) Generation of induced pluripotent stem cells from human foetal fibroblasts using the Sleeping Beauty transposon gene delivery system Differentiation 86, 30-37
19 Uren, A.G., Mikkers, H., Kool, J.et al (2009) A high-throughput splinkerette-PCR method for the isolation and sequencing of retroviral insertion sites Nat.Protoc. 4, 789-798
20 Yin, B. (2011) Isolation of genomic insertion sites of proviruses using Splinkerette-PCR-based procedures Methods Mol Biol 687, 25-42
21 Koudijs, M.J.M., Klijn, C., Weyden, L.V.D.et al (2011) High-throughput semiquantitative analysis of insertional mutations in heterogeneous tumors Genome 2181-2189
22 Rothberg, J.M., Hinz, W., Rearick, T.M.et al (2011) An integrated semiconductor device enabling non-optical genome sequencing Nature 475, 348-352
23 Quinlan, A.R. and Hall, I.M. (2010) BEDTools: a flexible suite of utilities for comparing genomic features Bioinformatics. 26, 841-842
24 Davis, R.P., Casini, S., van den Berg, C.W.et al (2012) Cardiomyocytes derived from pluripotent stem

cells recapitulate electrophysiological characteristics of an overlap syndrome of cardiac sodium channel disease Circulation 125, 3079-3091
25 Bellin, M., Casini, S., Davis, R.P.et al (2013) Isogenic human pluripotent stem cell pairs reveal the role of a KCNH2 mutation in long-QT syndrome EMBO J. 32, 3161-3175
26 Lan, F., Lee, A., Liang, P.et al (2013) Abnormal Calcium Handling Properties Underlie Familial Hypertrophic Cardiomyopathy Pathology in Patient-Specific Induced Pluripotent Stem Cells Cell Stem Cell 12, 101-113
27 Yusa, K., Rad, R., Takeda, J.et al (2009) Generation of transgene-free induced pluripotent mouse stem cells by the piggyBac transposon Nat Methods 6, 363-369
28 Elliott, D.A., Braam, S.R., Koutsis, K.et al (2011) NKX2-5(eGFP/w) hESCs for isolation of human cardiac progenitors and cardiomyocytes Nat.Methods 8, 1037-1040
29 Paige, S.L., Osugi, T., Afanasiev, O.K.et al (2010) Endogenous Wnt/beta-catenin signaling is required for cardiac differentiation in human embryonic stem cells PLoS.One. 5, e11134
30 Willems, E., Spiering, S., Davidovics, H.et al (2011) Small-molecule inhibitors of the Wnt pathway potently promote cardiomyocytes from human embryonic stem cell-derived mesoderm Circ.Res. 109, 360-364
31 Wang, H., Hao, J., and Hong, C.C. (2011) Cardiac induction of embryonic stem cells by a small molecule inhibitor of Wnt/beta-catenin signaling ACS Chem.Biol 6, 192-197
32 He, J.Q., Ma, Y., Lee, Y.et al (2003) Human embryonic stem cells develop into multiple types of cardiac myocytes: action potential characterization Circ.Res. 93, 32-39
33 Itzhaki, I., Rapoport, S., Huber, I.et al (2011) Calcium handling in human induced pluripotent stem cell derived cardiomyocytes PLoS One 6, e18037
34 Jonsson, M.K., van Veen, T.A., Goumans, M.J.et al (2009) Improvement of cardiac efficacy and safety models in drug discovery by the use of stem cell-derived cardiomyocytes Expert.Opin.Drug Discov. 4, 357-372
35 Burridge, P.W., Thompson, S., Millrod, M.A.et al (2011) A universal system for highly efficient cardiac differentiation of human induced pluripotent stem cells that eliminates interline variability PLoS.One. 6, e18293
36 Zhang, J., Wilson, G.F., Soerens, A.G.et al (2009) Functional cardiomyocytes derived from human induced pluripotent stem cells Circ.Res. 104, e30-e41
37 Zhang, J., Klos, M., Wilson, G.F.et al (2012) Extracellular matrix promotes highly efficient cardiac differentiation of human pluripotent stem cells: the matrix sandwich method Circ.Res. 111, 1125-1136
38 Lian, X., Hsiao, C., Wilson, G.et al (2012) Robust cardiomyocyte differentiation from human pluripotent stem cells via temporal modulation of canonical Wnt signaling Proc.Natl.Acad.Sci.U.S.A 109, E1848-E1857
39 Costa, M., Sourris, K., Hatzistavrou, T.et al (2008) Expansion of human embryonic stem cells in vitro Curr.Protoc.Stem Cell Biol Chapter 1, Unit-1C
40 Lian, X., Zhang, J., Azarin, S.M.et al (2013) Directed cardiomyocyte differentiation from human pluripotent stem cells by modulating Wnt/beta-catenin signaling under fully defined conditions Nat. Protoc. 8, 162-175
41 Mummery, C.L., Zhang, J., Ng, E.S.et al (2012) Differentiation of human embryonic stem cells and induced pluripotent stem cells to cardiomyocytes: a methods overview Circ.Res. 111, 344-358
42 Kattman, S.J., Witty, A.D., Gagliardi, M.et al (2011) Stage-specific optimization of activin/nodal and BMP signaling promotes cardiac differentiation of mouse and human pluripotent stem cell lines Cell Stem Cell 8, 228-240
43 Minami, I., Yamada, K., Otsuji, T.G.et al (2012) A small molecule that promotes cardiac differentiation of human pluripotent stem cells under defined, cytokine- and xeno-free conditions Cell Rep. 2, 1448-1460
44 Kobayashi, H., Yamada, K., Morita, S.et al (2009) Identification of the mouse paternally expressed imprinted gene Zdbf2 on chromosome 1 and its imprinted human homolog ZDBF2 on chromosome 2 Genomics 93, 461-472
45 Ieda, M., Tsuchihashi, T., Ivey, K.N.et al (2009) Cardiac fibroblasts regulate myocardial proliferation through beta1 integrin signaling Dev Cell 16, 233-244
46 Dambrot, C., Passier, R., Atsma, D.et al (2011) Cardiomyocyte differentiation of pluripotent stem cells and their use as cardiac disease models Biochem.J. 434, 25-35
47 Davis, R.P., van den Berg, C.W., Casini, S.et al (2011) Pluripotent stem cell models of cardiac disease and their implication for drug discovery and development Trends Mol.Med. 17, 475-484
48 Rajamohan, D., Matsa, E., Kalra, S.et al (2013) Current status of drug screening and disease modelling in human pluripotent stem cells Bioessays 35, 281-298

Supplementary Figures:

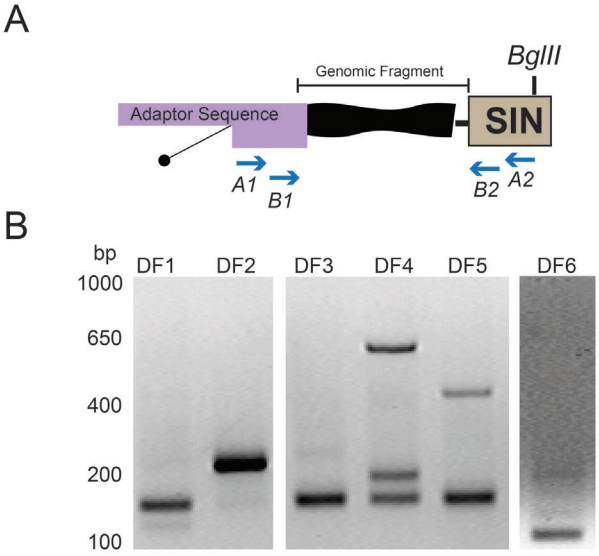

Figure S1. Genomic mapping of proviral integration sites in the lentiviral-generated DF-hiPSCs using restriction enzyme (RE)-splinkerette PCR. A) Schematic representation of the adaptor-ligated splinkerette products generated following RE digestion of the genomic DNA. The position of the BglII within the SIN sequence, as well as the splinkerette adaptor (purple box with angled black line) are shown. Also depicted are the PCR primers (A1, A2, B1, B2) used to amplify the junction sequences between the endogenous genome and the proviral insertion site (SIN, Self inactivating element). B) Agarose gel electrophoresis depicting the fragments generated following primary and secondary PCR amplification of the adaptor-ligated splinkerette products for DF-iPSC clones 1-6. Note, following cloning of these products into a TOPO vector, transformation into bacteria and subsequent sequencing, some fragments corresponded to non-specific amplification of genomic DNA sequences. Additionally, some bands corresponded to multiple proviral insertions. The column "RE-splinkerette; TOPO subcloning" in Table 1 lists the number of unique proviral inserts identified per clone using this technique.

F

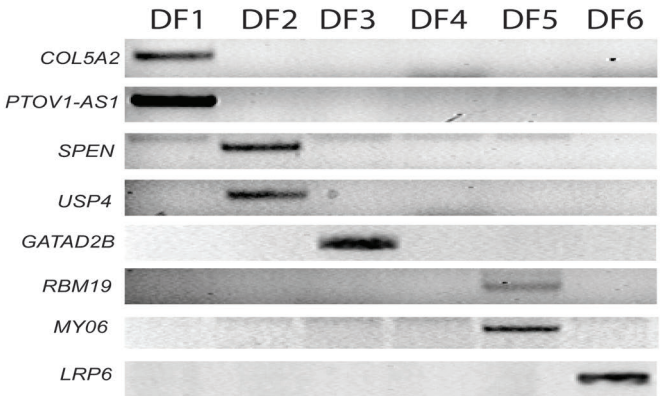

Figure S2. Validation of the lentivirus integration sites identified by Ion Torrent sequencing for the DF-hiPSC clones and listed in **Table S6**, using locus-specific PCR analysis. Results for clone DF4 are shown in **Figure 1F.**

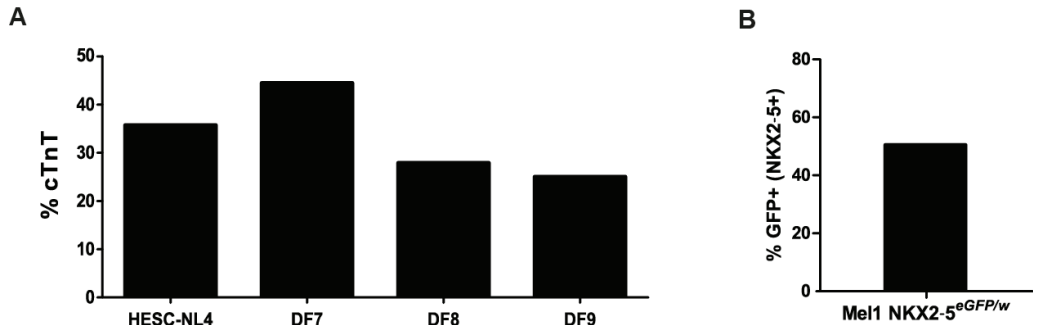

Figure S3. The monolayer differentiation protocol efficiently induces cardiogenesis for multiple hESC and hiPSC lines without requiring line-specific optimisation. A) The histogram shows the mean percentage of cells expressing cTnT at differentiation d14 for an additional 4 hPSC lines assessed for cardiac differentiation potential. B) The histogram shows the mean percentage of cells expressing eGFP (NKX2-5) at differentiation d14 from the hESC line, Mel1 NKX2-5eGFP/w. Data represents the mean ± SEM from 2 independent experiments.

Table S1: Pluripotent stem cell lines used in this study.

PSC type	Identification within article	Full name of clone*
hESC	*NKX2-5$^{eGFP/w}$* hESCs	HES3 *NKX2-5$^{eGFP/w}$* [1]
hESC	HESC-NL4	HESC-NL4 [2]
hESC	Mel1 *NKX2-5$^{eGFP/w}$*	Mel1 *NKX2-5$^{eGFP/w}$* [1]
hiPSC	DF1	LUMC0035iMyBPC01
hiPSC	DF2	LUMC0034iMyBPC02
hiPSC	DF3	LUMC0033iMyBPC09
hiPSC	DF4	LUMC0034iMyBPC12
hiPSC	DF5	LUMC0033iMyBPC13
hiPSC	DF6	LUMC0004ictrl03
hiPSC	DF7	LUMC0006ictrl01
hiPSC	DF8	LUMC0004ictrl10
hiPSC	DF9	FLB243#1 [3]

* Nomenclature of iPSC lines based on the standardized naming system proposed by Luong et al, Cell Stem Cell, 2011. Where applicable, the reference describing the initial characterization of the cell line is included in brackets.

[1] Elliott et al, Nat Methods, 2011; [2] Freund et al, Stem Cells, 2008; [3] Freund et al Neth Heart J, 2010

Table S2: Oligonucleotides used for splinkerette-PCR

Name	Sequence
Splink adaptor (Long)	*5'*-CCTCCACTACGACTCACTGAAGGGCAAGCAGTCCTAACAACCATGT-*3'*
Splink adaptor (Short)	*5*'-Phosphate-CATGGTTGTTAGGACTG-Amino-*3'*
PCR_A1	*5'*-CCTCCACTACGACTCACTGAAGGGC-*3*'
PCR_A2	*5*'-GCAGATCTTGTCTTCGTTGGGAGTT-*3'*
PCR_B1	*5*'-GGGCAAGCAGTCCTAACAACCATG-*3'*
PCR_B2	*5*'-CTTTCTAGAGAATAGGAACTTCGG-*3*'

Table S3: Amplification conditions for primary and secondary splinkerette-PCRs

Primary Amplification

Step	Temperature	Time	# of cycles
Initial denaturation	96°C	2'00"	1
Denaturation	92°C	0'40''	
Annealing	60°C (-1°C per cycle)	0'40''	10
Elongation	68°C	2'00"	
Denaturation	92°C	0'40''	
Annealing	63°C (-0.5°C per cycle)	0'40''	10
Elongation	68°C	1'00"	
Denaturation	92°C	0'40''	
Annealing	50°C	0'40''	25
Elongation	68°C	1'00"	
Elongation	68°C	10'00"	1

Secondary Amplification

Step	Temperature	Time	# of cycles
Initial denaturation	96°C	2'00"	1
Denaturation	92°C	0'40''	
Annealing	66°C (-1°C per cycle)	0'40''	6
Elongation	68°C	1'00"	
Denaturation	92°C	0'40''	
Annealing	59°C	0'40''	14
Elongation	68°C	1'00"	
Elongation	68°C	10'00"	1

Table S4: Adaptor and barcode sequences incorporated into primers PCR_B1 and PCR_B2 for sequencing splinkerette products using the Ion Torrent.

Adaptor	Sequence	Comment
Y	*5'*-CCTCTCTATGGGCAGTCGGTGAT -3'	Adaptor sequence attached to the 5'end of primer PCR_B1
Z	*5'*- CCATCTCATCCCTGCGTGTCTCCGACTCAGXXX-3'	Adaptor and barcode sequence attached to the 5'end of primer PCR_B2, where XXX refers to the barcodes listed below.

Barcode	Sequence	Barcode	Sequence
1	*5'*-ACGAGTGCGT-*3'*	14	*5'*-TCACGTACTA-*3'*
2	*5'*-ACGCTCGACA-*3'*	15	*5'*-CGTCTAGTAC-*3'*
3	*5'*-AGACGCACTC-*3'*	16	*5'*-TCTACGTAGC-*3'*
4	*5'*-AGCACTGTAG-*3'*	17	*5'*-TGTACTACTC-*3'*
5	*5'*-ATCAGACACG-*3'*	18	*5'*-ACGACTACAG-*3'*
6	*5'*-ATATCGCGAG-*3'*	19	*5'*-CGTAGACTAG-*3'*
7	*5'*-CGTGTCTCTA-*3'*	20	*5'*-TACGAGTATG-*3'*
8	*5'*-CTCGCGTGTC-*3'*	21	*5'*-TACTCTCGTG-*3'*
9	*5'*-TCTCTATGCG-*3'*	22	*5'*-TAGAGACGAG-*3'*
10	*5'*-TGATACGTCT-*3'*	23	*5'*-TCGTCGCTCG-*3'*
11	*5'*-CATAGTAGTG-*3'*	24	*5'*-ACATACGCGT-*3'*
12	*5'*-CGAGAGATAC-*3'*	25	*5'*-ACGCGAGTAT-*3'*
13	*5'*-ATACGACGTA-*3'*		

Table S5: Oligonucleotides used to confirm proviral genomic insert locations.

Gene	Forward Primer	Reverse Primer
RBM19	*5'*-TCTGGTTTCCCTTTCGCTTTCAGGT-*3'*	*5'*- GAGTTTGCATGCTCATCCTGTGG-*3'*
FINP1	*5'*-TCTGGTTTCCCTTTCGCTTTCAGGT-*3'*	*5'*-CTGTGAATTTTGGTTACATTTATG-*3'*
SPEN	*5'*-TCTGGTTTCCCTTTCGCTTTCAGGT-*3'*	*5'*-GCTGATCATCCAGCAAAAGCAG-*3'*
ASH1L	5'-AGCTCGGTACCTTTAAGACCAATGACTT-3'	*5'*-TCTGGTTTCCCTTTCGCTTTCAGGT-*3'*
USP4	*5'*-TCTGGTTTCCCTTTCGCTTTCAGGT-*3'*	*5'*-CAGCCTTGTGGTTTTTCACA-*3'*
COL5A2	*5'*-TCTGGTTTCCCTTTCGCTTTCAGGT-*3'*	*5'*- CATTTGGTTCATATATTTGACTTA-*3'*
PTOV1-AS1	*5'*- CGATGCTCTCGACACTTCC-*3'*	*5'*-TCTGGTTTCCCTTTCGCTTTCAGGT-*3'*
LRP6	*5'*-TCTGGTTTCCCTTTCGCTTTCAGGT-*3'*	*5'*-AAAAATAGCCATCTCCATGAGC-*3'*
MY06	*5'*-CTGTTCGGGCGCCACTGCTAG-*3'*	*5'*-GGAAAAAGTAACCGGGCACAGTGGC-*3'*
BAD	*5'*-CTGTTCGGGCGCCACTGCTAG-*3'*	*5'*-GCTTGTCTCAGTCCAGTTAGGGGTAG-*3'*
GATAD2B	*5'*-CTCCCGGCTGCCTTCCCTTTTT-*3'*	*5'*-CTGTTCGGGCGCCACTGCTAG-*3'*

Table S6: The proviral insert locations identified for the 6 DF-iPSC clones investigated in Figure 1.

Clone	Genome Location	Nearest Gene	Present in
DF1	chr19:503428454	*PTOV1-AS1*	Exon
	chr2:190023157	*COL5A2*	Intron
DF2	chr1:16232602	*SPEN*	Intron
	chr3:49349283	*USP4*	Exon
DF3	chr1: 153863681	*GATAD2B*	Intron
DF4	chr1:155370199	*ASH1L*	Intron
	chr11:64046798	*BAD*	Intron
	chr5:130981785	*FNIP1*	Intron
DF5	chr12:114276140	*RBM19*	Intron
	chr6:76532084	*MYO6*	Intron
DF6	chr12:12290145	*LRP6*	Intron

Table S7: Overview of the proviral insert locations determined by shear-splink and Ion Torrent sequencing for the 20 putative hiPSC clones.

Clone	Genome Location	Nearest Gene	Present in
1	N/A		
2	chr2: 207144375	*ZDBF2*	Exon
3	chr2: 207144375	*ZDBF2*	Exon
4	chr6: 44392477	*CDC5L*	Intron
5	chr12: 28151923	upstream of *PTHLH*	Non-coding region
6	chr1: 150653834	*GOLPH3L*	Intron
7	chr16: 71769591	*AP1G1*	Intron
8	chr1: 1330539	*CCNL2*	Intron
9	chr17: 79562823	*NPLOC4*	Intron
10	chr2: 207144375	*ZDBF2*	Exon
11	chr2: 207144375	*ZDBF2*	Exon
12	chr17: 74311940	*PRPSAP1*	Intron
	chr1: 10660290	*PEX14*	Intron
13	chr2: 207144375	*ZDBF2*	Exon
14	chr2: 42438527	*EML4*	Intron
15	chr19: 1121879	*SBNO2*	Intron
16	N/A		
17	chr5: 116648220	upstream of *CTC-504A5.1*	Non-coding region
18	chr1: 1330539	*CCNL2*	Intron
19	chr16: 71769591	*AP1G1*	Intron
20	chr17: 36589101	*ARHGAP23*	Intron