NIH Public Access
Author Manuscript
Stem Cells. Author manuscript; available in PMC 2015 October 01.

Published in final edited form as:
Stem Cells. 2014 October ; 32(10): 2657–2667. doi:10.1002/stem.1779.

Passage Number is a Major Contributor to Genomic Structural Variations in Mouse iPS Cells

Pengfei Liu[1,**], **Anna Kaplan**[2], **Bo Yuan**[1], **Jacob H. Hanna**[2], **James R. Lupski**[1,3,4], and **Orly Reiner**[2,*]

[1]Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX 77030, USA

[2]Department of Molecular Genetics, Weizmann Institute of Science, 76100 Rehovot, Israel

[3]Department of Pediatrics, Baylor College of Medicine, Houston, TX 77030, USA

[4]Texas Children's Hospital, Houston, TX 77030, USA

Abstract

Emergence of genomic instability is a practical issue in preparing neural stem cells (NSCs) and induced pluripotent stem cells (iPSCs). However, it is still not fully understood what the origins and mechanisms for formation are for the genomic alternations observed. Here, we studied the extent of genomic variation on the scale of individual cells originating from the same animal. We used mouse NSCs grown from embryonic cells and iPSCs generated from embryonic brain cells, B cells or fibroblasts, and performed comparative analysis with cultures of fibroblasts from the same mouse. In the first passage of these cell lines, aneuploidies were only observed for chromosomes 6, 11, 12, 19 and Y, which is overall at a rate lower than previously reported; *de novo* copy number variations (CNVs) were observed in 4.3% of neural iPSCs, 29% of B cell iPSCs, 10% of fibroblast iPSCs, and 1.3% of neurospheres. In contrast, propagation of these first passage cells to a later passage induced additional aneuploidies and CNVs. Breakpoint sequencing analysis suggested that the majority of the detected CNVs arose by replicative mechanisms. Interestingly, we detected identical *de novo* CNVs in different single cell colonies that appeared to have arisen independently from each other, which suggests a novel CNV formation mechanism in these cells. Our findings provide insights into mechanisms of CNV formation during

[*]Corresponding authors: Orly Reiner, Ph.D., Department of Molecular Genetics, Weizmann Institute of Science, 76100 Rehovot, Israel, Tel: +972-8-9344927, Fax: +973-8-9344108, orly.reiner@weizmann.ac.il. [**]Pengfei Liu, Ph.D., Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX 77030, USA, Tel: +1-713-798-3723, Fax: +1-713-798-5073, pengfeil@bcm.edu.

Disclaimer: O.R. is an incumbent of the Bernstein-Mason Professorial Chair of Neurochemistry. J.R.L. holds stock ownership of 23andMe and Ion Torrent Systems and is a coinventor on multiple United States and European patents for DNA diagnostics.

Author contributions:

Pengfei Liu: Conception and design, Collection and/or assembly of data, Data analysis and interpretation, Manuscript writing, Final approval of manuscript

Anna Kaplan: Collection and/or assembly of data, Data analysis and interpretation

Bo Yuan: Collection and/or assembly of data, Data analysis and interpretation

Jacob H. Hanna: Conception and design, Administrative support, Provision of study material or patients, Final approval of manuscript

James R. Lupski: Conception and design, Financial support, Administrative support, Provision of study material or patients, Data analysis and interpretation, Manuscript writing, Final approval of manuscript

Orly Reiner: Conception and design, Financial support, Administrative support, Provision of study material or patients, Data analysis and interpretation, Manuscript writing, Final approval of manuscript

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

reprogramming, and suggest that replicative mechanisms for CNV formation accompany mitotic divisions.

Keywords

DNA copy number variations; chromosomal aberrations; induced pluripotent stem cells; neural stem cells; DNA replication

Introduction

One of the main interests in generating multipotent stem cells from different individuals, various tissues and distinct stages of life lies in the promise of using these cells in regenerative medicine, especially brain diseases, where most of the affected cells are terminally differentiated neurons (reviewed in [1]). However, the findings that neural stem cells (NSCs), embryonic stem cells and induced pluripotent stem cells (iPSCs) may exhibit genetic instabilities raises practical safety concerns regarding the use of these cells in the clinic [2–6]. Mouse studies, which are considered as “gold standard” to compare with human studies, also demonstrated that genomic instability is frequently observed following preparation of iPSCs and stem cells [7–9].

In these scenarios, genetic aberrations were noted at several levels and included different types of genomic changes. One of these studies showed that in human iPSCs more CNVs were detected in earlier than in later passages [6]. However, recent studies conducted in human iPSCs suggested that about half of the genomic alterations could be detected as low level mosaicism in the parental fibroblasts from which the iPSCs were derived [10]. Most of the published studies were conducted on cells that have undergone multiple passages. For example, Martins-Taylor et al. used passage five as an early passage, and more than 90% (21/23) exhibited normal karyotypes [4]. A similar finding was reported by Laurent et al. using passage five to eight as early passage clones [11]. Another study examined the karyotype of more than 1,700 human iPSC and ESC long-term cultures. Approximately 12.5% of either iPSCs or ESCs had abnormal karyotypes [12]. The mere fact that cells are cultured over multiple passages is known to confer genetic/genomic alterations, some of which may lead to advantageous growth (reviewed in [13]).

Investigation of mechanisms for genomic rearrangements may provide insights into timing, location, frequency, and consequence of their emergence. Several studies suggested that replication stress can lead to genomic instability in induced stem cells [6–8], but few focused on precise molecular mechanism, which can be interpreted from rearrangement traces left at breakpoint sequences, for example, the presence of nucleotide sequence microhomology. Using next-generation sequencing, Abyzov et al. analyzed breakpoint junction sequences found in CNVs from iPSCs and concluded that non-homologous end joining (NHEJ) as the underlying mechanism [10]. However, Arlt et al. reported that in mouse embryonic stem cells, CNVs induced by replication stress are not predominantly formed by the NHEJ pathway. This suggests that the alternative pathway, i.e. replicative mechanism or fork stalling and template switching (FoSTeS) [14]/microhomology-mediated

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

break-induced replication (MMBIR) [15] (FoSTeS/MMBIR), may play a more prominent role [16].

We set out to pinpoint the critical factor(s) affecting gene dosage variation in cultured iPSCs. Our study focused on the analysis of mouse NSCs grown from a single embryonic cell and iPSCs generated from single embryonic brain cells, B cells or fibroblasts originating from the same animal. DNA was extracted from the cells at passage one and copy number variations (CNVs) were analyzed on a mouse genome-wide aCGH platform in comparison to DNA derived from fibroblasts of the same embryo (Figure 1). This approach successfully revealed aneuploidies and CNVs specific to single cells present in the tested material at an early passage. The recorded events of genomic instability varied according to the cell type. NSCs and iPSCs derived from neurons as well as fibroblasts exhibited fewer CNVs in comparison to iPSCs derived from B cells, but the overall detection of genomic instability in these early passage clones is lower than anticipated. A number of the iPSCs and neurospheres were reexamined at later passages. New changes appeared, and none of the previously noted changes disappeared except for two clones showing aneuploidy of chromosome Y, which was not evident at later passages. We analyzed breakpoint sequences at the nucleotide level, which enabled us to surmise that a majority of these events occurred through replicative mechanisms such as FoSTeS/MMBIR. Breakpoint sequence analysis also unexpectedly revealed a few identical CNVs detected in different iPS single cell expansions derived from the same embryo. These CNVs appear to be independent *de novo* events, but their recurrences cannot be explained by known recurrent genomic rearrangement mechanisms, suggesting that the loci of these mutational events were determined early in the development in the parental cells. Our studies suggest that cells derived from different tissues exhibit different degrees of genomic instability during reprogramming as well as cell propagation following reprogramming, and that passage number rather than the reprogramming process *per se* is the major factor underlying genomic instability. Our approach of CNV analysis using clones derived from single cells reveals a novel perspective regarding the process of *de novo* CNV generation.

Materials and Methods

All animal studies were approved by Weizmann Institute of Science ethics committee and carried out in accordance with the IACUC international guidelines. The source for iPSCs was from GFP-positive E13.5 embryos derived from F1 matings between ROSA26-M2rtTA mice [17] and Tg(Emx1-EGFP)FJ56Gsat/Mmucd obtained from the Jackson laboratories (Bar Harbor, ME).

Genotyping of ROSA26-M2rtTA mice

Two loci (Col1a1 and Rosa26) were checked by PCR with 3 primers for each loci: The primer sequences are: Col1a1-frtA-F, 5′GCACAGCATTGCGGACATGC 3′; Col1a1-frtA-R, 5′CCCTCCATGTGTGACCAAGG 3′; Col1a1-4F2A-R, 5′TTGCTCAGCGGTGCTGTCCA3′; Rosa26-A-F, 5′AAAGTCGCTCTGAGTTGTTAT3′; Rosa26-B-F, 5′GCGAAGAGTTTGTCCTCAACC3′; Rosa26-C-R, 5′GGAGCGGGAGAAATGGATATG3′. Primers for genotyping of mouse embryos for

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

EMX1-GFP: GFP-F, 5′TCACTCCGCTTCGGCGGCC3′; GFP-R, 5′TAGCGGCTGAAGCACTGCA3′. Primers for sex determination of mouse embryos: Sry-F, 5′TGGGACTGGTGACAATTGTC3′; Sry-R, 5′GAGTACAGGTGTGCAGCTCT3′; Control IL3-F, 5′GGGACTCCAAGCTTCAATCA3′; Control IL3-R, 5′TGGAGGAGGAAGAAAAGCAA3′.

Tissue culture media

Embryos were dissected in cold Leibovitz L15 medium supplied with 50μg/ml gentamicin (both from Biological Industries (Beit-Haemek, Israel)), 0.6% glucose and supplemented with oxygen. iPSCs were cultured on irradiated MEFs in ES medium (DMEM containing 15% FCS, 60μl leukemia inhibiting factor (LIF) (Millipore, Billerica, MA), 1mM sodium pyruvate, nonessential amino acids (1:100), 0.1mg/ml penicillin/streptomycin and 200mM L-glutamine purchased from Biological Industries (Beit-Haemek, Israel) and 1.2ml beta-mercaptoethanol from Gibco (Grand Island, NY). 2mg/ml of doxycycline (Dox) (Sigma, Rehovot, Israel) were used for induction of four reprogramming factors. MEFs were cultured in EF medium (DMEM containing 10% FCS, 1mM sodium pyruvate, 0.1mg/ml penicillin/streptomycin and 200mM L-glutamine purchased from Biological Industries (Beit-Haemek, Israel)). Cortical neurons were cultured in MEM medium (Sigma, Rehovot, Israel) containing 5% heat-inactivated horse serum, 5% fetal calf serum, 1 μl/ml B-27 supplement and 2 mM glutamax purchased from Gibco (Grand Island, NY), enriched with 0.6% glucose and supplemented with 20 μg/ml gentamicin (Biological Industries, Beit-Haemek, Israel). Neurospheres were prepared from isolated embryonic cortices and cultured in NB medium (Neurobasal media containing 2% of B27 (both from Gibco (Grand Island, NY)), 20 ng/ml Heparin (STEMCELL Technologies, Vancouver, Canada) 20 ng/ml human Epidermal Growth Factor (EGF) and 20 ng/ml basic Fibroblast Growth Factor (FGF-b) (Millipore, Billerica, MA) and 250 μM L-glutamine (Biological Industries, Beit-Haemek, Israel)).

Preparation of neurospheres

Neurosphere cultures were prepared from hippocampi of E14.5 mouse embryos as described [18]. Briefly, hippocampi were dissociated and resuspended in NB medium. After 7 days in culture, single neurospheres were picked and transferred to 96-well dish and neuronal progenitor cells (NPCs) were further propagated. Once neurospheres became confluent in duplicate 35mm dishes: cells from one dish were frozen in liquid N2 and from another dish neurospheres were subject for DNA extraction using DNeasy blood and tissue from Qiagen (Valencia, CA).

Preparation of MEFs

For MEF isolation, the head and internal organs were removed, and the remaining tissues were physically dissociated and incubated in trypsin at 37°C for 30 min after which cells were filtered through a sterile 70 μm cell strainer and resuspended in EF medium. 72 hours later, MEFs were expanded and re-plated either in EF medium for control DNA extraction or in Dox-containing ES medium (2 mg/ml) on gelatin–coated dishes for reprogramming experiments. Single iPS cell colonies were observed on day 21–35 after introducing Dox.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Preparation of cortical neurons

Primary cortical neuronal cultures were prepared from mouse E14.5 brains. Briefly, both cortices were mechanically dissociated and plated on poly-l–lysine/laminin coated dishes in MEM medium. After 7 days in culture, the plating medium was changed to Dox-containing ES medium (2 mg/ml). For neurons the originating iPS cell colonies were observed on day 21–35 after introducing Dox.

Preparation of B cell progenitors

Fetal liver from E14.5 embryos was used as a source of lymphoid progenitors [19]. The embryonic liver was dissected away from surrounding tissues and dissociated in ES medium followed by passing through a sterile 70 μm cell strainer. Liver cell preparations were plated on poly-l–lysine coated dishes in ES medium supplemented with 2mg/ml Dox. For lymphoid progenitors the reprogrammed iPS cell colonies were observed on day 14–20 after introducing Dox.

Mouse iPSCs preparation

Following Dox application within 25 to 40 days single iPS cell colonies reached a proper size to be manually picked up from their original dish by 20μl pipet. They were thoroughly dissociated by incubation for 10 min at 37°C in (1x) trypsin (Biological Industries, Beit-Haemek, Israel), followed by careful re-suspension and further re-plating in 96 wells dish on feeder cells. When iPSC cultures reached confluency while grown in 12-well dishes in duplicates: 1) cells from one well were frozen in liquid N2 and 2) from another well the cells were re-plated on gelatin-coated 35 mm dish and once confluent, DNA was extracted using DNeasy blood and tissue (Qiagen, Valencia, CA).

Immunostaining

iPSCs were plated on feeder cells-coated cover-slips and maintained in ES medium supplemented with doxycycline until confluent, fixed with 3% PFA for 20min at R.T., washed 3 times with PBS and permeabilized with 0.1 % Triton X-100 in PBS for 10min. After blocking with 0.1% BSA in PBS and incubation with primary antibodies to Oct 3/4 (Santa Cruz, San Diego, CA) and Nanog (Bethyl A300–398A) for 1 h in 0.1% BSA in PBS, cells were washed 3 times with PBS and incubated with fluorophore-labeled appropriate secondary antibodies from Jackson Immunoresearch (West Grove, PA, USA). Immunostained coverslips were visualized using wide-field microscopy (DeltaVision, Applied Precision, CA, USA). An example of immunostained iPSCs are shown in Figure S1.

Array CGH analysis

DNA extracted from iPSCs or neurospheres were comparatively hybridized with control DNA extracted from embryonic fibroblast DNA derived from the same mouse. The microarray analysis was initially performed on a genome-wide Agilent 8×60k array platform. DNA digestion, labeling, clean up, hybridization, slide wash, scanning and feature extraction were performed according to the manufacturer's instructions. The extracted raw data were analyzed using the Agilent Genomic Workbench software. CNVs were called

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

using the threshold of ADM2 4.0. Raw CNV calls were reviewed and filtered manually to eliminate low confidence calls. The filtered candidate CNV regions were collated for a follow up region-specific high-density aCGH design, which serves as a validation to the genome-wide array result and an approach to better define breakpoint intervals. The high-density design has probe spacing ranging from 100 bp to 500 bp for the targeted regions. After the high-density array analysis, the validated CNV regions were recorded and breakpoint PCRs were attempted for each of the region. In the region-specific array, in addition to the candidate CNV, smaller CNVs are occasionally identified in flanking regions. These findings are considered to be “incidental” because their identification was biased towards the regions interrogated in the validation array. Therefore, they were not included in any of the analysis in this work.

Breakpoint analysis

Based on the CNV boundary defined by the high-density array results, PCR assays were designed to amplify the breakpoint junctions. The primer design was based on the hypothesis of simple deletion, tandem duplication, or complex rearrangement when applicable. The PCR experiment was carried out using the Takara LA DNA polymerase and the manufacturer’s protocol. When PCR result is positive, the same assay is performed using the fibroblast DNA as a template to test the hypothesis that the same breakpoint junction is present in the somatic cells of the individual as a mosaic state. All PCR products were subsequently sequenced by the Sanger dideoxy method. The breakpoint sequences were aligned to reference sequences at proximal and distal breakpoints to infer potential mechanisms.

Results

Aneuploidy in NSCs and in iPSCs

Although it is known that aneuploidy can be observed in iPSCs [5], it is still not fully understood whether aneuploidies tend to occur during reprogramming and the following early divisions, or they tend to accumulate over the propagation process. To evaluate the abundance of aneuploidy in early cell divisions following reprogramming, we isolated single iPSC expansions generated from three different tissue sources from embryos, and performed aCGH on DNA extracted on the first passage of these iPSCs, which is considered as an early passage number. Oligonucleotide aCGH can readily detect whole chromosome aneuploidy even when present in a relatively low mosaic state (as low as 10%) [20]. We analyzed 71 iPSCs, including 23 neural iPSCs, 38 B cell iPSCs, and 10 fibroblast iPSCs. We identified aneuploidies for chromosomes 6 (1.4%, N=1), 11 (16.9%, N=12), 12 (5.63%, N=4), 19 (1.41%, N=1) and Y (7.04%, N=5) (Table S1). All of the aneuploidies observed were gains except for chromosome Y losses in four cases. These changes can be either *de novo* events induced by the reprogramming process or expansions from existing somatic mosaicism in the originating tissue.

In addition to iPSCs, we also assayed the degree of aneuploidy in the first passage of single cell colonies of neurosphere stem cells. None of the clones (N=76) exhibited any aneuploidy, suggesting that the genomes of the neurospheres are relatively stable and the

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

process of neurosphere single cell preparation may eliminate cells with chromosomal abnormalities present in the source tissue at a mosaic state.

The overall load of aneuploidy is lower than anticipated. For example, Rehen et al. investigated somatic aneuploidy in neurons using neuroblasts in the embryonic mouse brain, and reported that aneuploidy was observed as losses for all chromosomes at rates of 1.6–8.4% and as gains for all chromosomes except for 10, 16 and 18 at rates of ~2% [21]. In our neural iPSCs and neurospheres, we did not observe any aneuploid losses except for chromosome Y losses. For gains, the only chromosome with a higher rate of aneuploidy in our neural iPSC is chromosome 11 (30% vs. 0.9% from Rehen et al.), which is consistent with a previous report that trisomy 11 is one of the most common changes in both embryonic stem cells (7.8%) and iPSCs (10.2%) [7], suggesting a high prevalence of this mosaic change in the parental tissue or a growth advantage conferred by this change. Of note, we found that chromosome 11 gain is only prevalent in neural and fibroblast iPSCs but not in B cell iPSCs, suggesting tissue specific differences for this particular aneuploidy. We do not think the higher load of aneuploidy from Rehen et al is contradictory to our data because a large proportion of the aneuploid neurons may be selected against during the reprogramming and single cell selection process.

In a previous study of various mouse iPSCs cultured to high passage numbers, the most frequent aneuploidy was chromosome 8 gain [7]. However, this change is not observed in our iPSCs or neurospheres. In human neural iPSC studies, high rates of chromosomal losses and gains were also observed [22]. Therefore, these lines of evidence implicate that although chromosome level changes are not prominent in early passages of our iPSCs and neuropheres, additional instability may emerge in the following passages.

CNVs in NSCs and in iPSCs

Compared to aneuploidies, CNVs can contribute to genetic variation with a higher diversity and complexity. Investigating molecular mechanisms and characteristics of CNVs in iPSC and NSCs may provide insight into the timing and the process of their formation. Our genome-wide aCGH design has a median probe spacing of 33 kb. The median minimum CNV detection size is ~ 132 kb. All the potential CNVs reported from the genome-wide aCGH were validated by a custom designed high-density region specific aCGH. CNVs were detected in 4.3% (1/23) of neural iPSCs, 29% (11/38) of B cell iPSCs, 10% (1/10) of fibroblast iPSCs, and 3.9% (3/76) of neurospheres (Figure 2, Tables 1 and S1). The number of CNVs found in one sample ranged from 0 to 3. These CNVs included six deletions, eleven duplications, and three complex rearrangements. The genomic intervals affected ranged in size from 32 kb to 3.3 Mb and contained 0 to 31 genes per interval (0 to 35 when including CNVs arising after passage 1) (Table 2).

Most CNVs seemed to be scattered throughout the genome. The overall low number of CNVs identified in our study limited our ability to investigate potential regions of “CNV hotspot”. However, in two neurosphere clones derived from the same mouse, the same genomic region and thus the same six genes contained (*4930528P14Rik, 3110099E03Rik, BC052040, Mir1951, Meis2, and 2810405F15Rik*) are affected. The two CNVs are likely independent from each other because one is a deletion (21A) and the other is a duplication

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

(5A-1), and the breakpoints are grossly different (Table 2). This suggests that the locus may represent a region of instability in the mouse genome under our experimental or natural condition; alternatively, these CNVs may be selected because the genes affected are important for mouse neuronal differentiation. In an iPSC derived from a B cell, a CNV was identified on chromosome 5 affecting the *Auts2* gene. This gene resides in a hotspot with the most frequently occurring CNVs under aphidicolin-induced replication stress demonstrated in mouse ES cells [16]; the syntenic region is also a CNV mutation hotspot in human cells [23]. In two other loci, apparent identical CNVs are detected in different clones (G5, G30; G10-1, G23). The implications of CNV recurring in these two loci will be further discussed in a following section.

We PCR amplified and Sanger sequenced breakpoint junctions whenever possible. If the same breakpoint junction can be amplified from the MEF DNA extracted from the same embryo, the CNV was classified as a somatic mosaic event rather than a new mutation from reprogramming or culturing. Two of the three CNVs from neurospheres were classified as somatic mosaic events, whereas none of the CNVs from iPSCs in which breakpoint junction PCR was successful (N=10) fell into this category (Figure S2). Overall, we observed much lower frequency of somatic mosaic CNV events (0/10 for iPSCs or 2/13 for iPSCs plus neurospheres) than Abyzov et al (10/20 iPSCs) [10]. After excluding the somatic mosaic events, the percentage of CNV positive clones reduced to 1.3% for neurospheres, and remained the same for the other cell types (Table 1).

Breakpoint sequences were analyzed at the nucleotide level in 13 CNVs, including one from neural iPSC, eight from B cell iPSC, one from fibroblast iPSC, and three from neurospheres (Figure 3). Microhomology is a prevailing feature exhibited at these breakpoints, with seven breakpoints showing 2–4 bp and three breakpoints showing 1 bp microhomology. Two samples had templated insertions at the breakpoints, one of which had inserted sequence likely derived from part of an L1-LINE element (G19). These characteristics suggest that replicative mechanisms such as FoSTeS/MMBIR are the underlying mechanism. In addition, one deletion had short novel sequence inserted at the breakpoint (7D), which is indicative of an alternative mechanism potentially involving NHEJ in formation of this particular CNV.

Genomic Instability following additional passages

The low abundance of genomic changes observed in our samples led to the hypothesis that additional cell passages may result in more dramatic genomic instability. Therefore, five neuronal derived iPSC clones and three B cell derived iPSCs were propagated to six and twelve passages, four neurospheres cultures to five passages, in order to measure and compare their level of genomic instability with regards to passage one (Table 3). CNVs and aneuploidies existing in the beginning of the culturing (1st passage) largely remained stable, except that chromosome Y aneuploidy disappeared in one B cell iPSC and one neural iPSC. In contrast, new aneuploidies appeared in 2/4 neurosphere cells (Figures S3A–B), with one of the two being chromosome 8 aneuploidy, the most frequently observed chromosomal change in mouse NSCs [7]. New CNVs appeared at a later passage in 2/3 B cell iPSC (Figures S3C–F, Table 2), 1/5 neural iPSC (Figure S3G, Table 2). In particular, most of the

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

clones with new CNVs exhibited more than one event. One of the clones showed three new duplications on different chromosomes (Figure S3C–E), and in another clone two deletions in proximity with each other were apparent (Figure S3F). Of note, the new CNV detected in F29-P12 (Figure S3F) maps to the mouse CNV hotspot region at *Auts2*, partially overlapping with the CNV identified in F31 (Table 2 and Figure 4A).

These new CNVs obviously arose *de novo* during culturing (the fact that all clones were cultured from a single cell and that passage one cultures showed no CNV aberrations in these loci excludes the possibility that these CNVs were somatic mosaic events passed from parental tissues), and are therefore expected to be at least partially mosaic in the analyzed individual DNA samples. However, only two events appeared to be mosaic as evidenced by the $\log_2$ ratios of aCGH plots (Figures S3C and E). Therefore, it is probable that the other CNVs may have apparently reached relatively high level of homogeneity, possibly due to a conferred growth advantage.

De novo CNVs arising independently at the same locus from different single cell expansions

Among all the CNVs identified in this study, a few identical events were detected among the different clones. Surprisingly, evidence suggests that these events were not derived from a common ancestor clone, but rather arose independent from each other.

The first example came from two B cell iPSC clones of embryo 4. A deletion of ~275kb was identified in clone F31 on chromosome 5. Breakpoint PCR did not amplify the same breakpoint junction from fibroblast DNA isolated from the same embryo, suggesting that the deletion is either likely a *de novo* event that was produced during the reprogramming process or a somatic mutation occurring at the level of individual B cells. In another B cell iPSC clone, F29, a deletion with the same breakpoint was found in passage 12. PCR amplification of the breakpoint junction was negative in the DNA isolated from passages 1 and 6, strongly indicating that the deletion was a *de novo* event that occurred between the passages 6 and 12 (Figure 4A). Additionally, a small deletion was identified distal to this 275kb deletion in F29-P12, likely as part of a complex rearrangement event. This further strengthens the argument that the deletions in F29-P12 and F31 are not of the same origin.

Similarly, identical duplications of 1.4 Mb in size on chromosome 6 were found in passage 1 of F8 and passages 6 and 12 of F3, but was neither present in the fibroblast DNA of the same embryo nor in passage 1 of F3 (Figure 4B), consistent with the events being new mutations. Notably, F8 was prepared from neuron cells whereas F3 was prepared from embryonic liver cells; plus, the presence of two other new CNVs found in passages 6 and 12 of F3 excludes the possibility of F3-P6 and F8 being the same clone.

At two other loci, identical CNVs were found in two B cell iPSC clones at passage 1 for each locus (Figure 4C–D). In one of the loci, breakpoint junction PCR showed that the rearrangement was not present as a mosaic state in the fibroblast DNA (Figure 4D). This again suggests that the CNVs in the two samples can potentially be new mutation events associated with reprogramming, although for this locus, unlike the two loci mentioned above (the deletion on chromosome 5 and the duplication on chromosome 6), we cannot rule out

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

the alternative possibility that the CNV is mosaic in the somatic blood cells, i.e. the mutation occurred early in development and was passed onto the B cell lineage but not the fibroblast lineage. In the other locus, the breakpoint junction could not be amplified, probably due to breakpoint complexities. Therefore, the timing of the mutation cannot be inferred.

Breakpoint sequence analysis revealed that these apparent independent *de novo* events involve the same genomic region at the nucleotide sequence level (Figures 4A, B, and D). Further examination of the genomic structure surrounding these changes revealed an absence of known features in the genome architecture (e.g. inverted or direct repeats) that may explain the recurrence of these genomic instabilities.

Discussion

In this work, we explored the genomic instability in iPSC and neurosphere clones derived from mouse single cells. Our finding of aneuploidy in neurospheres and neural iPSC revealed a low degree of chromosomal abnormality in the first passage following preparation of iPSCs or neurospheres; genomic alterations accumulated at later passages; CNVs that appeared prior to passage one stayed stable upon further cell propagation (at least in the cells we assayed). Since the detection of CNV is heavily dependent on the chemistry and sensitivity of the technology used, we focused our comparison with other studies for the load of genomic instability only on the level of chromosome aneuploidy. We detected low level of chromosome-level instability in comparison with those previously reported in neuroblasts and differentiated neurons [21]. That study also observed that cultured neuroblasts reveal less aneuploidy than fresh ones. In a study examining early passage mouse iPSCs, Quinlan et al. observed only four structural variations in three lines [24]. Recently, McConnell et al. used a single cell approach to study large CNVs in neurons obtained from human iPSCs [22]. Genomic changes were observed in 27 out of 40 neurons studied; in contrast, the matching human iPSC-derived neural progenitor cells showed no obvious CNV aberration. This is consistent with our finding that the majority of genomic instabilities observed in neurons does not seem to be explained by reprogramming, but rather cell propagation or neuronal differentiation. In a previous investigation performed in human iPSCs, Hussein et al showed that CNVs tend to diminish during cell propagation [6]. The opposite trend observed in our study could reflect influence from species differences. Mouse embryonic stem cells are typically in a naive state, while human iPSC lines correspond to a primed state, which might affect cell properties and result in a heavier CNV load for low passage number cells in human. With a more rapid cell cycle and more robust growth in culture than human cells, mouse cells may have different CNV adaptation pattern from human cells. Additionally, our approach requires that the cells with mutation survive as a single cell, which may create a selective disadvantage of cells harboring mutations that compromise proliferation, therefore resulting in their underrepresentation in low passage cells.

We observed microhomologies, insertions, small complexities at rearrangement breakpoints. These features are reminiscent of rearrangements by FoSTeS/MMBIR, suggesting replicative mechanisms contributing to the majority of induced CNVs in iPSCs. Previous findings revealed that CNVs enrich in common fragile sites in human and mouse iPSCs [6–

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

8]. Furthermore, formation of *de novo* CNVs is not affected by depriving the mouse embryonic cells of the NHEJ repair pathway, a rearrangement mechanism alternative to FoSTeS/MMBIR [16]. Collectively, these pieces of evidence substantiate the hypothesis of replicative stress being triggered by reprogramming and new CNVs arising during subsequent mitosis.

With regards to distribution of CNVs, though we reported relatively low number of CNVs, we detected CNVs at the most frequently mutated mouse hotspot, the *Auts2* locus, in two separate samples. Arlt et al. reported that emergence of CNV at the *Auts2* locus can be induced by replication stress [16, 23]. Interestingly, one of our CNV was not present in cells from passages one and six, but only appeared in passage twelve; the emergence potentially associates with more cell division and DNA replication. These observations, together with our findings of breakpoint junction features reminiscent of replicative repair, further illustrate that replicative mechanism is potentially a major contributor to new CNVs found in iPSCs and ES cells. When comparing results from our mouse cells with that of human cells by Hussein et al. [6], we do not find overlap of syntenic regions, which could be explained by differences in species[16], technology, or cohort size.

It has been increasingly realized in both mouse and human that mosaicism, or inter -tissue genomic variations within an individual, contribute significantly to normal development and diseases [25, 26]. In this study, breakpoint PCR revealed a much lower percentage of somatic mosaic events compared to the reported frequency [10]. Although this discrepancy could be due to the different species analyzed (human vs. mouse) or technology used (next-generation sequencing vs. oligonucleotide aCGH), an equally possible explanation is the size differences in the ascertained CNVs. Even though our genome-wide array can detect CNVs as small as 33 kb (F33 in Figure 2) in genomic regions well interrogated by the array, in general, we focused on larger CNVs (> ~130 kb) because of the higher confidence for detecting these with our experimental design and arrays. Among the smaller CNVs (< 130 kb) reported by Abyzov et al., 67% were reported as somatic mosaicisms, whereas only 25% of the larger CNVs (> 130 kb) were somatic mosacisms [10]. These data, together with the limitations of the experimental design, may suggest that reprogramming tends to contribute to larger *de novo* CNVs. Investigation into smaller sized CNVs is warranted to confirm this hypothesis.

Surprisingly, we observed apparently identical CNVs arising independently in different clones. The *de novo* nature of these changes is demonstrated by their emergence in later passages, or absence in the somatic DNA of the same embryo. These identical CNVs were unanimously found in clones from the same individual. In other words, the “recurrence” is limited to be within an individual. So far, the only known intrinsic DNA sequence feature that can result in susceptibility to structural mutation recurring at the same nucleotide breakpoint is low-copy repeat (LCR) [27]; however, no LCRs exist in proximity to these CNV breakpoints in the mouse genome. Taken together, the identical CNVs seem to be occurring at a predetermined locus, and certain novel genetic or non-genetic features may be predisposing these loci to genomic instability in a precise manner. One possible predisposing factor may come from epigenetic modification to these loci. Recently, Lu et al. showed in human iPSCs that reprogramming lead to changes in CNV distribution associated

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

with replication-timing reshaping, suggesting that epigenetic remodeling is one of the consequences of reprogramming [28]. These epigenetic modifications imposed by reprogramming may destine a certain genomic locus to rearrangement. Interestingly, similar CNVs were reported involving the recurrent human *PAR1* deletion, wherein no LCRs exist flanking the deletion but the deletion occurs at the same nucleotide position in different individuals [29]. Further investigations are required to reveal the underlying mechanism contributing to these types of recurrent events not mediated by repeat sequences.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Our research has been supported in part by the Israel Science Foundation (grant no. 47/10 and 322/13), Minerva foundation with funding from the Federal German Ministry for Education and Research, a grant from the Chief Scientist Office at the Israeli Ministry of Health, under the frame of ERA-Net NEURON (DISCover, IMOS 3-00000-6785), the Benoziyo Center for Neurological diseases, the Kekst Family Center for Medical Genetics and the David and Fela Shapell Family Center for Genetic Disorders Research (to O.R.), National Institute of Neurological Disorders and Stroke (National Institutes of Health) grant R01NS058529., Texas Children's Hospital General Clinical Research Center grant M01RR00188, and Intellectual and Developmental Disabilities Research Centers grant P30HD024064 (to J.R.L). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

References

1. Marchetto MC, Brennand KJ, Boyer LF, et al. Induced pluripotent stem cells (iPSCs) and neurological disease modeling: progress and promises. Human molecular genetics. 2011; 20:R109–115. [PubMed: 21828073]
2. Baker DE, Harrison NJ, Maltby E, et al. Adaptation to culture of human embryonic stem cells and oncogenesis in vivo. Nat Biotechnol. 2007; 25:207–215. [PubMed: 17287758]
3. Ben-David U, Mayshar Y, Benvenisty N. Large-Scale Analysis Reveals Acquisition of Lineage-Specific Chromosomal Aberrations in Human Adult Stem Cells. Cell Stem Cell. 2011; 9:97–102. [PubMed: 21816361]
4. Martins-Taylor K, Nisler BS, Taapken SM, et al. Recurrent copy number variations in human induced pluripotent stem cells. Nature biotechnology. 2011; 29:488–491.
5. Mayshar Y, Ben-David U, Lavon N, et al. Identification and classification of chromosomal aberrations in human induced pluripotent stem cells. Cell Stem Cell. 2010; 7:521–531. [PubMed: 20887957]
6. Hussein SM, Batada NN, Vuoristo S, et al. Copy number variation and selection during reprogramming to pluripotency. Nature. 2011; 471:58–62. [PubMed: 21368824]
7. Ben-David U, Benvenisty N. High prevalence of evolutionarily conserved and species-specific genomic aberrations in mouse pluripotent stem cells. Stem cells. 2012; 30:612–622. [PubMed: 22328490]
8. Pasi CE, Dereli-Oz A, Negrini S, et al. Genomic instability in induced stem cells. Cell death and differentiation. 2011; 18:745–753. [PubMed: 21311564]
9. Chen Y, Guo L, Chen J, et al. Genome-wide CNV analysis in mouse induced pluripotent stem cells reveals dosage effect of pluripotent factors on genome integrity. BMC genomics. 2014; 15:79. [PubMed: 24472662]
10. Abyzov A, Mariani J, Palejev D, et al. Somatic copy number mosaicism in human skin revealed by induced pluripotent stem cells. Nature. 2012; 492:438–442. [PubMed: 23160490]
11. Laurent LC, Ulitsky I, Slavin I, et al. Dynamic changes in the copy number of pluripotency and cell proliferation genes in human ESCs and iPSCs during reprogramming and time in culture. Cell Stem Cell. 2011; 8:106–118. [PubMed: 21211785]

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

12. Taapken SM, Nisler BS, Newton MA, et al. Karotypic abnormalities in human induced pluripotent stem cells and embryonic stem cells. Nature biotechnology. 2011; 29:313–314.
13. Harrison NJ. Genetic instability in neural stem cells: an inconvenient truth? J Clin Invest. 2012; 122:484–486. [PubMed: 22269327]
14. Lee JA, Carvalho CM, Lupski JR. A DNA replication mechanism for generating nonrecurrent rearrangements associated with genomic disorders. Cell. 2007; 131:1235–1247. [PubMed: 18160035]
15. Hastings PJ, Ira G, Lupski JR. A microhomology-mediated break-induced replication model for the origin of human copy number variation. PLoS Genet. 2009; 5:e1000327. [PubMed: 19180184]
16. Arlt MF, Rajendran S, Birkeland SR, et al. De novo CNV formation in mouse embryonic stem cells occurs in the absence of Xrcc4-dependent nonhomologous end joining. PLoS genetics. 2012; 8:e1002981. [PubMed: 23028374]
17. Carey BW, Markoulaki S, Beard C, et al. Single-gene transgenic mouse strains for reprogramming adult somatic cells. Nature methods. 2010; 7:56–59. [PubMed: 20010831]
18. Chojnacki A, Weiss S. Production of neurons, astrocytes and oligodendrocytes from mammalian CNS stem cells. Nat Protoc. 2008; 3:935–940. [PubMed: 18536641]
19. Kantor AB, Herzenberg LA. Origin of murine B cell lineages. Annual review of immunology. 1993; 11:501–538.
20. Biesecker LG, Spinner NB. A genomic view of mosaicism and human disease. Nature reviews Genetics. 2013; 14:307–320.
21. Rehen SK, McConnell MJ, Kaushal D, et al. Chromosomal variation in neurons of the developing and adult mammalian nervous system. Proceedings of the National Academy of Sciences of the United States of America. 2001; 98:13361–13366. [PubMed: 11698687]
22. McConnell MJ, Lindberg MR, Brennand KJ, et al. Mosaic copy number variation in human neurons. Science. 2013; 342:632–637. [PubMed: 24179226]
23. Arlt MF, Ozdemir AC, Birkeland SR, et al. Hydroxyurea induces de novo copy number variants in human cells. Proceedings of the National Academy of Sciences of the United States of America. 2011; 108:17360–17365. [PubMed: 21987784]
24. Quinlan AR, Boland MJ, Leibowitz ML, et al. Genome sequencing of mouse induced pluripotent stem cells reveals retroelement stability and infrequent DNA rearrangement during reprogramming. Cell Stem Cell. 2011; 9:366–373. [PubMed: 21982236]
25. Lupski JR. Genetics. Genome mosaicism--one human, multiple genomes. Science. 2013; 341:358–359. [PubMed: 23888031]
26. Liang Q, Conte N, Skarnes WC, et al. Extensive genomic copy number variation in embryonic stem cells. Proc Natl Acad Sci U S A. 2008; 105:17453–17456. [PubMed: 18988746]
27. Dittwald P, Gambin T, Szafranski P, et al. NAHR-mediated copy-number variants in a clinical population: Mechanistic insights into both genomic disorders and Mendelizing traits. Genome Res. 2013; 23:1395–1409. [PubMed: 23657883]
28. Lu J, Li H, Hu M, et al. The Distribution of Genomic Variations in Human iPSCs Is Related to Replication-Timing Reorganization during Reprogramming. Cell reports. 2014; 7:70–78. [PubMed: 24685138]
29. Benito-Sanz S, Royo JL, Barroso E, et al. Identification of the first recurrent PAR1 deletion in Leri-Weill dyschondrosteosis and idiopathic short stature reveals the presence of a novel SHOX enhancer. Journal of medical genetics. 2012; 49:442–450. [PubMed: 22791839]

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

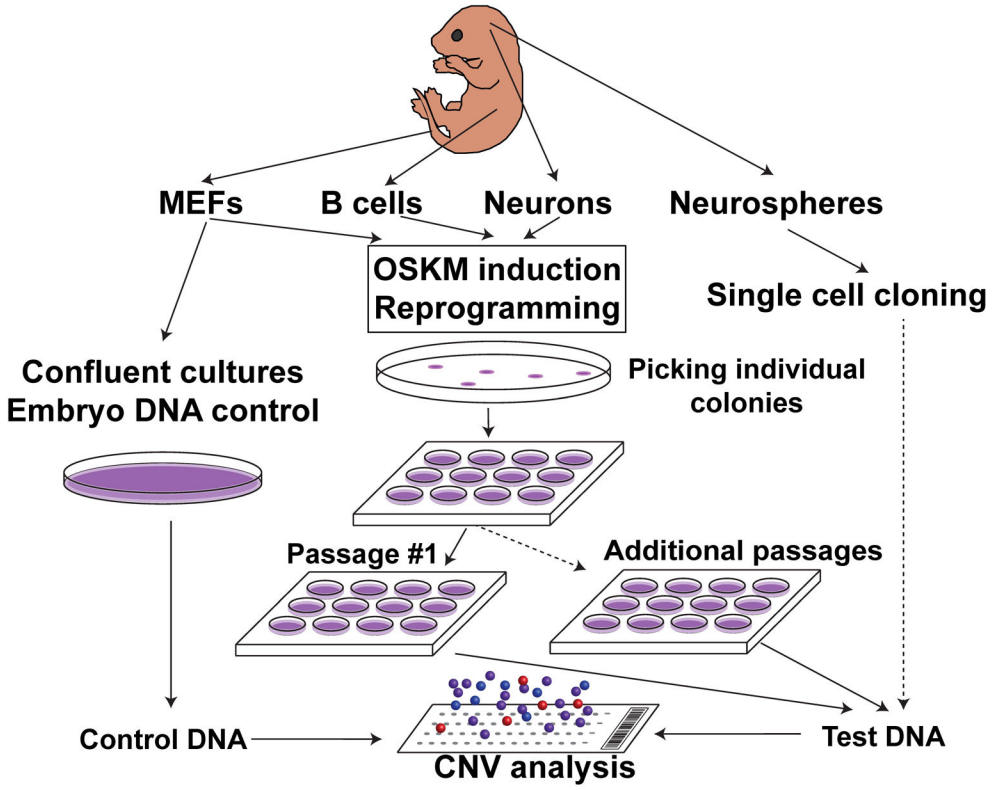

Figure 1.
Schematic flow chart illustrating the procedure of experiments in this study.

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

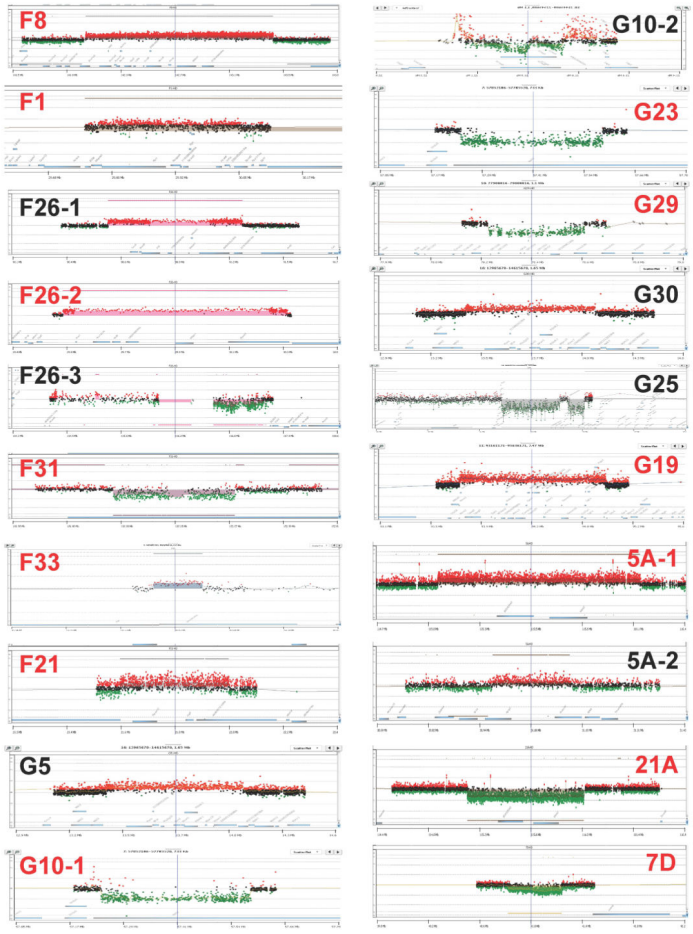

Figure 2. High-density aCGH results for CNVs identified

Samples whose breakpoint sequence was characterized at the nucleotide level are highlighted in red. The breakpoint sequences as well as junction PCR results can be found in Figures 3 and S2. The sample type and coordinates (NCBI37/mm9) of regions affected are listed Table 2. Dash one in the name (for example, F26-1) indicates that this is the first CNV in this sample; dash two indicates the second CNV, etc. Note that G10-1 has the same breakpoint sequence as G23; G5 and G10 have apparently the same breakpoint according to aCGH.

Figure 3. Breakpoint junction sequences characterized

The breakpoint sequences are aligned to reference sequences at the distal and proximal ends of the rearrangement. The red to blue color transition indicates the precise locus of the breakpoint junction. Microhomology is highlighted in purple. Seven breakpoints (F8, F26-2, F31, F33, F21, G29, and 21A) showed microhomology of 2–4 base pairs. Three breakpoints (G10-1, G23, and 5A-1) had one base pair microhomology. Two samples, F1 and G19, had templated insertions at the breakpoints. The inserted sequence in G19, highlighted in green, was likely derived from part of an L1-LINE element. The deletion in 7D had short novel sequence inserted at the breakpoint.

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

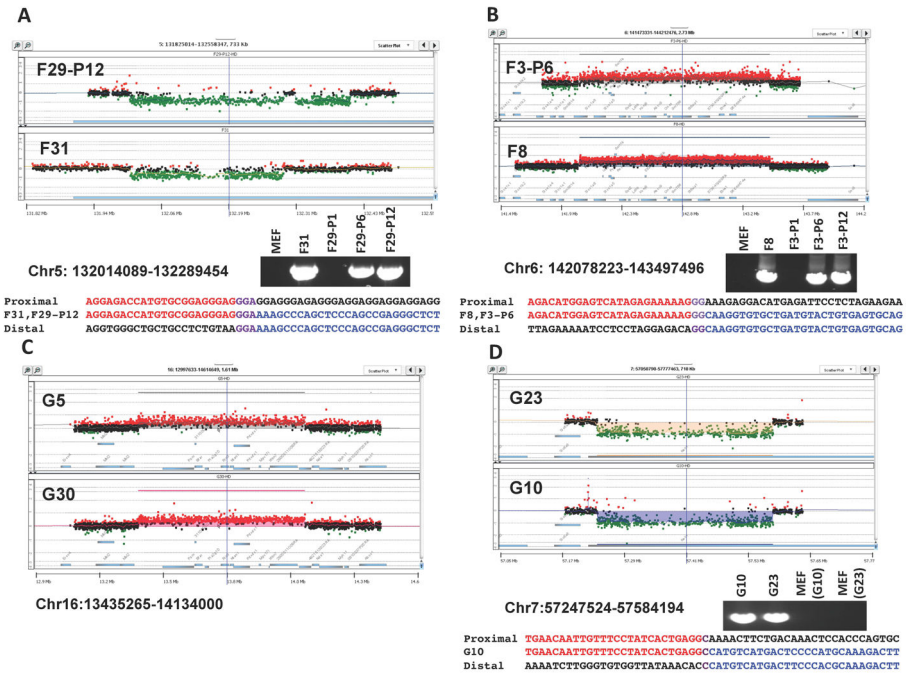

Figure 4. CNVs arising from predetermined loci

(A) The same 275 kb deletion was identified as de novo events in F31 and F29-P12 (passage 12 of F29). The top of the panel illustrates aCGH plots delineating the deletions. In F29-P12, the 275 kb deletion is part of a larger complex rearrangement event, which includes a smaller deletion in the distal region. Below the array image are coordinates of the CNV region and an agarose gel picture of breakpoint PCR results showing the presence or absence of the breakpoint junction in the assayed samples. The results suggest that the mutations leading to the deletions in F31 and F29-P12 are both de novo events. Below the gel image is alignment of breakpoint sequence to reference sequences. Red and blue colors indicate alignment of the breakpoint sequence to the reference sequences, whereas purple indicates the interval of microhomology. (B) The same 1.4 Mb duplications were identified in both F3-P6 and F8 as independent de novo events. (C) The apparently same 699 kb duplications were identified in G5 and G30. (D) The same 337 kb deletions were identified in G23 and G10.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Table 1

Number of neuronal, B cell, fibroblast iPSCs and neurospheres with CNVs

	Neuronal iPSC	B cell iPSC	Fibroblast iPSC	Neurosphere
Embryo 1	-	-	-	2/26
Embryo 2	-	-	-	0/27
Embryo 3	-	-	-	1/23
Embryo 4	1/12	4/13	-	-
Embryo 5	0/5	1/5	-	-
Embryo 6	0/3	5/13	0/1	-
Embryo 7	0/3	1/7	1/9	-
Total	1/23 (4.3%)	11/38 (29%)	1/10 (10%)	3/76 (3.9%)
Total excluding somatic mosaicism	1/23 (4.3%)	11/38 (29%)[a]	1/10 (10%)	1/76 (1.3%)

[a] the percentage may be overestimated here. This is because among the 14 CNVs identified by aCGH from the 11 B cell iPSC samples, breakpoint PCR was successful only for 9 CNVs. None of these 9 CNVs were detected to be present as a mosaic state in the corresponding fibroblast DNA by PCR. However, the possibility of the other 5 CNVs being somatic mosaicism cannot be ruled out.

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

Table 2

Summary of CNVs identified from the first and additional passage cells

Sample Name	Sample Type	Coordinate (estimated by aCGH)	Size of CNV	Type of CNV	# of UCSC genes
F8	neuronal iPSC	Chr6:142078223-143497496	1.4 Mb	Duplication	19
F1	B cell iPSC	Chr7:29781592-30059136	278 kb	Duplication	13
F26-1	B cell iPSC	Chr7:90622789-91300864	678 kb	Duplication	8
F26-2	B cell iPSC	Chr17:29585449-30341545	756 kb	Duplication	10
F26-3	B cell iPSC	Chr12:116647004-117237706	591 kb	Complex	1
F31	B cell iPSC	Chr5:132014089-132289454	275 kb	Deletion	1
F33	B cell iPSC	Chr1:107556718-107589113	32 kb	Duplication	2
F21	B cell iPSC	Chr6: 21678897-22044699	366 kb	Duplication	4
G5	B cell iPSC	Chr16:13435265-14134000	699 kb	Duplication	16
G10-1	B cell iPSC	Chr7:57247524-57584194	337 kb	Deletion	1
G10-2	B cell iPSC	Chr10:21750139-22267736	518 kb	Complex	10
G23	B cell iPSC	Chr7:57247524-57584194	337 kb	Deletion	1
G29	B cell iPSC	Chr10:78837653-79185916	348 kb	Deletion	15
G30	B cell iPSC	Chr16:13435265-14134000	699 kb	Duplication	16
G25	B cell iPSC	Chr14:42022784-45355691	3.3 Mb	Complex	17
G19	fibroblast iPSC	Chr11:93815253-94994961	1.2 Mb	Duplication	31
5A -1	neurosphere	Chr2:115093041-116171469	1.1 Mb	Duplication	6
5A -2	neurosphere	Chr9:30975300-31158250	183 kb	Duplication	4
21A	neurosphere	Chr2:115161358-116098550	937 kb	Deletion	6
7D	neurosphere	Chr16:40852266-41284924	433 kb	Deletion	0
F3-new-1 P6/P12	B cell iPSC/more passage	Chr2:76363317-77285102	922 Kb	Mosaic gain	9
F3-new-2 P6/P12	B cell iPSC/more passage	Chr6:142077895-143497398	1.4 Mb	Duplication	19
F3-new-3 P6/P12	B cell iPSC/more passage	Chr16:19625286-21735385	2.1 Mb	Mosaic gain	35
F29-new P12	B cell iPSC/more passage	Chr5:132014471-132408956	394 Kb	Complex	2
F13-new P6/P12	neuronal iPSC/more passage	Chr17:50579500-50917919	338 Kb	Duplication	2

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

Table 3

Comparison of aneuploidy and CNV load among the 1st, 6th, and 12th passages in the tested iPSCs and neurosphere clones

		Aneuploidy			# of CNVs		
		Passage 1	**Passage 6[a]**	**Passage 12**	**Passage 1**	**Passage 6[a]**	**Passage 12**
B cell iPSCs	F3	None	ChrY	ChrY	0	3	3
	F29	ChrY	None	None	0	0	2
	F31	None	None	None	1	1	1
Neuronal iPSCs	F7	ChrY	None	None	0	0	0
	F8	None	None	None	1	1	1
	F11	Chr11	Chr11	Chr11	0	0	0
	F13	Chr11	Chr11	Chr11	0	1	1
	F35	Chr11	Chr11	Chr11	0	0	0
Neuro-spheres	Z1	None	None	N/A	0	0	N/A
	Z2	None	ChrX	N/A	0	0	N/A
	Z3	None	None	N/A	0	0	N/A
	Z4	None	Chr8	N/A	0	0	N/A

[a] the neurosphere cells were examined at passage 5.