Dynamic Features of Chromosomal Instability during Culture of Induced Pluripotent Stem Cells

Casey O. Dubose, John R. Daum, Christopher L. Sansam and Gary J. Gorbsky*

Cell Cycle and Cancer Biology Research Program, Oklahoma Medical Research Foundation, Oklahoma City, OK 73104

*to whom correspondence should be addressed at:

Cell Cycle and Cancer Biology Research Program

Oklahoma Medical Research Foundation

825 NE 13th Street, MS 48

Oklahoma City, OK 73104

phone: 1-405-271-2032

email: GJG@omrf.org

Running Title: Chromosome Instability in iPSCs

Key Words: Aneuploidy, Mitosis, Trisomy, Chromosome 12, Optical Mapping, Structural Variants, Insertions, Deletions, Translocations

Abstract

Induced pluripotent stem cells (iPSCs) hold great potential for regenerative medicine. By reprogramming a patient's own cells, immunological rejection can be avoided during transplantation. For expansion and gene editing, iPSCs are grown in artificial culture for extended times. Culture affords potential danger for accumulation of genetic aberrations. To study these, two iPS cell lines were cultured and periodically analyzed using advanced optical mapping to detect and classify chromosome numerical and segmental changes that included deletions, insertions, balanced translocations and inversions. In one of the lines, a population trisomic for chromosome 12 gained dominance over a small number of passages. This appearance and dominance of the culture by chromosome 12 trisomic cells was tracked through intermediate passages by analysis of chromosome spreads. Mathematical modeling suggested that the proliferation rates of diploid versus trisomic cells could not account for the rapid dominance of the trisomic population. In addition, optical mapping revealed hundreds of structural variations distinct from those generally found within the human population. Many of these structural variants were detected in samples taken early in the culturing process and were maintained in late passage samples, while others were acquired over the course of culturing.

Introduction

Since their creation in 2007, induced pluripotent stem cells (iPSCs) have offered great promise in the field of regenerative medicine. By utilizing four key transcription factors, *Oct-4, Sox2, Klf4, and c-Myc*, the Yamanaka group was able to induce differentiated somatic cells to a pluripotent state (Takahashi et al., 2007). These iPSCs display similar characteristics to embryonic stem cells (ESCs), namely the ability to self-renew and differentiate into a wide range of somatic cells. iPSCs however have the potential to provide an alternative source of pluripotent cells while avoiding possible immunological rejection because they can be derived from the patient being treated. However, to be used in this manner, iPSCs must be created and expanded in culture. The requirement for growth in culture opens the possibility of the accrual of chromosomal variants over time.

Candidate cells must be cultured under artificial conditions during the reprogramming process and expanded after they are successfully transformed into iPSCs and potentially used for gene editing. During this extended period of cell culture, chromosome abnormalities may arise (Liu et al., 2014; Martins-Taylor et al., 2011; Mayshar et al., 2010; Rebuzzini et al., 2016; Taapken et al., 2011) . These changes may promote the iPSCs to become tumorigenic and develop other genetic or epigenetic abnormalities that make them risky for therapeutic use (Andrews et al., 2017; Ben-David & Benvenisty, 2011; Halliwell et al., 2020). A number of genomic abnormalities can arise. One of the most significant genome changes is numerical aneuploidy. A previous investigation of over 200 iPSC lines found that 12.5% of the cultures examined had an abnormal karyotype, while a study of 125 ESC lines found that 34% of cell lines contained abnormal karyotypes; both studies demonstrate the widespread occurrence of chromosomal aberrations (Amps et al., 2011; Taapken et al., 2011). In particular, chromosome 12 in iPSCs has been shown to have a high propensity for trisomy, representing as much as 32% of all chromosome aberrations detected in iPSCs (Mayshar et al., 2010; Taapken et al., 2011). The accrual of multiple chromosome 12p arms has also been shown to repeatedly occur in human embryonal carcinoma cells (Draper et al., 2004). The appearance of large chromosome changes correlated with altered expression of genes on the chromosome. Lines with an additional 12p segment overexpress pluripotency genes *NANOG* and *GDF3* (Mayshar et al., 2010). Smaller segmental changes, termed structural variants, are comprised of deletions, insertions, inversions, duplications and translocations of at least 50 base pairs (Mills et al., 2011). These may also have significant impact on gene expression. In the overall human genome, on average, single nucleotide polymorphisms (SNPs) contribute to 0.1% genetic variation between individuals, while structural variants contribute 1.5% (Pang et al., 2010).

A wide variety of disorders have been associated with structural variants, including cancer, diabetes, and cognitive disease (de Vries et al., 2005; Marshall et al., 2008). Structural variants implicated in disease can affect a single gene coding region, multiple genes, or can affect gene regulators at a distance (Lupski, 1998; Weischenfeldt et al., 2013). Although structural variants create significant genetic diversity and are implicated in various diseases, they are difficult to discover using short-read sequencing and therefore remain poorly mapped compared to SNPs. One solution to this problem is long-read sequencing, which aims to produce reads of thousands of base pairs, thereby easing the process of mapping and increasing structural variant sensitivity.

In this study, we examined the effects of cell culture on the genomic integrity of iPSCs by culturing two related iPS cell lines in parallel for 50 passages. The cell lines were examined at various time points

throughout the experiment by optical mapping supplemented by chromosome counts. Optical mapping creates marked DNA fragments which can be assembled into whole genomes quickly and efficiently (Müller & Westerlund, 2017). This technology can detect chromosome and structural variants with high accuracy. We detected hundreds of structural variants in both iPS cell lines, including many not previously mapped as existing human alleles. We identified both preexisting variants in these lines and those acquired during culture. We documented the gain of an additional chromosome 12 in one line. Of the structural variants acquired over the course of culturing, many disrupted protein coding sequences.

Results and Discussion:

iPSC Growth Rate and Aneuploidy

A concern with multiple passages of any cell line in culture is that some cells may acquire genetic changes that allow favorable adaptation to the specific culture conditions (cite). An advantageous mutation may increase the proliferation rate or decrease the rate of apoptosis, eventually leading the variant progeny to become dominant in the cell population (Morata and Ripoll 1975;Bowling et al., 2019). Our study involved two human iPSC lines. The first line is denoted WTC-11 iPSC and is the cell line from which the second iPSC line, AICS-0012, was derived (Kreitzer et al., 2013). AICS-0012 contains an mEGFP tag on the N-terminus of the coding sequence of the α-tubulin gene TUBA1B, created with CRISPR-Cas9 technology at the Allen Cell Institute, and is referred to in this study as Tuba1-GFP iPSC. Note that for both WTC-11 iPSC and Tuba1-GFP iPSC passage numbering reflects the number of passages performed during the course of this study, starting with passage number zero and ending with fifty. By using two iPSC lines from the same donor, we avoid potential differences in gene expression that may arise from differences in individual donors (Kilpinen et al., 2017). We counted the number of living cells present during each passage of both iPSC lines, allowing us to monitor the doubling times of both cell lines over time (Figure 1). Neither cell line demonstrated significant changes in doubling time over the course of the experiment, indicating that genetic variants with reduced doubling time did not overtake the lines during the culture periods. One benefit of our chosen optical mapping platform, the Bionano Genomics Saphyr, is its ability to detect copy number changes throughout a sample's genome down to the subchromosomal level. By analyzing the same cell line at multiple time points, this capability allows us to examine the change in gene dosage of the cell population, potentially identifying adaptations to cell culture conditions. One significant change was the gain of an additional chromosome 12 in Tuba1-GFP iPSC line, first identified in passage 32 via optical mapping (Figure 2).

As noted above, trisomy 12 is a common aneuploidy in iPSC lines and ESC lines (Amps et al., 2011; Draper et al., 2004; Taapken et al., 2011). Conventionally this has been hypothesized to occur through increased expression of growth promoting genes that provide a growth advantage to the variants (Baker et al., 2007; Draper et al., 2004). However, none of the four samples analyzed (excepting the copy number increase in the Tuba1-GFP iPSC Passage 32 trisomy) displayed an increase in copy number or structural variation events, such as duplications for NANOG or GDF3. A broader study of embryonic stem cell (ESC) and iPSC lines similarly found that genes such as *NANOG* do not disproportionately acquire structural variations during cell culture (Amps et al., 2011). Whatever changes of expression occur, an explanation for the dominance of the chromosome 12 trisomic population is that cells with this trisomy have a growth advantage, allowing them to outgrow the normal diploid cells. However, we did not detect a significant change in rate of proliferation after the trisomy became dominant. A recent publication reported that certain hESC, chromosome variant lines could actively induce apoptosis in cells of the parental line in

mixed cultures (Price et al., 2021). However, again this mechanism of overgrowth required that the chromosomal variant lines have a significantly higher proliferation rate (Price et al., 2021). We did not detect trisomies or large structural changes in the WTC-11 line even at later passages. While this might be explained simply by stochastic variation, an alternative explanation could be that the Tuba1-GFP iPSC line is more susceptible to numerical aneuploidy due to the additional time spent in culturing conditions, necessitated by the GFP tagging process. Compared to the WTC-11, the Tuba1-GFP iPSC line had spent approximately 26 more passages under culturing conditions at the time we received it. During each passage, we froze samples, allowing us to reexamine chromosome content in more detail at later times. To map the acquisition of trisomy in the Tuba1-GFP iPSCs, we thawed selected passages and performed chromosome spreads, focusing on the passages during which the trisomy arose and became dominant (Figure 3).

The chromosome number of the Tuba1-GFP iPSC population shifts from 46, the expected number for diploid human cells, to 47 over the course of 5 passages. The cell population containing the trisomy is initially modest, reflecting only 12.5% at passage 21 which may reflect, at least in part, the error rate inherent in counting chromosome spreads. However, the trisomic population rapidly increases to 80% by passage 26. The aberrant genotype then persists in the cell population, as indicated by 84% of the population possessing 47 chromosomes at Passage 40. By combining this information with the recorded doubling time information, we can approximate that the trisomy 12 genotype shifted from 12.5% to 80% of the population in approximately 20 doublings. This is a surprisingly short time period given that the post-trisomy 12 population does not divide at a significantly faster rate compared to the diploid population. The doubling times for the predominantly diploid Tuba1-GFP cultures, passages prior to passage 21, and for the predominantly chromosome 12 trisomy cultures, passages post passage 26, though not found to be statistically significantly different, were determined to be 18.3 ± 1.8 and 17.9 ± 1.1, respectively.

To understand the potential mechanisms that might explain the rapid increase in chromosome 12 trisomy, we created a series of mathematical model based on the measured growth of the early passage, diploid cells and late passage trisomic Tuba1-GFP cells depicted in Figure 1. We compared conditions where the initial population consisted of 10% trisomic cells. Based on the measured growth rates and simple competition, we calculated that after approximately 360 hours occurring during 5 passages the proportion of trisomic cells would only rise to 12.2% (Figure 4A). In contrast, experimental observations derived from chromosome spreads indicated that the proportion of trisomic cells rose from ~10% to ~85% over the same duration of 5 passages (passages 21 through 26, figure 2). We then modulated other parameters to model potential drivers for rapid increase of trisomy from 10 to 90% in360 hours. We determined that trisomic cells would have to exhibit a doubling time of approximately 12 hours rather than the 18 hours observed and thus proliferate at approximately 1.5 times the rate of the diploid cells to achieve 90% culture dominance in 360 hours (Figure 4B). For preferential cell death of diploid to account for the change, our model indicates that to achieve rapid dominance, the diploid cells would have to show 20% cell death per division where the trisomic cell death rate would be 0% (Figure 4C). Although we have not yet specifically tested differences in cell death, this level of difference would have been unlikely to escape our notice during cell culture. Finally we modeled the situation where chromosome missegregation leading to trisomy was not restricted to a single founder cell but could occur in multiple diploid cells during each division. We found that to account for the change in dominance, the missegregation would require that 10% of the diploid cells would become trisomic for chromosome 12

during each division (Figure 4D). Because we have stocks of cells frozen during the critical passages, we are currently investigating whether any of these scenarios, or indeed others not yet modeled, may account for the rapid rise of trisomic cells.

Structural Variant Detection, Filtration, and Characterization

We analyzed four samples via Bionano Genomics (BNG) optical mapping, two for each cell line. These were taken once early in an early passage and once in a late passage. Optical mapping in this way works by labeling a single known sequence on large fragments of genomic DNA, often larger than 150 kbp. Labeled fragments are then aligned to create a de novo assembly which, for human samples, is compared to a known reference map. Doing so allows for the detection of structural variants, some of which may contain repeated sequence and be missed in next generation sequencing. All samples exceeded recommended molecule quality requirements, and the metrics for each sample can be seen in Supplemental Table 1. Each sample was assembled de novo and then mapped to human reference hg38 using BNG solve (version 3.4) and access software (version 1.4.2). Structural variant calls were filtered with recommended thresholds excepting minimum required molecule coverage, which was doubled. To identify structural variants unique to our iPSC samples, we filtered our results to exclude all structural variants found in BNG control database, which is comprised of samples from the 1000 Genome Project and donors from San Diego Blood Banks. In the four samples, we identified 169 deletions, 81 insertions, 47 duplications, and 97 inversions, all of which were absent from the 1000 genomes project low frequency alleles database, as determined by using the Ensemble Variant Effect Predictor (Auton et al., 2015; McLaren et al., 2016). Given that the four cell populations analyzed came from two cell lines which both share a common ancestor, we examined how many variants were present in 2 or more samples. A variant was considered unique if it had less than 50% reciprocal overlap with any other variant. Of the total 394 variants detected, 233 (59%) were present in more than one sample. The distribution of structural variants across all samples can be seen in Figure 5.

Both WTC-11 and Tuba1-GFP iPSC lines demonstrate maintained, lost, and acquired structural variants from early to late passage samples. The total structural variant frequency increased in both lines over the course of culturing. Inversion calls from both lines show a low rate of maintained variants compared to other structural variant types. However, this may be due to detection limits due to reduced sensitivity of the optical mapping technology for inversions below 30 kbp. The size distribution of structural variants varied by variant type and cell line, and can be seen in figure 6. The chromosome locations and sizes for maintained, lost, and acquired variants are indicated in Supplemental Table 2. Among both cell lines, duplications were on average the largest structural variants, inversions an intermediate size, while deletions and insertions were similar in size and smaller than other types of structural variants.

Structural Variation Impact on Gene Function

Acquired structural variations can impart negative or, perhaps more rarely, positive consequences for cell survival. If a structural variant completely supplants a gene, the loss or gain in copy number can directly affect expression. Changes in gene expression may have negative downstream consequences, impacting expression of other genes or compromising proteostasis through over or under expression of proteins normally produced as balanced components of protein complexes (Kane et al., 2021; Pavelka et al., 2010; Torres et al., 2007). Similarly, if a structural variant encroaches partially into protein coding sequences of a gene, truncations or gene fusions may arise, leading to a loss or gain of function (Collins et

al., 2020; Lupski, 1998; Shaffer LG and Lupski JR 2000). In order to understand the potential consequences, both in the genome and phenotypically, we used the Ensembl Variant Effect Predictor to predict the impact of our structural variant set (McLaren et al., 2016). To identify potentially high impact effects of our detected structural variants, we focused directly on protein coding consequences and identified 121 genes whose coding sequence overlaps with structural variants from the four samples (Table 1). In the case of both cell lines, the sample collected later in the culture experiment contained more genes whose protein coding sequences were affected by structural variants. Of the 121 genes affected, 28 were present in both cell lines, perhaps owing to their common parental line. Interestingly, one gene, DAPL1, which was unaffected in the early passage samples of both cell lines, was modified in late passage samples by an insertion variant into an intron segment of the protein coding region. DAPL1 has been implicated in in epithelial differentiation, apoptosis, and suppressor of cell proliferation in the retinal pigment epithelium (Ma et al., 2017). Notably, DAPL1 has been identified as a potential susceptibility locus for age-related macular degeneration in females (Grassmann et al., 2015).

Next, we utilized the functional annotation toolset DAVID to determine if the structural variant impacted genes disproportionately affected particular gene ontology groups or were linked to any disease associations (Huang et al., 2009). Gene lists from each sample, as shown in table 1, were divided into two groups - genes affected by duplications and thus potentially enriched, and non-duplication structural variants which could potentially interfere with gene expression. In the WTC-11 iPSC line many gene ontology groups were disrupted by deletions, insertions, and inversions. The most statistically significant disruptions in were driven by SVs acquired over the course of culturing. Conversely, no gene ontology groups were significantly affected by duplications (Figure 7A). The Tuba1-GFP iPSC line was impacted by both disruptive and duplicative SVs. Similar to WTC-11, the most statistically significant impacts arose from SVs which were acquired during culture (Figure 7B). Notably, large structural variants can impact several genes, which may lead to a particular gene ontology term being deemed enriched if the neighboring genes are homologues or perform similar functions. For instance, the gene group "gonadal differentiation" in the WTC-11 passage 45 sample was likely enriched due to six genes on the Y chromosome being impacted by a single large deletion (Figure 7A).

Conclusions

Over the course of 150 days of continuous culture, or 50 passages, we observed a substantial change in the genomes of two iPSC lines. Most notably, the Tuba1-GFP iPSC cell line experienced the appearance and rapid dominance of a population trisomic for chromosome 12, a frequently observed aneuploidy in stem cell lines. Given that the proliferation rate of the cells did not significantly increase after acquiring the trisomy, it is unlikely that the rapid dominance of the trisomic population can be attributed to simple competition or differences in cell death rates. Further studies actively pursuing the mechanism or mechanisms underlying the rapid conversion of the population may uncover a novel source of stem cell aneuploidy.

We detected hundreds of structural variants not found in the general population using long-read optical mapping technology. However, 59% of structural variants were found in more than one sample, suggesting that those variations may be due to the unique genetic constitution present in the donor genome. More significantly, both iPSC lines acquired numerous structural variants over the course of culturing. After ascertaining the genes whose protein coding sequences were affected by structural variants, we were able to identify several enriched gene ontology and disease clusters. While it is unclear

if these changes might compromise use of iPSCs in therapy, the accumulation of variants suggests that culture times be minimized in therapeutic practice.

Methods

Cell culture

Induced Pluripotent Stem Cells (iPSCs) were cultured in 25 square centimeter flasks coated in Growth Factor Reduced Matrigel (Corning 354230) diluted in DMEM/F12 at 1:30 ratio (Caisson Labs DFL14-500ML). Flasks were coated with Matrigel and used within one week, with care taken to avoid evaporation or drying of the Matrigel solution. Cells were maintained with mTesR1 and mTesR plus media (Stemcell Technology, 85850 and 05825) supplemented with penicillin-streptomycin (Thermo Fisher Scientific, 15-140-122). With each passage, 300,000 iPS cells were transferred to a new flask following detachment via Accutase (Thermo Fisher Scientific, A1110501). Culturing media was supplemented with 10uM Y-27632 ROCK inhibitor (MedChem Express HY-10583) for approximately 24 hours following passaging, then changed to media containing no ROCK inhibitor. Cells were maintained at 37C in 5% CO_2 in a water-jacketed incubator. Cells were passaged approximately every 72 hours and were maintained continuously for 150 days. iPSC colonies maintained normal morphology with minimal differentiation and had an average death rate of approximately three percent.

Chromosome spreads

Cells were treated with Nocadazole at a concentration of 100 ng/ml for 4 hours, then treated with Accutase to bring cells into suspension. Cells were collected, then washed with media by centrifuging at 200 × *g* for 3 min. Cells were resuspended in 500 µl of warmed swelling buffer (70% deionized water + 30% mTesR media). Cells were incubated in a 37C water bath for 20 minutes. Cells were then fixed by adding 1 ml of freshly prepared 3:1 methanol to acetic acid and then incubated at room temperature for 15 minutes. Cells were centrifuged for 5 minutes at 200 × *g*, washed with 1 ml of fixative and pelleted again. The cells were resuspended in 150-200 µl of fixative, then 50-60 µl of cell suspension was dropped form a height of 70 cm onto a 22-mm^2 coverslip. The coverslips were then placed inside a 150-mm dish on top of wet filter paper. The coverslips were then allowed to dry overnight. Next the coverslips were stained with 4', 6-diamidino-2-phenylindole (DAPI) (100 ng/ml) and SYBERGold nucleic acid dye (Thermo Fisher Scientific S11494) at a 1:20,000 dilution of stock. Imaging was done with a Zeiss Axioplan II microscope platform using a 100× objective, Hamamatsu Orca II camera, and Metamorph software.

Predictive mathematical modeling

Mathematical model created using Microsoft Excel predicted fractions of cells within culture containing a mixed population of two cell types, Tuba1-GFP diploid and Tuba1-GFP chromosome 12 trisomy, exhibiting unique characteristics. Inputs permitted cell type specific manipulation of doubling times, cell death rates, and conversion rates from one cell type to another so that theoretical cell type fractions within culture could be predicted over time with respect to these varying input characteristics. For probing cell type specific death rates, the formulae employed a percentage loss of newly created daughter cells at each doubling time. Conversion rates were calculated similarly, with certain percentages of diploid cells converting to chromosome 12 trisomy cells at each diploid doubling time. Microsoft Excel stacked area graphs were utilized to provide illustrations of predictive results.

Bionano Genomics Techniques

Genomic DNA Isolation

iPS cells were collected from culture and counted using the Countess II FL Automated Cell Counter (Thermo Fisher Scientific, AMQAF1000). 1×10^6 and 1.5×10^6 cell aliquots (corresponding to 6 µg and 9 µg of DNA) were targeted for each sample preparation. Samples were prepared immediately after being collected, following Bionano guidelines (Bionano Genomics, Bionano Prep Cell Culture DNA Isolation Protocol, Doc. 30026). Briefly, cells were pelleted, resuspended in cold Cell Buffer (Bionano Genomics), and suspended in agarose plugs in order to minimize DNA shearing (Bio-Rad CHEF Mapper XA system, 1703713). The DNA-agarose plugs were subjected to a series of Proteinase K Digestions (Qiagen, 158920) followed by an RNase A digestion (Qiagen, 158922). Following digestion treatments, the plugs were washed with 1x Wash Buffer (Bionano Genomics) and Tris-EDTA Buffer (Invitrogen, AM9849). The DNA-agarose plugs were then digested with agarase (Thermo Fisher Scientific, EO0461) at 43C for 45 minutes. The DNA solution was then purified via drop dialysis, using Millipore filters floated on TE buffer for 1 hour. Following purification, the DNA was quantified through the use of a Qubit 4 Fluorometer (Thermo Fisher Scientific, Q33238). DNA with a concentration of 35-200 ng/µl and a coefficient of variation (standard deviation/mean) of less than 0.25 was deemed acceptable.

DNA Labeling

DNA labeling was achieved using the Direct Label Staining (DLS) method from Bionano Genomics (Bionano Prep Direct Label and Stain Protocol, 30206 D), whose protocol instructions were strictly followed. Briefly, 750 ng of purified genomic DNA was added to a master mix including the DLE-1 Labeling Enzyme and incubated for 2 hours at 37C. Afterward, the sample was treated with Proteinase K (Qiagen, 158920) and incubated at 50C for 30 minutes. Next, the samples were cleaned up via membrane adsorption using reagents supplied by Bionano Genomics. Following cleanup, the labeled DNA was stained and homogenized in preparation for loading the sample onto the Bionano chip. After staining the samples were incubated at room temperature overnight. The following day the stained DNA samples were quantified using a Qubit 4 Fluorometer and its complimentary dsDNA High Sensitivity Assay Kit. Samples with concentrations between 4 and 12 ng/µl were chosen to be loaded onto the Bionano Chip.

Chip Loading and Analysis

Once the labeled DNA was quantified it was loaded onto Bionano Saphyr chip. The chip is then loaded into the Saphyr instrument where the labeled DNA is linearized into nanochannels using electrophoretic principles to guide the DNA. Once loaded, the Saphyr instrument begins imaging the labeled DNA. Each flow cell can process 320-480 Gbps of DNA in the span of 24 to 36 hours. The data output was analyzed using Bionano Solve 3.4. The compiled data was then visualized through the Bionano Access software, version 1.4.2.

Structural Variant Analysis Software

Structural variant analysis involving co-localization with other detected structural variants and comparing variant coordinates across samples to determine if the structural variants were “maintained”, “lost”, or “unique” was performed using bed-tools intersect (version 2.26.0). Structural variant impact, such as effect on protein coding genes, was determined using the Ensembl Variant Effect Predictor (VEP)

(McLaren, W et al 2016). Gene ontology analysis of structural variants was performed using the Database for Annotation, Visualization and Integrated Discovery (DAVID) version 6.8 (Huang DW, et al. 2009).

Data Availability

Bionano optical mapping files .cmap, .smap, and .bnx data can be found at NCBI Supplementary Files data base, submission SUB10823566. (Hyperlinks to the data files will be added when NCBI completes finalization of the uploads.)

Acknowledgements

We thank Paul Gorbsky for assistance with the mathematical modeling. Support was obtained from a grant from The Oklahoma Center for Adult Stem Cell Research to GJG and from grants 5R35GM126980 to GJG and 1R01GM121703 to CLS from the National Institute of General Medical Sciences.

Author Contributions

Conceptualization: CD, GG; Investigation: CD; Formal Analysis: CD, JD, CS; Writing – Original Draft: CD; Writing Review & Editing: JD, GG, CS; Supervision: GG

References

Amps, K., Andrews, P. W., Anyfantis, G., Armstrong, L., Avery, S., Baharvand, H., Baker, J., Baker, D., Munoz, M. B., Beil, S., Benvenisty, N., Ben-Yosef, D., Biancotti, J. C., Bosman, A., Brena, R. M., Brison, D., Caisander, G., Camarasa, M. v., Chen, J., … Zhou, Q. (2011). Screening ethnically diverse human embryonic stem cells identifies a chromosome 20 minimal amplicon conferring growth advantage. *Nature Biotechnology*, *29*(12), 1132–1144. https://doi.org/10.1038/nbt.2051

Andrews, P. W., Ben-David, U., Benvenisty, N., Coffey, P., Eggan, K., Knowles, B. B., Nagy, A., Pera, M., Reubinoff, B., Rugg-Gunn, P. J., & Stacey, G. N. (2017). Assessing the Safety of Human Pluripotent Stem Cells and Their Derivatives for Clinical Applications. *Stem Cell Reports*, *9*(1), 1–4. https://doi.org/10.1016/j.stemcr.2017.05.029

Auton, A., Abecasis, G. R., Altshuler, D. M., Durbin, R. M., Bentley, D. R., Chakravarti, A., Clark, A. G., Donnelly, P., Eichler, E. E., Flicek, P., Gabriel, S. B., Gibbs, R. A., Green, E. D., Hurles, M. E., Knoppers, B. M., Korbel, J. O., Lander, E. S., Lee, C., Lehrach, H., … Schloss, J. A. (2015). A global reference for human genetic variation. In *Nature* (Vol. 526, Issue 7571, pp. 68–74). Nature Publishing Group. https://doi.org/10.1038/nature15393

Baker, D. E. C., Harrison, N. J., Maltby, E., Smith, K., Moore, H. D., Shaw, P. J., Heath, P. R., Holden, H., & Andrews, P. W. (2007). Adaptation to culture of human embryonic stem cells and oncogenesis in vivo. In *Nature Biotechnology* (Vol. 25, Issue 2, pp. 207–215). https://doi.org/10.1038/nbt1285

Ben-David, U., & Benvenisty, N. (2011). The tumorigenicity of human embryonic and induced pluripotent stem cells. In *Nature Reviews Cancer* (Vol. 11, Issue 4, pp. 268–277). https://doi.org/10.1038/nrc3034

Bowling, S., Lawlor, K., & Rodríguez, T. A. (2019). Cell competition: The winners and losers of fitness selection. *Development (Cambridge)*, *146*(13). https://doi.org/10.1242/dev.167486

Collins, R. L., Brand, H., Karczewski, K. J., Zhao, X., Alföldi, J., Francioli, L. C., Khera, A. v., Lowther, C., Gauthier, L. D., Wang, H., Watts, N. A., Solomonson, M., O'Donnell-Luria, A., Baumann, A., Munshi, R., Walker, M., Whelan, C. W., Huang, Y., Brookings, T., … Talkowski, M. E. (2020). A structural variation reference for medical and population genetics. *Nature*, *581*(7809), 444–451. https://doi.org/10.1038/s41586-020-2287-8

de Vries, B. B. A., Pfundt, R., Leisink, M., Koolen, D. A., Vissers, L. E. L. M., Janssen, I. M., van Reijmersdal, S., Nillesen, W. M., Huys, E. H. L. P. G., de Leeuw, N., Smeets, D., Sistermans, E. A., Feuth, T., van Ravenswaaij-Arts, C. M. A., Geurts Van Kessel, A., Schoenmakers, E. F. P. M., Brunner, H. G., & Veltman, J. A. (2005). Diagnostic Genome Profiling in Mental Retardation. In *Am. J. Hum. Genet* (Vol. 77).

Draper, J. S., Smith, K., Gokhale, P., Moore, H. D., Maltby, E., Johnson, J., Meisner, L., Zwaka, T. P., Thomson, J. A., & Andrews, P. W. (2004). Recurrent gain of chromosomes 17q and 12 in cultured human embryonic stem cells. *Nature Biotechnology*, *22*(1), 53–54. https://doi.org/10.1038/nbt922

Grassmann, F., Friedrich, U., Fauser, S., Schick, T., Milenkovic, A., Schulz, H. L., von Strachwitz, C. N., Bettecken, T., Lichtner, P., Meitinger, T., Arend, N., Wolf, A., Haritoglou, C., Rudolph, G., Chakravarthy, U., Silvestri, G., McKay, G. J., Freitag-Wolf, S., Krawczak, M., … Weber, B. H. F. (2015). A Candidate Gene Association Study Identifies DAPL1 as a Female-Specific Susceptibility Locus for Age-Related Macular

Degeneration (AMD). *NeuroMolecular Medicine*, *17*(2), 111–120. https://doi.org/10.1007/s12017-015-8342-1

Halliwell, J., Barbaric, I., & Andrews, P. W. (2020). Acquired genetic changes in human pluripotent stem cells: origins and consequences. In *Nature Reviews Molecular Cell Biology* (Vol. 21, Issue 12, pp. 715–728). Nature Research. https://doi.org/10.1038/s41580-020-00292-z

Huang, D. W., Sherman, B. T., & Lempicki, R. A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols*, *4*(1), 44–57. https://doi.org/10.1038/nprot.2008.211

Kane, A. J., Brennan, C. M., Xu, A. E., Solís, E. J., Terhorst, A., Denic, V., & Amon, A. (2021). Cell adaptation to aneuploidy by the environmental stress response dampens induction of the cytosolic unfolded-protein response. *Molecular Biology of the Cell*, *32*(17), 1557–1564. https://doi.org/10.1091/mbc.E21-03-0104

Kilpinen, H., Goncalves, A., Leha, A., Afzal, V., Alasoo, K., Ashford, S., Bala, S., Bensaddek, D., Casale, F. P., Culley, O. J., Danecek, P., Faulconbridge, A., Harrison, P. W., Kathuria, A., McCarthy, D., McCarthy, S. A., Meleckyte, R., Memari, Y., Moens, N., … Gaffney, D. J. (2017). Common genetic variation drives molecular heterogeneity in human iPSCs. *Nature*, *546*(7658), 370–375. https://doi.org/10.1038/nature22403

Kreitzer, F. R., Salomonis, N., Sheehan, A., Huang, M., Park, J. S., Spindler, M. J., Lizarraga, P., Weiss, W. A., So, P.-L., & Conklin, B. R. (2013). Original Article A robust method to derive functional neural crest cells from human pluripotent stem cells. In *Am J Stem Cell* (Vol. 2, Issue 2). www.AJSC.us/

Liu, P., Kaplan, A., Yuan, B., Hanna, J. H., Lupski, J. R., & Reiner, O. (2014). Passage number is a major contributor to genomic structural variations in mouse iPSCs. *Stem Cells*, *32*(10), 2657–2667. https://doi.org/10.1002/stem.1779

Lupski, J. R. (1998). Genomic disorders: structural features of the genome can lead to DNA rearrangements and human disease traits. *Trends Genet. 14(10),* 417-22. Https://doi: 10.1016/s0168-9525(98)01555-8.

Ma, X., Li, H., Wang, Y., Wang, J., Zheng, Q., Hua, J., Yang, J., Pan, L., Lu, F., Qu, J., & Hou, L. (2017). DAPL1, a susceptibility locus for age-related macular degeneration, acts as a novel suppressor of cell proliferation in the retinal pigment epithelium. *Human Molecular Genetics*, *26*(9), 1612–1621. https://doi.org/10.1093/hmg/ddx063

Marshall, C. R., Noor, A., Vincent, J. B., Lionel, A. C., Feuk, L., Skaug, J., Shago, M., Moessner, R., Pinto, D., Ren, Y., Thiruvahindrapduram, B., Fiebig, A., Schreiber, S., Friedman, J., Ketelaars, C. E. J., Vos, Y. J., Ficicioglu, C., Kirkpatrick, S., Nicolson, R., … Scherer, S. W. (2008). Structural Variation of Chromosomes in Autism Spectrum Disorder. *American Journal of Human Genetics*, *82*(2), 477–488. https://doi.org/10.1016/j.ajhg.2007.12.009

Martins-Taylor, K., Nisler, B. S., Taapken, S. M., Compton, T., Crandall, L., Montgomery, K. D., Lalande, M., & Xu, R.-H. (2011). Recurrent copy number variations in human induced pluripotent stem cells. *Nature Biotechnology*, *29*(6), 488–491. https://doi.org/10.1038/nbt.1890

Mayshar, Y., Ben-David, U., Lavon, N., Biancotti, J. C., Yakir, B., Clark, A. T., Plath, K., Lowry, W. E., & Benvenisty, N. (2010). Identification and classification of chromosomal aberrations in human induced pluripotent stem cells. *Cell Stem Cell*, *7*(4), 521–531. https://doi.org/10.1016/j.stem.2010.07.017

McLaren, W., Gil, L., Hunt, S. E., Riat, H. S., Ritchie, G. R. S., Thormann, A., Flicek, P., & Cunningham, F. (2016). The Ensembl Variant Effect Predictor. *Genome Biology*, *17*(1). https://doi.org/10.1186/s13059-016-0974-4

Mills, R. E., Walter, K., Stewart, C., Handsaker, R. E., Chen, K., Alkan, C., Abyzov, A., Yoon, S. C., Ye, K., Cheetham, R. K., Chinwalla, A., Conrad, D. F., Fu, Y., Grubert, F., Hajirasouliha, I., Hormozdiari, F., Iakoucheva, L. M., Iqbal, Z., Kang, S., … Peterson, J. L. (2011). Mapping copy number variation by population-scale genome sequencing. *Nature*, *470*(7332), 59–65. https://doi.org/10.1038/nature09708

Müller, V., & Westerlund, F. (2017). Optical DNA mapping in nanofluidic devices: principles and applications. *Lab on a Chip*, *17*(4), 579–590. https://doi.org/10.1039/c6lc01439a

Pang, A. W., MacDonald, J. R., Pinto, D., Wei, J., Rafiq, M. A., Conrad, D. F., Park, H., Hurles, M. E., Lee, C., Craig Venter, J., Kirkness, E. F., Levy, S., Feuk, L., & Scherer, S. W. (2010). Towards a comprehensive structural variation map of an individual human genome In *Genome Biology* (Vol. 11). http://genomebiology.com/2010/11/5/R52

Pavelka, N., Rancati, G., Zhu, J., Bradford, W. D., Saraf, A., Florens, L., Sanderson, B. W., Hattem, G. L., & Li, R. (2010). Aneuploidy confers quantitative proteome changes and phenotypic variation in budding yeast. *Nature*, *468*(7321), 321–325. https://doi.org/10.1038/nature09529

Price, C. J., Stavish, D., Gokhale, P. J., Stevenson, B. A., Sargeant, S., Lacey, J., Rodriguez, T. A., & Barbaric, I. (2021). Genetically variant human pluripotent stem cells selectively eliminate wild-type counterparts through YAP-mediated cell competition. *Developmental Cell*, *56*(17), 2455-2470.e10. https://doi.org/10.1016/j.devcel.2021.07.019

Rebuzzini, P., Zuccotti, M., Redi, C. A., & Garagna, S. (2016). Achilles ʼheel of pluripotent stem cells: genetic, genomic and epigenetic variations during prolonged culture. In *Cellular and Molecular Life Sciences* (Vol. 73, Issue 13, pp. 2453–2466). Birkhauser Verlag AG. https://doi.org/10.1007/s00018-016-2171-8

Taapken, S. M., Nisler, B. S., Newton, M. A., Sampsell-Barron, T. L., Leonhard, K. A., McIntire, E. M., & Montgomery, K. D. (2011). Karyotypic abnormalities in human induced pluripotent stem cells and embryonic stem cells. In *Stem Cell Rev. Rep* (Vol. 318, Issue 6).

Takahashi, K., Tanabe, K., Ohnuki, M., Narita, M., Ichisaka, T., Tomoda, K., & Yamanaka, S. (2007). Induction of Pluripotent Stem Cells from Adult Human Fibroblasts by Defined Factors. *Cell*, *131*(5), 861–872. https://doi.org/10.1016/j.cell.2007.11.019

Torres, E. M., Sokolsky, T., Tucker, C. M., Chan, L. Y., Boselli, M., Dunham, M. J., & Amon, A. (2007). Effects of aneuploidy on cellular physiology and cell division in haploid yeast. *Science*, *317*(5840), 916–924. https://doi.org/10.1126/science.1142210

Weischenfeldt, J., Symmons, O., Spitz, F., & Korbel, J. O. (2013). Phenotypic impact of genomic structural variation: Insights from and for human disease. In *Nature Reviews Genetics* (Vol. 14, Issue 2, pp. 125–138). https://doi.org/10.1038/nrg3373

Figures

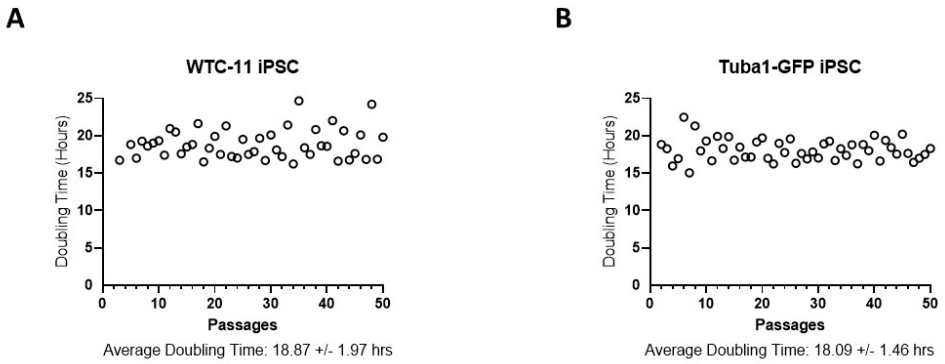

Figure 1: Doubling times of iPSCs vary little of the course of culturing. (A) WTC-11 cell line. **(B)** Tuba1-GFP cell line. Cell count were taken with each passage using an automated cell counter, and the time between each passage was recorded in order to calculate population doubling times. Passage numbers reflect documented passages following the creation of WTC-11 iPSC line and begin at the passage number for which they were thawed for these experiments. There was no significant change in doubling times between the first ten passages and the last ten passages in either cell line (unpaired t-test).

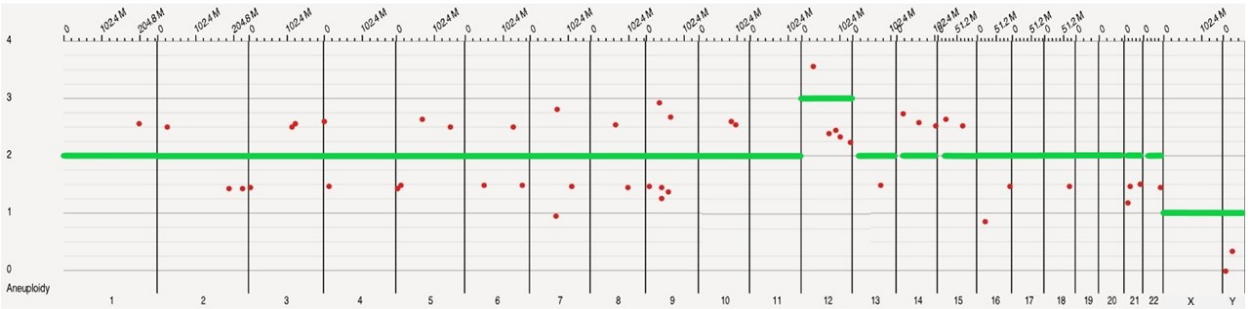

Figure 2: The late passage of Tuba1-GFP iPSC displays a gain of chromosome 12. Optical mapping data of Tuba1-GFP iPSC passage 32 detected a copy number of three for chromosome 12 in majority of cells, indicating an acquired aneuploidy. The green horizontal bars indicate the ploidy of each chromosome in the sample population, while the red dots indicate specific subchromosomal regions that vary in copy number from the called ploidy of the chromosome. In order to call a whole chromosome aneuploidy, it is necessary for 80% or more of the chromosome length to be categorized as higher or lower than the copy number baseline. Note that the X and Y chromosomes report a copy number of one, as the sample is derived from a male donor. Data visualized using Bionano Access software.

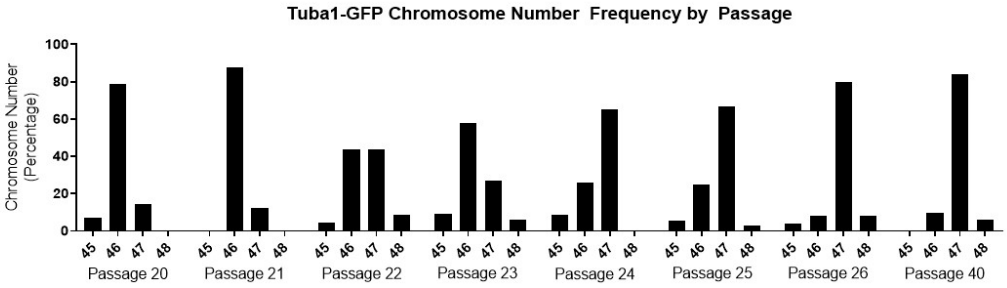

Figure 3: The presence of an aneuploidy event in the Tuba1-GFP iPSC population is confirmed via chromosome spreads. Each sample reflects the ploidy of the passage population. Passages 22-26 indicate a progressive shift from normal ploidy to aneuploidy as the chromosome number of the populations shift from 46 to 47. The aneuploidy of the population then persists in the population, as indicated in the passage 40 sample. Each passage reflects >20 cells quantified.

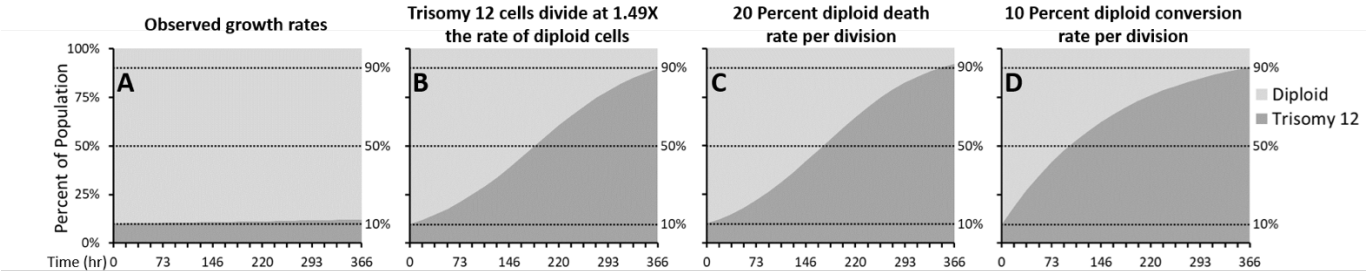

Figure 4. Mathematical modeling of various growth parameters. Mathematical simulations modeling the effects of various parameters of the conversion of iPSCs from predominantly diploid to predominantly trisomy 12 populations in the 20 doublings that occur during 5 passages. (A) Simulation using the observed growth rates of majority diploid iPSCs (passages 1-21) and majority trisomy 12 iPSCs (passages 26-50) fails to model rise of trisomy 12 dominance. (B) Modeling trisomy 12 cells replicating with doubling time of 12.5 hours, at 1.49 times the rate of diploid cells. (C) Modeling a very high (20%) death rate for diploid cells and a very low (1%) rate for trisomy 12 cells. (D) Modeling our hypothesized mechanism of a 10% conversion rate of diploid to trisomy 12 cells per doubling.

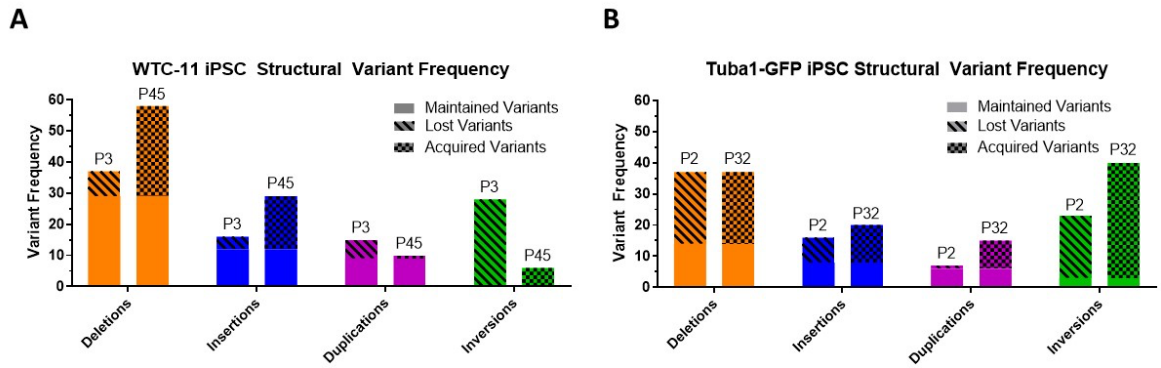

Figure 5: Structural variant frequency varies from early passage sample to late passage in both iPSC lines. Using optical mapping, two samples from each cell line were analyzed, one early passage sample and one late passage sample. Structural variants were organized into three categories: maintained variants, which were detected in both early and late passage samples; lost variants, which were detected in the early passage samples but not in the late passage samples; and acquired variants, which were not detected in the early passage samples but subsequently appeared in the late passage samples. **(A)** Structural variant frequencies from WTC-11 iPSC passage 3 and passage 45 are shown. Deletions and insertions demonstrate an increase in frequency from early to late passage samples, while total duplications and inversions decrease. **(B)** Structural variant frequencies from Tuba1-GFP iPSC passage 2 and 32 are shown. All variant types exhibit an increase in frequency from early to late passage samples with the exception of deletions, which maintain the same total instances.

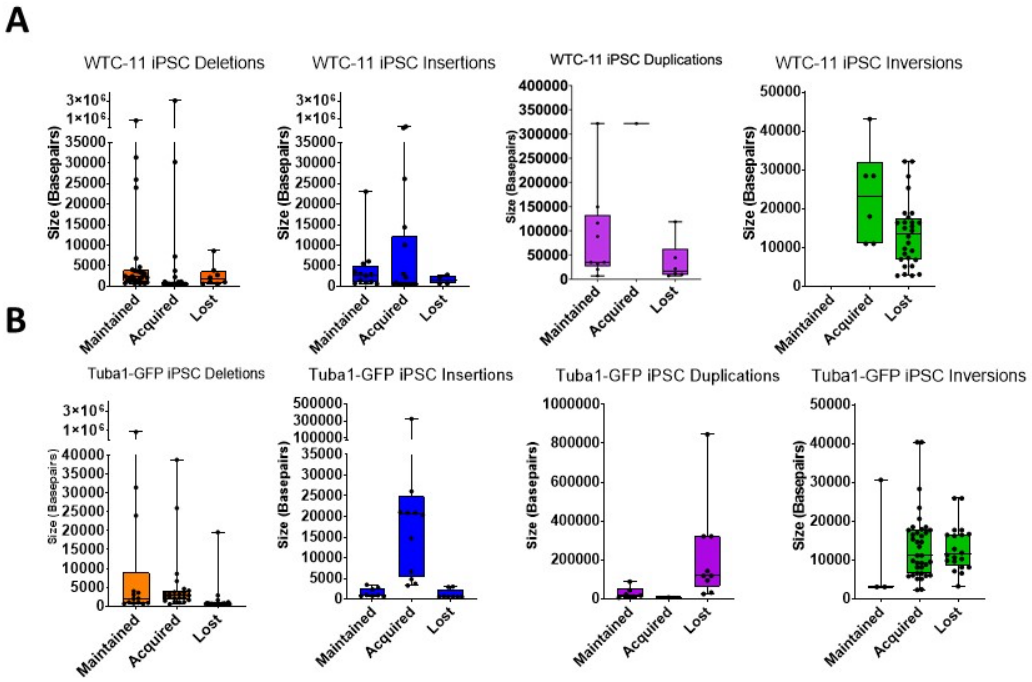

Figure 6: Structural variants vary in size distribution. A/B. All detected structural variants are represented on box and whisker plots, where the box indicates the first, median, and third quartiles. Notably, duplications were the least frequent and on average largest structural variants in both iPSC lines.

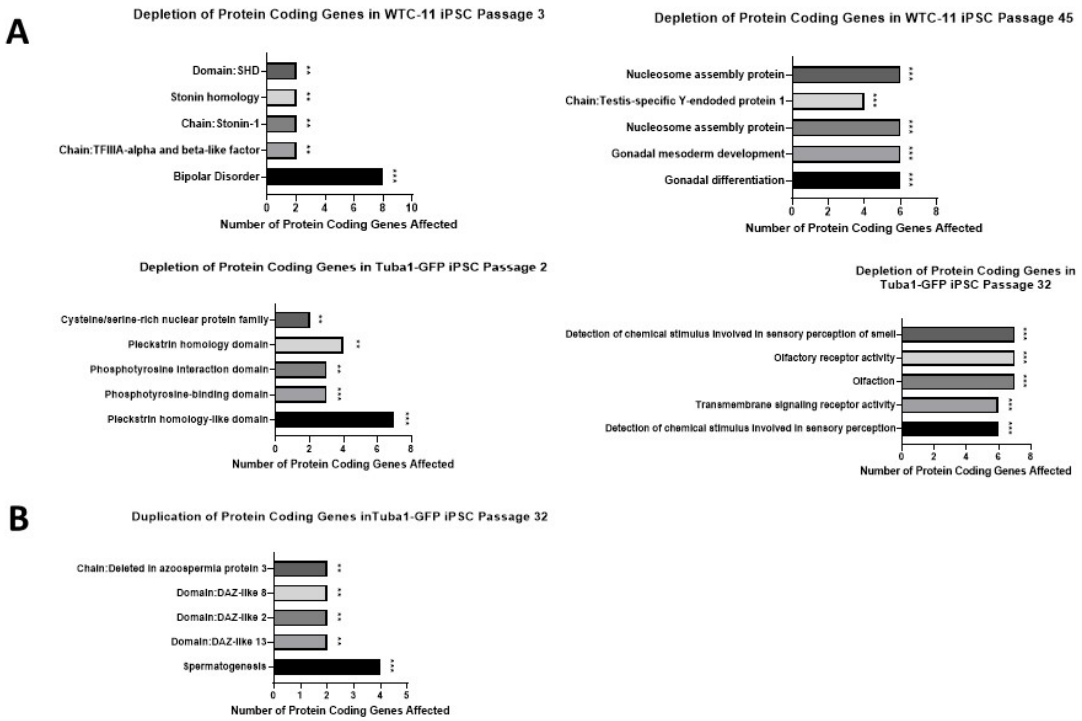

Figure 7: Many gene ontology groups are disproportionately affected by structural variants. **A** Many gene ontology groups were depleted as a consequence of a deletion, insertion, or inversion interfering with the coding sequence of a gene. **B** Only the late passage sample of the Tuba1-GFP iPSC line displayed a significant enrichment of gene ontology groups affected by duplication structural variants, which can lead to overexpression of the gene product.

WTC-11 Passage 3		WTC-11 Passage 45		Tuba1-GFP Passage 2		Tuba1-GFP Passage 32	
AHRR	RPRD1B	A4GNT	RPRD1B	A4GNT	TFCP2	ABR	MARCHF5
AMER1	SLC9A3	AHRR	SLC9A3	AHRR	TGM2	AHRR	MECOM
ANKRD11	STON1	AMELY	SRGAP2C	AKIRIN2	TIAM2	AUTS2	MSC
ANKRD46	STON1-GTF2A1L	ANKRD46	STON1	ASAP2	TMCC3	BACE1	MX1
ARL13B	TFCP2	AUTS2	STON1-GTF2A1L	BTNL2	TMEM50B	BPY2C	MX2
AUTS2	TGM2	BACE1	TANGO6	CBLN2	WASHC2C	BTNL2	OR11H2
BACE1	TMEM50B	CD8A	TBL1Y	CHMP2B	XAGE1B	C8orf44-SGK3	OR4K1
CEP164	TSBP1	CEP164	TGM2	CNOT6L		CDY1	OR4K2
CNOT6L	TSPAN8	CNTNAP2	TSPAN8	CSRNP2		CEP164	OR4K5
CNTNAP2	TYW1B	DAPL1	TSPY1	CSRNP3		CMSS1	OR4M1
CSRNP2	XAGE1B	DGKI	TSPY10	DNM2		CNOT6L	OR4N2
DCUN1D3	ZNF484	DMXL2	TSPY2	DOCK9		CNTNAP2	OR4Q3
DGKB		DOCK9	TSPY3	DSP		COL26A1	PEX7
DOCK9		DTWD2	TSPY4	DZIP1L		CSRNP2	PREP
DTWD2		DZIP1L	TSPY8	EPS8		DAPL1	PRKCE
EFCAB5		EFCAB5	TSPY9P	EXOC3		DAZ3	RAB31
ERI2		ESR1	TYW1B	FAM214A		DAZ4	RPRD1B
EXOC3		EXOC3	UBP1	GULP1		DCUN1D3	RRP7A
FSTL1		FBXL2	XAGE1B	HYDIN		DOCK9	SELENOS
GRM3		FSTL1	ZNF385D	IFNGR2		DTWD2	SERHL2
HTR5A		IGF2BP2		ITGB1BP1		EFCAB5	SGK3
IFNGR2		JAM3		JAZF1		ERI2	SLC9A3
IGHV3OR16-9		JAZF1		LENG8		EXOC3	SNRPA1
JAM3		KCNIP1		MAP3K9		FAM90A1	SRGAP2
JAZF1		LYRM1		NINJ2		FILIP1L	STON1
LYRM1		MOBP		OR4F5		FSTL1	STON1-GTF2A1L
MICU1		MSC		ORC3		FSTL5	TBC1D22A
MSC		NEB		PREP		IFNGR2	TFCP2
NINJ2		OR4F5		RPRD1B		JAM3	TGM2
OR4F5		PREP		SLC9A3		JAZF1	TMEM50B
PEX7		PRKY		STRAP		LYRM1	TYW1B
PREP		RNF138		TANGO6		MAP3K9	XAGE1B

Table 1: Structural Variants Affect Coding Sequences of Many Genes. List of genes whose protein coding sequences are impacted by structural variants. Impacts may be serve, such as complete deletion or truncation or may be more moderate, such as intron variants. Deletions, insertions, duplications, and inversions are indicated by orange, blue, purple, and green respectively.