Perturbation of Replication Dynamics In Human Pluripotent Stem Cells Links Structural and Numerical Chromosomal Instability

By:

Jason Alexander Halliwell

A thesis submitted in partial fulfilment for the requirements for the degree of Doctor of Philosophy

The University of Sheffield,

Faculty of Science,

Department of Biomedical Science.

February 2020

Acknowledgements

Without the help and support of several people, the work presented in this thesis would not have been possible. I would like to start by thanking Peter Andrews. When we met during the UKRMP project, I never imagined that I would be completing a PhD after 4 thoroughly enjoyable years. Before joining your laboratory, I was considering leaving my career in science after several failed attempts to secure a PhD position. However, you gave me an opportunity that I will be forever grateful for. Your enthusiasm and encouragement over the years has inspired me to reach my goals and I am now more passionate than ever.

To Ivana Barbaric, thank you for all the advice, support and mentorship throughout this process. I feel incredibly fortunate to have been supervised by you. You always encouraged me to discuss my ideas and helped me refine them finding solutions for problems that I would never have thought of. I look forward to collaborating with you in the future.

To the members of the lab, you all made coming to work worthwhile even during the times when every experiment I turned my hand to would fail. I've never known such an energetic, intelligent and passionate group of people, who I know have such a bright future ahead of them. I would especially like to thank Tom Frith, who took me under his wing on day one and has been an endless source of encouragement. Also, to Paul Gokhale, who has always been open to discussing ideas with me over a coffee or a beer.

I would also like to thank my family. To my parents, Anna and Michael Halliwell and my brothers and sister-in-law, Chris, Robert, Kimi and Gemma, thank you for always supporting me throughout my education. Whilst undertaking a PhD, I have felt your love, support and pride more than ever. Without you, I would not have had the motivation to continue.

Lastly, to my best friend Kyle Lynd. The term 'best friend' is often misused by those that have yet to experience true friendship, I however, am one of the fortunate few. You sacrificed so much to allow me to move to Sheffield, putting yourself in an

2

uncomfortable position so that I could be happy. This was truly something to behold and an act of generosity that I feel undeserving of, yet I am truly grateful for. Thanks for all the conversations and the good times we've shared, I look forward to our countless future endeavours. Thank you, brother.

Abstract

Human pluripotent stem cells acquire genetic changes on prolonged culture which pose a safety concern for the translation of human pluripotent stem cell-based medicine. The recurrent nature of these genetic changes suggests that first a mutation must occur and then selection and enrichment of that variant should it provide the variant cell with a growth advantage. It is now understood that the mechanism of selection is responsible for the recurrent appearance of these genetic changes. However, the origins and the mechanisms through which these mutations arise in the first place is poorly understood. To ensure the successful translation of human pluripotent stem cell-derived therapies, it will be important to develop culture conditions that enable the expansion of these cells whilst minimising mutation. Current efforts to achieve this have been impeded by a lack of in-depth knowledge of the factors responsible for mutation in these cells.

In this body of work, the breakpoint sequence of a frequent tandem duplication that affects chromosome 20 has been elucidated. Following this, it was possible to infer that these breakpoint regions are susceptible to lesions from replication stress. In contrast to somatic cells, human pluripotent stem cells have an increased susceptibility to DNA damage and mitotic errors that are caused by persistent replication stress. Importantly, the addition of exogenous nucleosides to cell culture medium is sufficient to alleviate replication stress, DNA damage and errors that occur during mitosis. Finally, nucleosides also improved survival of human pluripotent stem cells, demonstrating that replication stress in these cells was a major cause of death during S phase and also responsible for mitotic catastrophe.

Overall, our findings have significant implications, such as allowing the expansion of large numbers of human pluripotent stem cells that are required for medical applications whilst minimising the occurrence of genetic change. These findings will facilitate the safe translation of human pluripotent stem cell-based regenerative medicine applications.

Abbreviations

ES	Embryonic stem (cell)
iPSC	induced pluripotent stem cell
PSC	Pluripotent stem cell
HR	Homologous recombination
NHEJ	Non-homologous end joining
Rb	Retinoblastoma
EdU	5-Ethynyl-2'-deoxyuridine
CldU	5-Chloro-2'-deoxyuridine
IdU	Iododeoxyuridine
RNR	Ribonucleotide reductase
dNTP	Deoxynucleoside triphosphate
DMEM	Dulbecco's Modified Eagle Medium
DMSO	Dimethylsulfoxide
MEF	Mouse embryonic fibroblasts
KOSR	Knock out serum replacement
PBS	Phosphate-buffered saline
DNA	Deoxyribose nucleic acid
RNA	Ribose nucleic acid
SD	Standard deviation
SEM	Standard error of the mean
qPCR	quantitative polymerase chain reaction
SAC	Spindle assembly checkpoint
GSK3β	Glycogen synthase kinase 3 beta
TGFβ	Transforming growth factor beta
FGF	Fibroblast growth factor
RFP	Red fluorescent protein
GFP	Green fluorescent protein
ISCI	international Stem Cell Initiative
Ct	Cycle threshold

Contents

Table of figures

1 Introduction

1.1 A brief introduction and the objectives of this report

Human pluripotent stem cells (PSC), specifically human embryonic stem (ES) cells and human induced pluripotent stem cells (iPSC) are unique in their ability to self-renew indefinitely in culture and retain the capacity to differentiate into any cell type of the human body (Thomson et al., 1998, Takahashi et al., 2007). Unlike human iPSC, human ES cells are found in the early embryo. Specifically, human ES cells are derived from the inner cell mass of the blastocyst stage embryo, where they exist transiently before going on to form all somatic and germ cell lineages (Thomson et al 1998). Human iPSC are generated directly from adult cells through a process known as reprogramming. Over the last 20 years, human PSC have been extensively researched for their application in regenerative medicine, disease modelling and developmental biology (Wu and Hochedlinger, 2011, Murry and Keller, 2008, Pera and Trounson, 2004). To fully and safely exploit human PSC for these applications, it is first necessary to expand these cells into large number whilst maintaining their genetic integrity. However, it was not long after human PSC were first derived that it was discovered that they are subject to genetic changes *in vitro*. These changes provided the genetically 'variant' cells with a selective advantage in culture which enables them to outcompete their genetically 'normal' counterparts (Draper et al., 2004, Olariu et al., 2010).

Through the course of this study, the origins and mechanisms of mutation in human PSC will be explored. In human PSC, advantageous mutations are selected for during culture and it is this mechanism that is responsible for the recurrent nature of these genetic changes. However, very little is known of the origins and mechanisms for how these mutations are acquired. To achieve the clinical translation of genetically stable human PSC therapeutics, a comprehensive understanding of the origins of *in vitro* mutation will enable the optimisation of culture conditions to reduce genomic damage and mutation.

1.2 Human Pluripotent Stem Cells

1.2.1 Research leading to the derivation of human embryonic stem cells

1.2.1.1 Mouse embryonal carcinoma cells and embryonic stem cells

Embryonal carcinoma (EC) cells are the stem cells of teratocarcinomas. These tumours exhibit differentiated tissue of the three primary germ layers that are otherwise associated with the early embryo. It was discovered that the mouse 129 strain frequently developed testicular teratocarcinoma and not only did they contain differentiated tissue of the three germ layers but also pluripotent EC cells (Stevens and Little, 1954). The culture mouse EC cells was first achieved in 1967 by maintaining the cells in Dulbecco's modified eagle media (DMEM) supplemented with 15% calf serum in Petri dishes that had been coated with a feeder layer of irradiated mouse embryo fibroblasts (MEF) (Finch and Ephrussi, 1967). Under these conditions, the cells were able to be maintained for long periods in an undifferentiated state (Finch and Ephrussi, 1967).

At this point, the similarity of mouse EC cells with the inner cell mass cells of the blastocyst stage, early mouse embryo were apparent (Jacob, 1978). The development of culture conditions able to support mouse EC cells then led to the successful derivation of mouse ES cells. After several unsuccessful attempts to derive pluripotent cells from mouse embryos, it was established that previous failures had been a result of harvesting the improper stage of the embryo, insufficient cell numbers being explanted in vitro and culture conditions that were unsupportive of the pluripotent cells (Evans and Kaufman, 1981, Martin, 1981). In 1981, these failures were overcome by harvesting 2.5-day old blastocysts and culturing for 4-6 days. After 2 days it was noted that the trophectoderm had grown out and the inner cell mass had developed and begun to represent large egg-like structures. These were then picked and re-plated, cells that resembled EC cells attached and began to proliferate. They presented pluripotent characteristics, forming teratomas when injected into syngeneic mice and formed embryoid bodies when removed from the MEF feeder layer and most importantly, generated chimeric mice with germline contribution (Evans and Kaufman, 1981).

22

1.2.1.2 Human embryonal carcinoma and embryonic stem cells

Like with mice, teratocarcinomas can form in humans and commonly affect young adults in the form of testicular and ovarian tumours, amongst others (dos Santos Silva and Swerdlow, 1991). By adapting the techniques learnt in mouse studies, it was found possible to propagate human EC cells *in vitro* (Hogan et al., 1977, Andrews et al., 1980). Upon characterisation of human EC cells, distinct differences in cell surface antigen expression were identified, such as SSEA1, SSEA3, SSEA4, TRA-1-60 and TRA-1-81, which differed between species (Andrews et al., 1980, Andrews et al., 1996). However, both human and mouse EC cells were later found to express pluripotency-associated transcription factors NANOG, OCT4 and SOX2 (Mitsui et al., 2003, Schöler et al., 1989, Gubbay et al., 1990). Further, human and mouse EC cells also acquire similar of karyotypic abnormalities that are associated with germ cell tumours. Amplifications to chromosome 12 and 17 are frequently observed in human EC cells (Skotheim et al., 2002, Atkin and Baker, 1982, Rodriguez et al., 1993). In mouse EC cells, chromosome 11 is frequently amplified, which is syntenic with human chromosome 17q, suggesting these variants are selected for tumour growth (Andrews et al., 2005).

Several years later, Thomson et al published their seminal work describing the derivation of human ES cells *in vitro*. Fresh or frozen IVF cleavage stage embryos were cultured to the blastocyst stage before transferring the cells of the inner cell mass to a layer of MEF feeder cells (Thomson et al., 1998). These human ES cells proliferated in an undifferentiated state for prolonged periods and produced teratomas, containing tissue pertaining to the three germ layers (Thomson et al., 1998). Further, they expressed the same cell surface antigens as human EC cells, including SSEA3, SSEA4, TRA-1-81 and Alkaline Phosphatase (Thomson et al., 1998), again highlighting that EC cells were likely the malignant counterparts of ES cells (Andrews et al., 2005).

As the field progressed, much research turned to understanding the signalling pathways responsible for lineage specification. Early work in this area resulted in protocols which were rarely reproducible often suffering from the undefined

combinations of growth factors in different lots of serum. However, with the development of serum-free media, the directed differentiation of human ES cells began to be optimized (Wiles and Johansson, 1999). With these discoveries, the protocols used to generate clinically relevant cell types, such as dopaminergic neurones, amongst many others, have become more defined (Perrier et al., 2004). Twenty years since the *in vitro* derivation of human ES cells, the promise of stem cells in regenerative medicine has begun to be realised. The first clinical trial to treat Parkinson's disease with human PSC derived dopaminergic neurones is now underway in Japan (Barker et al., 2017).

1.2.2 Human induced pluripotent stem cells

The use of human ES cells has been controversial due to ethical and religious objections to the manipulation of human embryos. Furthermore, the foreseen issues of using non-autologous cells in cell therapy and the lack of available ES cells donors has motivated research into generating an alternative source of pluripotent stem cells.

In 2006, Shinya Yamanaka and his colleagues presented their work showing the ability to reprogram mouse adult cells back to a pluripotent state that they termed induced Pluripotent Stem Cells (iPSC) and in 2007 they repeated this using human cells (Takahashi and Yamanaka, 2006, Takahashi et al., 2007). Human dermal fibroblasts were retrovirally transfected to ectopically express four key pluripotency-associated genes now known as the Yamanaka factors; *OCT3/4, Sox2, Klf4* and *c-MYC*. The cells presented a human ES cell-like morphology and when tested presented indistinguishable pluripotent characteristics. Like human ES cells, human iPSC did not express the surface antigen SSEA1 but did express SSEA3, SSEA4, TRA-1-60, TRA-1-81, Alkaline Phosphatase and the transcription factor NANOG. They formed embryoid bodies that expressed genes associated with the three germ layers and when injected into SCID mice they formed teratomas. Lastly, they demonstrated that direct differentiation protocols to midbrain dopaminergic neurones and cardiac myocytes, that were previously established in human ES cells, could be replicated in human iPSC (Takahashi et al., 2007). This discovery presented a breakthrough for the future of autologous cell therapy but has also rapidly expanded

the field of disease modelling. Human iPSC technology has allowed researchers to generate an unlimited source of pluripotent cells which are capable of being differentiated into disease-relevant target cells for their study *in vitro* (Kumar et al., 2018).

Despite these advantages, it is important to highlight some of the problems associated with using human iPSC for clinical use. Early work relied on retroviral integration of the Yamanaka factors which would restrict the use of human iPSC for clinical applications as these retroviruses integrate randomly within the genome with variable copy numbers (Takahashi et al., 2007). However, it was not long before a plethora of transgene-free alternatives were reported and included; lentiviral integration (Carey et al., 2009), non-integrating DNA plasmid (Okita et al., 2008), Sendai virus (Fusaki et al., 2009), micro-RNA (Judson et al., 2009) and the most promising, mRNA based reprogramming, which is clinically relevant, non-viral and non-integrating (Warren et al., 2010).

Another issue relates to the ectopic expression *OCT3/4, Sox2, Klf4* and *c-MYC* used during the reprogramming process. By comparing the enrichment patterns of gene sets associated with pluripotency to the expression profiles of tumours; transcripts of *OCT4*, *SOX2* and *c-MYC* were all highly enriched within poorly differentiated tumours (Ben-Porath et al., 2008). Further, when analysing the ability of iPSC to generate mouse chimaeras, it was found that approximately 20% of the F_1 mice had developed tumours caused by the reactivation of *c-MYC* expression in the iPSC derived tissue (Okita et al., 2007). However, it should be noted that the transplant of human PSC products will be terminally differentiated so the risk of tumour formation from non-proliferative cells is minimised. Despite this, the generation of human iPSC has been fundamental in overcoming ethical restrictions, donor availability and immunological barriers to using pluripotent cells in regenerative medicine.

1.3 Genetically variant human pluripotent stem cells

The application of human PSC in cell-based regenerative medicine will require the generation and expansion of genetically normal undifferentiated cells. However, numerous reports of genetically variant human PSC have been made to the

literature that raise concerns over human PSC therapeutic application. Genetically variant human PSC have been shown to display changes in behavioural characteristics which can confuse experimental conclusions or alter the capacity for the cells to differentiate (Fazeli et al., 2011, Werbowetski-Ogilvie et al., 2009). These findings have raised particular concern for the safe application of human PSC in cell-based regenerative medicine, particularly as some genetically variant cells show signs of abnormal and excessive growth (Werbowetski-Ogilvie et al., 2009).

The origin of these mutations can be broadly classified as source cell mutations or acquired genetic variants. Source cell mutations relate to those mutations found in the starting cell, put differently these are either mutations present in the early embryonic cells or adult cells used to derive ES and iPSC or mutations induced during their derivation or reprogramming. Acquired genetic mutations arise during the process of prolonged *in vitro* cell culture of established PSC lines.

Mutations that arise in the early embryo are generally not viable but if tolerated can become fixed into the differentiated lineages these cells go on to form. In the same way, it is possible to clonally expand variant ES cells from donated IVF embryos. However, these mutations will be present in 100% of the population and so can be easily detected and excluded from future work if the mutation is deemed to be problematic for the intended application. The choice of adult cell used in reprogramming to iPSC should be carefully considered. In one study, fibroblasts and endothelial progenitor cells were acquired from patients of different ages and reprogrammed to iPSC cells to analyse mutations during reprogramming from monoclonal or polyclonal cell sources. Overall, the fibroblasts showed a mutational profile that was consistent with the mosaic nature of a polyclonal source of cells (Rouhani et al., 2016). The monoclonal endothelial progenitor cells had a lower frequency of mutation, although some of these mutations had arisen as a result of their *in vitro* expansion, prior to reprogramming (Rouhani et al., 2016). Further, the same study showed that the process of reprogramming itself was mutagenic at the nucleotide level, which is consistent with reports elsewhere (Rouhani et al., 2016, Cheng et al., 2012b). For these reasons, when generating clinical-grade human iPSC, haematopoietic stem cells could provide a good source of starting material as

they are highly amenable to reprogramming, have a low mutation rate and are also not exposed to environmental mutagens such as UV radiation (Wang et al., 2019).

The methodology used to generate human iPSC could cause potentially dangerous mutations that result from the delivery of reprogramming factors. It was demonstrated, in a dose-dependent manner, that the transfection of reprogramming factors induced replication stress and DNA damage in the source cell lines that could lead to mutations in the daughter human iPSC (Ruiz et al., 2015). Multiple studies have compared the different reprogramming approaches and their relative impact on the genetic integrity of human iPSC (Sugiura et al., 2014, Cheng et al., 2012a, Bhutani et al., 2016, Schlaeger et al., 2014, Ruiz et al., 2015).
Upon sequencing, the studies reported a high incidence of SNP mutations that were not present in the parental source cell line, and so, must have been introduced during reprogramming. The mRNA reprogrammed lines were deemed to be the safest as they possessed fewer overall SNPs in exonic regions (Sugiura et al., 2014, Cheng et al., 2012a, Bhutani et al., 2016, Schlaeger et al., 2014) due to the non-integrating and short half-life of mRNA transfections (Warren et al., 2010).

Acquired genetic mutations present a much larger problem to the human PSC field. These mutations are not present at the stage of derivation but arise during subsequent culture. Both single nucleotide polymorphisms and structural and numerical chromosomal instabilities afflict human PSC and arise during culture in a non-random nature. Assuming that the initial mutations occur randomly across the genome, this non-random nature of observed mutants most likely reflects particular genetic mutations providing the variant cell with a growth advantage (Olariu et al., 2010). As these mutations could confer a behavioural change in the pluripotent cell or the differentiated derivatives, they pose a potential threat for clinical use. Regular screening of PSC cultures is necessary to monitor for the appearance of genetic change.

1.3.1 Acquired genetic mutations in human pluripotent stem cells

The appearance of large karyotypic changes in cultures of human ES cells was first reported in 2004 (Draper et al., 2004, Cowan et al., 2004, Rosler et al., 2004). It is not an uncommon observation that dividing cells acquire karyotypic changes. However, the non-random retention and enrichment of certain genetic variants over prolonged periods of *in vitro* culture showed that they imparted a growth advantage to the genetically variant cell (Baker et al., 2007, Olariu et al., 2010).

With the potential implications of using genetically variant human PSC for medical applications, the International Stem Cell Initiative characterised both the prevalence and types of mutations in a large cohort of unrelated and ethnically diverse human PSC (Amps et al., 2011). Of 122 human ES cell lines analysed by G-banding karyology, 34% were found to be variant. When the cohort was divided into early and late passaged paired cell lines it was observed that only 14% of the early passage lines had acquired changes whereas 33% of the late passage were karyotypically variant, indicating that prolonged culture increased the likelihood of acquiring karyotypic changes (Amps et al., 2011). This aligned with a previous report of a karyotypically diploid cell line that became aneuploid for chromosome 17 in 76% of cells after 22 passages. When the same cell line was passaged a further 17 times it was present in 95% of the cells (Draper et al., 2004)

When the types of changes were categorised, it was noted that chromosomal amplifications were far more prevalent than deletions. Over half of all the karyotypic changes noted were amplifications affecting chromosome 1, 12, 17, 20 or X (**Figure 1.1**). To a lesser degree, deletions to chromosome 10, 18 and 22 were also observed (**Figure 1.1**). Strikingly, all chromosomes were affected to some degree in the profiled lines, other than chromosome 4. As karyology can only sensitively detect structural changes above 5 MB, SNP karyotyping was employed to detect structural variants down to 1 kb in length. Based on this analysis the frequency of amplification to chromosome 20q became apparent (**Figure 1.1**). Of the karyotypically normal cell lines analysed, 25% were found to have acquired an amplification to chromosome 20 that encapsulated the q11.21 region, making this the most prevalent structural variant categorised by this study (Amps et al., 2011).
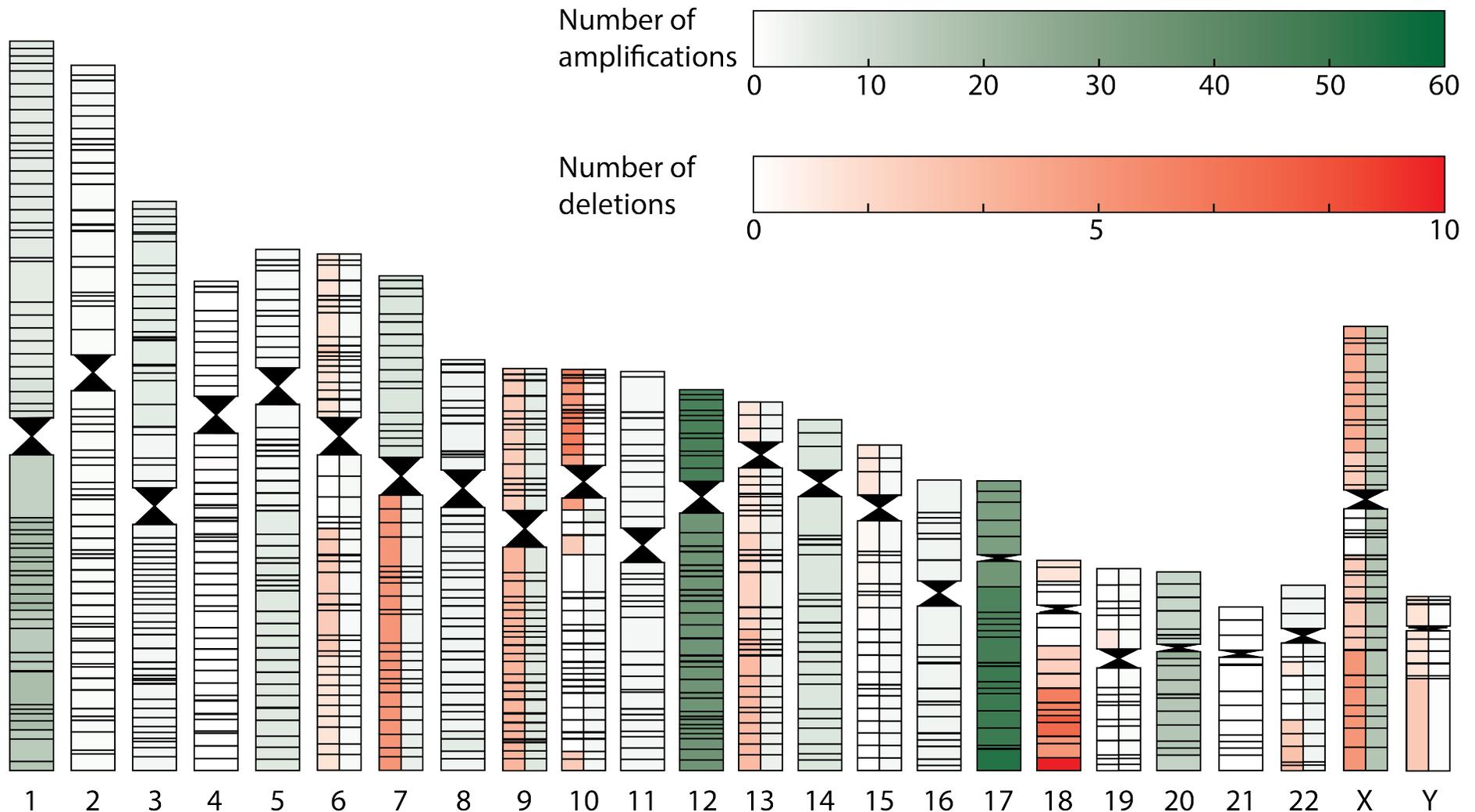
Figure 1.1 | Heat map ideogram of the reported chromosome changes in human PSC. Heat map depicting the frequency of amplification (green) or deletions (red) to each chromosome that has been reported to the literature. White coloured regions are rarely amplified or deleted regions.

Following on from this study, the types of changes have continued to be characterised and provided an intriguing insight into the development of more complex karyotypes. Through routine cytogenetic analysis of 30 diploid human ES cells over a period of 18 months, 16 of these lines acquired karyotypic changes. Of these 16 variant lines, 15 were observed to have amplified chromosome 17. Yet, in only 4 of these was chromosome 17 seen as the sole change, with the other 11 variant lines having a combination of chromosome 17 with, on the most part, chromosome 12 or X. Chromosome 12 was observed in 9 out of 16 variant lines and only once was it observed as the sole karyotypic change (Baker et al., 2007). What is striking about this observation is the similarity of these chromosomal changes to the karyotypes of human EC cells as the addition of chromosome 12 and 17 are also observed in teratocarcinoma (Skotheim et al., 2002, Atkin and Baker, 1982, Rodriguez et al., 1993). It has been suggested that genetically variant human PSC may fit into a spectrum of transformation where EC cells represent complete abnormality at one end and newly derived genetically stable human PSC lies at the other (Andrews et al., 2005).

With the advent of next-generation sequencing, the ability to probe the genome with greater resolution has been achievable. With this technology, it has been possible to detect protein-coding mutations throughout the genome that arise during extended culture of human PSC. From whole-exome sequencing of a cohort of 140 human PSC, it was discovered that *TP53* was the only gene that was subject to more than one mutation (Merkle et al., 2017). Out of the 140 human ES cell lines tested, 5 contained 6 mutations to the DNA binding region of p53 and all presented a dominant-negative phenotype that had previously been characterised to cause cancer (Merkle et al., 2017). The same group then obtained one of the affected lines at an earlier passage and monitored that the frequency p53 variant cells increased over subsequent passages (Merkle et al., 2017). This result suggested that, like large karyotypic changes, *TP53* mutations were providing a selective advantage to the variant cells.

A great deal of resources has been dedicated to understanding what drives selection so that selective pressures can be minimised in cultures of human PSC and thereby suppress the selection of variant cells to maintain stable and safe cultures for

downstream applications. Normal human PSC cultures require passaging at a ratio of 1:2 to 1:4 every 4-5 days. Considering the cell cycle time of human PSC is less than 24 hours this would suggest that approximately 90% of cells are lost between passages (Olariu et al., 2010). This huge loss in cells as a result of passaging is an example of the restrictive bottlenecks placed on human PSC through *in vitro* culture. Time-lapse microscopy experiments have been used to examine the re-plating bottlenecks in more detail. By tracking single cells from paired isogenic lines that were either diploid or karyotypically variant, it was found that the variant cells had a survival rate of 82% compared to only 50% in the diploid cells following re-plating (Barbaric et al., 2014). Likewise, in a separate study, the survival of chromosome 20 variant cell lines with their diploid counterparts was assessed. It was found that 59% of the variant line survived compared to only 30% of diploid counterparts, with the diploid line more likely to express cleaved caspase-3, a marker of early apoptosis (Avery et al., 2013). Together, these data provided evidence that the survival of human PSC, particularly after re-plating was the first selective bottleneck encountered (Barbaric et al., 2014). Further analysis showed that of those cells that did survive, far fewer diploid cells re-entered the cell cycle and survived following the first or second cellular division demonstrating that the second and third restrictive bottleneck post-plating was the capacity of these cells to re-enter the cell cycle post-plating and survive post division (Barbaric et al., 2014).

These studies would suggest that the appearance of anti-apoptotic mutation during culture provides a strong selective advantage to genetically variant cells during passaging and may provide a mechanism by which the recurrent nature of chromosomal changes occurs in human PSC.

1.3.1.1 Characterised and candidate apoptosis-associated driver genes found on recurrently acquired mutations in human pluripotent stem cells

One challenging aspect of research into advantageous karyotypic changes in human PSC is the identification of the driver genes in the mutated regions. Often, the size of the genomic region effected is so large that identifying a single driver gene or combination of genes is difficult. However, in certain cases, the persistent amplification or deletion of a common small region has narrowed the window of

possible driver genes allowing for its identification. In other cases, candidate driver genes have been suggested based on their predicted gain or loss of function. The driver genes and chromosome locations are discussed below.

1.3.1.2 Chromosome 20

One of the most frequently observed and best characterised karyotypic change that affects cultures of human PSC is the amplification of a proximal region of the long arm of chromosome 20. In the ISCI study, over 20% of the tested lines possessed an amplification of part or all of chromosome 20 (Amps et al., 2011). Due to the nature of this study, it was possible to identify a common amplicon of 0.55MB in the 20q11.21 region. This region is also commonly amplified in a number of cancers that include breast carcinoma, lung cancer and gastric cancer, to name just a few (Tanner et al., 1996, Guan et al., 1996, Tonon et al., 2005, Jin et al., 2015).

Thirteen annotated genes lie within this minimal amplicon but only three of these are expressed in human PSCs; *HM13, ID1* and *BCL2L1* (Amps et al., 2011). Of these genes *BCL2L1*, that is expressed as the BCL-XL anti-apoptotic splice variant in human PSC, was the most likely candidate driver gene. By overexpressing *BCL2L1* in wildtype cells and comparing it to their isogenic chromosome 20 variant counterparts it was shown that these cells had comparable growth characteristic and ability when overtaking diploid cells in culture. Further, by using a small molecule inhibitor of BCL2L1 the clonal advantage of the CNV cells was removed (Avery et al., 2013). *BCL2L1* was shown to provide a survival advantage to the variant cells through resistance to apoptosis that was particularly beneficial during re-plating.

1.3.1.3 Chromosome 17q

Amplifications to chromosome 17 present as a gain of the whole chromosomes or as an amplification of the long arm. The human chromosome 17 is highly syntenic with mouse chromosome 11, which is commonly amplified in mouse ES and EC cells and would suggest a common driver gene across these species. Due to the size of the amplified region, it has been difficult to determine which gene or genes are providing the variant cells with a growth advantage. It's been suggested that acquired genetic

changes in human PSC help them to adapt to *in vitro* culture in a similar way to how the genetic changes of human EC cells facilitate malignant transformation. However, even when normal human PSC are injected into mice they form teratomas, consisting of only differentiated tissue whereas teratocarcinomas are malignant and contain undifferentiated EC cells (Oosterhuis and Looijenga, 2005). This would suggest tumour formation by ES cells was independent of EC cells and the ability of ES cells to form tumours was due to an embryonic-like differentiation and expansion of primary tissue (Blum et al., 2009). If this is the case, there must be a common gene expressed in human ES cells and teratomas but not in differentiated embryoid bodies, which unlike the differentiated tissue in teratomas, becomes cystic and ceases to divide much sooner (Blum et al., 2009). Blum et al, discovered the *SURVIVIN* gene was expressed in ES and ES-teratomas but downregulated in mature embryoid bodies and identified this as the likely candidate (Blum et al., 2009). Interestingly*, SURVIVIN* is located on chromosome 17q25 that is within the commonly acquired genetic variant seen in human PSC. Inhibition of SURVIVIN leads to apoptosis in human PSC and cancer cells, yet it did not affect normal somatic cells (Blum et al., 2009, Mesri et al., 2001, Ma et al., 2006, Yang et al., 2004). Amplification of *SURVIVIN* would provide a selective advantage by minimising apoptosis and potentially enhancing tumorigenicity in these cells (Blum et al., 2009). As *SURVIVIN* is an oncofetal gene, it could be argued that the presence of this variant would present little risk to the patient in cell-based therapy as the intended cell product in regenerative medicine would be terminally differentiated adult tissue that does not express this gene.

1.3.1.4 Chromosome 12p

Amplifications to chromosome 12 always include the short arm. The driver gene or genes in this region have yet to be identified, although one particular candidate has often been discussed. *NANOG* is the gene that encodes the homeobox protein transcription factor that was identified as a key component in promoting self-renewal (Amps et al., 2011). In the inner cell mass, NANOG deficient cells lead to a failure of these cells to go on and generate the epiblast and, NANOG deficient human ES cells lost pluripotency and differentiated into extraembryonic endoderm defining it as a critical factor for maintaining pluripotency (Mitsui et al., 2003). Gene manipulation

that results in the amplification of *NANOG* presents cells with a growth advantage through its capacity to remove the dependency on factors present in conditioned media or those released from the MEF feeder layer (Darr et al., 2006). However, It was also noted that through the amplification of *NANOG*, *LECTIN-1* involved in promoting apoptosis, is downregulated and enhanced cell survival (Darr et al., 2006).

1.3.1.5 Chromosome 1q

Through mapping 74 unique amplifications and 32 translocations involving chromosome 1 in human ES cells a common minimal amplicon was observed to include the 1q32.1-32.2 region (Unpublished data; McIntire et al. WiCell, Madison, WI, USA). Of the expressed genes in this region, interphase FISH identified that the *MDM4* gene was amplified in all cases tested (Unpublished data; McIntire et al. WiCell, Madison, WI, USA).

MDM4 is a p53 regulator and interacts with the p53 transcription factor binding domain to suppress its cellular response to stress (Haupt et al., 2019). MDM4 is critical for development in the early embryo and its loss results in lethality which can be rescued through the elimination of p53 (Parant et al., 2001, Finch et al., 2002, Migliorini et al., 2002). It has been previously shown that SNP mutations to p53 act to infer a growth advantage in human PSCs (Merkle et al., 2017) and so it could be reasoned that dysregulation of other genes that regulate p53 would confer a similar growth advantage. In the study of retinoblastomas, CNVs of MDM4 result in a distinct growth advantage through increased cell proliferation and survival (Danovi et al., 2004, Laurie et al., 2006) making it an obvious candidate gene on chromosome 1 that warrants future exploration.

1.3.1.6 Chromosome 18

The pro-apoptotic *NOXA* gene maps to 18q21.32 and is part of the BCL2 family. It has been previously shown that the pro-apoptotic family members of the BCL2 family are expressed in human PSC at levels far greater than seen in primary cells and of these, *NOXA* showed the highest gene expression with a relative ratio of 50:1

compared to other primary cell lines (Madden et al., 2011). However, it should be noted that the anti-apoptotic gene, *BCL-2* is also found within the deleted region on chromosome 18 and, although the expression of this gene was also moderately amplified, the net gain from a deletion could be to suppress apoptosis (Madden et al., 2011).

At this stage, we can only speculate what the driver gene on chromosome 18 may be but given the high expression of the pro-apoptotic gene, *NOXA*, this would be a worthy candidate.

1.3.1.7 Chromosome 10 and 22

Despite deletions to chromosome 10 and 22 being classified as recurrent in human PSC culture, their infrequent nature has meant discussion about the mechanisms or driver gene function have been neglected.

The appearance of deletions to chromosome 10 in human PSC are particularly rare. The literature has shown that recurrent deletions incorporate the distal regions of the p arm, 10p15.3 (Baker et al., 2016). In cancer, a loss of heterozygosity has been previously reported to include this region, yet very little is known about functions of genes in this area or even whether their loss could present a growth advantage to *in vitro* cultured human PSC (DeScipio et al., 2012). In the context of cancer, *ZMYND11* acts as a tumour suppressor gene by repressing a transcriptional program that is essential for tumour growth (Masselink and Bernards, 2000, Velasco et al., 2006, Wang et al., 2014). The second gene in this location is *DIP2C*, which has been noted for its role in breast cancer (Jiao et al., 2012). Gene expression analysis performed on cell lines with either one or two copy deletions of the *DIP2C* gene showed gene set enrichment that indicated a function in epithelial to mesenchymal transition, apoptosis and angiogenesis (Larsson et al., 2017).

Candidate genes on chromosome 22 could also include other regulators of apoptotic function. As discussed earlier, the relative gene expression of pro-apoptotic *BCL2* family member, *BIK*, is much higher in human PSC than in somatic cells and is found on Chromosome 22 (Madden et al., 2011). Its deletion would confer a growth

advantage through the suppression of apoptosis during re-plating and culture of human PSC. Although, at this time further research is required to elucidate the mechanisms of selective advantage possessed by these genetic variants.

1.3.1.8 Single nucleotide polymorphisms to *TP53*

Upon classifying mutations to the *TP53* gene in human PSC, Merkle et al discovered that all mutations were missense and mapped to the most frequently disrupted residues in human cancer that involve p53 (Merkle et al., 2017). Each mutation acts in a dominant-negative fashion and eliminated the p53 function in regulating apoptosis, cell cycle progression and genomic instability in human cancers (Willis et al., 2004).

In response to DNA damage, human PSC increase p53 expression leading to apoptosis or loss of pluripotency-associated transcription factors, NANOG and OCT4 resulting in spontaneous differentiation (Qin et al., 2007, Grandela et al., 2007). Knockdown of *TP53* has been shown to impart a growth advantage in human PSC through enhanced proliferation by promoting G1 to S cell cycle progression, resistance to apoptosis, resistance to differentiation and higher cell survival with greater DNA damage (Amir et al., 2017). The loss of p53 function and the resulting resistance to DNA damage may be of particular advantage to human PSC as these cells are subject to high levels of DNA damage during routine culture (Vallabhaneni et al., 2018).

1.3.2 Methods to detect genetic variants in human PSC

It is crucial that cultures of human PSC are monitored for the appearance of genetic variants over time to ensure the validity of results attained during research and to safeguard from potentially harmful mutations in cell-based therapy. Different methods of detection are available, all have their limitations, with some more appropriate to certain applications than others. For instance, when screening cell lines intended for clinical applications, it is critical that the approach used is highly sensitive and can detect variant cells even when they are present in very small sub-populations.

The assays used to detect genetically variant human PSC can be broadly categorised as i) indiscriminate, whole-genome probing or ii) probed targeting of known loci that are commonly mutated. G-banding analysis of metaphase spread is an example of an indiscriminate approach and was employed during the first report of genetic variant human PSC (Draper et al., 2004). Colcemid condensed chromosomes that have been Giemsa stained present distinct banding patterns. These bands represent light and dark patches of highly condensed heterochromatin and less condensed euchromatin respectively. Chromosome number and structural variations are detected through the differences in the Giemsa staining. The G-banding approach is labour intensive and often requires the skills of a trained cytogeneticist. Its use in the detection of small sub-populations of variant cells in mosaic cultures has been called into question. Routinely, a small sample size of 30 metaphases are chosen as a representative of the whole culture, with so few cells sampled variant cells may be missed (Baker et al., 2016). Further, it has been shown that three genetically variant human PSC with amplification to chromosome 12 and/or chromosome 17 presented condensation defects when compared to their genetically normal counterparts (Lamm et al., 2016). During analysis, appropriate metaphases are chosen partly on the quality of chromosome condensation. Condensation defects in variant cells may bias analysis and under-represent the variant population. However, through mixing experiment of isogenic variant and normal cells, results confirmed sensitivity matched what was expected from random sampling (Baker et al., 2016). The sensitivity of G banding for detecting mosaic sub-populations of variant cells was 18% when 30 metaphases were analysed (Baker et al., 2016). When the analysis was expanded to 500 metaphases, the sensitivity of detecting down to 1% was possible, although sampling so many metaphases was impractical and time-consuming (Baker et al., 2016).

Small CNVs, such as the tandem duplications, frequently occur on chromosome 20. These often require alternative methods of detection as G-banding can fail in detecting changes below 5Mb in size (Steinemann et al., 2013). The detection of small amplifications and deletions typically require techniques such as array Comparative Genome Hybridisation (aCGH) and SNP arrays (Amps et al., 2011). These approaches present a global screening approach that can detect CNVs down

to 1Kb in size. In the case of SNP arrays, the detection CNVs are revealed by an increase or decrease in nearby SNP markers or, in aCGH through a comparison of copy numbers to a reference (Rassekh et al., 2008). The detection of mosaic populations of cells is ultimately limited by the detection of the variant signal above the background of the diploid cell signal. The detection limit of these approaches when detecting small CNVs was shown to be approximately 10-15% from mixing experiments (Cross et al., 2007, Valli et al., 2011).

Interphase FISH allows for the detection of chromosomal variants in interphase cells through the hybridisation of fluorescent probes to highly complementary nucleic acid sequences. The size of the CNV does not restrict its detection, although when 100 interphase nuclei were analysed, the presence of false-negative signals limits the detection of mosaic sub-populations of variant cells to 5% (Baker et al., 2016). False-negative signals are infrequently observed when amplifications occur as translocations or aneuploidies, because of the good spatial separation of the amplified chromosomal region to the wild type regions. However, false-negative signals still arise when two signals overlap as a consequence of imaging the nucleus, a 3-D object, in 2-D. Comparatively, false-negative signals arise frequently in amplifications that present themselves as tandem duplications. In this scenario, the amplification occurs in tandem and the distance to the amplification is small, increasing the likelihood that the two signals will overlap or that the proximity is unable to be resolved by fluorescence microscopy. As such, the presence of false-negatives are more frequently observed with small tandem duplications to chromosome 20 in human PSCs. However, when detecting aneuploidies involving chromosome 17 a sensitivity of 1% can be achieved when 1000 interphase nuclei are analysed (Baker et al., 2016).

Recently, a rapid and affordable qPCR-based approach was developed. It requires no specialist cytogenetic skills and can be completed within a day (Baker et al., 2016, Laing et al., 2019). Primers are designed for specific chromosomal loci known to be commonly amplified or deleted. Using gDNA extracted from cultures of human PSC relative to a diploid calibrator cell line the copy number can be calculated. The approach is particularly useful for rapidly screening cell lines and can detect the

38

presence of variant sub-populations above 10% (Baker et al., 2016). However, this approach is limited as it requires prior knowledge of the CNV that is to be detected.

Next-generation sequencing approaches would represent the future of both CNV and SNP detection in human PSC. Before the advent of next-generation sequencing, Sanger sequencing approaches offered an alternative to conventional cytogenetic techniques. Although Sanger sequencing is capable of detecting heterogeneous subpopulations, its limit of detection is 20% (Zagordi et al., 2010). Next-generation sequencing revolutionised the sensitivity of sequencing approaches due to its vastly high throughput nature. However, due to this throughput, NGS has a notoriously high error base calling rate of between 0.1-1% (Salk et al., 2018). The technical advancements of computational and biochemical approaches have enabled the detection of low-level genetic variants present in less than 1% of the sequenced DNA (Zagordi et al., 2010, Ley et al., 2008). Although NGS has not yet been applied to the detection of genetic mosaicism in human PSC, these studies have demonstrated the potential sensitivity, that as yet is unrivalled by other approaches.

Single-cell NGS may still further improve the sensitivity of detecting low-level sub-populations of genetic variants. It could be envisioned that sample sizes of 10,000 cells or greater, could be sequenced per sample making detection limits as low as $0.1X10^{-3}$%. Despite recent advancements in single-cell omics technology, there are still several technical challenges that need to be overcome before it can be utilised in detecting genetic heterogeneity of human PSC cultures. One challenge presented by single-cell sequencing is the isolation and accurate amplification of DNA from a single cell. Each cell contains approximately 6pg of DNA which is insufficient for NGS and so requires amplification. MDA amplification uses Phi29 DNA polymerases that offer high fidelity and low error rates. It provides the best genome coverage of ~75% of the current techniques (Dean et al., 2001) but its non-uniform amplification distorts genome copy numbers making it unsuitable for copy number variant detection (Voet et al., 2013). Alternatively, MALBAC detects copy number variants with high efficiency but has a high false-positive rate that makes it impractical for single nucleotide variant detection (Lasken, 2013). Possibly the best approach for accurate copy number variant detection is DOP-PCR and despite amplification only

providing 40% genome coverage it has been applied to the determination of rearranged cancer genomes (Navin et al., 2011).

Currently, the expense of performing both whole sample and single-cell NGS for the routine detection of genetic variants is too high and computational analysis requires a highly trained bioinformatician. Further, the extent of genetic heterogeneity between cells in a single culture will make it difficult for the stem cell community to understand which mutations present a safety issue and which can be ignored. The application of NGS technology is continuing to be explored and with rapid advancements in its technology, its future application in the detection of low-level genetic mosaicism in human PSC could soon become a reality.

Table 1.1 provides a summary of the various approaches available for the detection of genetically variant human PSC and various parameters that should be considered when choosing one approach over another.

Table 1.1 Summary table of the common approaches for the detection of genetically variant human PSC

Technique	Sensitivity	Suitable for non-dividing cells	Genome-wide / region-specific	CNV / SNP detection
Karyology (G-banding)	10% (30 metaphases)	No	Genome-wide	CNV >5MB in size
Interphase FISH	10-20% (100 interphases)	Yes	Region-specific	CNV
qPCR	20%	Yes	Region-specific	CNV & SNP
Array CGH	10-15%	Yes	Genome-wide	CNV >1kb
SNP array	10-15%	Yes	Genome-wide	CNV >1kb
NGS	0.1-1% *	Yes	Genome-wide	CNV & SNP
Single-cell NGS	0.1×10^{-3}% *	Yes	Genome-wide	CNV & SNP

*Sensitivity not yet tested using human PSC

1.4 Origins of mutation in human pluripotent stem cells

Although much is known about the selective pressures inflicted on human PSC, relatively little is known about the origins and mechanisms that generate acquired mutations during culture. By understanding the mechanisms of selection, it may be possible to slow the rate at which variants cells overtake a culture, yet it is unlikely to stop mutations from arising in the first place. However, as many of the substrates of genetic instability, such as DNA damage, replication stress and mitotic error also lead to apoptosis it is conceivable that by reducing these it may lead to a reduction in mutation whilst also alleviating the selective pressure brought about through cell death.

The slow progress in this field is a consequence of poor sensitivity of detecting variant human PSC when they arise. As discussed in 1.3.2, most methods used to detect variant human PSC are unable to detect a mosaic population of cells below 10% (Baker et al., 2016). Monte Carlo simulations have shown that if a culture of human PSC contains 0.2% variant cells, it takes on average more than 10 passages before the variant population reaches 10%, by which time any conclusions on what may have caused the mutation will not be very informative (Olariu et al., 2010). Alternatively, the frequency of new mutation or mutation rate could be used to test the origins of mutation, although these involve the measurement of single nucleotide polymorphisms that arise through a different mechanism than structural and numerical instabilities, such as those acquired by human PSC. Further, the ability to perform these experiments is confounded by the selection of those genetic variants that provide a growth advantage and as such would bias the mutation rate towards that sub-populations of cells. Despite this, a recent report has estimated that the mutation rate in several human ES cells and human iPSC was between $0.8\text{-}1.7\text{X}10^{-9}$ mutations per base pair per division (Rouhani et al., 2016). This would suggest the mutation rate in human PSC is appreciably lower than in the soma (Milholland et al., 2017).

To accurately determine the mutation rate in human ES cells, a cloning and whole-genome sequencing approach has been employed by our group. To look at the effect of different growth conditions on the mutation rate, the mShef11 line was

cloned and a diploid clone chosen and subsequently re-cloned into either standard conditions, standard conditions with the addition of Rho Kinase inhibitor (commonly included in human PSC culture to reduce cell loss through passaging) or whilst grown in hypoxic conditions (5% O_2). The subsequent cells were then re-cloned and 20 clones from each condition expanded and whole-genome sequenced before calculating the number of SNP mutations that had occurred during the expansion. From this experiment, the mutation rate of human ES cells grown in standard conditions was calculated at 0.28 and 0.37×10^{-9} single nucleotide polymorphisms per day per base pair (Thompson et al., 2020). This value was comparable to previous estimations (Rouhani et al., 2016) and importantly it was substantially lower than what had been predicted in somatic cell lines (Milholland et al., 2017). Interestingly, they found that culturing human ES cells with Rho Kinase inhibitor did not affect the mutation rate, but the culture of these cells in hypoxic conditions did reduce the number of mutations two-fold. Further, the mutational signature of human ES cells showed a high proportion of C to A transversions, a marker of oxidative damage and *in vitro* culture. When the cells were cultured in low oxygen they showed a reduction in C to A transversions (Thompson et al., 2020). This would suggest that growing human PSC at low oxygen could present a practical approach to minimising the appearance of genetic variants. These findings show that, despite the frequent reports of karyotypic variants in human PSC, the rate of mutation is relatively low.

As demonstrated by Thompson et al, complex experimentation is required to study mutation rate independent of the influence of selection and, due to the relative infrequency of chromosomal instabilities, it is often impracticable to study directly the factors that may affect the origins of mutation (Thompson et al., 2020). However, what has been more informative is understanding the similarities between cancer and pluripotency which overlap with striking similarity. Cancer stem cells reactivate gene networks that are associated to pluripotency and are required for them to become immortal, enable limitless self-renewal and acquire the ability to differentiate into unrelated tissue (Ben-Porath et al., 2008, Pezzolo et al., 2011). The capacity of cancer to endlessly renew and its link to genetic stability has been extensively studied through comparisons to somatic cells. However, little is known whether

human PSC ability to endlessly self-renew could also be driving genetic instability *in vitro*.

Like human PSC, the ability of cancer cells to endlessly renew likely comes from atypical regulation of the cell cycle that leads these cells to rapidly proliferate. The entire cell cycle takes between 24 and 32 hours in somatic cells, whereas in certain cancers and human PSC it is considerably shorter taking between 15 and 18 hours in the latter (Becker et al., 2006, Barbaric et al., 2014, Calder et al., 2013, Sherr, 1996). Nearly 50% of all cancers show inactivation of the p16^{INK4a} that would normally inhibit cyclin D-CDK4/6 phosphorylation of Rb-E2F and is required for cellular senescence (Gonzalez and Serrano, 2006). Human PSC are incapable of senescence and as such, it is no surprise that silencing of p16^{INK4a} is crucial when reprogramming adult cells to iPSC and is responsible for their truncated G1 phase and rapid proliferation (Li et al., 2009a, Becker et al., 2006). In cancer, the silencing p16^{INK4a} and subsequent relaxation of Rb-E2F checkpoint and S phase entry leads to a susceptibility to replication stress, DNA damage and genetic instability (Gadhikar et al., 2018). A susceptibility to DNA damage that is associated with replication has also been reported in human PSC, although its direct cause has yet to be determined (Simara et al., 2017, Vallabhaneni et al., 2018). However, how cancer and human PSC respond to replication stress and DNA damage is very different. For instance, human PSC do not activate CHK1 in response to DNA replication stress and instead opt for apoptosis (Desmarais et al., 2012), whereas CHK1 is required by cancer cells to permit their survival in the face of high replication stress and inhibition of ATR and CHK1 is a cancer-specific synthetic lethal treatment (Sanjiv et al., 2016).

In the following sections the similarities and differences in cell cycle dynamics, replication stress and the DNA damage response in somatic, cancer and human PSC will be discussed in greater detail.

1.4.1 Cell cycle dynamics

The cell cycle is composed of a series of events that lead to cellular division creating two daughter cells *(Nurse, 1997)*. It is divided up into four main compartments that

are controlled by the cycling action of proteins. During G1, the cell grows in size and begins to prepare for DNA synthesis. The proceeding S phase is characterised by the synthesis of DNA where the cell duplicates its genome. Upon completion of DNA synthesis, the cell enters the G2 phase where it then prepares to divide into two daughter cells and finally enters mitosis to complete cell division.

The sequential passing of cells from one phase to another is controlled by a classification of proteins known as the cyclins and their kinase partners, CDK. During G1 the cell begins to build the necessary mRNA and proteins that are required for the subsequent steps. Importantly, cells also spend G1 phase monitoring genomic quality before committing to the synthesis of the DNA. The initiation of the G1 phase is led by the expression of cyclin D-CDK4/6 and pushes the cell beyond a point known as the restriction point. The restriction point is crucial for a cells decision to continue proliferating or to differentiate. Beyond the restriction point, the cell enters the late phase of G1 which is marked by the downregulation of cyclin D-CDK4/6 and upregulation of cyclin E-CDK2. The activity of cyclin E-CDK2 phosphorylates the Retinoblastoma protein (Rb) thereby removing its suppression on the E2F transcription factor family. E2F can then activate the gene transcription of its downstream targets that are required for entry into S phase. DNA synthesis proceeds and its end is marked by the downregulation of cyclin E-CDK2 and the upregulation of cyclin A-CDK1 in G2 phase to initiate the necessary protein synthesis for cell division to occur. During mitosis, cyclin B-CDK1 drives the cells to condense, align and separate the chromosomes to opposite poles of the cell before division takes place (Nurse, 1997).

In human PSC the cell cycle time is considerably shorter than a typical somatic cell and is characterised by a truncated G1 phase (Becker et al., 2006). It was noted that although the G1 phase in these cells was significantly shorter, 2.5 to 3 hours, the time spent in S and G2/M phase was unchanged in comparison to reports from somatic cell lines (Becker et al., 2006). As a consequence of the truncated G1 phase, a high proportion of cells reside in S phase (Filipczyk et al., 2007). How the cells control the progression through the G1 phase is still widely debated with many reports presenting conflicting results. Becker et al, were the first to report on the pluripotent cell cycle and described high levels of cyclin D2-CDK4/6, yet a low

expression of cyclin D1, D3 and cyclin E when observing mRNA expression (Becker et al., 2006, Becker et al., 2010). Later, by immunofluorescence, it was shown that cyclin E was seen in all cells and implied it was constitutively expressed throughout the cell cycle (Filipczyk et al., 2007). The same group found there to be no expression of cyclin D in any of the cells analysed (Filipczyk et al., 2007). Later, studies presented yet more conflicting results. By differentiating human PSC into embryoid bodies, a direct comparison of cell cycle gene expression was made between differentiated and undifferentiated isogenic cells. The group found high expression of *cyclin E1, A2, B1, D1, D3* and high expression of *CDK4* and *6* during G1 phase (Neganova and Lako, 2008). However, the rapid proliferation of human PSC could also be controlled by a lack of expression of CDK inhibitors and not due to elevated expression of G1 related cyclins and CDK. The expression of the INK and CIP/KIP inhibitors, that includes p16^{INK4a}, suppress CDK4/6 activity, were found to be lowly or not expressed at all in human PSC (Zhang et al., 2009). This could explain the unrestricted progression through G1 phase, the relaxed Rb-E2F pathway and the non-phasic expression of cyclin E that's been reported elsewhere (Filipczyk et al., 2007). Further, p16^{INK4a} silencing was found to be necessary for reprogramming and was continued to be silenced in the pluripotent state of ES and iPSC (Li et al., 2009a). It is also important to note the p16^{INK4a} is crucial for cells entering senescence, a state not observed in human PSC (Noda et al., 1994). Although these reports conflict, it seems likely that a combination of these findings is most likely true. Silenced p16^{INK4a} would increase the activity of cyclin D-CDK4/6 and lead to constitutive mono-phosphorylation of the Rb-E2F checkpoint. Subsequent gene targets of the active transcription factor E2F include cyclin E, which may explain reports of its non-phasic expression (Filipczyk et al., 2007).

Despite current efforts failing to fully elucidate cell cycle control in pluripotent cells, it is understood that the differences between pluripotent and somatic cells must be intrinsically linked to the pluripotent state. The pluripotency transcription factor, NANOG, is crucial for maintaining pluripotency and suppressing differentiation, with these functions thought to be linked to the cell cycle. *NANOG* overexpressing human PSC lines showed a greater proportion of S phase cells due to a shortening of the time required to enter S phase (Zhang et al., 2009). When analysis of the transcriptional targets of NANOG was performed, cell cycle genes *CDK6* and

CDC25A were identified and found to be upregulated in the *NANOG* overexpressing lines (Zhang et al., 2009). CDC25A is important in G1 phase progression and acts to remove suppression on cyclin E and CDK6 and would support claims that the lengthening of G1 phase precedes differentiation of human PSC after the loss of NANOG (Stead et al., 2002, Faast et al., 2004, Calder et al., 2013). However, the atypical regulation of the cell cycle may also be responsible for genetic instability that results from extended culture (Ahuja et al., 2016).

Manipulation of the cell cycle by atypical expression of cell cycle components is utilised by cancer to induce replication stress and genetic instability that, within the context of the selective pressures put on the cancer cells, selects for mutations that promote tumour progression. Replication stress can arise from many different sources and so, the exact definition for it can be ambiguous, although it has previously been defined as the slowing or stalling of replication fork progression (Zeman and Cimprich, 2014). Oncogenes and tumour suppressors, such as c-Myc, p16^{INK4a} and cyclin E, can deregulate S phase entry through Rb-E2F and cause replication stress through several mechanisms that alter origin usage during S phase (Hills and Diffley, 2014).

In somatic cells, the timing of initiation of DNA replication from numerous origin sites in the genome is a highly flexible process (Cayrou et al., 2011). The mechanism of oncogene-induced replication stress includes origin over usage, origin under usage and origin re-usage. The aberrant activation of the Rb-E2F pathway has been shown to lead to increased replication initiation and origin firing, the consequences of which are depletion of nucleotide pools, replication factors and an increase in collisions between replication and transcription complexes (Bester et al., 2011, Jones et al., 2013, Halazonetis et al., 2008, Toledo et al., 2013). The addition of exogenous nucleosides to cancers increases dNTP pools and reduces replication stress, DNA damage and genetic instability (Bester et al., 2011). Origin under usage is a mechanism of replication stress and occurs in cancer as oncogenic signals can compromise the origin checkpoint allowing S phase entry with a reduced number of origins (Sherr and McCormick, 2002). The exact mechanism as to how under origin usage leads to replication stress is unknown but it may be due to an increased likelihood of fork stalling due to the greater distance each fork has to cover. This, in

turn, could lead to under-replicated regions which if persist into mitosis may lead to chromosome segregation defects (Burrell et al., 2013, Hills and Diffley, 2014). Elevated CDK levels can cause origins to fire more than once within a single cell cycle. Origin re-usage is a problem unique to cancer cells and is caused by deregulation of Cdt1 when other pre-replication complexes components are overexpressed (Gonzalez et al., 2005). Origin re-usage is thought to generate ssDNA gaps which can become obstacles for the re-replication forks, leading to fork collapse or an increase in the number of active origins during S phase which will deplete replication metabolites and factors much in the same way as origin over usage (Neelsen et al., 2013).

The cell cycle control of cancer cells enables endless proliferation, yet also induces replication stress and DNA damage. Therefore, it seems more than a mere coincidence that rapid proliferation punctuated by a short G1 phase and a susceptibility to DNA damage are unlinked events in human PSC (Simara et al., 2017, Vallabhaneni et al., 2018). However, what is known is that human PSC and cancer cells respond distinctly to replication stress and DNA damage and this likely reflects their different characteristics.

1.4.2 Response to Replication stress

Replication stress not only results from atypical cell cycle control but can also be influenced by endogenous and exogenous sources that include nicks, gaps, stretches of ssDNA, DNA lesions, fragile sites and secondary DNA structures such as those caused by *Alu* elements (Zeman and Cimprich, 2014). As these factors can affect any cell type it is important that cells can respond to replication stress to stop further detrimental effects on the cell. The cell employs various checkpoints throughout the cell cycle that are responsible for monitoring cellular processes and restricting the cell's progression should the integrity of these events be lost (Barnum and O'Connell, 2014).

Checkpoint regulation in human PSC is different from that observed in somatic cell types. In somatic cells, the CHK1 checkpoint is important in the surveillance of DNA replication, (Zhang and Hunter, 2014). ATR-CHK1 is activated in response to ssDNA

formed at stalled forks and results in proteasomal degradation of CDC25A, leading to downregulation of *cyclin E*-CDK2 and stalling of the cell cycle in S phase (Falck et al., 2001, Mailand et al., 2000, Zhao et al., 2002, Sørensen et al., 2003). This, in turn, regulates late origin firing and maintains the stability of collapsed replication forks (Lopes et al., 2001, Feijoo et al., 2001). Complete loss of CHK1 results in apoptosis as a consequence of mitotic catastrophe (Huang et al., 2005), with its inhibition leading to increased origin firing (Katsuno et al., 2009). These findings are surprising in the context of human PSC as replication stress has been shown not to activate CHK1 (Desmarais et al., 2012, Desmarais et al., 2016). When replication was blocked by thymidine or cisplatin, human PSC stalled in S phase and showed an increased sub-G1 population that was attributed to apoptotic cells. It was found that this response to replication stress was due to the inactivation of CHK1 (Desmarais et al., 2012, Desmarais et al., 2016). Looking upstream it was noted that RPA did not organise into foci within the nucleus as the cells failed to generate ssDNA when replication was perturbed. The recruitment of RPA to ssDNA is a crucial event needed for the activation of ATR and CHK1 (Desmarais et al., 2012, Desmarais et al., 2016). Therefore, to overcome the threat of genetic instability in the face frequent DNA lesion, human PSC, unlike somatic cells, undergo extensive apoptosis as a consequence of them lacking cell cycle checkpoints that would normally act in promoting cell cycle stalling and DNA repair. This unique mechanism could reflect the demands of the early embryo where genomic damage or mutation could be catastrophic for the whole embryo.

In cancer, uncontrolled proliferation can promote unfaithful DNA replication that leads to genetic instability and DNA damage (Bartkova et al., 2006, Di Micco et al., 2006). It is common for cancer cells to lack DNA damage response proteins including ATM and p53 (Jiang et al., 2009), this increases the reliance of these cells on the ATR-CHK1 activity to mediate survival from replicative stress (Choi et al., 2011). This point is exemplified by the success of CHK1 inhibitors in cancer therapy. In fact, cancer cell-specific lethality can be further improved by inhibiting CHK1 and ATR simultaneously (Sanjiv et al., 2016). When CHK1 is inhibited, the cancer cells are no longer able to fire from dormant origins and instead rely on ATR to stabilise stalled forks until it is possible for them to restart (Toledo et al., 2013).

The difference in the way that somatic, pluripotent and cancer cells respond to replicative stress is likely due to different needs and characteristics of these cells. In somatic cells, the limited number of divisions reduces the risk of mutation from DNA replication. A full complement of options are available to address replication stress including senescence, replication restart or cell death. Human PSC and cancer cells cannot senesce, but human PSC balance the need to expand rapidly whilst preserving the genetic integrity of the cells by activating apoptosis instead of a replication stress response. Death of cancer cells would limit the capacity for the disease to progress and instead they strongly rely on the replication stress response to allow rapid proliferation to continue in the face of high replicative stress and DNA damage.

1.4.3 Response to DNA damage

Like with replication stress, the cell employs various checkpoints throughout the cell cycle to manage cellular processes in response to DNA damage. In human PSC during G1 phase, CHK2 activates p53 and p21 to cause cell cycle arrest in response to DNA damage to protect the cells from apoptosis (el-Deiry et al., 1993, Dulić et al., 1994). Despite p53 being recruited to, and mRNA expression of p21 being upregulated in response to DNA damage, it was shown that translation of p21 was inhibited through the microRNA family, miR302 (Dolezalova et al., 2012). This lack of p21 protein could account for the propensity of human PSC to undergo apoptosis and it is particularly interesting that p21 knockout MEF lines show a similar inability to activate the G1 checkpoint (Deng et al., 1995, Brugarolas et al., 1995). However, CHK2 checkpoint was functional in G2/M phase in response to IR and strong doses of UV radiation (Momcilović et al., 2009, Momcilovic et al., 2010, Hyka-Nouspikel et al., 2012).

CHK2 is a tumour suppressor and in response to DNA damage activates and phosphorylates many downstream components that function in checkpoint control, DNA repair and apoptosis (Matsuoka et al., 1998, Zhang et al., 2004, Stevens et al., 2003, Hirao et al., 2000). Mutations in CHK2 that result in loss of function occurs in a range of human tumours including breast, colon and prostate (Bartek and Lukas, 2003). Although CHK2 is thought to be important in tumour initiation in these cancer

sub-sets, the actual mechanism of tumorigenesis has not been well defined. However, studies in mice have shown that CHK2 deficiency resulted in resistance to ionizing radiation and defects in p53 mediated DNA damage-induced apoptosis (Takai et al., 2002, Hirao et al., 2000).

The DNA damage response (DDR) pathway is driven by a family of PIKK kinases that include DNA-PKc, ATM and ATR (Falck et al., 2005). It is a highly complex signalling cascade that orchestrates cellular responses to various DNA lesions. Double strand breaks are a highly toxic DNA lesion and can be caused by ionizing radiation, free radicals and the collapse of replication forks. If these lesions are not repaired or if repaired erroneously, they can lead to gross chromosomal instabilities like those observed during the prolonged culture of human PSCs. Further, DSB can activate apoptosis providing selective pressure for mutations that abrogate cell death and enhance the tumorigenicity of cancers (Bartkova et al., 2005, Gorgoulis et al., 2005), a mechanism that is not too dissimilar to the selection of genetically variant human PSCs.

In defence against double-strand breaks, cells possess two mechanisms of double-strand break repair, non-homologous end joining (NHEJ) and homologous recombination (HR). The choice between mechanisms is dependant primarily on the cell cycle stage. NHEJ repair ligates the two ends of a DNA sequence that have been broken by a double-strand break. It is primarily active in G1 phase although not exclusively so. The presence of double-strand breaks is first detected by a protein complex heterodimer Ku70/80 (Smith and Jackson, 1999). This ring-like structure of Ku70/80 binds to the end of the break to recruit and stabilise the binding of DNA-PKc (Walker et al., 2001). If the broken ends are incompatible for ligation then Mre11/Rad50/Nbs1 (MRN) complex of proteins detects and tether the two ends (Lukas et al., 2003, Lavin, 2007) whilst Artemis nuclease opens the DNA hairpin loop and cleaves the protruding 5' and 3' ends (de Jager et al., 2001, Williams et al., 2008, Ma et al., 2002). With each end suitably processed the final step simply involves ligating the two ends together. DNA-PKc phosphorylates XRCC4 that is recruited to the break and stabilises and activates Ligase IV to complete the ligation of the two ends (Calsou et al., 2003, Matsumoto et al., 2000, Leber et al., 1998).

50

A main cause of DSB during S and G2 phase is the collapse of replication forks. As a complementary template strand on the sister chromatid is available during S and G2 phase repair by HR is favoured. The MRN complex of proteins senses these DSBs and activates ATM (Lee and Paull, 2005). ATM then phosphorylates its downstream substrates; H2AX, BRCA1, CHK2 and p53 to aid in chromatin remodelling, DNA repair, cell cycle arrest and apoptosis respectively (Lavin, 2007). CtIp nuclease, activated by the S/G2 CDK activity, initiates the resection of the DSB (de Jager et al., 2001) with more extensive end resection performed by EXO1 exonuclease that forms a section of ssDNA (Mimitou and Symington, 2008). RPA protein then binds the length of the ssDNA to inhibit the formation of secondary structures and causes a switch from ATM to ATR kinase activity (Shiotani and Zou, 2009). ATR acts to prevent the collapse of replication forks through its regulation of several replication processes (Byun et al., 2005) including the restraint of further replication origin firing (Costanzo et al., 2003), avoidance of replication factor pool exhaustion (Couch et al., 2013, Ragland et al., 2013), increases dNTP pools for DNA synthesis (Pfister et al., 2015, Buisson et al., 2015) and activates CHK1 checkpoint to slow or stall the cell cycle (Meyer et al., 2000, Guo et al., 2000, Liu et al., 2000, Bartek et al., 2004). Meanwhile, BRCA1, previously activated by ATM, then displaces RPA and replaces it with Rad51 DNA dependant ATPase (San Filippo et al., 2008). Rad51 forms nucleoprotein filaments that search for similar DNA sequences on the sister chromatid and once found, the 3' strand invades forming a D loop into the dsDNA of a sister chromatid and forms base pairs with the complementary strand (Shinohara et al., 1992, Shinohara et al., 1993, Wyman et al., 2004). Using the invaded strand as a template, DNA polymerases are then recruited to extend the 3' end. At this point the DNA cross over is resolved through three independent mechanisms; double-strand break repair (DSBR), synthesis dependant strand annealing (SDSA) or break-induced replication (BIR). In DSBR the second 3' overhang that wasn't involved in HR also form a Holliday junction. This double Holliday junction is cleaved with nicking endonuclease to cleave each of the DNA double strands, often resulting in cross over between chromosomes (Mimitou and Symington, 2009). During mitotic and meiotic dividing cells, SDSA is often performed and results in non-crossover products. In a process referred to as branch migration the newly extended DNA sequence is released from the invaded strand, leaving the resultant ssDNA overhangs to anneal to their complementary strand on the same

chromosome (Allers and Lichten, 2001). BIR is important for DNA repair at the site of collapsed DNA forks. Fork collapse can be catastrophic and so BIR stimulates repair when only one broken end can invade the homologous template. BIR is initiated with 5' to 3' resection and then strand invasion and DNA synthesis that can copy large portions of the chromosome even up to the telomere. The replication by BIR is distinct from that of regular S phase replication as both the leading and lagging strand are synthesised separately, leaving long stretches of ssDNA (Saini et al., 2013). Multiple rounds of strand invasion and dissociation are performed with lagging strand synthesis performed during each round of leading strand displacement (Smith et al., 2007). However, BIR can be induced by oncogene overexpression or at fragile sites that can be responsible genomic instability similar to those leading to cancer (Minocherhomji et al., 2015, Costantino et al., 2014). Chromosome aberrations can form when strand invasion occurs incorrectly elsewhere on the same chromosome or another chromosome and is often associated with homology to repeated sequences at multiple different sites, such as Alu retrotransposons, a class of repetitive element that makes up more than 11% of the human genome (Umezu et al., 2002)

DNA is also subject to a range of other lesions that are not repaired using NHEJ or HR. Insertion/deletions and base mismatches can occur during DNA synthesis and are removed by first degrading and then resynthesizing of the DNA strand by a process known as mismatch repair (MMR). Oxidation, alkylation and deamination are all forms of chemical damage to DNA bases that are removed and replaced using base excision repair (BER), whereas lesions caused by UV radiation or transcriptional stalling are corrected by removal and re-synthesis of the damaged strand through nucleotide excision repair (NER) (Lindahl and Barnes, 2000).

The repair of double-strand breaks in human PSC is thought to be performed in an HR dominant fashion. The BRCA1 and Rad51 proteins, integral to the HR repair are highly expressed in human PSC (Adams et al., 2010, Vallabhaneni et al., 2018). To further explore the reliance of human PSC on HR or NHEJ, individual knockdowns of ATR and DNA-PKc was performed. When ATR gene expression was reduced an increase in replication stress-induced DNA damage was caused by a loss of HR repair (Adams et al., 2010). Contrastingly, when DNA-PKc was knocked down no

difference was observed in the resolution of DNA damage marked by γH2AX or 53BP1 and suggested NHEJ was not required for DSB resolution in human PSC (Adams et al., 2010). However, when cells were transfected with an I-SceI-DsRed repair cassette, that reports on DNA repair by NHEJ, positive human PSC were detected and suggested an alternative non-DNA-PKc, NHEJ repair was in use in these cells (Adams et al., 2010). They confirmed this by knocking down *XRCC4*, required for NHEJ, and found that the reporter was no longer activated (Adams et al., 2010). This has led to the suggestion that human PSC use a backup NHEJ mechanism for when HR fails, although further work is required to substantiate these claims (Wu et al., 2008, Windhofer et al., 2007, Audebert et al., 2004).

By challenging human PSC in vitro with radiation or genotoxic agents it's possible to monitor the cellular responses to a range of DNA lesion. In response to γ radiation, ionizing radiation, UV and other DNA damaging agents the cells triggered extensive apoptosis, indicating a high sensitivity to DNA damage (Simara et al., 2017, Momcilović et al., 2009, Luo et al., 2012, Maynard et al., 2008, Hyka-Nouspikel et al., 2012). Of the cells that survived, the repair of UV induced lesions by NER was proficient (Hyka-Nouspikel et al., 2012). It has also been reported that human PSC show increased expression of genes related to the pathways of MMR and BER although this does not necessarily correlate to increased protein or enzymatic levels (Maynard et al., 2008, Momcilovic et al., 2010).

Contrastingly, cancer cells are known to depend on DNA repair mechanisms to ensure survival in the face of high replicative stress and DNA damage. Cancer cells can become addicted to repair pathways, and as such these pathways can become effective targets that can be inhibited by drugs to induce apoptosis as a result of stalled replication. There are many different examples of cancer reliance on DNA repair but one of the most well-known are the inherited mutations in HR components, BRCA1 and BRCA2, in breast and ovarian cancers (Bryant et al., 2005). As HR is required in the repair and restart of stalled replication forks, these cells switch to SSB repair or NHEJ repair of stalled forks to ensure replication can continue (Bryant et al., 2005). Further to this, the HR pathway is crucial in the resolution of DNA crosslinks via the Fanconi anemia repair pathway (Michl et al., 2016). Mutation in the Fanconi anemia components, such as *FANC2*, causes a

predisposition to cancer that may result as FANC2 is normally required to activate the S phase checkpoint (Taniguchi et al., 2002). There is significant cross-talk between HR and NHEJ repair of double-strand breaks. Mutations in the NHEJ pathway that include the KRAS mutations are commonplace in acute leukaemia that results in enhanced reliance on the NHEJ pathways and genomic instability and carcinogenesis (Nussenzweig and Nussenzweig, 2007). Additionally, KRAS mutations are classified as an early genetic event in cellular transformation, allowing the proto-oncogene RAS to remain in its active state (Schubbert et al., 2007). The continual proliferation of cancer cells can lead to the incorporation of oxidized DNA bases that can form DNA crosslinks that stall the DNA replication fork. It is not uncommon for cancers to overexpress BER pathway components to remove oxidized bases from the DNA strand. One component commonly overexpressed in human cancer is the rate-limiting BER component, APE-1, that cleaves the phosphodiester backbone to allow for further processing by the BER pathway. Without this, single-stranded breaks form at the sites of oxidized bases, which stall replication forks and lead to double-stranded breaks that cause mitotic catastrophe and apoptosis (Dianov et al., 2003). Atypical DNA repair has also been associated with a predisposition to cancer. The NER pathway processes DNA lesions that result from exogenous sources, such as UV radiation (Nouspikel, 2009). Patients with the autosomal recessive inherited disease, xeroderma pigmentosa, are predisposed to cancer (Rubin, 1998). These patients are highly sensitive to UV induced mutations and chromosomal instability that leads to carcinogenesis (Cleaver, 1968). Cancer is an umbrella term for a host of diseases, and as such, it is no surprise that there is a diverse way in which DNA repair can be both utilised by cancer to enhance survival and, through mutations, may lead to susceptibility to cancer development.

Both human PSC and cancer cells show atypical regulation of DNA repair pathways. Current data would suggest that human PSC respond to DNA damage with a preference of HR mediated repair or apoptosis. In contrast, cancer cells enhanced capacity to repair DNA helps to mediate survival in the face of extensive DNA damage that results from extensive proliferation, although mutations in DNA repair pathways may also initiate tumorigenesis. However, both cancer and pluripotent cells show distinct differences to somatic cells types that reflect the specific needs of these cells. As human PSC represent the cells of the inner cell mass, loss of genetic

integrity could be catastrophic and so the mechanisms employed by these cells could reflect the necessity of remaining genetically stable.

1.5 A reflection on the early embryo

The contrasting control mechanisms discussed here may reflect the early embryonic characteristics of pluripotent cells. The cells of the early embryo have to balance the need to expand rapidly whilst preserving the genetic integrity of the cells, as failure to do so could result in pregnancy failure or developmental defects. Following fertilisation of the egg, the cleavage cell divisions divide the cell cytoplasm while doubling the nuclear mass with each division so that the cell size can be quickly reduced from approximately 120μm of the egg to 10-20μm size of pluripotent cells found within the inner cell mass (Kiessling et al., 2003). To achieve this the cell cycle has limited gap phases, normally responsible for cell growth, and instead the cells cycle through successive rounds of DNA synthesis and cell division. Time-lapse imaging of human embryos up to the blastocyst stage have shown how quickly these cell divisions occur. The second mitotic division takes 11 hours which is followed by two successive synchronised divisions to the 3 and 4 cell stage that last only 1 hour (Wong et al., 2010).

Monitoring cells in the human blastocyst stage embryos have shown that, despite their rapid proliferation, cell numbers of the inner cell mass plateau and this steady state of cells is maintained through apoptosis (Hardy, 1997, Hardy et al., 2003). Why cell numbers in the inner cell mass is regulated in this way has not been fully substantiated in humans, although it is speculated to be crucial for successful development by removing defective or genetically compromised cells and to regulate the growth of the inner cell mass so that oxygen and nutrients can still diffuse efficiently (Hardy, 1997, Hardy et al., 2003). These observations would match what has been observed *in vitro*. Human pluripotent stem cells have a cell doubling time of 30 to 36 hours yet a cell cycle time of approximately 15-18 hours which is accounted for by frequent cell death (White and Dalton, 2005, White et al., 2005, Becker et al., 2006). The extent of cell death in human PSC has been crudely calculated, normal human PSC cultures require passaging at a ratio of 1:2 to 1:4 every 4-5 days. Considering the cell cycle time of human PSC is between 15-18

hours this would suggest that a maximum of 90% of cells are lost between passages (Olariu et al., 2010). As human PSC are sensitive to the activation of apoptosis, it is striking that most if not all the driver genes affected by recurrent mutations, regulate apoptosis.

1.6 Summary and research aims

The study of human PSC has highlighted their capacity to proliferate endlessly and differentiate into any tissue of the human body. These unique characteristics have made human PSC excellent candidates for the generation of cell-based regenerative medicine. However, genetic changes that arise in human PSC as a result prolonged culture (Draper et al., 2004, Olariu et al., 2010) threaten to preclude their use in cell-based therapy and could cast doubt on the reliability of results from human PSC research. Illustrating this concern, a PSC-based clinical trial was recently terminated in Japan upon the discovery of genetic changes that had arisen during culture (Chakradhar, 2016). For these reasons, we must understand the origins of mutation in cultures of human PSC and optimise the *in vitro* conditions to preclude the appearance of potentially harmful mutations.

In this study, I planned to address the mechanisms by which mutations arise in human PSC. In chapter 3, the routine methods for detecting genetically variant human PSC were tested and in addition, a novel assay was developed that improves the sensitivity of interphase FISH. In chapter 4, copy number variant lines identified in the previous chapter were sequenced to define the breakpoints and infer the mechanisms responsible for this mutation. From hypothesis driven by the data accumulated so far, in chapter 5 replication stress as investigated as the underlying mechanism influencing genetic change. This led to a discovery, presented in chapter 6, of a way in which replication stress can be moderated, allowing for new culture methods that enhance the growth dynamics and genomic stability of human PSC.

2 Methods

2.1 Culture of human pluripotent stem cells on MEF, Matrigel and Vitronectin coating.

2.1.1 Preparation of Vitronectin, Geltrex and Matrigel coating. Human PSC were cultured on Vitronectin (VTN-N) recombinant human protein (ThermoFisher Scientific A14700). Culture vessels were coated with 200μL/cm^2 of Vitronectin that had been diluted to 6μg/ml with PBS and incubated at 37°C for at least 1 hour. Alternatively, Geltrex (Gibco 15180617) or Matrigel (Corning 354277) was thawed on ice, diluted 1:100. Culture vessels were coated at 100μL/cm^2 and set at 37°C for 1 hour.

2.1.2 Plating of mitotically inactivated mouse embryonic fibroblasts (MEF). Culture vessels were coated with 0.1% gelatin in PBS and incubated at 37°C for a minimum of 20 minutes. MEF were thawed into DMEM that had been supplemented with 10% FBS (HyCLone SV30160.03). MEF were seeded at a density of 10,000 cells/cm^2 and allowed to attach in an incubator set to 37°C and 5% CO_2 overnight.

2.1.3 Knock out serum Replacement medium (KOSR). Human PSC were fed with KOSR when grown on MEF. KOSR was prepared using Knockout DMEM medium (Thermo Fisher Scientific, 10829018) supplemented with 4ng/mL (Peprotech, 100-018B), 10μM L-Glutamine (Thermo Fisher Scientific, 25030081), 1X Non-Essential Amino Acids (Thermo Fisher Scientific, 11140050), 200nM 2-Mercaptoethanol (Thermo Fisher Scientific, 31350010) and 20% KnockOut Serum Replacement (Thermo Fisher Scientific, 10828028).

2.1.4 Human pluripotent stem cell culture. Human PSC cultured in MEF free conditions were batch fed daily with Essential 8 (in house) (see 2.1.7) or mTeSR™1 (STEMCELL Technologies, 85850) cell culture media that had been pre-warmed to 37°C in a water bath.

2.1.5 Passaging of human pluripotent stem cells. 1ml ReLeSR (STEMCELL technologies, 05873) was added to hPSC cultures. Excess ReLeSR was aspirated

after 1 minute. The flask was agitated until the cells begin to detach. Pre-warmed media was added to collect the detached cells. The cell solution was then split between flasks pre-coated with Vitronectin (VTN-N) recombinant human protein (ThermoFisher Scientific A14700).

2.1.6 Dissociating human pluripotent stem cells to single cells. 1ml of TrypLE (Thermo Fisher Scientific, 12504013) cell dissociation enzyme was pre-warmed to 37°C and added to the culture vessel. The flask was agitated until the cells had detached. The cells were washed with DMEM F12 (Sigma, D6421) and centrifuged for 3 minutes at 1100RPM.

2.1.7 Essential 8 (E8) media preparation. E8 media was prepared from a recipe adapted from a previously published manuscript (Chen et al., 2011) details of the components can be found in **Table 2.1**. A concentrated X50 solution of the E8 supplement was constituted in large batches and aliquoted into 10mL aliquots before being frozen at -20°C. To generate X1 E8, an aliquot was thawed overnight at 4°C and added to 490mL DMEM F12 (Sigma, D6421) before filter sterilisation using 0.22μm filter (Millipore).

Table 2.1. E8 media components.

Component	50X concentrate	Company	Catalogue number
DMEM F12	-	Sigma	D6421
L-ascorbic acid	3200mg/L	Sigma	A8960
Sodium selenium	700μg/L	Sigma	S5261
Insulin	970mg/L	Thermo Fisher Scientific	A11382IJ
NaHCO$_3$	27.15g/L	Sigma	S5761
Transferrin	535mg/L	Sigma	T0665
Glutamax	50X	Thermo Fisher Scientific	35050038
FGF2	5mg/L	Peprotech	100-18B
TGFB1	100μg/L	Peprotech	100-21

<u>2.1.8 Nucleoside supplementation.</u> Embryomax Nucleosides 100X (Merck, ES-008-D) were added to mTeSR cell culture media at a final concentration of 0.5X. All experiments were performed after 72 hours in culture with the supplementation of nucleosides.

2.2 Differentiation of human pluripotent stem cells to mesoderm lineage.
Human PSC were grown for 5 days in E8 (Chen et al., 2011) without FGF-2 and TGF-β but supplemented with 10μM CHIR99021 (Tocris, 4423). Loss of pluripotency was confirmed by RT-qPCR panel of self-renewal, mesoderm, endoderm and ectoderm genes (**Table 2.4**) and by immunofluorescence staining and imaging of NANOG.

2.3 Fibroblast cell culture. Fibroblasts (ATCC, CRL2429) were grown in Iscove's Modified Dulbecco's Medium (Thermo Fisher Scientific, 12440053) with 20% FBS (HyClone, SV30160.03). Cells were passaged using TrypLE cell dissociation enzyme (see 2.1.6) (Thermo Fisher Scientific, 12504013). Cells were maintained at 37°C and 5% CO_2 in a humidified incubator.

2.4 Freezing of cell lines. Cells were harvested with ReLeSR (see 2.1.5) (STEMCELL technologies, 05873) and neutralised in DMEM F12 (Sigma, D6421). The solution centrifuged for 3 minutes at 1100 RPM. The supernatant was removed and the cell pellet resuspended culture media supplemented with 10% DMSO. The cell solution was then aliquoted into cryovials and placed into a Mr Frosty (Nalgene) in a -80°C freezer overnight. The following day the cryovials were transferred into liquid nitrogen for long term storage.

2.5 Thawing of cell lines. Cryovials were removed from liquid nitrogen and placed into a 37°C water bath. The cell suspension was diluted with pre-warmed DMEM-F12 (Sigma, D6421) and centrifuged for 3 minutes at 1100 RPM. The supernatant was removed and the cell pellet was resuspended with pre-warmed media and added to culture vessels that had been supplemented with 10μM Y-27632.

2.6 Single-cell deposition. hPSC were subcloned using single-cell deposition by FACS into MEF coated plates (see 2.1.2). Flasks of cells were dissociated into

single cells using TrypLE cell dissociation enzyme (Thermo Fisher Scientific, 12504013) (see 2.1.6). Single cells were deposited directly into the 96 well plate containing KOSR media (see 2.1.3) supplemented with $10\mu M$ Y-27632 and $50\mu g/ml$ Gentamycin (LifeTechnologies) using a BD FACS Jazz. Single cells were centrifuged for 3 minutes at 1100RPM and 48 hours the media was exchanged to remove the Y-27632 and allowed to form colonies for the next 12 days. The resultant colonies were passaged into larger plates until sufficient material was present to screen for genetically variant or wildtype colonies confirmed by qPCR (see 2.9).

2.7 Genetic diagnostics.

<u>2.7.1 G-banding.</u> G-banding was performed by the Sheffield Diagnostics Genetic Service. Normally, 30 G-banded metaphases were analysed per sample by a clinical cytogeneticist.

<u>2.7.2 Fluorescence *in situ* hybridisation (FISH) for the detection of chromosomal variants.</u> FISH detection of chromosomal variants was performed by Sheffield Diagnostics Genetic Service. Analysis was performed on 100 interphase nuclei per sample that had been probed with BCL2L1 (Chromosome 20q), MDM4 (Chromosome 1q), MPO (Chromosome 17q) or BCL2 (Chromosome 18q) FISH probes.

<u>2.7.3 qPCR for the detection of genetic variants</u>**.** Full details of the protocol can also be found in previous publications (Baker et al., 2016, Laing et al., 2019)

<u>2.7.3.1 gDNA extraction and digestion.</u> gDNA was extracted from cell pellets using the DNeasy Blood and Tissue kit (Qiagen, 69504). DNA quantity and quality was measured using a NanoDrop spectrophotometer (Thermo Fisher Scientific). $1\mu g$ of DNA was digested with $1\mu l$ FastDigest EcoR1 enzyme and $10\mu l$ FastDigest buffer (Thermo Fisher Scientific, FD0275), the solution was made up to $100\mu l$ with H2O and incubated at $37°C$ for 5 minutes followed by deactivation of the enzyme by incubating at $80°C$ for a further 5 minutes.

2.7.3.2 qPCR reaction set up. To perform the qPCR, 10µl reactions were set up in triplicate in 384 well plates. Each reaction contained 1X TaqMan Fast Universal Master Mix (ThermoFisher, 4352042), 100nM of forward and reverse primers (**Table 2.2**), 100nm of probe from the Universal Probe Library and 10ng of genomic DNA. For each experiment a calibrator sample (gDNA harvested from a karyotypically normal cell line) and a positive control sample. The plate was analysed using the QuantStudio 12K Flex (Life Technologies), cycled as follows; 50°C (2mins), 95°C (10mins), 95°C (15secs), 60°C (1min) for 40 cycles.

2.7.3.3 Calculation of copy number from Ct values. The average Ct for the reference gene (RELL1, 4p) was calculated for the test sample by averaging values of the three technical replicates. The dCt was then calculated for the test locus within a DNA sample by subtracting the average Ct value of the reference gene (RELL1, 4p) from the Ct value of the test sample. The dCt was then calculated for the calibrator sample as was done with the test sample. The dCt of the test and calibrator sample were subtracted giving the ddCT. The relative quantity was then calculated by raising 2 to the power of –ddCt. Finally, the relative quantity was multiplied by 2 and averaging the values of the 3 replicates. A cell line was determined as possessing a variant population when above 3 times the standard deviation of the target gene for the calibrator sample.

Table 2.2. qPCR primers for the detection of genetic variants

Gene	Location	Sense	Anti-sense	UPL Probe	Amplicon size (bp)
NPHP4	1p36	ccggcctatcgtactttt	gccggtgtgtgcagaact	8	60
MDM4	1q32.1	gcccccagacctaaatcaat	tcggtatgacagcaatgtctctt	13	76
RELL1	4p14	tgcttgctcagaaggagctt	tgggttcaggaacagagaca	12	64
DPPA3	12p13.31	cgtagcgtcgttgcatca	tcctttttaccgttcctgaca	60	63
LGR5	12q21.1	gatatgttggggattgacacg	tgctcaaagaggacaaccttc	6	60

FLCN	17p11.2	tgcagtccacaatgac aagtg	ccatgagagccgaag actgt	68	74
TK1	17q23.2- q25.3	ggtgacagctgcttac agcttag	actggttgccaccttctc ag	60	64
BCL2L1	20q11.21	tctgcagaaggctacc ccta	tgctgtgtctaagacct ctttcat	44	75
BCL2	18q21	tcaagcattgcccttag ctt	ccttaaagcatcacttc catc	25	88

2.8 Colony FISH (C-FISH).

2.8.1 Cell seeding to produce clonal colonies. 15,000 Single cells (see 2.1.6) were passed through a 20μM pluriStrainer (pluriSelect, 43-50020-03) and seeded into pre-warmed mTeSR™1 (STEMCELL Technologies, 85850) supplemented with 10μM Y-27632 onto Superfrost Plus Adhesion microscope slides (Thermo Fisher Scientific, J1800AMNT) that had pre-coated with 5mL of Vitronectin (Thermo Fisher Scientific A14700) (see 2.1.1) in a Nunc 4 well Rectangular Dish (Thermo Fisher Scientific, 267061). To ensure the clonality of the colonies produced the cells are that separates out the single cells. One partial media change was performed after 48 hours retaining 10μM Y-27632 throughout the expansion.

2.8.2 Fluorescence in situ hybridisation on cell colonies. The slides were fixed with Carnoy solution (60% ethanol, 30% chloroform and 10% glacial acetic acid) in a Coplin jar for 30 minutes. The slides are dried before immersing the slides twice for 2 minutes each in PBS. Dehydrated the slides through an ethanol series (70%, 95% and 100%) for 2 minutes and air dry. Add 5μL of FISH probe (**Table 2.3**) in the dark and mount with a coverslip (Sigma, C9802). Seal the edges of the coverslip with rubber solution (Weld-tite) to ensure they don't dry out during the hybridisation steps. Denature the slides in the dark at 72°C for 2 minutes on a flat metal block on a standard thermocycler. Remove immediately and placed in a humidified hybridisation oven set to 37°C over-night or for a minimum of 16 hours.

2.8.3 Post hybridisation. Heat 50mL of 0.4xSSCT (Sigma, S6639) post hybridisation wash solution pre-warmed to 73°C in a Coplin jar suspended in a water-bath. Remove the rubber solution and coverslip(s) from the slide and transfer to the heated 0.4xSSC (Sigma, S6639) post hybridisation wash solution for 2 minutes. Remove the slides and transfer to the 2xSSC for 30 seconds to 2 mins at room temperature. Dehydrate the slides through an ethanol series; 70%, 95% and 100% ethanol for 2 minutes. Air-dry the slides in the dark and mount with Fluoromount G semi-permanent mountant (Thermo Fisher Scientific, 00-4958-02) supplemented with Hoechst 33342 (Thermo Fisher Scientific, H3570; diluted 1:1000) beneath a coverslip. Images were acquired using an IN Cell Analyzer 2200 (GE Healthcare) high content microscope and automated image analysis was performed using CellProfiler (Carpenter et al., 2006), further details of which can be found in section 3.3.

Table 2.3. FISH probes for C-FISH

FISH Probe	Supplier	Catalogue number
BCL2L1/CCp20 FISH BAC probe	Cytotest	CT-PAC119
LSP BCL2 5' FISH BAC probe	Cytotest	CT-PAC206
LSP MDM4 (1q32)/1cen) BAC probe	Leica Biosystems	KI-10736

2.9 Cell cycle time analysis.

2.9.1 Total cell cycle time measured from time-lapse analysis. Cells were seeded at 500 cells/cm^2 onto multi-well plates pre-coated with Vitronectin (VTN-N) recombinant human protein (ThermoFisher Scientific A14700) (see 2.1.1). Images were acquired every 10 minutes for 48-72 hours using 20X objective using a Nikon Biostation CT. Images were compiled in CL Quant (NIKON) and analysed using FIJI (ImageJ).

2.9.2 EdU and pulse-chase analysis for cell cycle phase quantification. The following protocol was adapted from a previous publication (Begg et al., 1985). Cells were seeded as before into multi-well plates (see 2.1.1). Once at 60% confluency, the cells were pulse labelled with 10μM EdU for 45 minutes (Click-iT EdU Alexa Fluor 647 Flow Cytometry Assay Kit, Thermo Fisher Scientific, C10424). At hourly intervals for 24-30 hours a single well was harvested using TrypLE cell dissociation

enzyme (see 2.1.6) (Thermo Fisher Scientific, 12504013), pelleted and fixed using Click-iT fixative. After washing with 10% FCS and 1% BSA the cells were permeabilised with the Click-iT saponin based wash buffer and stained with the Click-iT reaction cocktail and counterstaining with Hoechst 33342 (Thermo Fisher Scientific, H3570; diluted 1:1000). A minimum of 10,000 dual labelled cells were recorded by FACs cytometry, relative movement and mid-S phase movement was calculated using FLOWJO single-cell flow cytometry analysis software (Becton Dickinson).

2.10 RNA Extraction and reverse transcriptase qPCR. RNA was extracted using the Qiagen RNeasy kit. cDNA synthesis was performed using high capacity reverse transcription kit (Thermo Fisher Scientific, 4368814). qPCR was performed in 384 well plates with 10μL reactions consisting of 1X TaqMan Fast Universal Master Mix (ThermoFisher, 4352042), 100nM of forward and reverse primers (**Table 2.4**), 100nm of probe from the Universal Probe Library (Roche) and 2μL of 5ng/μL cDNA. PCR reactions were analysed using QuantStudio 12K Flex Thermocycler (Life Technologies 4471087). All reactions were performed in triplicate with comparative Ct normalized to GAPDH or B-ACTIN expression. For primer sequences see **Table 2.4**;

Table 2.4. Primer sequences for gene expression analysis

Gene	sense	anti-sense	Probe
OCT4	agcaaaacccggaggagt	ccacatcggcctgtgtatatc	35
NANOG	agatgcctcacacggagact	tttgcgacactcttctctgc	31
SOX17	cgccgagttgagcaagat	ggtggtcctgcatgtgct	13
TFAP2A	acatgctcctggctacaaaac	aggggagatcggtcctga	62
TH	tcagtgacgccaaggaca	gtacgggtcgaacttcacg	42
NEUROD1	acctcgaagccatgaacg	cttccaggtcctcatcttcg	55
SOX7	ttcctcaccagccaggtc	atttgcgggaagttgctcta	30
AFP	tgtactgcagagataagtttagctgac	tccttgtaagtggcttcttgaac	61
FOXA2	cgccctactcgtacatctcg	agcgtcagcatcttgttgg	9

GATA6	aatacttcccccacaacacaa	ctctcccgcaccagtcat	90
MIXL1	gacacagatgaggggcagtt	cccgttttcagctaccattc	6
BRACHYURY	aggtacccaaccctgagga	gcaggtgagttgtcagaataggt	23
DESMIN	ggagattgccacctaccg	ggtctggatggggagattg	55
PECAM	ggtctggatggggagattg	ttcaagtttcagaatatcccaatg	37
GAPDH	agccacatcgctcagacac	gcccaatacgaccaaatcc	60
B-ACTIN	ccaaccgcgagaagatga	ccagaggcgtacagggatag	64
CCND1	tgtcctactaccgaatcaca	cagggcttcgatctgctc	55
CCND2	agctgctggctaagatcacc	acggtctgctgcaggctat	68
CCND3	ggtcacctgacgaggaggta	ggtagcgatccaggtagttca	68
CCNE1	ggccaaaatcgacaggac	catcatcttctttgtcaggtgtg	32
CCNE2	gaaagaagagaatgtcaagacgaa	tcttggcctggattatctgg	20
CDK4	ggccctcaagagtgtgagag	ccacctcacgaactgtgct	63
CDK6	tgatcaactaggaaaaatcttggac	ggcaacatctctaggccagt	2

2.11 Western blotting. Protein was isolated from cellular extracts, Laemili buffer (4% SDS, 20% Glycerol, 0.125M Tris HCl, 0.004% bromphenol blue) was added to cell pellets and sonicated for 10 seconds. The protein lysate was incubated for 10 minutes at 95°C. Protein concentration was determined by NanoDrop spectrophotometer (Thermo Fisher Scientific). Protein was separated on 10% ProtoGel (National Diagnostics) run at 120V for 1.5 hours and transferred onto PVDF membrane (Millipore, #IPVH00010). Primary antibodies were incubated overnight at 4°C; α-Tubulin (Cell Signalling Technology, 2144; diluted 1:1000), Cyclin E1 (D7T3U) (Cell Signalling Technology, 20808; diluted 1:500), Cyclin E2 (Cell Signalling Technology, 4132; diluted 1:500), Cyclin D2 (Cell Signalling Technology, 3741; diluted 1:500), RRM2 (Abcam, ab57653; diluted 1:100). The blot was washed and incubated with anti-rabbit IgG or anti-mouse IgG secondary antibody for 1 hour (Promega, W401 & W402). Immunoreactivity was visualised with ECL prime (GE Healthcare, RPN2232) on a CCD-based camera.

2.12 Immunofluorescence staining and imaging. Cells were fixed for 10 minutes at room temperature in 4% paraformaldehyde and blocked with 10% goat serum (Thermo Fisher Scientific, 16210072), 3% BSA (Sigma) and 0.3% Triton-X (Sigma, T8787) for 1 hour. Primary antibodies were incubates overnight at 4°C: Anti-Phospho-Histone H2A.X (Ser139) (Cell Signalling Technologies, 9718; diluted 1:400), Anti-gamma H2A.X (Phospho S139) (Abcam, ab26350; diluted 1:500), Anti-Nanog (Cell Signalling, 4903; diluted 1:500), Anti-Nanog (Cell Signalling, 4893; diluted 1:500) and Anti-Ki67 (Abcam, ab238020; diluted 1:100). The secondary antibodies were incubated for 1 hour at room temperature: Alexa Fluor 488-conjugated anti-rabbit IgG (Life Technologies, A11034; diluted 1:400), Alexa Fluor 647 AffiniPure Goat anti-Rabbit IgG (H+L) (Jackson Immuno Research, 111-605-003; 1:1000) and Alexa Fluor 647 AffiniPure Goat anti-Mouse IgG (H+L) (Jackson Immuno Research, 115-605-003; 1:400) and counterstained with Hoechst 33342 (Thermo Fisher Scientific, H3570; diluted 1:1000). Images were acquired using the IN Cell Analyzer 2200 (GE Healthcare) high content microscope taking ≤25 randomized images per well.

2.13 Immunofluorescence data analysis. CellProfiler (Carpenter et al., 2006) was utilised to analyse high content imaging. Expression levels of a protein were calculated or counted above a threshold set by a secondary only control. The counterstained DAPI nuclei were used to segment individual. Cell cycle stage was calculated using CellProfiler Analyst (Carpenter et al., 2006) based on the integrated intensity of the DAPI stain.

2.14 DNA fibre assay. DNA fibre assay was performed as previously described (Groth et al., 2010).

2.14.1 DNA fibre Labelling. Cells grown for a minimum of 72 hours were sequential pulse labelled with a stock of 2.5mM CldU (Sigma, C6891; 1:100) and 2.5mM IdU (Sigma, I7125; 1:10) for 20 minutes each. The cells were washed with ice-cold PBS and dissociated using TrypLE cell dissociation enzyme (Thermo Fisher Scientific, 12504013) (see 2.1.6) and diluted to 3.5×10^5 cells/ml in cold PBS.

2.14.2 DNA fibre Spreading. Labelled fibres were spread onto glass slides by adding 2μL of cell suspension and allowing it to dry for 5-7 minutes before adding 7μL of spreading buffer (200mM Tris-HCL PH7.4, 50mM EDTA, 0.5% SDS). The cell solution was mixed with a pipette tip and incubating for 2 minutes. Slides were tilted at an angle of 10° and timed to ensure the droplet ran to the bottom edge of the slide within 3-5 minutes. Slides were air-dried and fixed with 3:1 methanol/acetic acid.

2.14.3 DNA fibre immunofluorescence. The glass slides were first washed twice with H$_2$O for 5 minutes each, denatured with 2.5M HCL for 1 hour and then blocked in 1% BSA (Sigma) and 0.1% Tween20 (Sigma). Primary antibodies were incubated for 1 hour at room temperature: Rat anti-BrdU, clone BU1/75 (Novus Biologicals NB500-169) (AbD Serotec; diluted 1:400) or Anti-BrdU clone BU1/75 (ICR1) (Abcam, ab6326; diluted 1:400) and Mouse anti-BrdU (Clone B44) (Becton Dickinson, 347580; diluted 1:250). The slides were fixed for 10 minutes using 3% paraformaldehyde PH8.0. Secondary antibodies were incubated for 2 hours at room temperature: Alexa Fluor 555 goat anti-rat IgG (Thermo Fisher Scientific, A21434; diluted 1:500) and Alexa Fluor 488 F (ab')2-Goat anti-Mouse IgG (Thermo Fisher Scientific, A-11017; diluted 1:500). Slides were mounted with Fluoroshield (Sigma, F6182), and images were acquired using an Olympus FV1000 confocal microscope.

2.15 Neutral comet assay.

2.15.1 Comet assay slide preparation. 150μL of 0.6% agarose (Sigma, A9539) was set on a fully frosted glass slide, sandwiched beneath a coverslip. Once dried, 12,000 cells suspended in 75μL ice-cold PBS were mixed with 75μL of 1.2% low melting agarose (Sigma, A4018). The coverslip was removed, and the cell/low melting agarose mixture was mounted on top of the original agarose layer, beneath a new coverslip. The agarose was set for 30 minutes in the fridge. The slides were immersed in pre-chilled lysis buffer (2.5M NaCl,10mM Tris-HCL,100mM EDTA PH8.0, 0.5% Triton-X, 3% DMSO) for 1.5 hours at 4°C.

2.15.2 Comet assay electrophoresis. The slides were washed in H$_2$O and equilibrated in electrophoresis buffer (300mM sodium acetate, 100mM Tris-EDTA and 1% DMSO) for 1 hour. Electrophoresis was performed at 25V for 1 hour in a

comet assay electrophoresis tank. Slides were stained with SYBR green (Sigma, S9430; diluted 1:10,000) and quantified using a fluorescence microscope and Comet Assay IV (Instem) live video measurement system.

2.16 Oxford nanopore long-read next-generation sequencing. Cells were expanded to high confluency under standard conditions and harvested gently using ReLeSR (STEMCELL technologies, 05873) (see 2.1.5). The harvested cells were pelleted by centrifugation for 3 minutes at 1100RPM. gDNA from the pelleted cells was extracted using the DNeasy Blood and Tissue kit (Qiagen, 69504). Sequencing of the samples was performed in collaboration with the Sanger Institute. Bioinformatics analysis was performed using adapted previously published pipelines (Cretu Stancu et al., 2017) on the Sheffield Iceberg high-performance computing system. Further details are described in section 4.2.2.

2.17 Antibody staining for flow cytometry analysis of pluripotency-associated surface markers. Cells were dissociated to single cells using TrypLE cell dissociation enzyme (see 2.1.6) (Thermo Fisher Scientific, 12504013). Cells were counted and resuspended in FACS buffer (PBS, 10% FBS) at a density of $1X10^7$ cells/ml. After resuspension, 100μL of cell suspension ($1X10^6$ cells) was added to a FACS tube (Falcon, 352053) and incubated with the primary antibody for 15 minutes at 4°C (**Table 2.5**). Cells were washed once with 2mL of FACS buffer and centrifuged for 3 minutes at 1100RPM. The supernatant was aspirated and resuspended before adding the secondary antibody in 100μL FACS buffer (Alexa Fluor 647 AffiniPure Goat anti-Mouse IgG (H+L), Jackson Immuno Research, 115-605-003; 1:200). Again, the cells were washed with FACS buffer and centrifuged for 3 minutes at 1100RPM before the supernatant was aspirated and the cells resuspended in 300μL of FACS buffer. Analysis was performed on BD FACS Jazz with baseline fluorescence set using the control antibody P3X which does not show any expression on human cells.

Table 2.5. Primary antibodies for FACS analysis.

Antibody	Type	Dilution	Reference
P3X	Mouse monoclonal IgG	1:10	(Köhler and Milstein, 1975)
TRA-1-85	Mouse monoclonal IgG	1:10	(Williams et al., 1988)
SSEA3	Rat monoclonal IgM	1:10	(Shevinsky et al., 1982)
SSEA4	Mouse monoclonal IgG3	1:100	(Kannagi et al., 1983)

2.18 Apoptosis assay. Cell media was not replaced on the day of the experiment but was collected prior to the experiment to harvest the cells that have detached as a result of apoptosis. The remaining cells were dissociated to single cells using TrypLE cell dissociation enzyme (see 2.1.6) (Thermo Fisher Scientific, 12504013) and resuspended in cell media collected in the previous step. The entire cell solution was added to a FACs tube (Falcon, 352053), before being centrifuged for 5 minutes at 1400RPM. The supernatant was aspirated off and the cells were resuspended in 200μL of 4% PFA for 15 minutes. The fixed cells were then centrifuged at 1400RPM for 5 minutes and resuspended in 200μL of permeabilization buffer (PBS, 0.5% Triton X) for 5 minutes. The cells were centrifuged again for 5 minutes at 1400RPM and resuspended in 200μL of blocking buffer (PBS, 1% BSA, 0.3% Triton X) with anti-cleaved caspase-3 antibody (Cell Signalling Technology, 9661; diluted 1:400) for 1 hour with periodic, gentle agitation. The samples were washed with 1mL of blocking buffer before being centrifuged for 5 minutes at 1400RPM before being stained with Alexa Fluor 647 AffiniPure Goat anti-Rabbit IgG (H+L) (Jackson Immuno Research, 111-605-003; 1:1000) in blocking buffer for 1 hour in the dark. Finally, the cells were washed with 1 ml of blocking buffer before being centrifuged again as before. The pelleted cells were resuspended in 400μL of blocking buffer

and analysed above the baseline (secondary antibody only stained sample) on a BD FACS Jazz system.

2.19 Embryoid body formation. 3.5mL of APEL media was divided amongst the inner 60 wells of the 96 well plate (50μL per well) with the outer wells filled with 100μL PBS. hPSC were cultured to approximately 70-80% confluency and dissociated to single cells using TrypLE cell dissociation enzyme (see 2.1.6) (Thermo Fisher Scientific, 12504013). The cells were pelleted and resuspended in APEL media so that 50μL of cell suspension contains 3000 cells. 50μL of the cell suspension was added to each well of the 96 well plate. To aggregate the cells, the 96 well plates were centrifuged at 1100RPM for 3 minutes and allowed to grow in the incubator for 10 days. After 10 days, the EBs were imaged on a standard light microscope and harvested with a pipette, pooling each condition into a single 15mL conical tube.

2.20 GFP or RFP-tagging construction. Cells were transfected with pCAG-GFP-PURO or pCAG-H2B-RFP-IRES-PURO vector using electrophoresis. Cells were expanded to give 2.5million cells per electroporation, including cells for a no plasmid control. In preparation for the electroporation, a 6 well plate was coated with Vitronectin (VTN-N) recombinant human protein (ThermoFisher Scientific A14700) (see 2.1.1). Transfections were performed using the Neon Transfection system (Thermo Fisher Scientific, MPK10025), 3mL of E2 buffer was added to the neon tube and placed in the electroporation device. The cells were dissociated to single cells using TrypLE cell dissociation enzyme (see 2.1.6) (Thermo Fisher Scientific, 12504013) and a cell count was taken. The cells were centrifuged for 3 minutes at 1100RPM and the cell pellet resuspended to a concentration of 2.5 million cells per 120μL of resuspension buffer. Next, 120μL of cell suspension was mixed with up to 5μg of plasmid, keeping the volume of the plasmid to less than 5μL. Pre-warmed media, either Essential 8 or mTeSR™1 (STEMCELL Technologies, 85850) supplemented with 10μM Y-27632 was added to the 6-well plate in preparation of cell seeding. To perform the electroporation, 100μL of the plasmid/cell solution was taken up using a Neon 100μL tip. The neon tip was placed into the E2 buffer and electroporated using the following conditions; 1600V, 20msec, 1 pulse. The cells are resuspended in the pre-warmed cell culture media and returned to the incubator.

Cells successfully transfected were selected for by its puromycin resistance, Puromycin concentration was increased gradually over 5 days to a final concentration of 0.375μg/mL before flow sorting for the brightest population of GFP or RFP.

2.21 Time-lapse imaging and analysis.

2.21.1 Time-lapse analysis of mitotic errors. MIFF1 transfected with pCAG:H2B-RFP:PURO plasmid (see 2.25) were passaged from a T12.5 flask 1:10 to a 35mm IBIDI μ-dish (ibidi, 81156) pre-coated with Vitronectin (VTN-N) recombinant human protein (ThermoFisher Scientific A14700) (see 2.1.1) containing either mTeSR™1 (STEMCELL Technologies, 85850) or mTeSR™1 supplemented with exogenous nucleosides. The plates were transferred to a Nikon dual-camera widefield Live-Cell system fitted with 100X oil lens. Time-lapse images were acquired every minute for 2 hours. Analysis was performed using FIJI.

2.21.2 Time-lapse analysis to generate lineage trees. Plates of cells were set up as in 2.21.1 with cells grown in either mTeSR™1 (STEMCELL Technologies, 85850) or mTeSR™1 supplemented with exogenous nucleosides. Using FIJI, individual cells were traced, recording when the cell divided or died. Using this recorded data, lineage trees were generated manually using Adobe Illustrator (Adobe).

2.22 Clonogenic assay. Clonogenic assays were performed on vitronectin coated 24 well plates. Cells dissociated using TrypLE cell dissociation enzyme (Thermo Fisher Scientific, 12504013) (see 2.1.6), counted and resuspended in mTeSR™1 (STEMCELL Technologies, 85850) or mTeSR™1 supplemented with exogenous nucleosides. The cells were plated at a density of 500 cells/cm^2. To allow the cells to attach they were plated with 10μM Y-27632 for 24 hours before the media was replaced to remove Y-27632. After 72 hours the cells were washed and fixed with 4% PFA and stained for Nanog as previously described.

2.23 Cell growth assay.

2.23.1 Cell plating. The inner 60 wells of a 96 well plate (Greiner, 655090) were coated with 60μL Vitronectin (VTN-N) (ThermoFisher Scientific A14700) (see 2.1.1). The Vitronectin was replaced with 50μL mTeSR™1 (STEMCELL Technologies, 85850) or mTeSR™1 supplemented with exogenous nucleosides and 20μM Y-27632. Cells were dissociated to single cells using TrypLE cell dissociation enzyme (Thermo Fisher Scientific, 12504013) (see 2.1.6) and resuspended to a density of $1X10^6$ cells/mL in the respective condition. Cells were seeded at 10,000 cells per cm2. After 24 hours the media was replaced to remove the Y-27632 and grown for a further 72-96 hours.

2.23.2 Fixation and cell staining. The majority of media was then removed gently and the cells were fixed with 100μL of 4% PFA supplemented with Hoescht 33342 (Thermo Fisher Scientific, H3570; diluted 1:1000) for 15 minutes, shielding the plate from light. The wells were gently washed 3 times with PBS for 5 minutes each before images were captured on InCell analyser (GE Healthcare) high content imager, imaging the entire well. Automatic image analysis using CellProfiler (Carpenter et al., 2006) was performed, counting the number segmented DAPI or Hoescht 33342 stained nuclei.

2.29 Statistical analysis. Statistical analysis and graphical representations were performed using GraphPad Prism version 8.0.0 (La Jolla California USA, www.graphpad.com). Two independent sets data sets, such as differences between cells grown in mTeSR™1 or mTeSR™1 supplemented with exogenous nucleosides were analysed using an unpaired students t-test to determine the significant difference between the means of the two data sets. A significant difference was determined when $P<0.05$.

3 Testing and developing routine screening assays for the detection of genetically variant human pluripotent stem cells

3.1 Introduction

The spectrum of karyotypic changes that arise during the culture of human PSC can be broadly classified as large structural or whole chromosome amplifications, large or whole chromosome deletions and small amplifications including tandem duplications. The most commonly acquired genetic changes include structural and numerical deletions of chromosomes 10,18 and 22 and amplifications of chromosomes 1, 12, 17 and 20, with those effecting chromosome 20 appearing commonly as small tandem duplications (Amps et al., 2011, Baker et al., 2007, Cowan et al., 2004, Draper et al., 2004). It is important for the safe application of human PSC in regenerative medicine that we minimise the appearance of these variants, yet we will be unable to do so unless we can first sensitively detect them. Currently, no single routine approach is capable of detecting all these aberrations simultaneously and therefore genetic monitoring of human PSC relies on a set of complementary tests that have been optimised for specific types of genetic change.

Karyotyping by G-banding detects chromosomal changes through variation in the banding pattern of metaphase spreads, allowing the user to detect structural and numerical changes to every chromosome without the need for direct probing of specific regions. However, when it was applied to the detection of populations of variant human ES cells with large chromosomal amplifications, the sensitivity is in the range of 10-14%, when the typical 20-30 metaphases are practicably screened (Baker et al., 2016). Further, due to the 400-500 Giemsa stained bands that are produced per haploid genome, the resolution of this assay is limited to structural changes greater than 5Mb in size (Simons et al., 2013). As such, G-banding is unable to detect small amplifications, particularly the tandem duplication that frequently afflict chromosome 20. Exemplifying this point, the first reports of karyotypic changes, detected by G-banding, did not include amplifications to chromosome 20 (Draper et al., 2004, Cowan et al., 2004).

Only during the course of SNP array screening of a large cohort of human PSC was the frequent amplification of a small amplicon on chromosome 20q11.21 first realised (Amps et al., 2011). SNP arrays, comparative array CGH and qPCR based analyses utilise DNA taken from cultures of human PSC to detect copy number changes relative to other markers along the genome or to a reference DNA sample (Rassekh et al., 2008). These techniques are robust and capable of detecting both large and small amplifications and deletions down to 1kb in size or lower. Currently, only the sensitivity of qPCR has been tested in human ES cells, where it was possible to detect copy number changes in greater than 10% of the population (Baker et al., 2016). However, the sensitivity of arrays is often limited as they rely on being able to detect the signal of variant cells above that of diploid cells and fail in providing a count of the absolute numbers of genetically variant cells within a culture.

When it is required to discern the absolute numbers of variant cells, interphase FISH can be performed. Interphase FISH involves the molecular hybridisation of fluorescently labelled DNA probes to complementary DNA sequences. By counting the number of signals present in each cell it is possible to detect when a specific genomic region has been amplified or deleted (Trask, 2002) (**Figure 3.1a**). It is a rapid approach that only requires two days to complete, can be performed in non-specialist laboratories and is capable of detecting both large and small amplifications and deletions, aneuploidies and is routinely employed to detect small tandem duplications, like those commonly found on chromosome 20 (Trask, 2002). The versatility of Interphase FISH makes it an attractive assay for screening cultures of human PSC. However, in practice, inconsistent results can arise that are a consequence of false-negative signals. High false-negative rates result from the co-localisation of signals that is an inherent problem of visualising a 3-D object, such as a cell, in 2-D (Kearney, 2001) (**Figure 3.1b**). The frequency of false-negative signals also depends on the structure of the variant being detected. Translocations and aneuploidies infrequently result in false-negative signals because there is usually a good spatial separation of the amplified and wild-type chromosomal regions. However, a high incidence of false negatives can arise when detecting tandem duplications where the distance between the amplification is small, increasing the likelihood that the two signals will overlap when viewed by fluorescence microscopy

(Gozzetti and Le Beau, 2000) (**Figure 3.1c**). Therefore, the sensitivity of detecting small tandem duplications will likely be compromised but this has not currently been tested. When applied to the detection of low-level mosaicism of human ES cell cultures it was possible to detect down to 1% variant cell when 1000 interphase nuclei were scored, although it is highly labour-intensive to do so and would be impractical for regular screening of cell lines (Baker et al., 2016). Consequently, 100 interphase cells are routinely scored which limits its sensitivity to 10% when detecting large amplifications to chromosome 17 (Baker et al., 2016).

In summary, the assays currently employed to routinely screen human PSC cultures tend to offer poor sensitivity are unable to detect the presence of low-level genetic mosaicism when present in less than 10-15% of a culture. Additionally, current reports to the literature have failed to test the sensitivity of array-based approaches and have only provided limited results from mixing experiments using human ES cells that present large structural and numerical amplifications.

a

b

c

20q11.21 FISH probe

x
y

x z
y

Proximity of probes
un-resolvable by fluoresence
microscopy

20q11.21   20q11.21

Tandem duplication
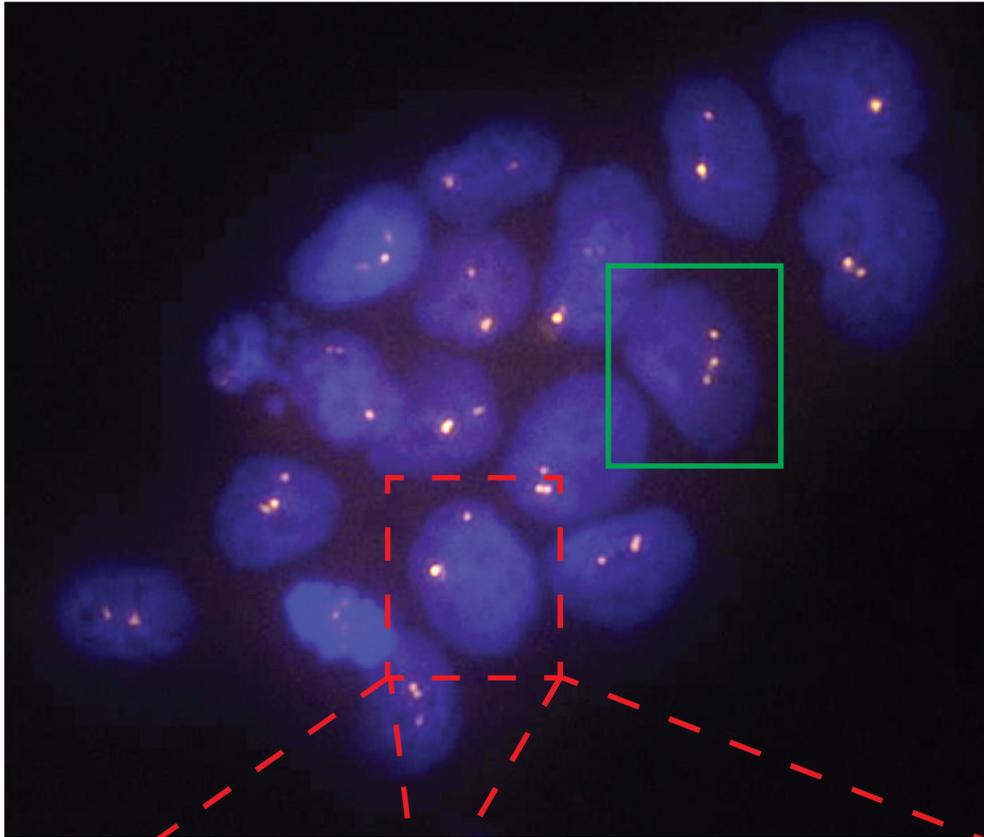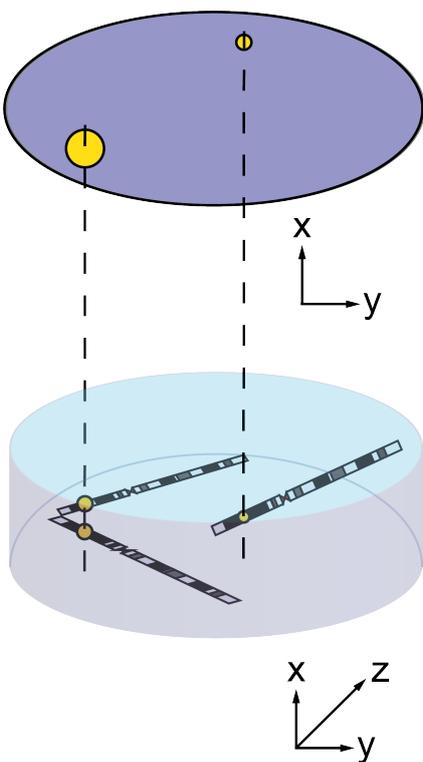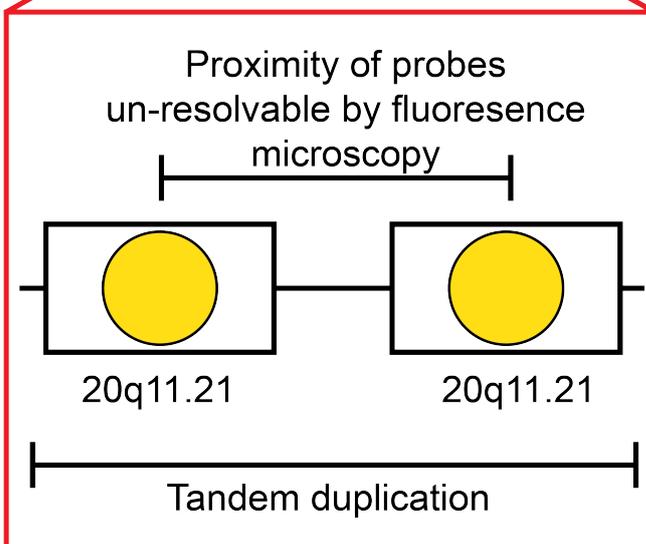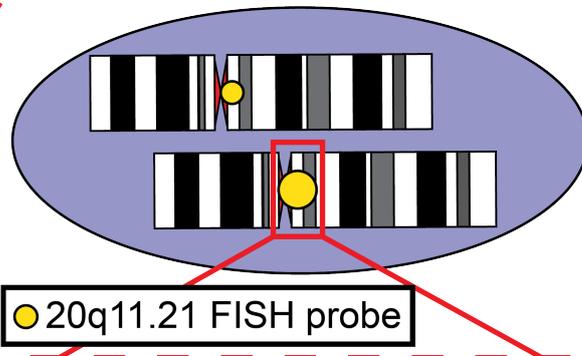
Figure 3.1 | Sources of false negative signals from interphase FISH analysis. a, Representative image of a chromosome 20q variant human PSC line probed with interphase FISH 20q (BCL2L1) probe. Green box indicates a cell displaying a true positive result and the dashed red box indicates a false negative cell. **b,** Schematic illustrating the co-localisation of signals in aneuploid cell as a result of viewing the 3-D nucleus of a cell in 2-D. **c,** Schematic illustrating the close proximity of signals that are unresolvable by fluorescence microscopy resulting in signal overlap and a false negative result. Note, the signal intensity on the amplified chromosome is greater than that observed on the unaffected chromosome.

The results presented in this chapter come from robust testing of the sensitivity of all the approaches we have discussed here, expanding on previous studies by performing mixing experiments using human iPSC and where necessary testing the sensitivity of detection of small tandem amplifications on chromosome 20 and chromosomal deletions. Further, we present data on the development of a novel interphase FISH approach, which we have termed colony-based interphase FISH (C-FISH) to address the poor sensitivity of interphase FISH. In this approach, test cells are allowed to form small colonies of about 10 to 15 cells, which we then score by averaging the signal from all the cells within a colony and show it is possible to discount the impact of false negative signals in assessing the genetic status of the parent cell of the colony (**Figure 3.2**). By also employing a scanning microscope technique it was possible to automate the scanning of a large number of colonies. Through this development of the C-FISH assay, I was able to improve the sensitivity of detecting specific common genetic variants, a necessary step for achieving the translational promise of human PSC.
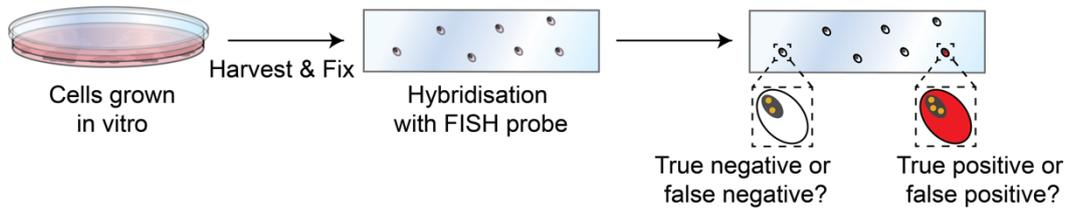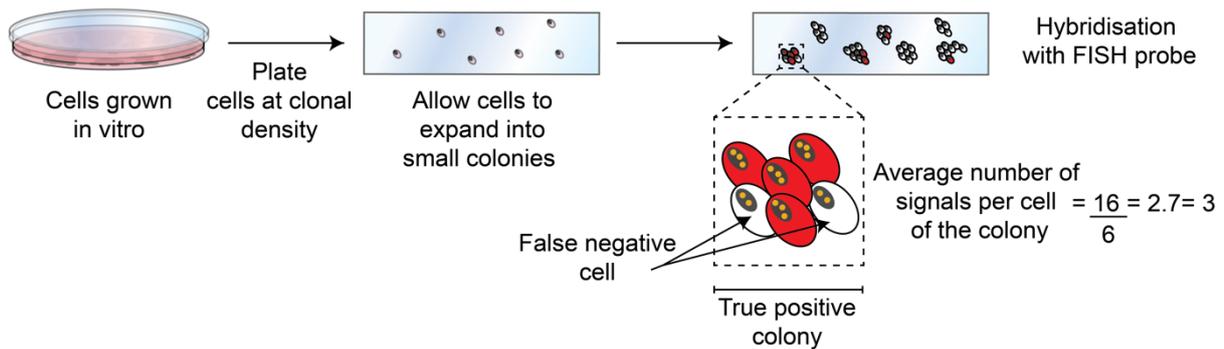
Figure 3.2 | Workflow schematic of conventional interphase FISH and C-FISH. Conventional interphase FISH (top), cells are harvested before being fixed and hybridised with a FISH probe on glass slides. Analysis is performed on single cells where confirmation of a positive or negative result is not possible. C-FISH (bottom), live single cells are cultured for 72 hours on glass slides allowing the single cells to form clonally derived colonies of roughly 10 to 15 cells. The slides are fixed and hybridised with a FISH probe corresponding to the genomic region of interest. False negative results are reduced by assessing the average number of signals per cell per colony. An example of the colony calculation is displayed.

3.2 Results

3.2.1 Generating a panel of genetically variant and diploid isogenic human induced pluripotent stem cells

The sensitivity of detecting low level mosaicism by G-banding, qPCR and Interphase FISH has previously been assessed (Baker et al., 2016). However, the previous study focused on the assessment of human ES cells, whereas for the purpose of the present study it was important to confirm these results are consistent in cultures of human iPSC. Further, in the following chapter, I needed highly sensitive assays to identify variant cell lines to explore the breakpoint regions in the chromosome 20 amplicon, to gain an insight into the mechanism through which it arises.

To test the sensitivity of G banding, qPCR, array CGH and interphase FISH, our benchmark panel of assays, isogenic human iPSC lines containing commonly observed karyotypic changes and diploid counterparts were generated. These clonal lines would provide the tools to investigate the sensitivity of detecting low level mosaicism through mixing experiments. Prior to the initiation of this project, the human iPSC line MIFF1 had acquired a mosaic population of variant cells that possessed an amplification on chromosome 1 and 20. A single cell cloning approach was used to isolate the diploid and genetically variant populations, confirmed by G banding or Interphase FISH and qPCR (**Figure 3.3a,b**). The MIFF1 clone Cl33 would provide the genetic background to test the sensitivity of detecting small mosaic populations of human iPSC with large (duplication of chromosome 1) and small structural amplifications (duplication of chromosome 20) through mixing experiments. However, deletions to chromosome 10, 18 and 22 are also known to commonly afflict human PSC (Baker et al., 2007). I sought to identify a cell line with a population of cells that contained a deletion to chromosome 10, 18 or 22. Cell lines donated from our colleagues on the TechnoBeat consortium were screened by G banding. The UCB144 cell lines was identified as having a small population of cells with a deletion on chromosome 18. The mosaic UCB144 cells were thawed from a cryopreserved bank, expanded and single cells were flow sorted into a 96 well plate containing MEF. After 2-3 weeks, cells that had successfully formed colonies were harvested with half of the material screened by qPCR for the chromosome 18

deletion and the other half re-plated and expanded. Clones with the chromosome 18 deletion and a selection of the isogenic diploid clones were cryopreserved. To confirm the genotype of the selected clones, a sample culture at the point of banking was karyotyped by G banding (**Figure 3.3c,d**).

By fluorescently labelling the diploid MIFF1 cell line it was possible to accurately distinguish between the labelled diploid and unlabelled variant cells combined during mixing experiments. The MIFF1 diploid line was stably transfected with a constitutively active GFP, driven by a pCAG promoter. The pCAG promotor simultaneously drives antibiotic resistance to neomycin allowing for the positive selection of successfully transfected cells following electroporation. The pCAG promotor is not silenced in human PSC, which is important when monitoring proportions of cells (Liew et al., 2007). To confirm that no further genetic changes had occurred through this process G-banding, interphase FISH and qPCR confirmed the diploid genotype of the MIFF1-GFP line (**Figure 3.4**). A summary of these cell lines and their respective karyotypes are listed in **Table 3.1.**

Additionally, the H7 human ES and the TC113 human iPS cell line clones were utilised whilst developing the C-FISH approach. These cell lines were previously generated in much the same way as described above and provided additional tools to test the detection of large amplifications in the form of a trisomy, in the case of H7s6 and small tandem duplications with the H7s14-20q and TC113-E1 and E6 lines (**Table 3.1**)

Figure 3.3 | Isogenic panel of human iPSC, genotype confirmed by G banding, Interphase FISH and qPCR. a, MIFF1 was confirmed diploid (46, XY) by G banding (top), interphase FISH (middle) and qPCR (bottom). **b,** the MIFF1 isogenic clone cl33 derived from single cell FACS sorting was found to have an amplification on chromosome 1 and 20 when screened by G banding (top), interphase FISH (middle) and qPCR (bottom). **c,** The diploid UCB144 clone, F6. Karyology result (top) and qPCR for the determination of copy number (bottom). **d,** The D14 clone of cell line UCB144 contains a homozygous population with a deletion on chromosome 18 confirmed by G banding (top) and qPCR (bottom). **a-d,** qPCR results are the mean copy number from three technical replicates ±s.d.

a MIFF1

46,XY [30]
FISH: 20q No evidence
of duplication

b MIFF1 - Cl33
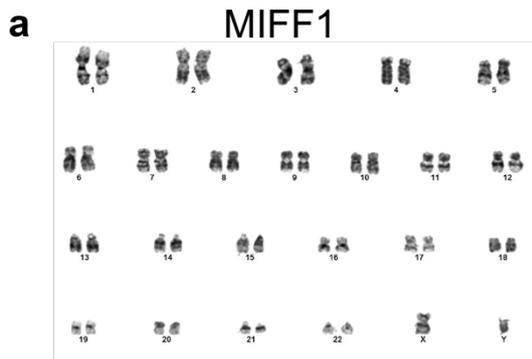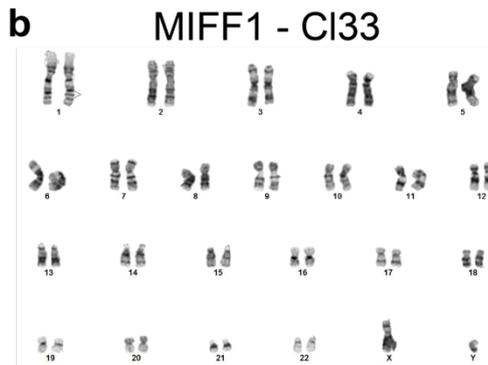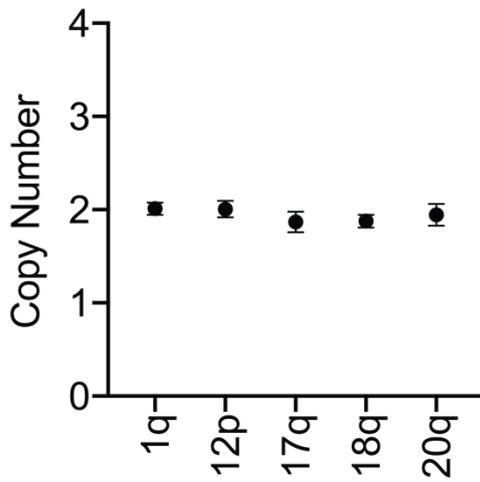
46,XY,dup(1)(q32q42), inv(9)
(q21q34) [20] FISH: 20q
duplication detected

c UCB144 F6

46,XY [50]

d UCB144 D14
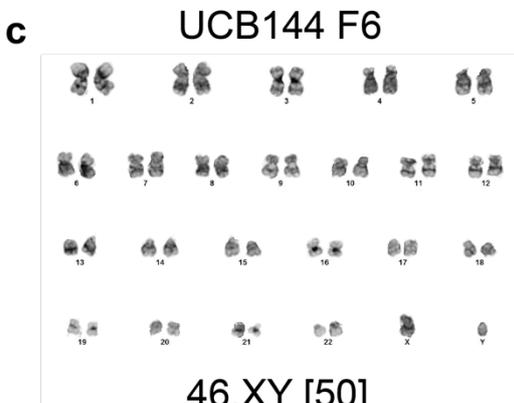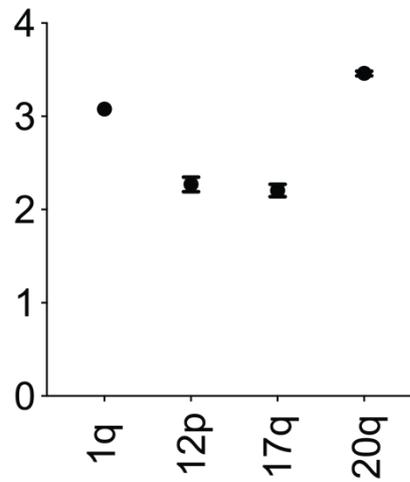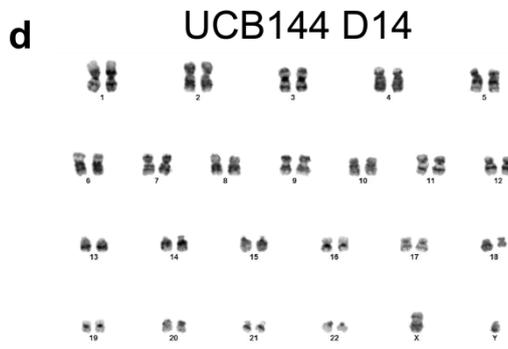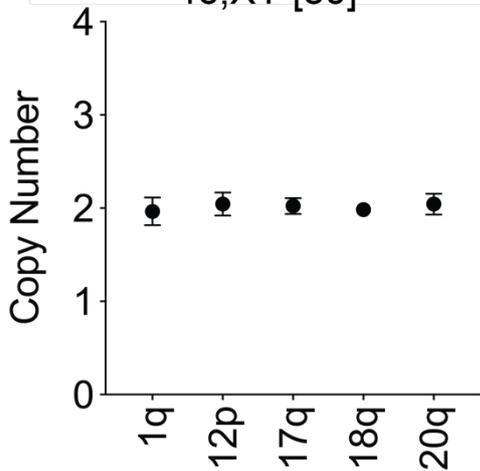
46,XY,del(18)(q21) [50]
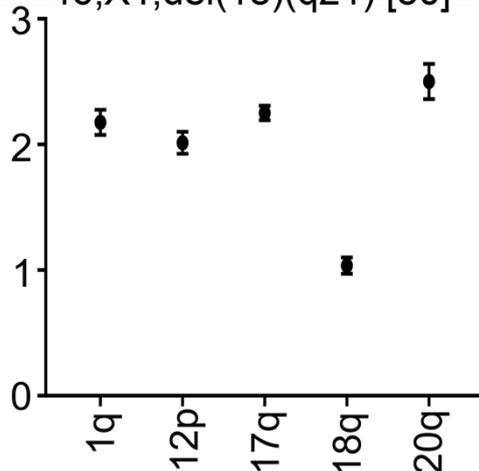
Table 3.1. Cell lines used to test the sensitivity of detection of common genetic variants in human iPSC. Colour coding shows the cell lines with small amplifications such as tandem duplication (yellow), large amplifications including trisomy (green) and deletions (red).

Cell Line	Karyotype/FISH	Comment
MIFF1	46,XY	Normal karyotype. No evidence of a duplication to *BCL2L1* (chromosome 20q11.21) as detectable by interphase FISH and qPCR.
MIFF1-GFP	46,XY	Normal karyotype. No evidence of a duplication to *BCL2L1* (chromosome 20q11.21) as detectable by interphase FISH and qPCR.
MIFF1-Cl33	46,XY,dup(1)(q32q42),inv(9)(q21q34)	Abnormal karyotype: An amplification to chromosome 1 between the regions q32 to q42. Chromosome 9 has displayed an inversion to the q arm between q21 and 34, an unusual genetic change rarely seen in human iPSC. There is also a gain of the chromosome 20q11.21 amplicon, including BCL2L1, detectable by interphase FISH and qPCR, but not G-banding.
UCB144-F6	46,XY	Normal karyotype. No evidence of a duplication to *BCL2L1* (chromosome 20q11.21) as detectable by interphase FISH and qPCR.
UCB144-D14	46,XY,del(18)(q21)	Abnormal karyotype. A deletion to chromosome 18q21. No evidence of a duplication to *BCL2L1* (chromosome 20q11.21) as detectable by interphase FISH and qPCR.
H7s14	46,XX	Normal karyotype. No evidence of a duplication to *BCL2L1* (chromosome 20q11.21) as detectable by interphase FISH and qPCR.
H7s14-20q	46,XX	Normal karyotype. Evidence of a duplication to *BCL2L1* (chromosome 20q11.21) as detectable by interphase FISH and qPCR but not detectable by G-banding
H7s6	47,XX,+del(1)(p22p22),der(6)t(6;17),(q27;q1),t(12;20)(q13;13.3),i(20)(q10)	Abnormal karyotype. Trisomy of chromosome 1 that has an interstitial deletion at 1p22. A gain of chromosome 17q via an unbalanced translocation with chromosome 6, trisomy of chromosome 12 and duplication of the entire chromosome 20q via an isochromosome.
TC113-G2	46,XY	Normal karyotype. No evidence of a duplication to *BCL2L1* (chromosome 20q11.21) as detectable by interphase FISH and qPCR.
TC113-E1	46,XY	Normal karyotype. Evidence of a duplication to *BCL2L1* (chromosome 20q11.21) as detectable by interphase FISH and qPCR but not detectable by G-banding
TC113-E6	46,XY	Normal karyotype. Evidence of a duplication to *BCL2L1* (chromosome 20q11.21) as detectable by interphase FISH and qPCR but not detectable by G-banding

a MIFF1 pCAG GFP

46,XY [30]
FISH: 20q No evidence
of duplication

b
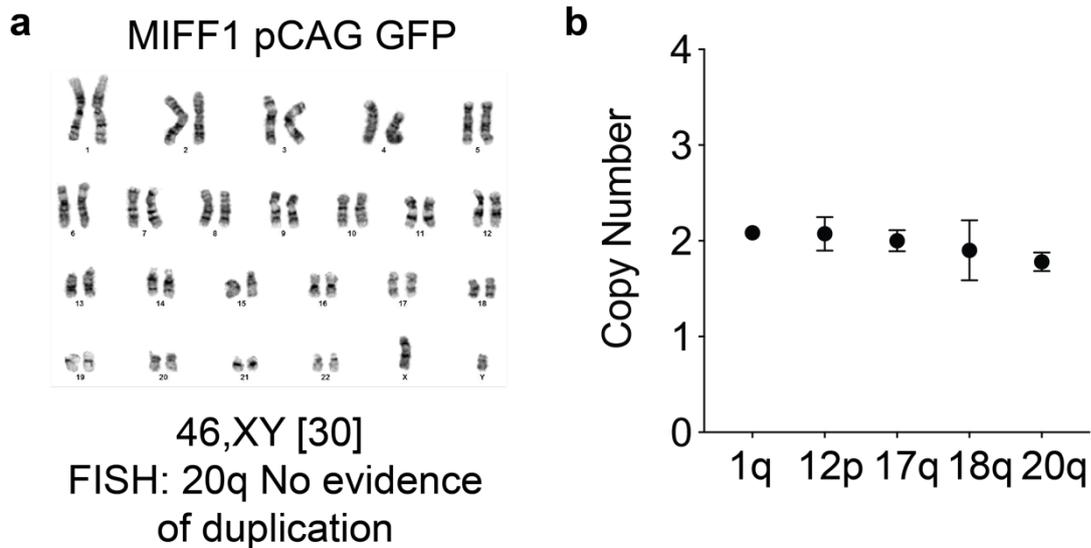
Figure 3.4 | Fluorescently labelled MIFF1 remained diploid following stable transfection of constitutively active GFP. a, G banding and interphase FISH confirmed the GFP labelled MIFF1 line remained diploid following transfection. **b,** qPCR copy number analysis of the MIFF1-GFP cell line. qPCR results are mean copy number from three technical replicates ±s.d.

3.2.2 The sensitivity of G banding in detecting mosaic populations of variant human iPS cells

To assess the sensitivity of G banding for the detection of low-level mosaicism in cultures of human iPS cells we performed mixing experiments of the diploid MIFF1-GFP with MIFF1-CL33. The cell lines were mixed at increasing ratios (ranging from 1% to 28%) with each mix split across two identical cell culture flasks. The first flask was sacrificed to accurately determine the proportion of MIFF1-GFP cells in the mixing experiment by flow cytometry. The sister flask was treated with colcemid to arrest the cells in metaphase and processed for G banding. The prepared chromosome spreads were analysed and increasing numbers of unique metaphases (between 5 and 100) were sampled, in each case, the presence or absence of the chromosome 1 amplification was counted. Each experiment was performed in triplicate by two independent cytogeneticists, which was accomplished in collaboration with Duncan Baker and the team at the Sheffield Children's Hospital Diagnostic Genetics Service. Each cytogeneticist analysed the same slide of

prepared metaphases in each experiment to exclude the potential of analyst bias. A second slide was also prepared and analysed by analyst 1 to confirm the results from the first slide. Generally, the detected numbers of abnormal cells fell within the expected confidence levels (**Figure 3.5**). At the ratio of 22%, two values fell below the statistically predicted lower limit when 100 cells were scored. Overall, the number of metaphases scored as positive for an amplification to chromosome 1 matched what was anticipated from random sampling theory (Baker et al., 2016). Detecting low-level mosaicism of 1% was not possible from our assay even when 100 metaphases were scored but we readily detected the 14% mix even at low levels of sampling.

Figure 3.5 | Sensitivity of G banding in the detection of the chromosome 1 duplication in human iPSC. a-d, MIFF1 (diploid-GFP) and MIFF1-Cl33 (46,XY,dup(1)(q32q42),inv(9)(q21q34)) were mixed at the ratios indicated and these numbers were confirmed by fluorescence-activated cells sorting (FACS) based on the expression of GFP. The mosaic cultures were analysed by G-banding and metaphases were scored for the presence of the chromosome 1 duplication (q32q42) y-axis. Increasing numbers of metaphases were scored (5-100) by two cytogeneticists, x-axis. The individual data points indicate the blind analysis performed as follows: cytogeneticist 1 scoring slide 1 (blue circle), cytogeneticist 1 scoring slide 2 (red circle) and cytogeneticist 2 scoring slide 2 (green circle). The dotted lines represent the statistically determined expected numbers, upper and lower limits (CI=95%).

a 1%

b 14%

c 22%

d 28%

Variant Cells Detected

Number of Cells Scored

Upper and Lower Limit
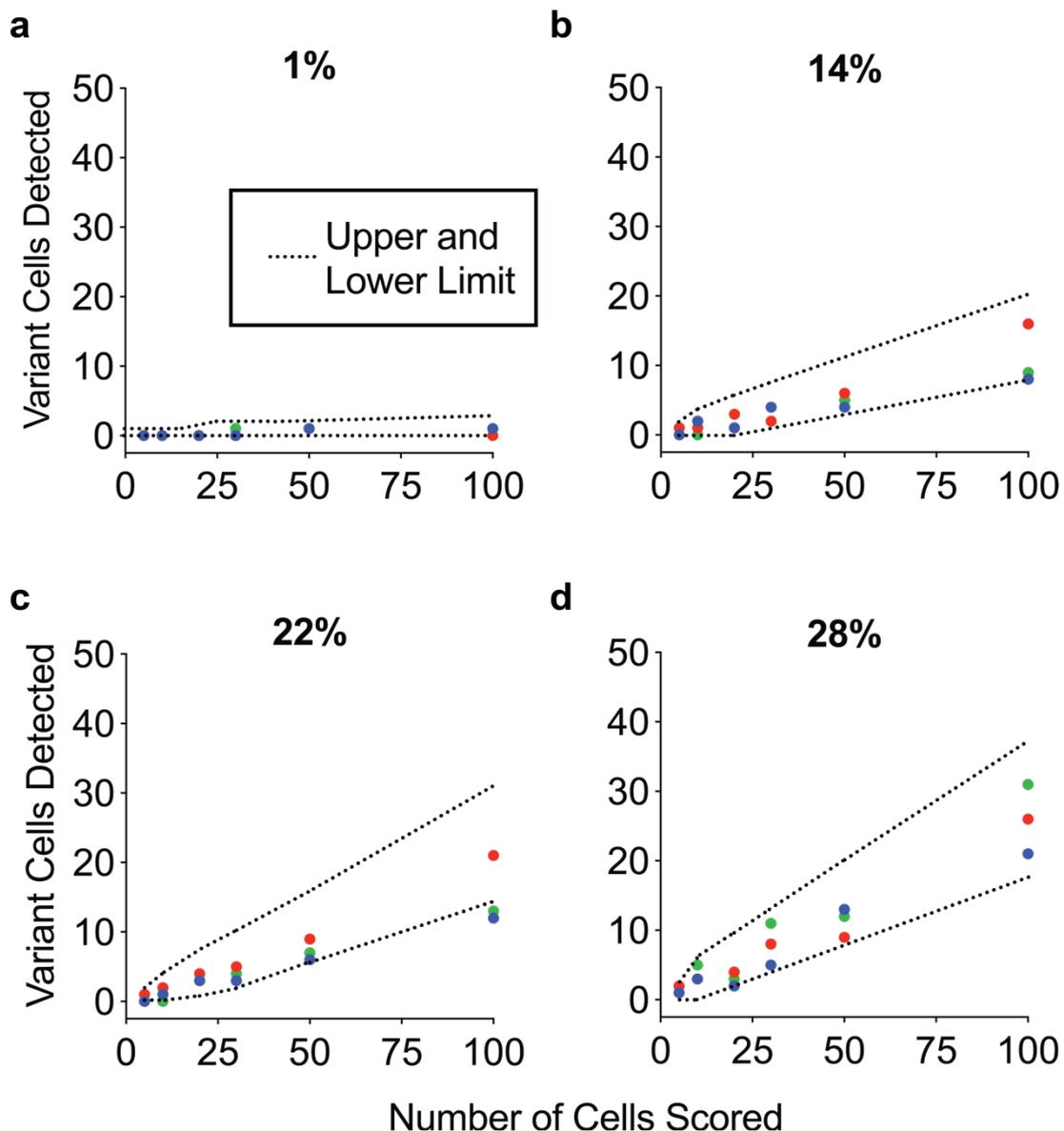
3.2.3 Sensitivity assessment of array CGH for the detection of chromosome 1 and chromosome 20 amplifications

Previously, the sensitivity of array-based techniques had not been tested for the detection of low-level mosaicism in human PSC cultures (Baker et al., 2016). To test the lower limits of detection of array CGH, genomic DNA was first prepared from cultures of MIFF1 and MIFF1-CL33. The DNA sample concentrations were measured and diluted so that the concentrations matched. MIFF1 gDNA was then spiked with increasing ratios (0%, 1%, 5%, 10%, 20%, 30%, 50% and 100%) of MIFF1-CL33 DNA. The spiked DNA was sent for blind cytogenetic analysis with the analyst returning results of a positive or negative array for amplifications on chromosome 1 and 20. In both cases, the analyst reported a positive array when the

MIFF1 DNA had been spiked with greater than 20% of MIFF1-Cl33 DNA (**Figure 3.6**).

a
Chromosome 1 arrayCGH - 20-30% mosaicism
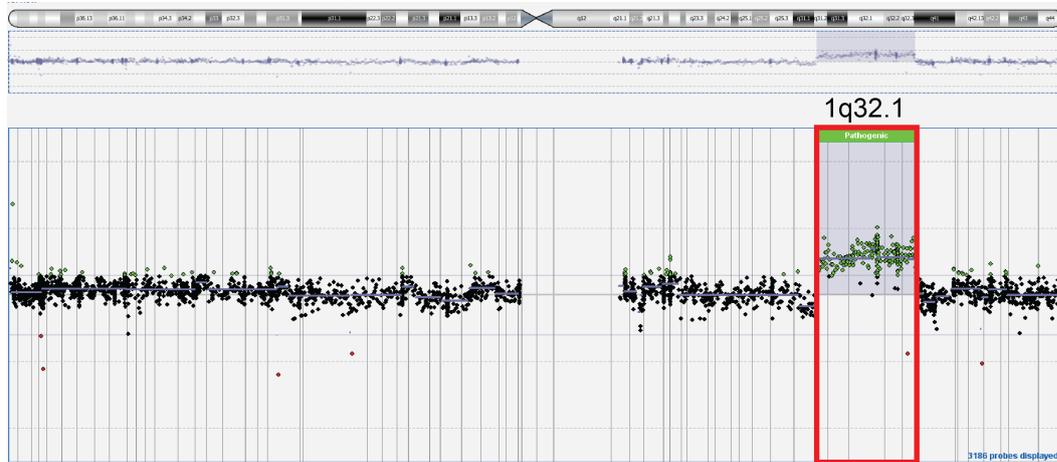
b
Chromosome 20 arrayCGH - 20-30% mosaicism

Figure 3.6 | Array CGH can detect a 20-30% mosaic populations of variant human PSC. a, A positive chromosome 1 array of MIFF1 DNA that was spiked with 20% of MIFF1-CL33 DNA. **b,** A positive chromosome 20 array of MIFF1 DNA that was spiked with 20% of MIFF1-CL33 DNA.

3.2.4 Assessment of the sensitivity of qPCR for the detection of low-level mosaicism in human iPSC

The qPCR detection method was developed as a complementary alternative to G banding to allow routine screening of human PSC cultures (Baker et al., 2016). A

target panel of primers located on commonly amplified regions allows for the detection of copy number changes based on relative quantification to a reference locus on chromosome 4. Chromosome 4 was chosen as it is very rarely afflicted by genetic changes and so acts as a diploid internal control locus. The design of the primers is critical for the accurate detection of copy number variants. During the initial G-banding screen of the clone MIFF1-CL33, we detected an amplification on chromosome 1 that spanned the q32q42 region (**Figure 3.7a)**. However, a subsequent screen by qPCR analysis did not reveal this duplication (**Figure 3.7b)**. The original primer was designed for the gene *CHD1L* on locus q12 of chromosome 1 which falls outside of the duplication in the MIFF1-CL33 cell line (Baker et al., 2016). Before the sensitivity of qPCR could be assessed it was necessary to redesign this primer so that it targeted a gene within the common minimal amplicon. Recent unpublished data has indicated that *MDM4* may be the driver gene found on chromosome 1, which lies within the 1q32 band (Unpublished data; McIntire et al. WiCell, Madison, WI, USA). Accordingly, a new primer set was designed that mapped to the intronic region of *MDM4* following the previously described method (Laing et al., 2019, Baker et al., 2016). The newly designed primer set successfully detected the chromosome 1 amplification that was reported by G-banding (**Figure 3.7c**).
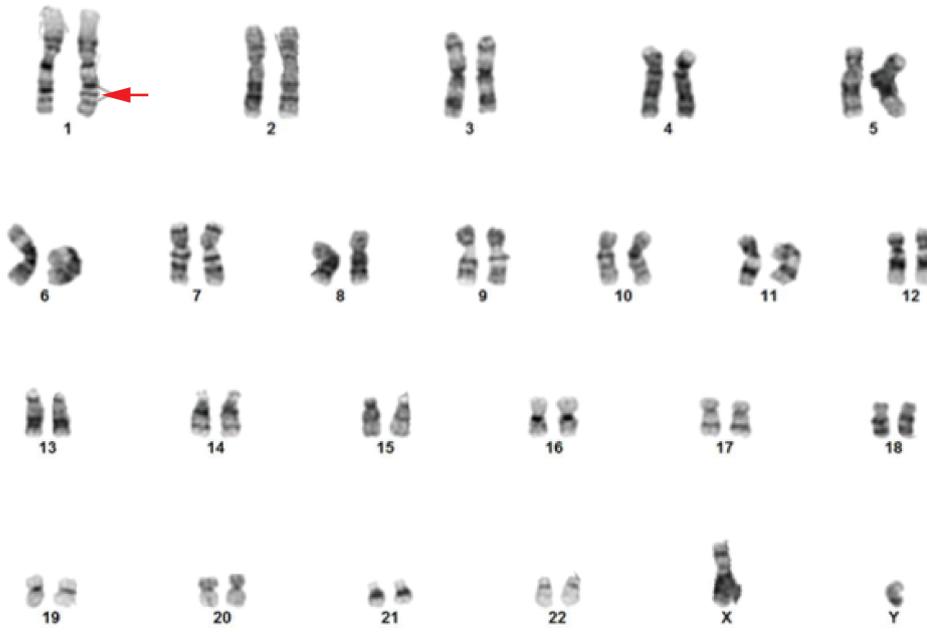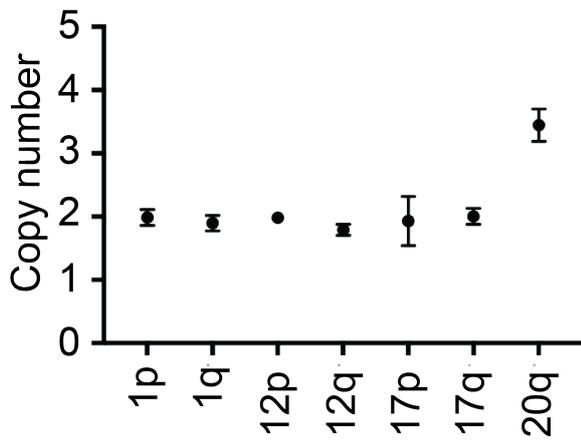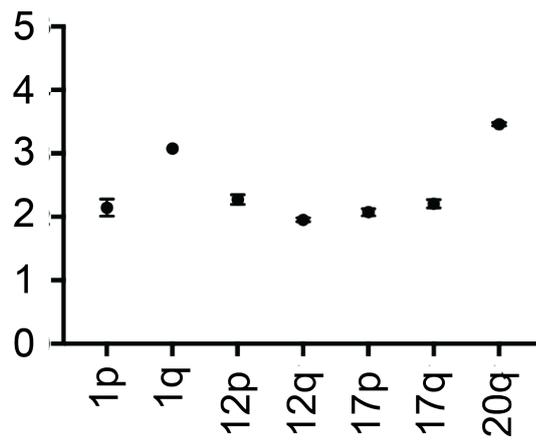
Figure 3.7 | Development of appropriate primers for the detection of the minimal amplicon on chromosome 1. a, A representative karyotype of MIFF1-CL33 produced by G banding. The red arrow indicates an amplification of the q32q42 region. **b,** qPCR assay that included a primer for the gene *CHD1L* (1q) was unable to detect the duplication on chromosome 1 in the cell line MIFF1-CL33. The small tandem duplication on chromosome 20 was readily detected by the assay using primers designed for the gene *BCL2L1* (20q). **c,** The newly designed primers for gene *MDM4* sensitively detect the chromosome 1 amplification in the MIFF1-CL33 cell line. qPCR results are mean copy number from three technical replicates ±s.d.

With the re-designed primers now suitable for the detection of minimal amplifications on chromosome 1, the sensitivity of qPCR and interphase FISH was tested using the same samples of cells. To determine the sensitivity of detecting amplifications to chromosome 1 and 20, cultures of the cell lines MIFF1 and MIFF1-CL33 were harvested, counted and then mixed with an increasing proportion (0%, 1%, 5%, 10%, 20%, 30%, 50% and 100%) of MIFF1-CL33. To test the relative sensitivity of detecting deletions to chromosome 18, we designed a primer specific to the PHLPP1 gene of in the q21.33 region following the protocol previously described (Baker et al., 2016, Laing et al., 2019). The primer accurately detected the chromosome 18 deletion in the clonal line, UCB144-D14 (**Figure. 3.3d**). To test the sensitivity of the qPCR approach for the detection of chromosome 18 deletions UCB144-F6 and UCB144-D14 were mixed with the same increasing proportion of the variant line, UCB144-D14. After mixing the cells, each sample was split into two tubes with one tube processed for gDNA extraction that was used in the qPCR assay and the other fixed and processed for interphase FISH (see section 3.2.5).

Using the previously described calculation (Laing et al., 2019, Baker et al., 2016), copy numbers were calculated from the qPCR Ct values. As expected, the control samples containing 0% and 100% of the variant MIFF1-CL33 returned a copy number of 2 and 3 respectively, whereas, the control sample containing 0% and 100% of the variant UCB144-D14 cells returned a copy number of 2 and 1 respectively. Across the three independent experiments, a significant difference was

calculable between the 0% control and the 20% mix, confirming a sensitivity of 20% (**Figure 3.8a-c**).

a

Mosaicism of chromosome 1q32 amplification

b

Mosaicism of chromosome 20q11.21 amplifications

c

Mosaicism of chromosome 18q21.33 deletions

Abnormal cells
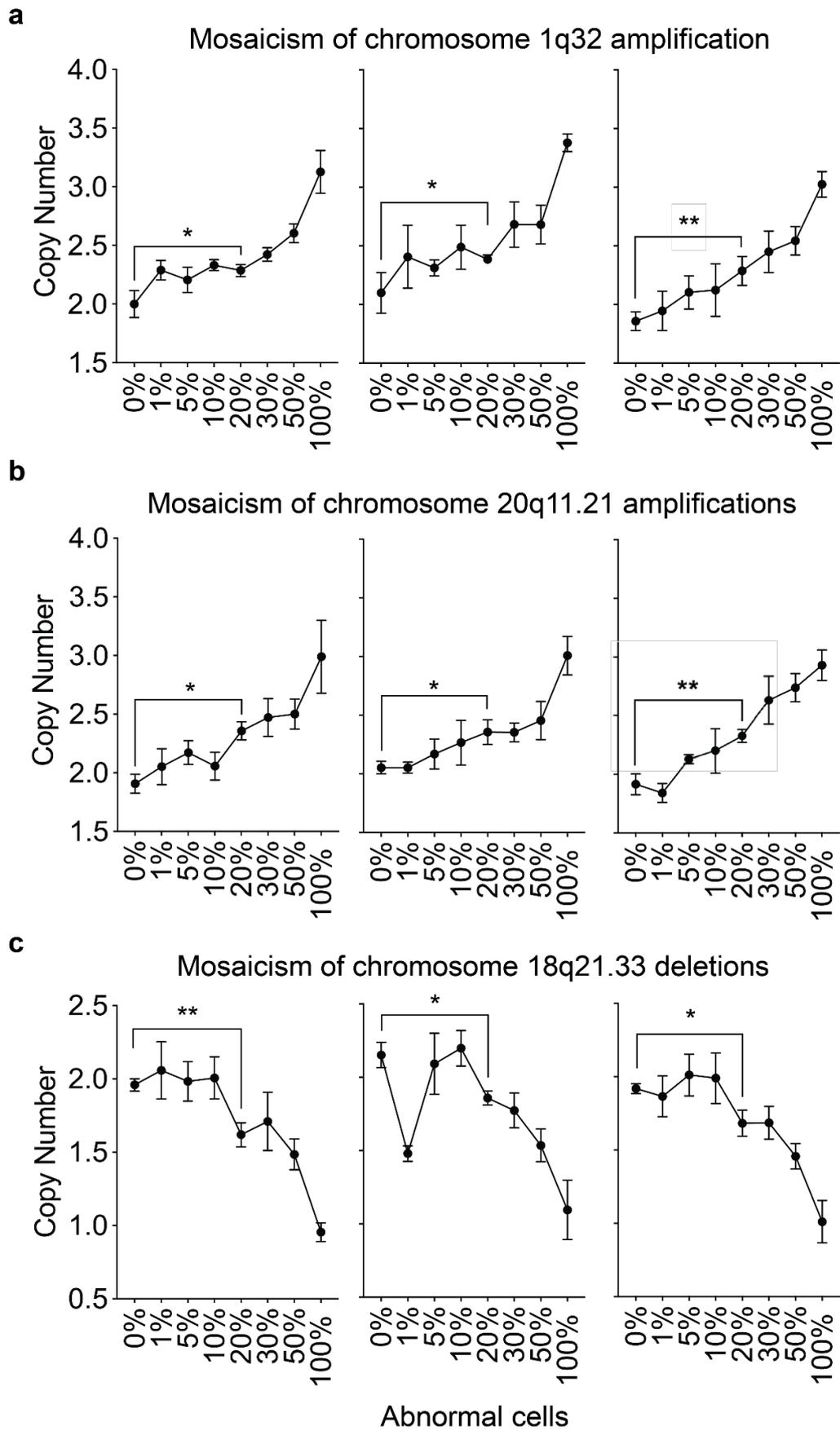
Figure 3.8 | qPCR-based assay can detect genetic low-level mosaicism in human IPSC. a-c, Mixing experiments of increasing proportions of isogenic variant lines (x-axis) in a background of diploid lines. Copy numbers are calculated from qPCR Ct values and plotted along the y-axis. Results of three independent experiments are displayed from left to right. **a,** qPCR detection of spiked populations of the chromosome 1 variant, MIFF1-CL33 in a background of the diploid MIFF1 cell line by qPCR. **b,** Results from qPCR variant detection of spiked MIFF1-CL33 in diploid MIFF1 cells for the amplification on chromosome 20. **c,** Detecting genetically mosaic cell samples with an increasing proportion of the variant chromosome 18 deletion cell line UCB144-D14 in a background of diploid UCB144-F6. Data in **a-c**, are mean $\pm$ s.d., two-tailed *t*-test, *P<0.05, **P<0.01, ***P<0.001, ****P<0.0001 (*n* = 3 experiments).

3.2.5 Testing the sensitivity of interphase FISH for the detection of amplifications to chromosome 1 and 20 and deletions to chromosome 18

Using the paired fixed samples that we generated in section 3.2.4, FISH probes that complement the *MDM4* gene on chromosome 1, *BCL2L1* on chromosome 20 and *BCL2* on chromosome 18 were hybridised to each of the relevant samples. One hundred interphase nuclei were scored per mixing experiment across three independent tests. The lower limit of detection was calculated by performing an unpaired t-test comparing the difference between the 0% control sample with the mixed spiking experiments. Amplifications on chromosome 1 were reliably detected in the MIFF1-CL33 variant cells only when they were present at proportions greater than 5% (**Figure 3.9a**). Unsurprisingly, the sensitivity was diminished when interphase FISH was applied to the detection of the small tandem duplication on chromosome 20 in the MIFF1/MIFF1-CL33 mixing experiments. The amplification was consistently detected when present at greater than 10% of the total population (**Figure 3.9b**). This result is consistent with increased numbers of false-negative results that arise when probing small tandem duplications by interphase FISH (Gozzetti and Le Beau, 2000). This deduction is further substantiated by the results attained from the 100% control samples where the amplification of chromosome 1 and the deletion of chromosome 18 were detected in 100% and 98% of the scored cells (**Figure 3.9a,c**). In contrast to this, the detection of the chromosome 20 variant

only returned an average result of 92% positive cells, possibly due to the difficulty of resolving the third signal from tandemly duplicated regions (**Figure 3.1a**). Unexpectedly, the detection of chromosome 18 deletions was also only reliably detected above 10% (**Figure 3.9c**). This could be explained by the acquisition of a population of cells with a deletion on chromosome 18 in the diploid UCB144-F6, although this would seem unlikely as the cell line was karyotyped at the end of the experiment and returned a diploid result. A more likely explanation would be a false-negative result caused by co-localisation of signals from imaging cells in a 2-D plane (**Figure 3.1b**). Overall, the sensitivity of the current panel of assays cannot reliably detect low-level mosaicism when present below 5-20% of a culture (**Table 3.2**). Given the potential consequence of genetically variant human PSC use in regenerative medicine, it is vital that sensitive and robust techniques are developed to accurately screen human PSC products intended for the clinic. The following sections in this chapter will present a novel approach developed to address the limited sensitivity of interphase FISH, particularly when detecting small tandem duplications.

Figure 3.9 | Sensitivity of interphase FISH for the detection of amplifications to chromosome 1 and 20 and deletion to chromosome 18. a-c, Mixing experiments were performed to determine the sensitivity of interphase FISH. The proportion of variant cells spiked in each experiment is plotted on the x-axis with frequency of detection plotted on the y axis. Three independent experiments were performed and displayed as individual graphs (left to right). **a,** Interphase FISH reliably identified mosaic populations of chromosome 1 duplications in the MIFF1-CL33 cell line when present in greater proportions than 5%. Statistical significance was found between 0% and 5% (unpaired t-test, P<0.05 for the biological triplicate of 0% versus 5% sample). **b,** Chromosome 20 duplications were detectable when more than 10% of variant MIFF1-CL33 were spiked into a background of diploid MIFF1 cells. Statistical significance was found between 0% and 10% (unpaired t-test, P<0.05 for the biological triplicate of 0% versus 10% sample). **c,** A mosaic population of the UCB144-D14 cell line was detected by interphase FISH with probes specific to the chromosome 18 deletion when present in greater proportions than 10%. Statistical significance was found between 0% and 10% (unpaired t-test, P<0.01 for the biological triplicate of 0% versus 10% sample).

a

Mosaicism of chromosome 1q32 amplification

b
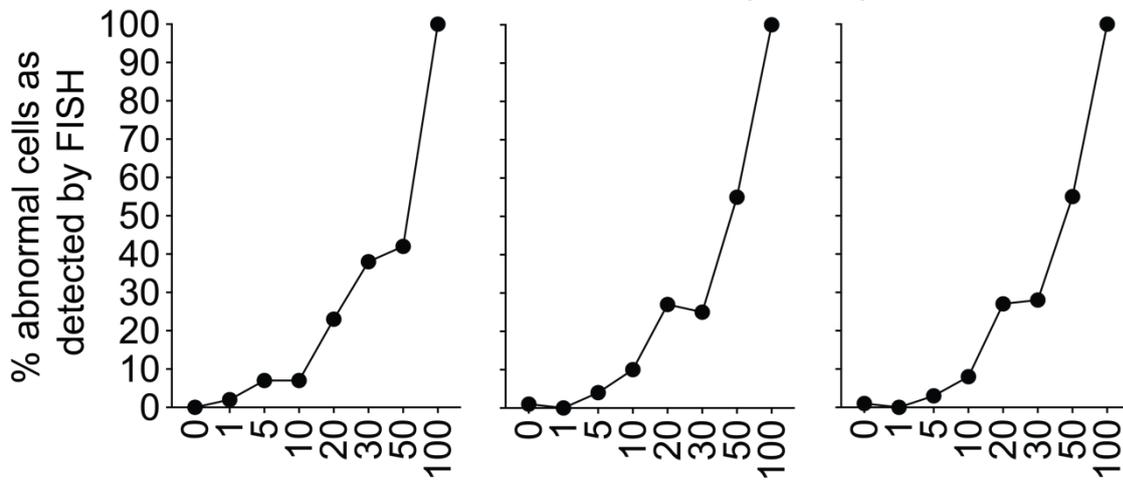
Mosaicism of chromosome 20q11.21 amplification

c
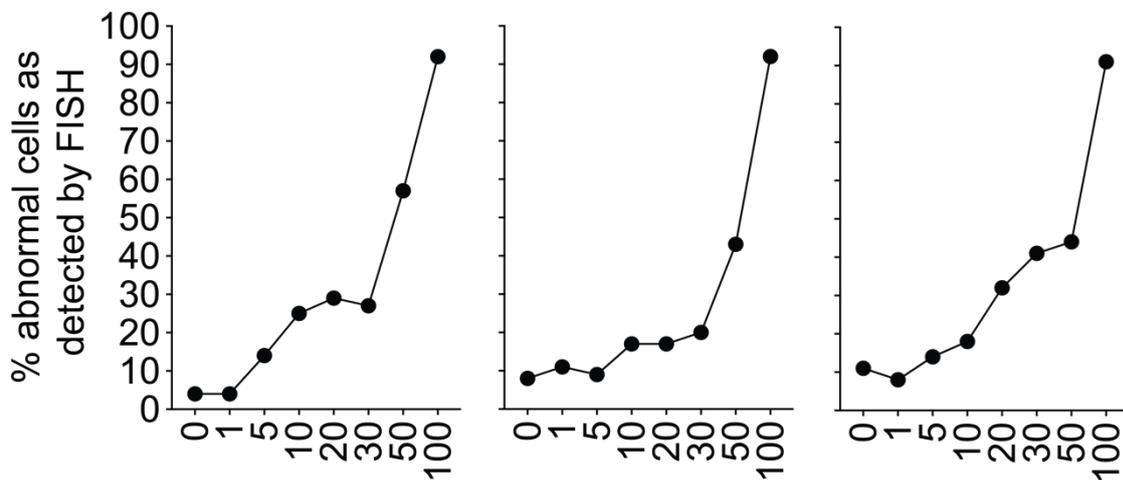
Mosaicism of chromosome 18q21.33 deletions
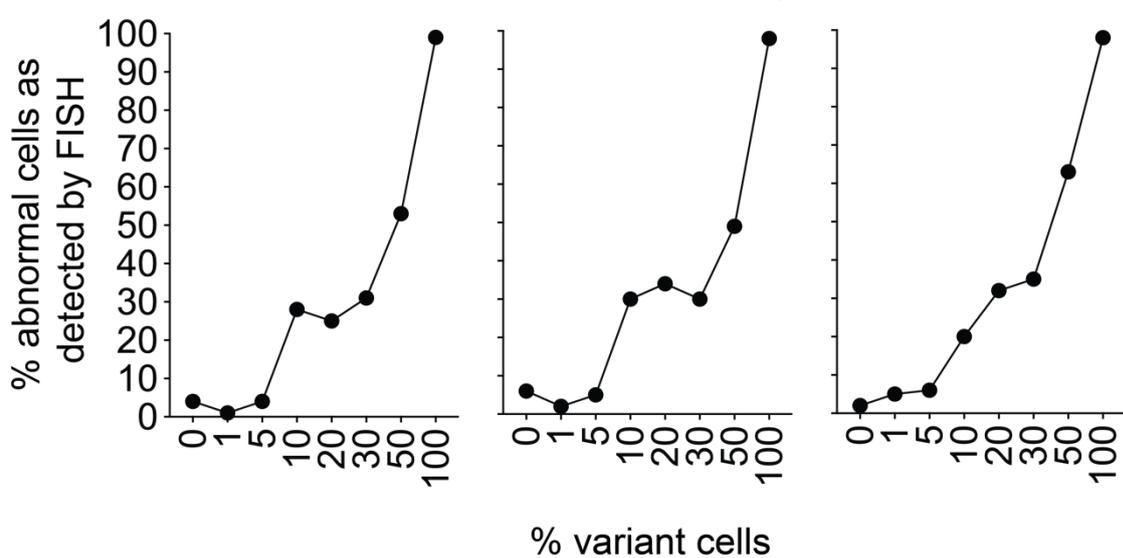
% variant cells

Table 3.2 Summary table for the common techniques for detecting genetic variants, their sensitivity and assay limitations.

Technique	Sensitivity	Suitable for non-dividing cells?	Genome wide/region specific
Giemsa Banding	10% (scoring 30 metaphases)	No	Genome wide
Interphase FISH	10-20% (scoring 100 interphase cells)	Yes	Region specific
qPCR	20%	Yes	Region specific
Array-CGH	20-30%	Yes	Genome wide

3.2.6 Development of the colony-based FISH (C-FISH) protocol

The C-FISH assay requires the analysis of individual colonies that originate from single cells grown on glass slides. Previous data has shown that 95% of single variant cells are capable of migrating a maximum of 43μm over 12 hours post-plating. When human PSC are plated at a density of 4300 cells/cm^2 the average distance between cells was 142μm and they failed to make contact with another cell, ensuring the colonies were clonal (Barbaric et al., 2014). In addition to this, a range of plating densities from 25 to 7500 cells/cm^2 were tested to ensure clonal colonies could be grown using the C-FISH experimental conditions. At low plating densities, there was a linear relationship between the plating density and the number of colonies formed after 72 hours (**Figure 3.10**). At cell densities greater than 5000 cells/cm^2 the number of colonies plateaued or decreased, indicating they had begun to merge and were no longer clonal (**Figure 3.10**). Based on these data, a seeding density of 250 cells/cm^2 was chosen for further experiments.
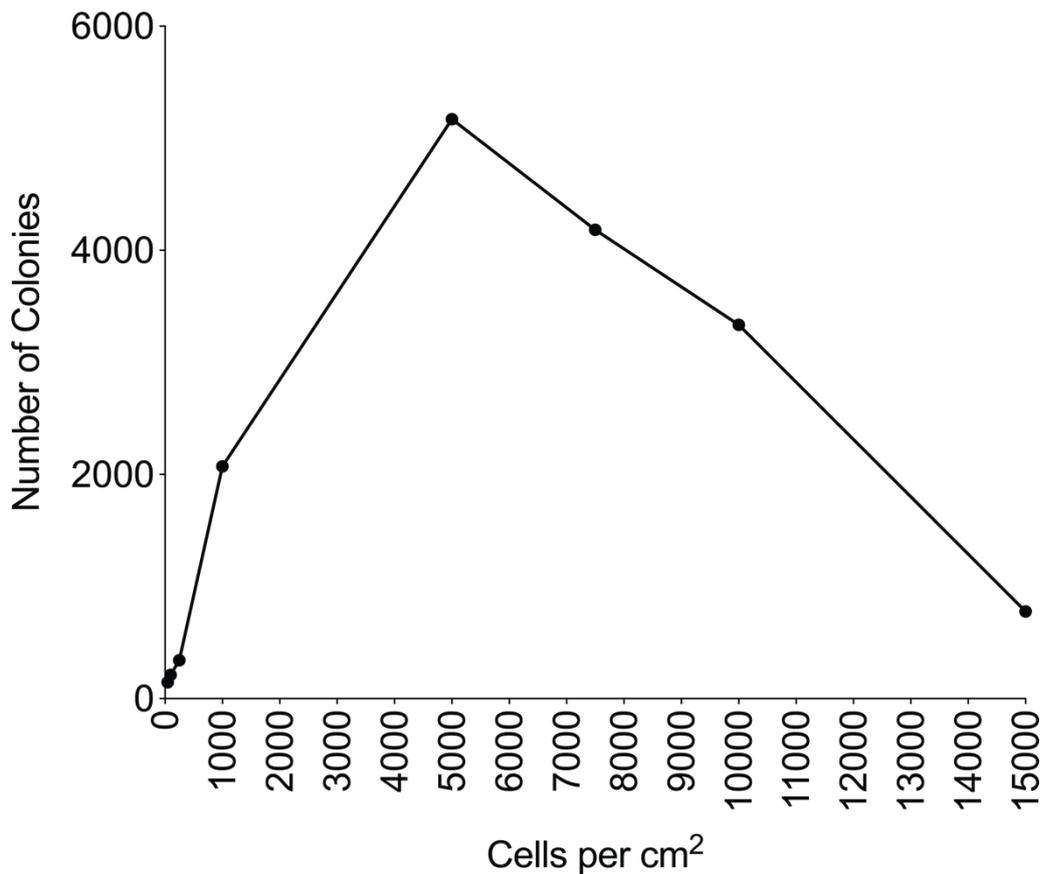
Figure 3.10 | Low plating densities result in a linear relationship to number of colonies. Plot is of number of cells seeded to colonies formed after 72 hours of culture. At a seeding density of less than 5000 there is a linear relationship to the number of colonies formed. Seeding densities greater than 5000 resulted in colonies merging and a loss of clonality.

Interphase FISH requires labour-intensive analysis to score for the presence of positive cells. This introduces variability between analysts assigned to score each assay. In an effort to reduce this variability and also to increase the throughput of C-FISH, a high-throughput image analysis protocol was developed using the CellProfiler open-source software (Carpenter et al., 2006).

To enable the development of the analysis protocol a test C-FISH experiment was performed by plating the chromosome 1 aneuploid H7s6 cell line and the diploid H7s14 cell line at 250 cells/cm^2. The cells were grown for 72 hours and the resulting colonies were fixed and hybridised using a probe for the *MDM4* gene on chromosomal 1q32 region according to the C-FISH protocol. The resulting colonies were imaged using a high content microscopy platform (**Figure 3.11a**). The

remaining cells that were not used in the C-FISH experiment were fixed and processed by conventional interphase FISH with analysis performed manually using a fluorescent microscope.

The image analysis consisted of a nuclei segmentation step based on the DAPI staining (**Figure 3.11b**). The segmented nuclei were then dilated to merge the individual nuclei, allowing them to be segmented as a single object, the colony (**Figure 3.11c**). The segmented nuclei image was also used to mask over the FISH probe image, which eliminated any non-nuclear signal from non-specific binding or debris and improved the accuracy of analysis (**Figure 3.11d**). This masked image was then segmented to count the number of FISH probe signals (**Figure 3.11e**) and a relationship was formed between the FISH probes, nuclei and colony for which they each belonged (**Figure 3.11f**). The copy number of the *MDM4* gene for the parent cell of each colony was determined by averaging all of the cells within the colony and rounding to the nearest whole number (**Figure 3.11g**).

Figure 3.11 | High throughput automated C-FISH image analysis. **a,** Representative high content image acquisition of DAPI stained nuclei. **b,** Segmentation of nuclei from a representative high content DAPI image. **c,** Pixel expansion of the segmented nuclei to generate a pseudo-colony. **d,** Representative high content FISH probe image that is pre-processed to enhance the signal contrast and masked with the segmented nuclei to remove non-specific signals or debris. **e,** FISH probe signal segmentation. **f,** Object relationship formed to attribute the correct probe signals back to the nuclei and colony which they belong. **g,** Copy number of the parent cell was determined as an average per cell within the colony rounded to the nearest whole number.

a High content image acquisition → **b** Nuclear segmentation → **c** Expand to a pseudo-colony

d Preprocessing and nuclear mask of FISH probe → **e** Probe segmentation

nuclear mask

Inset

f Relate colony to nucleus to probe

g Copy number determination
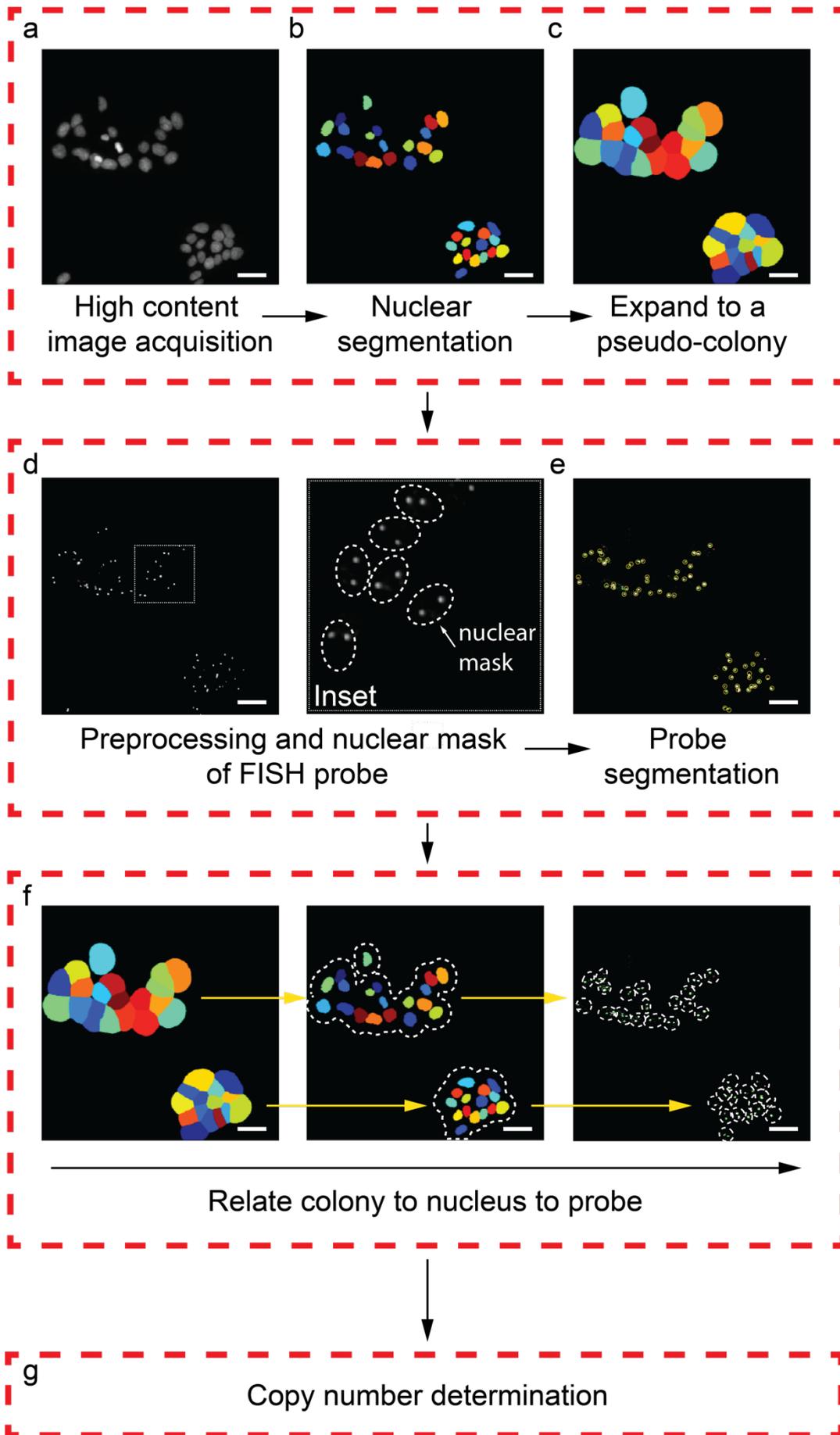
With the assay developed, its accuracy, sensitivity and specificity on a range of different structural and numerical chromosomal variants was tested in comparison to conventional interphase FISH using ROC curve analysis. When developing a novel diagnostic assay, its accuracy at distinguishing a true positive (TP) result from a true negative (TN) result can be compared to other available assays by plotting the sensitivity (true positive rate, TP) against the specificity (false positive rate, FP). This analysis is known as a receiver operating characteristic (ROC) curve and was first developed by radar engineers during World War II to test a radars sensitivity, specificity and accuracy when detecting enemies on the battlefield (Zweig and Campbell, 1993). The sensitivity can be defined as how good the test is at detecting a variant cell and can be calculated from a ROC curve using the following equation (an explanation of terms used in these equations can be found in **Figure 3.12)**:

$$Sensitivity = TP / (TP + FN)$$

Whereas, the specificity or the likelihood that the test will correctly identify the diploid cells can be calculated from the ROC curve using the equation below:

$$Specificity = TN / (TN + FP)$$

Figure 3.12 | Receiver operator characteristic curves assess the accuracy of diagnostic tests. a, Schematic of a ROC curve where a test that is 100% accurate. The test is able to completely distinguish between true negative (TN) and true positive (TP) results (left histogram) and produces a ROC curve that passes through the upper left-hand corner of the ROC plot (right). The area under the curve is equal to 1. **b,** An example schematic of a ROC curve (right) when the test is mostly able to distinguish between true positives and true negative, yet there are some false positive (FP) and false negative (FN) results (left). **c,** An example of a test where the results are random and the test is unable to distinguish between true positive and true negative results (left). The ROC plot passes from the lower left corner the upper right corner with the area under the curve equalling 0.5.

a

TN TP

AUC = 1

True positive rate

False positive rate

b

TN TP

FP

FN

AUC = 0.75

True positive rate

False positive rate

c

AUC = 0.5

True positive rate

False positive rate
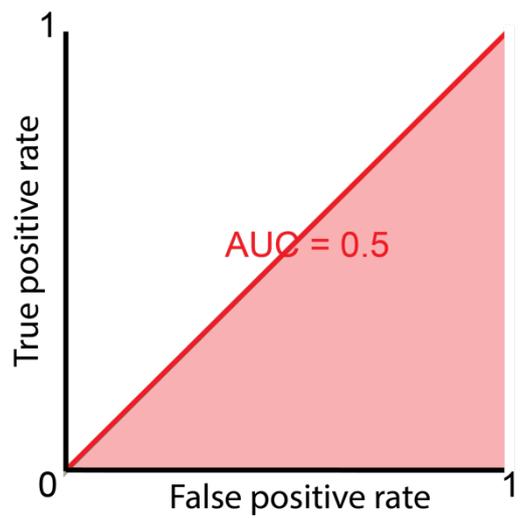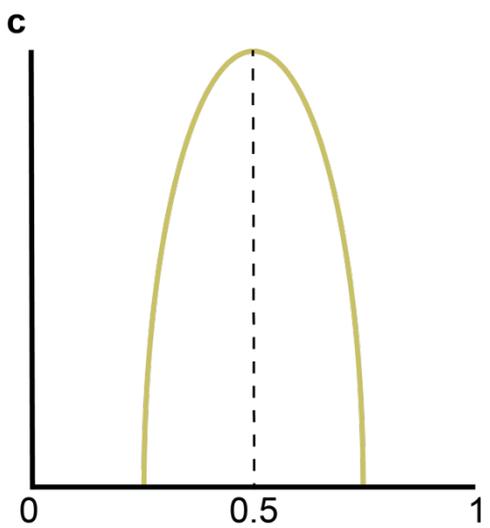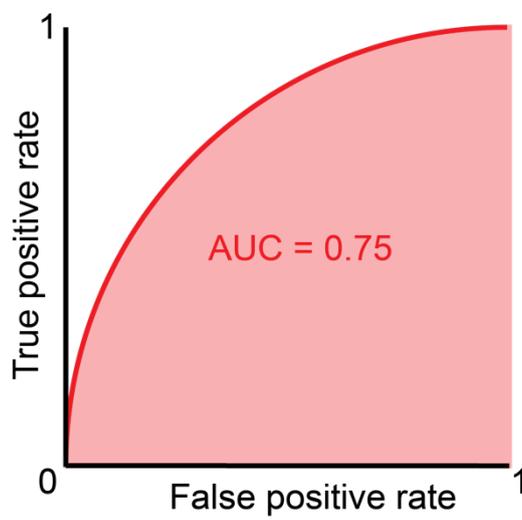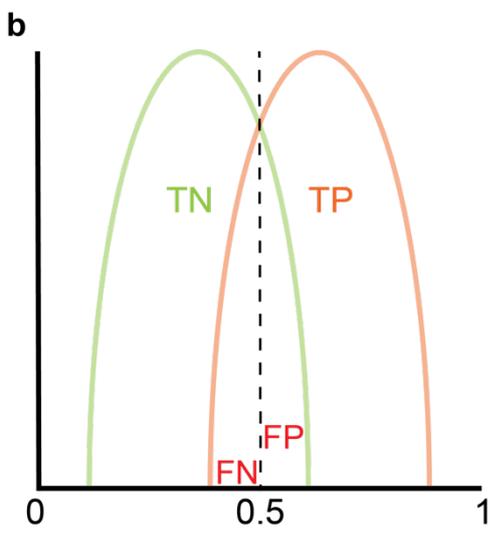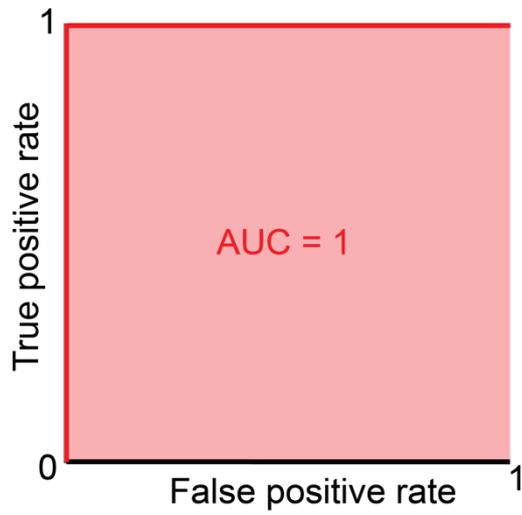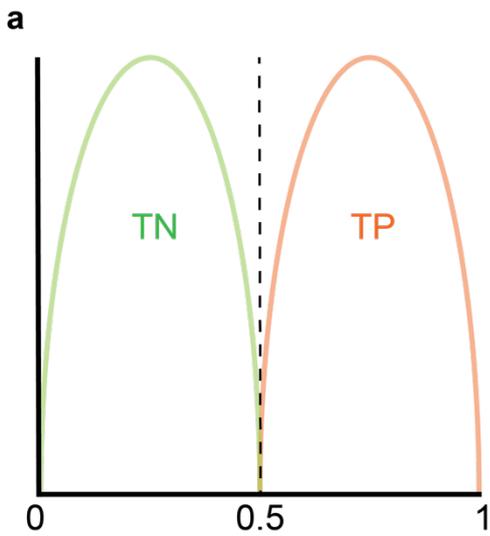
True negative (TN)
True positive (TP)
False negative (FN)
False positive (FP)

Area under the curve (AUC)

The numerical values returned from the sensitivity and specificity equations represent the probability that the assay tested will identify a cell that is in fact variant or diploid respectively. For example, a sensitivity of 95% means, that the test performed on a cell that is certainly variant, there is a 95% chance it will be identified correctly within the specified confidence interval. Determining the accuracy of a test using the ROC is a valuable statistic particularly when making a comparison between multiple assays as it factors in both the sensitivity and specificity values. A test that is 100% accurate will have a sensitivity (true positive rate, TP) of 1 and a specificity (false positive rate, FP) of 0 and will pass through the upper left corner of the ROC curve (**Figure 3.12a**) whereas a test that is returning a random result, will have a sensitivity (true positive, TP) of 0.5 and a specificity (false positive, FP) of 0.5 (**Figure 3.12c**) and will present a line that passes from the bottom left corner to the top right corner of a ROC curve. When results from multiple assays are obtained ROC plots can be graphed together with the plot closer to the upper left corner indicating an assay with greater accuracy (Zweig and Campbell, 1993). To quantify the accuracy of a test the area under the curve can be measured, where a value of greater than 0.5 and less than 1 is expected (**Figure 3.12a-c**).

3.2.7 Sensitivity of C-FISH for the detection of large amplifications to chromosome 1

The accuracy, sensitivity and specificity of C-FISH was tested in comparison to interphase FISH using the same samples of cells across both assays. To ensure the C-FISH assay was suitable for the detection of the full spectrum of karyotypic changes that arise during culture we have tested the assay using cloned cell lines with large amplifications in the form of a trisomy, deletions and tandem duplications.

To test whether C-FISH reduced the number of false negatives when detecting a trisomy, three independent C-FISH experiments were performed using the trisomic chromosome 1 variant line, H7s6. Any remaining cells were sent for conventional interphase FISH analysis by a trained cytogeneticist who analysed 100 interphase nuclei per sample. As the H7s6 variant cell line was clonally derived it was assumed that any cell that was scored as diploid was a false negative. The frequency of true positive and false negative signals detected by conventional interphase FISH was quantified from the three independent experiments (**Figure 3.13a**). These results

were compared to two independent types of analysis performed on the C-FISH data. First, each cell of the colony was scored as variant or diploid independently using the image analysis software, this analysis was termed high content FISH (**Figure 3.13b**). Second, the same data was analysed, although this time the colony calculation was applied to get the average copy number per colony (**Figure 3.13c**). These analyses were chosen to determine the improvement made by performing the colony analysis over the analysis of single cells whilst using the same equipment. By performing interphase FISH in parallel we compared this conventional assay to the C-FISH assay. This revealed that the colony-based approach decreased the frequency of false negatives in comparison to the high content FISH approach by 23%. The frequency of false negatives was 8% over the three experiments which was comparable to the 5% seen with conventional interphase FISH.

Next, to determine the accuracy of C-FISH in comparison to interphase FISH, three independent C-FISH and interphase FISH experiments were executed on the diploid H7s14 line that was the isogenic diploid parent line to H7s6. The true-positive and the false-positive rates from both the conventional interphase FISH and C-FISH experiments were plotted as a ROC curve (**Figure 3.13d**). The ROC analysis showed that the C-FISH approach was more accurate than conventional interphase FISH with an area under the curve (AUC) value of 0.99 versus 0.98 (**Figure 3.13d**). These areas were deemed to be significantly different by an unpaired t-test (P<0.05). A summary of the sensitivity and specificity of these assays can be found in **Table 3.3a**.

a

Conventional Interphase FISH

b

High content FISH

c

Colony Based FISH

d
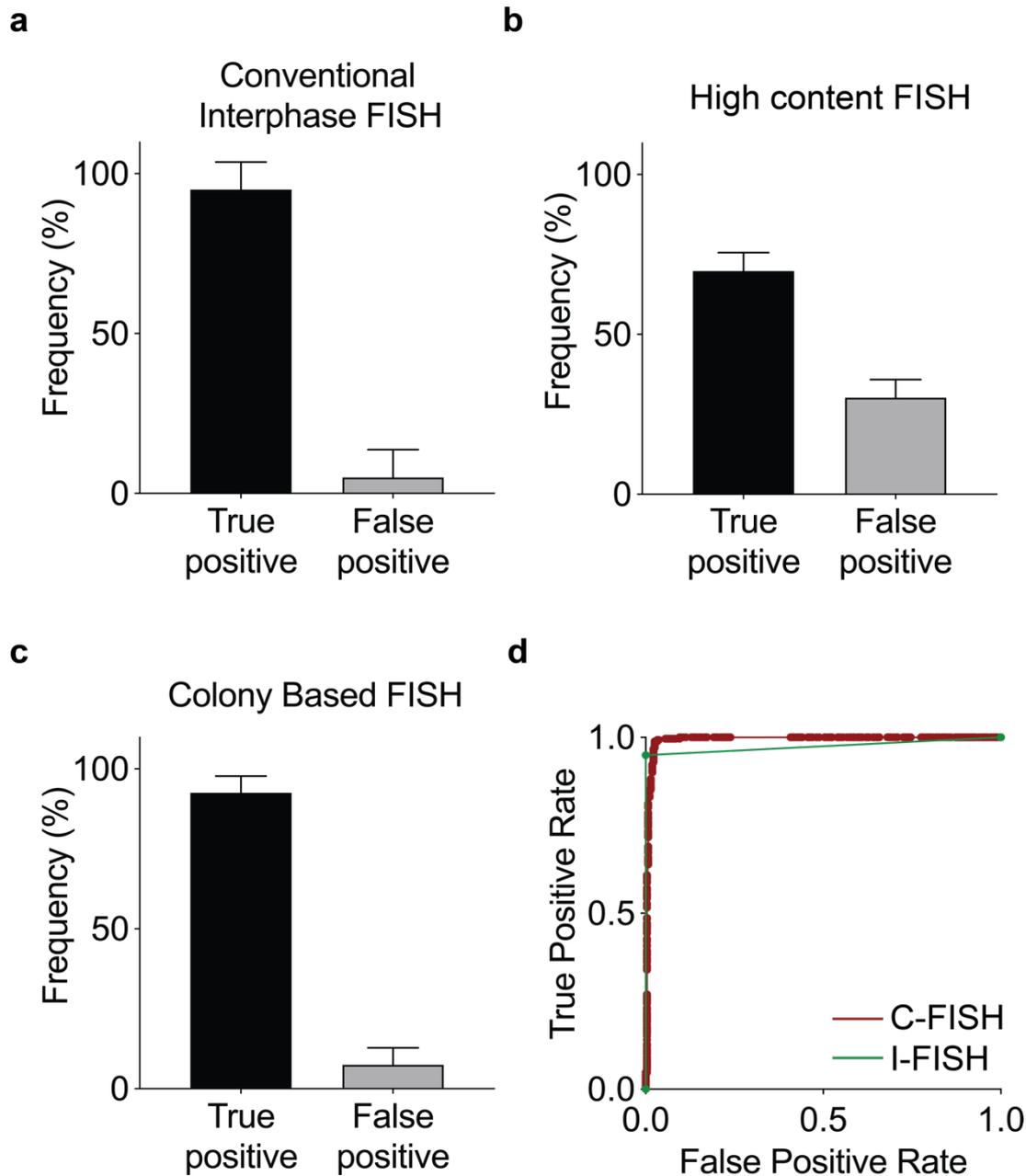
Figure 3.13 | C-FISH improved the accuracy of detecting amplifications to chromosome 1. a-c, The frequency of true positive and false positive results returned when the same culture of H7s6 1q variant cells was analysed by (**a**) conventional FISH (**b**) high content FISH or (**c**) colony-based FISH. Results are the mean of three independent experiments, ± s.d. **d,** ROC curve analysis of the true positive and false positive rate of C-FISH (red) and conventional interphase FISH (green) for the detection of amplifications to chromosome 1q in the variant H7s6 and diploid H7s14 cell lines. Accuracy of detection was determined from the area under the curve, 0.99 and 0.98 respectively.

Table 3.3 Summary table of sensitivity and specificity values determined from ROC analysis

a

	Large amplification or trisomy	
	Sensitivity (±95% confidence interval)	Specificity (±95% confidence interval)
Interphase FISH	95% (92% to 97%)	100% (99% to 100%)
C-FISH	99% (96% to 100%)	97% (95% to 99%)

b

	Deletions or monosomy	
	Sensitivity (±95% confidence interval)	Specificity (±95% confidence interval)
Interphase FISH	100% (99% to 100%)	100% (99% to 100%)
C-FISH	99% (97% to 100%)	99% (97% to 100%)

c

	Tandem duplications					
	Sensitivity (±95% confidence interval)			Specificity (±95% confidence interval)		
	<350kb	350kb to 2Mb	>2Mb	<350kb	350kb to 2Mb	>2Mb
Interphase FISH	59% (53% to 64%)	73% (67% to 78%)	73% (68% to 78%)	100% (99% to 100%)	94% (91% to 97%)	100% (99% to 100%)
C-FISH	67% (61% to 72%)	93% (89% to 95%)	95% (92% to 97%)	90% (85% to 93%)	92% (89% to 94%)	94% (90% to 978%)

3.2.8 Sensitivity of C-FISH for the detection of deletions to chromosome 18

Next, the C-FISH approach was applied to the detection of deletions to chromosome 18. For this analysis, the isogenic UCB144 cell lines generated in 3.2.1 was utilised. Again, the accuracy of the C-FISH approach was compared with conventional interphase FISH using ROC curve analysis (**Figure 3.14**). Three independent experiments were performed and in each experiment, the cells harvested from a single flask of chromosome 18 deletion, UCB144-D14 or the diploid UCB144-F6 were assayed by C-FISH and the remaining cells were fixed and processed by conventional interphase FISH. The individual sensitivity and specificity values determined from the ROC curve for both C-FISH and interphase FISH when detecting deletions are shown in **Table 3.3b.** By conventional interphase FISH, no

false-negative calls were reported, and the assay was deemed perfect returning an AUC value of 1.00. For C-FISH we did detect some errors, although far greater numbers of colonies were analysed, and the accuracy determined from the AUC value to two decimal places was also 1.00.

Deletion to chromosome 18
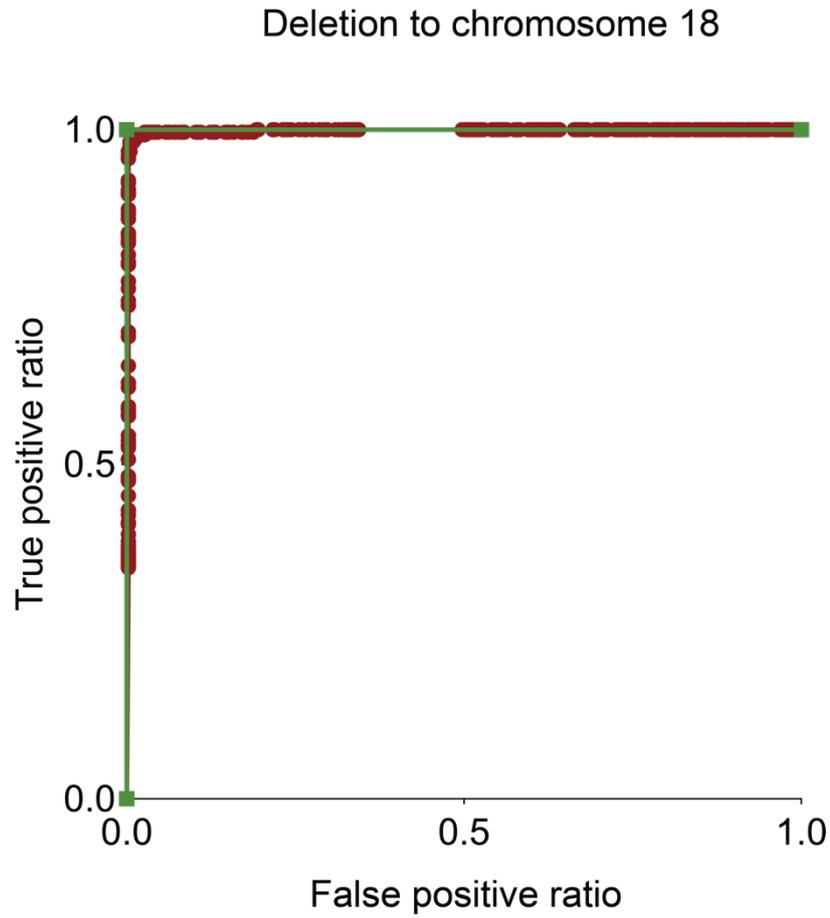
Figure 3.14 | ROC curve analysis showed comparable accuracy when detecting deletions to chromosome 18 by C-FISH and conventional interphase FISH. ROC curve analysis of the true positive and false positive rate of C-FISH (red) and conventional interphase FISH (green) for the detection of deletion to chromosome 18 in the variant UCB144-D14 and diploid UCB144-F6 cell lines.

3.2.9 Applied intensity threshold improves the resolution of overlapping signals in cell lines with tandem duplications

Next, C-FISH was tested when detecting tandem duplications to chromosome 20. As before, the efficacy of averaging the number of FISH signals per colony, to improve

the sensitivity of the assay was tested. To do this, the number of signals per cell of a colony was averaged and compared to an analysis of each cell when analysed individually. For these analyses, the clonally derived chromosome 20 variant line, TC113-E1, was utilised. The number of correctly called true positive cells or colonies and incorrectly called false-negative cells or colonies from each approach across three independent experiments was determined. The colony based approach reduced the frequency of false negatives calls from 55% to 46% when compared to the high content FISH analysis of the same cells when analysed individually (**Figure 3.15a,b**). However, conventional interphase FISH outperformed the colony-based approach, reporting only 31% false negatives (**Figure 3.15c**). I surmised that this was due to high powered objectives that are capable of resolving overlapping signals more readily in conventional interphase FISH.
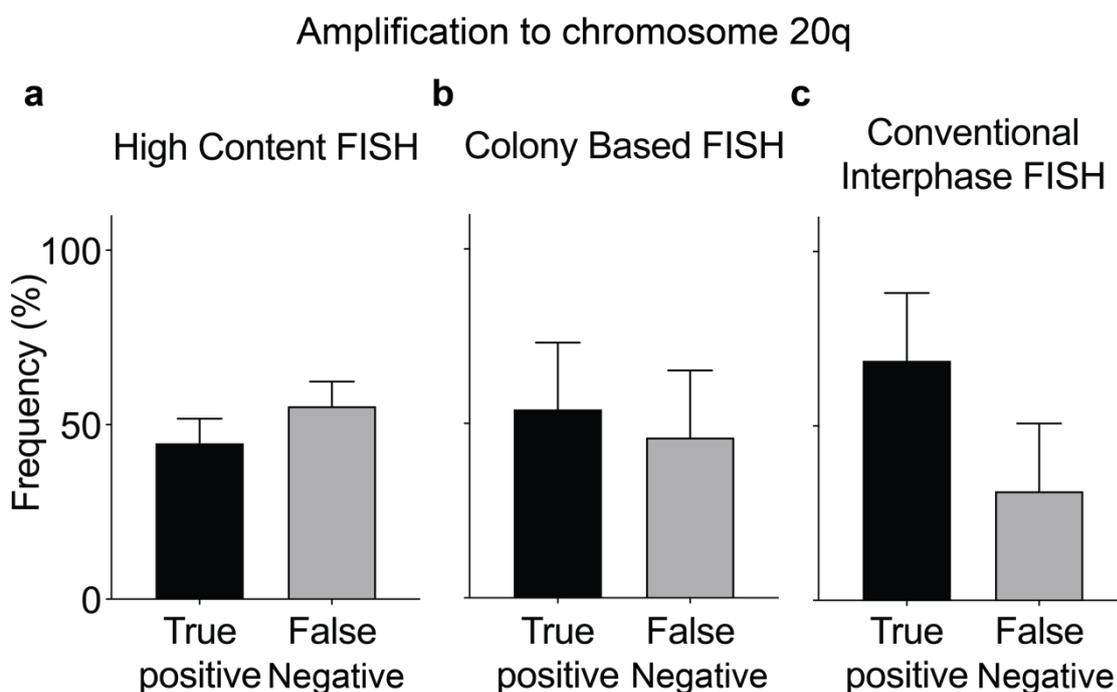
Figure 3.15 | C-FISH failed to improve the detection of chromosome 20 tandem duplications. a-c, The frequency of true positive and false positive results returned when the same culture of chromosome 20 variant, TC113-E1 cell line, was analysed by (**a**) high content FISH, (**b**) colony-based FISH and (**c**) conventional interphase FISH. Displayed are the mean results from 3 independent experiments, ± s.d.

When the average C-FISH copy number of the diploid, TC113-G2 and the chromosome 20 variant, TC113-E1 cell lines were compared it was found that the results largely overlapped (**Figure 3.16a**) showing that the C-FISH assay was unable to distinguish between the diploid and variant cells due to the prevalence of overlapping signals in the variant cell line. It was reasoned that it may be possible to detect the presence of overlapping signals in false-negative cells *in silico.* Whilst performing the C-FISH approach on the chromosome 20 variant line TC113-E1, it was noted that those cells with two FISH signals, i.e. false negatives, have a stronger signal intensity on the chromosome harbouring the 20q duplication (for an example see the red box, **Figure 3.1a**). To test this theory, the C-FISH protocol was performed on the diploid cell line TC113-G2 and its 20q CNV clonal variant subline, TC113-E1. The ratio between signal intensities was measured in all cells where only two probe signals were detected i.e. a false negative (**Figure 3.16b**). The signal intensity ratio in the diploid and variant lines were 0.75 and 0.59 respectively. Using this ratio, I applied what I termed an 'intensity correction threshold', whereby any cell with 2 probes that had an intensity ratio of <0.55 was classified as false negatives and corrected *in silico*. This modification to the C-FISH protocol greatly improved the separation of the respective normal and variant populations (**Figure 3.16c**), a necessary step to improve the accuracy of the C-FISH approach.

Figure 3.16 | Intensity correction threshold greatly improves the separation of the chromosome 20 variant and diploid populations. a, Histogram plot of the mean number of signals per cell per colony of diploid TC113-G2 (46,XY) (green) and the 20q variant line TC113-E1 (46,XY,dup(20)(q11.21)) (red). This plot shows the inability of C-FISH to distinguish between variant and diploid cells. **b,** Signal intensity ratio, ratio between the signal intensities of all cells where only two probe signals were detected. Data plotted is from TC113-G2 (green) and TC113-E1 (red). An intensity ratio between the two signals present within each cell in the diploid and variant populations was 0.75 and 0.59, respectively (****$P<0.0001$; unpaired *t*-test). **c,** Histogram plot of the data displayed in (**a**) after the application of the signal intensity correction threshold. The signal intensity correction threshold greatly improved the separation of the respective normal and variant populations.
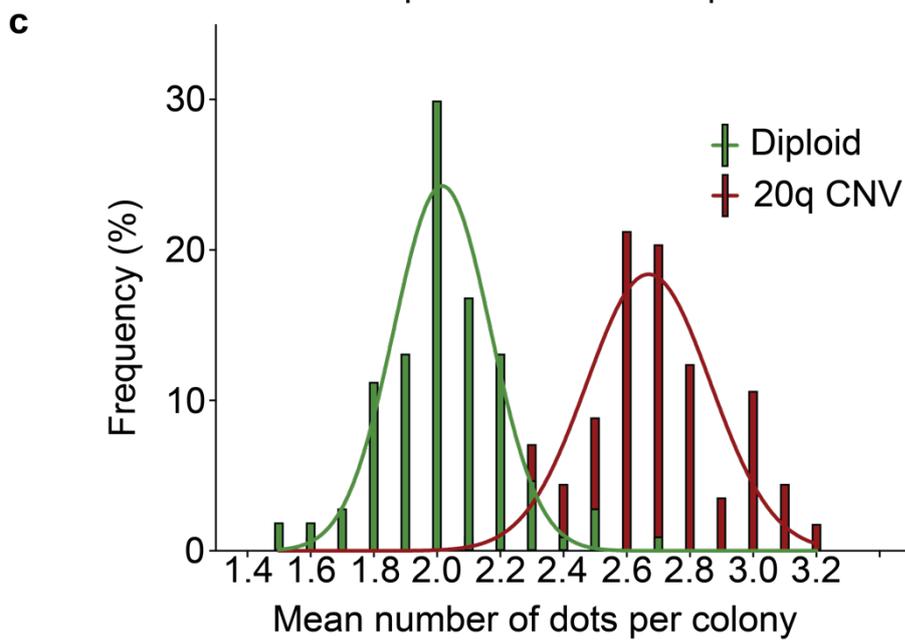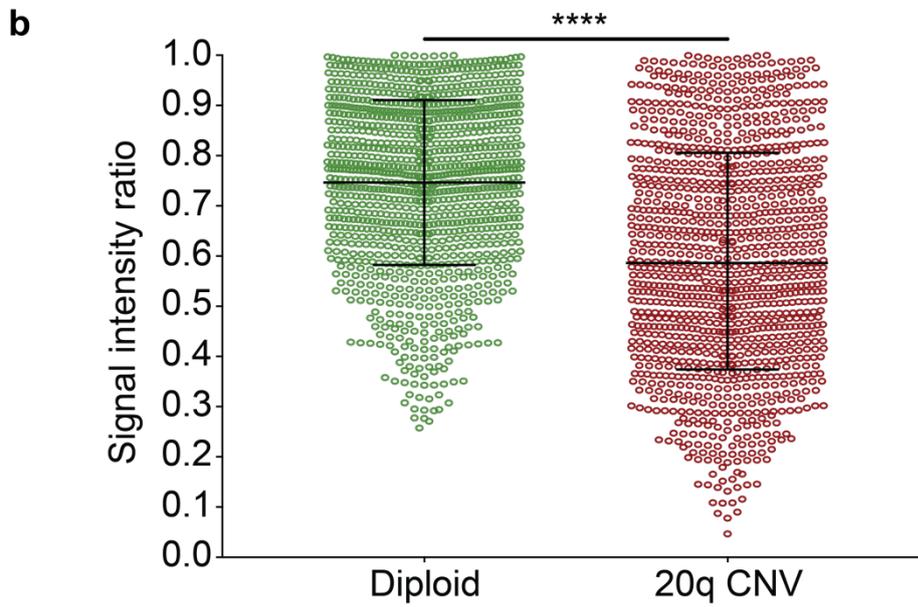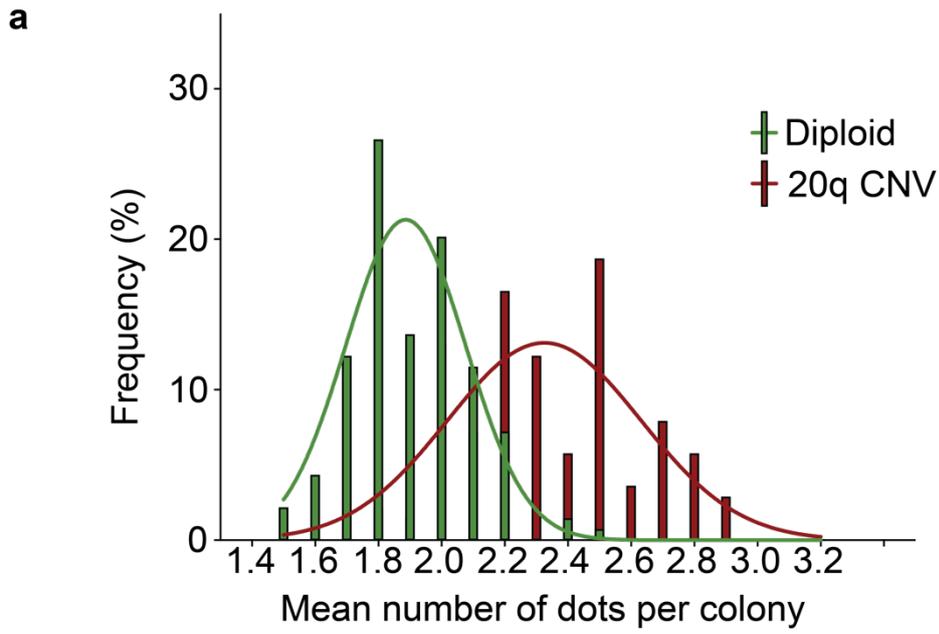
3.2.10 Sensitivity of C-FISH for the detection of amplifications to chromosome 20, including small tandem amplifications

The C-FISH and conventional interphase FISH assays were performed on four variant lines that harbour a chromosome 20 duplication and their parent diploid counterparts. Three of these variant lines were selected as the duplication in each line were of different lengths, as determined by qPCR and G-banding (**Figure 3.17**). The cell line H7s14-20q had the smallest duplication measuring approximately 350kb (**Figure 3.17a**), TC113-E1 and TC1130-E6 have a duplication and a triplication respectively, measuring 2Mb (**Figure 3.17b**) and H7s6 has a whole q arm duplication present as an isochromosome as detected by qPCR and G-banding (**Figure 3.17c**). For the ROC based analysis, these lines were paired with their isogenic diploid lines, H7s14 and TC113-G2.

Figure 3.17 | Chromosome 20 duplication length in the three selected variant cell lines. a-c, Chromosome 20 duplication length, determined by qPCR for the cell lines S14-20q (**a**), TC113-E1 (**b**) and H7s6 (**c (i)**). The length of the duplication was 350kb, 2MB and whole 20q arm duplication respectively. qPCR results are mean copy number from three technical replicates ±s.d. **c (ii),** G-banding result for H7s6 which presents a duplication of the entire q arm of chromosome 20 in the form of an isochromosome.

a

S14-20q

b

TC1133 E1

c (i)

H7s6

Primer gene/position on chromosome 20
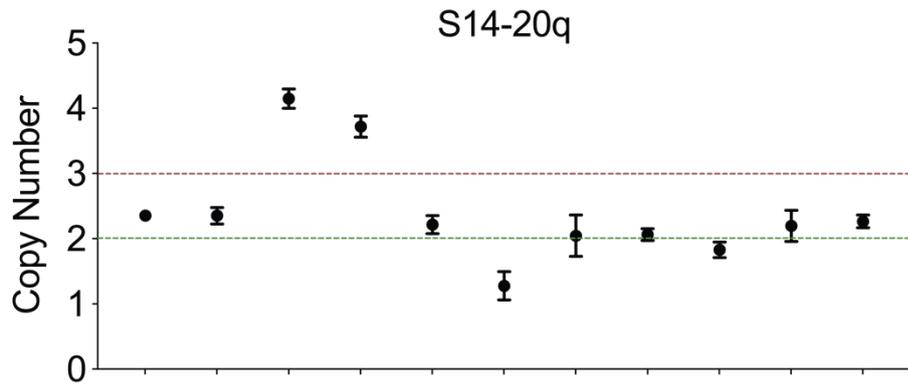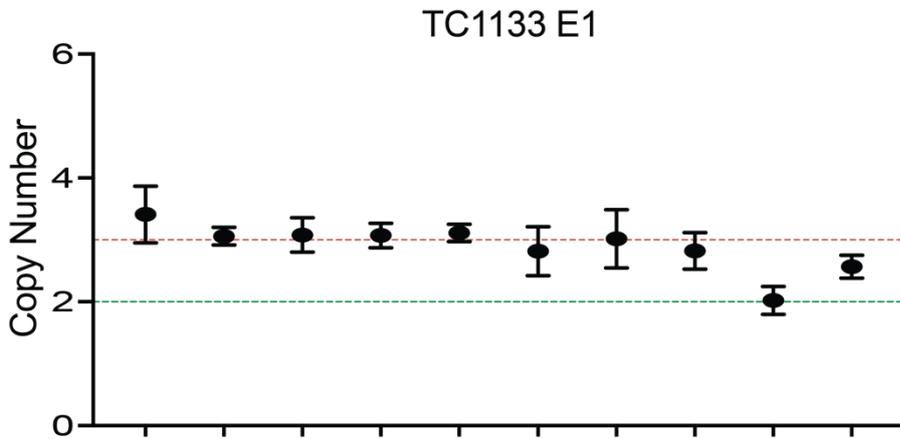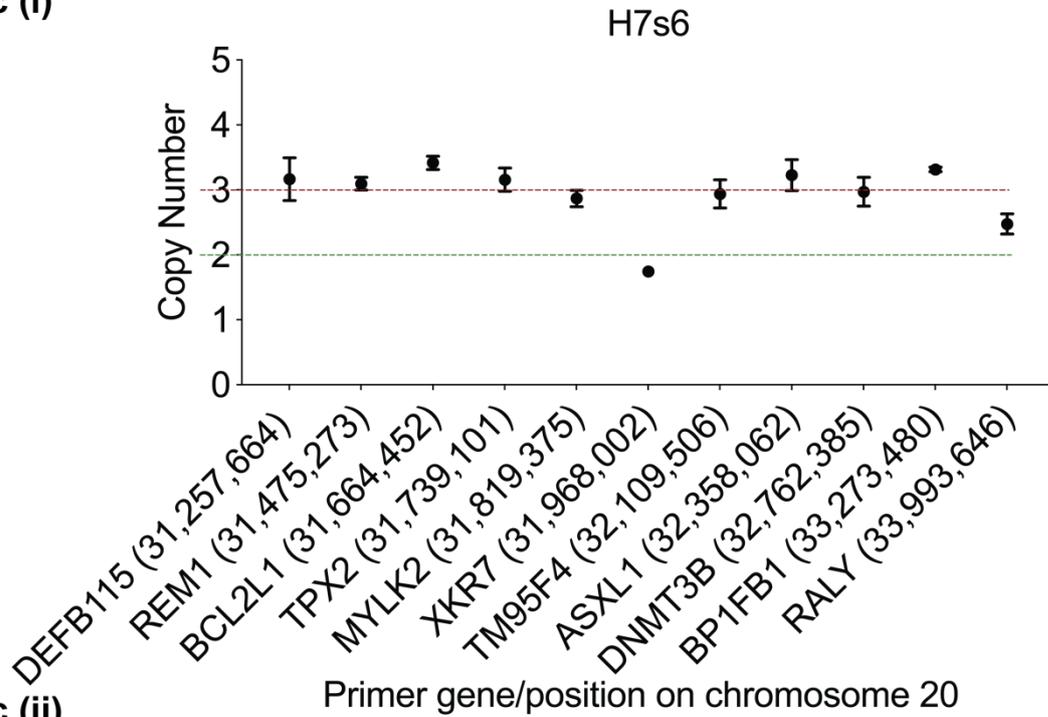
c (ii)

Karyology (G banding result)

Name	H7S6 P+3
Result	47,XX,+del(1)(p22p22),der(6)t(6;17)(q27;q1),t(12;20)(q13;q13.3),i(20)(q10)[30]

For each of the variant and diploid paired cell lines, the C-FISH experiment was performed in triplicate with paired samples analysed using conventional interphase FISH by a trained cytogeneticist. The performance of C-FISH was compared to the results attained from the paired interphase FISH experiment using ROC curve analysis (**Figure 3.18**). In all cases, C-FISH was able to more accurately detect the variant cells than conventional interphase FISH. The AUC values, a measure of assay accuracy, for the H7s14-20q/H7s14 pair, C-FISH was 0.89 whereas interphase FISH was 0.79 (**Figure 3.18a**). For the cell pair TC113-E1/TC113-G2, the AUC values were 0.966 and 0.835 for C-FISH and interphase FISH respectively (**Figure 3.18b**). The C-FISH approach was also tested on the clonal cell line TC113-E6 that possesses a triplication of the same 2MB region on chromosome 20q11.21. As expected, the accuracy increased to 0.995 (**Figure 3.18b**) as we found fewer false negatives and the signal intensity ratio was greater due to the three overlapping signals of the triplication that enabled better *in silico* resolution. Lastly, for the cell pair H7s6/H7s14, the AUC value was 0.985 and 0.865 for C-FISH and interphase FISH, respectively (**Figure 3.18c**). Both sensitivity and specificity values of C-FISH and interphase FISH when detecting tandem duplications <350kb, between 350kb and 2Mb and greater than 2Mb can be found in **Table 3.3c**. The analysis from these cell lines has shown that the accuracy of both C-FISH and interphase FISH increased as the size of the duplication increased. With larger duplications the distance between probes increased and the likelihood of overlapping signals was reduced, therefore increasing the accuracy of the analysis.

Figure 3.18 | Comparative ROC curve analysis of Interphase FISH and C-FISH assays on 4 independent paired diploid and chromosome 20 variant cell lines. a-c, Comparison ROC curve analysis (left) for interphase FISH (green) and C-FISH (red) using a range of chromosome 20 variant cell lines that each display a different sized duplication. The length of the duplication is depicted in the chromosome 20 ideogram (right). The cell line pairs assayed were (**a**) H7s14-20q and H7s14 when detected by C-FISH and interphase FISH. (**b**) The ROC curve analysis for the cell line pair TC113-E1 and TC113-G2. The cell line TC113-E6 has a tandem triplication and was paired with the diploid TC113-G2 (blue) which was more readily detected by C-FISH (**c**) The largest amplification in the cell line H7s6 presents a whole q arm duplication as an isochromosome and was the most accurately detected by C-FISH.
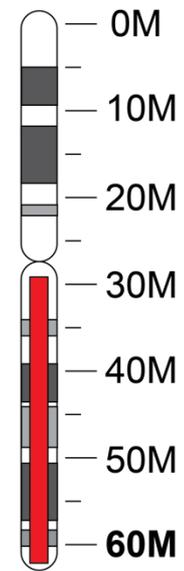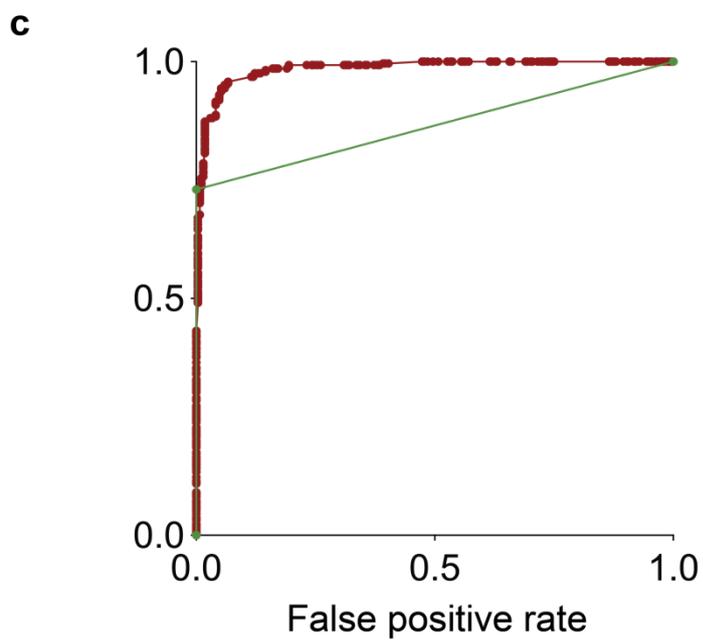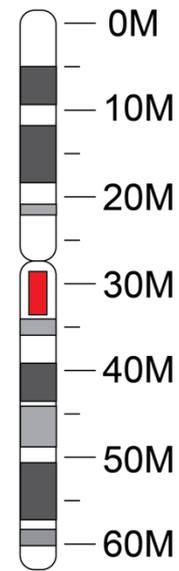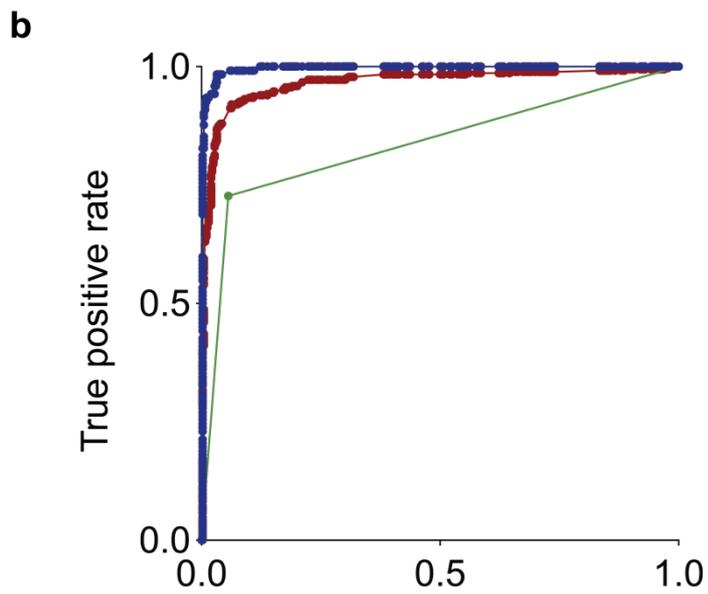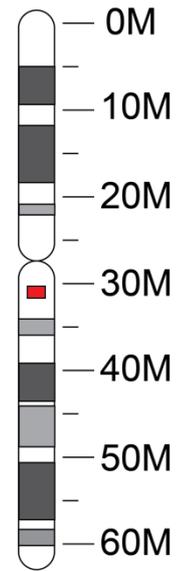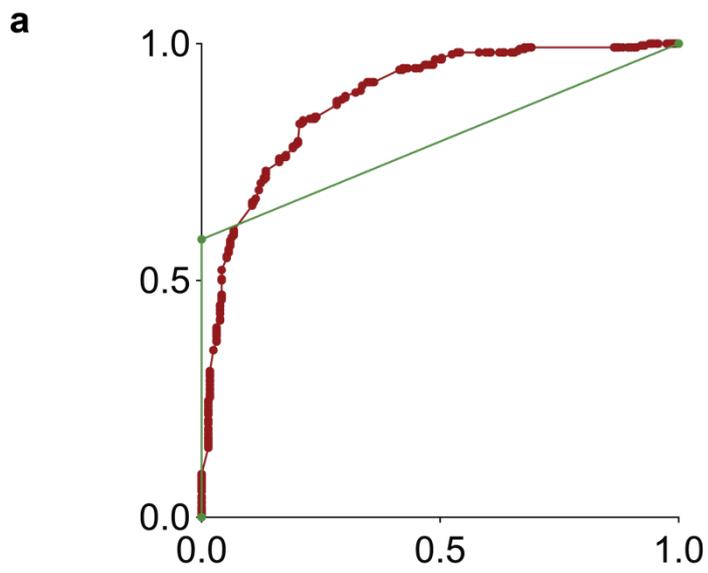
3.3 Discussion

Currently, the detection of genetically variant human PSC relies on a battery of assays that collectively can detect the full complement of karyotypic variants, including large amplifications including trisomy, deletions that include monosomy and small amplifications such as tandem duplications. Interphase FISH is routinely employed to detect small duplications on chromosome 20, although it is insensitive and unable to detect these variants when present in less than 10% of a culture (Baker et al., 2016). The sensitivity of interphase FISH is hindered by frequent false-negative results that are caused by overlapping signals or the inability to separate signals from two probes in close proximity to one another (Kearney, 2001). Within this study, a novel and high throughput interphase FISH assay was developed that was termed C-FISH, for the accurate detection of chromosomal instabilities that afflict cultures of human PSC. C-FISH improves on the sensitivity of interphase FISH by discounting the impact of false-negative results by averaging the number of FISH signals per cell of a clonal colony. The copy number of the colonies parent cell is then deduced from this number.

For routine screening of human PSC cultures by interphase FISH, the analysis of more than 100 interphase nuclei is impractical. However, it was previously shown that it is possible to increase the sensitivity of detecting chromosome 17 variant cells by interphase FISH to 1% when the number of interphase cells scored was increased from 100 to 1000 (Baker et al., 2016). With this in mind, the throughput of C-FISH was increased and also analyst subjectivity was reduced by utilising a high content and automatic image acquisition platform, coupled this with our in-house developed C-FISH image analysis pipeline. These steps allowed for the scoring of hundreds of colonies in only a few hours. Further, the indexed output from our analysis pipeline allows the user to review the images from putative results and further improves the accuracy of the C-FISH assay.

First, C-FISH and interphase FISH were tested when detecting cell lines with a chromosome 1 aneuploidy and a deletion to chromosome 18 and found the assay improved or performed comparably with these readily detected chromosomal variants. However, with the automation of C-FISH, it can be performed

independently of a trained cytogeneticist making it more appropriate for routine laborious laboratory screening of cultures of human PSC. The main aim when developing this assay was to enhance the accuracy with which small tandem duplications are detected. Initially, when the chromosome 20 tandemly duplicated line TC113-E1 was screened, C-FISH reported a higher frequency of false negatives compared to conventional interphase FISH. This lack of improvement was likely due to the 40X microscope objective used in C-FISH compared the 160X objective used in the conventional approach. The high powered, 160X objective used in conventional interphase FISH was more able to resolve overlapping signals in chromosome 20 CNV lines than the 40X objective we used in C-FISH. However, the lower power objective was necessary to ensure that C-FISH was higher throughput and automated, unlike conventional interphase FISH. Therefore, I chose to resolve closely bound signals *in silico*. False-negative cells often displayed one signal which had a much higher signal intensity. This higher intensity was caused by two probe signals that have merged due to the proximity on the amplified chromosome. This observation was capitalised on to detect false negatives *in silico*. The *in silico* intensity correction threshold improved the precision of the assay when detecting false-negative cells in the variant line and, importantly had a minimal negative effect on the precision of the assay for correctly calling diploid cells. This development greatly improved the accuracy of C-FISH.

I attempted to directly test the sensitivity of the C-FISH approach using our GFP labelled diploid cells in mixing experiment with our chromosome 20 variant lines. It was postulated that the sensitivity of C-FISH could be confirmed on a colony by colony basis, confirming the genotype of the colony based on the presence or absence of the GFP signal. It was discovered that the GFP protein was destroyed during the Carnoys fixation, a process required when hybridising cells with FISH probes. To get around this, attempts were made to image the slide whilst the cells were still alive and before the GFP signal was lost by fixing the cells. This image could then be related to the final images of the probed colonies following fixation. However, this was technically very difficult to achieve, as I was unable to image through both the cell culture flask and the glass slide. However, using ROC curve analysis it was possible to circumvent these issues and ascertain the accuracy,

sensitivity and specificity of C-FISH in comparison to interphase FISH when both assays were performed on the same samples of cells.

Using four chromosome 20 variant lines, that present increasing lengths of the duplication and their isogenic diploid counterpart lines, the accuracy of C-FISH and interphase FISH were compared using ROC curve analysis. In each case, the C-FISH assay greatly improved sensitivity, specificity and accuracy. Interestingly, the variant lines were more accurately detected when they presented larger amplifications. This improvement to sensitivity was due to fewer false-negative results from the greater separation of the FISH probe signals present on the same, amplified chromosome. Additionally, using C-FISH the TC113-E6 cell line was tested, TC113-E6 has a tandem triplication which is the same length as the tandem duplication found in the TC113-E1 cell line. The TC113-E6 tandem triplication line was detected with greater accuracy than the duplication. This suggests that this improvement came from several different sources. First, the triplication presents an additional signal to the duplicated line, where this was detected it increased the average copy number per colony than the three signals produced from the duplicated line. Second, the presence of four signals decreased the likelihood of a false negative on a cell by cell basis. Lastly, when we did detect a false negative, the signal intensity between the two signals was even greater than in the duplicated line and as such was more readily detected by our *in silico* detection protocol.

Collectively, the novel C-FISH assay developed here provides greatly improved assay accuracy than conventional interphase FISH. The protocol requires equipment accessible by most laboratories and with the developed automated image analysis pipeline, does not require a cytogenetic specialist. Per test, the material required for C-FISH cost less than $100 which is considerably cheaper than outsourcing for interphase FISH that can cost upward of $850 (Baker et al., 2016). Further, the throughput of the assay is not limited by the laborious analysis and can potentially be expanded to analyse a limitless number of colonies. Moreover, the development of improved assays that can sensitively detect low-level mosaicism in cultures is crucial, progressing with genetically compromised cells could have catastrophic consequences for patients of cell-based therapy and to a lesser point, compromise the integrity of results attained from research on human PSC.

The development of C-FISH was imperative for future work in this study. Sensitive assays for the detection of chromosome 20 variant lines were crucial in the identification of subject lines that we will use in the following chapter, where long-read next-generation sequencing coupled with bioinformatic analysis was used to explore the breakpoint positions involved in this commonly occurring tandem duplication. With a better understanding of the breakpoints involved in the generation of this tandem duplication, it will be possible to elucidate the mechanisms that lead to it arising and target this to stop its appearance, ensuring the safety of human PSC based therapies.

4 Elucidation of the breakpoints in chromosome 20 variant human PSC cell lines

4.1 Introduction

Chromosome 20 is one of two small metacentric chromosomes and in Karyology is given the F classification based on its small size. Evolutionarily, the Human chromosome 20 shares its organization with the African great apes, both of which diverged after a common pericentric inversion event took place involving the entire p arm of the chromosome (Misceo et al., 2004). Diseases with chromosomal abnormalities affecting chromosome 20 have, comparatively, not drawn much attention in the past. Trisomic duplication of chromosome 20 is rarely viable and partial duplications of the p and q arms are rare and often only present as part of a recognisable syndrome (Simpson, 1988).

During the culture of human PSC, copy number variants are known to arise on chromosome 20 and represent one of the more frequently observed karyotypic changes (Amps et al., 2011, Seth et al., 2011). The majority of copy number variants appear as tandem duplications and always include the selectively advantageous driver gene, *BCL2L1* (Avery et al., 2013). The proximal end of 20q11.21 duplication lies within the unmapped centromeric region, whereas the distal breakpoint has been reported to vary in position down the q arm of the chromosome (Amps et al., 2011). The chromosome 20 variant has been well characterised, with a great deal of resources expended on understanding the mechanism through which this variant possesses a selective advantage (Avery et al., 2013). However, little is known about the mechanism through which these tandem duplications arise.

Genomic architecture plays a role in a region's susceptibility to acquiring a copy number change (Lupski, 1998). Through the sequencing of breakpoint regions, it is possible to infer the mechanism for the generation of copy number variants. In mitotic cells, it's been observed that breakpoint regions lie in repetitive sequences with microhomology (Kidd et al., 2008, Perry et al., 2008). Likewise, the induction of replication stress can lead to copy number variations that contain breakpoints in regions of microhomology (Arlt et al., 2009). This indicates that replication may be

responsible for inducing genome-wide copy number changes. Homologous recombination plays a major role in restarting collapsed forks. One ended double-strand breaks are left in the wake of collapsed forks and are repaired by break-induced replication (Llorente et al., 2008). However, as break-induced replication relies on the accurate alignment of genomic regions, misalignment of repetitive sequences can lead to genetic amplification.

One of the most abundant repetitive elements are the *Alu* retrotransposons that make up 11% of the human genome and have a copy number that is greater than 1 million (Lander et al., 2001). These repetitive elements are approximately 300bp in size, derived of a dimer of 7SL RNA genes that are divided by a short A-rich region (Deininger et al., 2003). *Alu* elements have been established as a significant source of genomic instability and are responsible for copy number variation in tumours (Deininger and Batzer, 1999, Kolomietz et al., 2002, Elliott et al., 2005). In one study, a high frequency of translocations was observed when a double-strand break was introduced adjacent to the *Alu* sequence (Elliott et al., 2005), whereas another showed that double-strand breaks are likely to form at *Alu* elements due to their secondary structures that causes fork stalling (Voineagu et al., 2008). It is then likely that these elements have not only contributed to genetic diversity but are also a major player in driving genetic instability.

This chapter will describe the experimental results from long-read next-generation sequencing to map the breakpoint of a chromosome 20 tandem duplication detected in a human ES cell line. Short read sequencing is often limited when it comes to sequencing repetitive regions such as those found in the centromere that remain unmapped. Therefore, the long-read strategy chosen here allows for the sequencing of breakpoints, even those flanked by repetitive unmapped regions and permits sequence alignment analysis to infer the mechanism responsible for this copy number variant.

Due to the complex nature of DNA sequencing terminology, a table of terms used in the following chapter can be found on the following page (**Table 4.1**).

Table 4.1 Appendix of terms for Chapter 4

Term	Description
Microhomology	Short regions of the genome with the same (homologous) sequence of bases.
Next-generation sequencing	DNA sequencing method, whereby millions of DNA sequencing reactions are carried out in parallel. Each sequencing reaction generates a single read.
Read	A single uninterrupted series of nucleotides that represents a portion of the template DNA sequence that is being sequenced.
Split read	One portion of an NGS read maps to the to one location and the other portion of the sequencing read maps to a different location in the genome.
Reference sequence	A sequence file that is used as a reference. Generally speaking, these are nucleotide sequences relating to the genome of the species being sequenced and allows for the detection of variation.
hg38	The assembly of the human genome that was released in December 2013.
minimap2	DNA sequence alignment tool that aligns DNA sequences produced during next-generation sequencing against a reference database.
Sambamba	Bioinformatic toolbox for handling SAM and BAM sequencing files.
Samtools	A suite of bioinformatic programs that allow the reading, writing and editing

	interactions with high-throughput sequencing data.
nanoSV	A bioinformatic program that utilises split read mapping as a basis for structural variant discovery.
IGV viewer	Visualization tool for the interactive exploration of genomic datasets that include next-generation sequence data.
Soft clipping	Allows for the masking of portions of sequencing reads that do not align to the genome from end to end, desirable for structural variant detection.
BLAT	Sequence alignment tool that finds sequence homology between the input sequence and a designated reference sequence database such as the human hg38.
BLAT score	A statistical score of significance applied to an alignment that takes into account the region of the query sequence that matches the reference, size of the query sequence and the percentage of the query that matches the reference sequence.
Dfam database	A database containing the DNA sequence and genomic positions of repetitive elements for five model organisms.
AluSz6	An evolutionary old family member of the Alu SINE retrotransposon elements.

4.2 Results

4.2.1 Detection of chromosome 20 variant cell lines and qPCR mapping of breakpoints

First, to determine the breakpoint sequence of the chromosome 20 tandem duplication, several cell lines were screened using the approaches developed and tested in the previous chapter. A mosaic population of chromosome 20 variant cells was detected in the mShef7 human ES cell line by interphase FISH. This line was subsequently cloned to generate a homogenic chromosome 20 clonal variant line suitable for sequencing.

Based on these preliminary screens the mShef7 line was chosen for long-read next-generation sequencing using the Oxford Nanopore technology. Before sequencing, the approximate position of the distal breakpoint in the mShef7 line was determined using an adapted protocol for the qPCR detection method. Primers were designed for genes that spanned the length of the chromosome 20q11.21 region. As with the conventional qPCR assay the copy number is attained for each of these genomic loci. The approximate breakpoint position was detected when the copy number dropped from 3 to 2 in adjacently located primers. The breakpoint for mShef7 was detected to lie between *TM9SF4* and *ASXL1* (**Figure 4.1**). These data confirm the mShef7 clone chosen for sequencing has a typical tandem duplication on chromosome 20 that commonly occurs after prolonged culture.
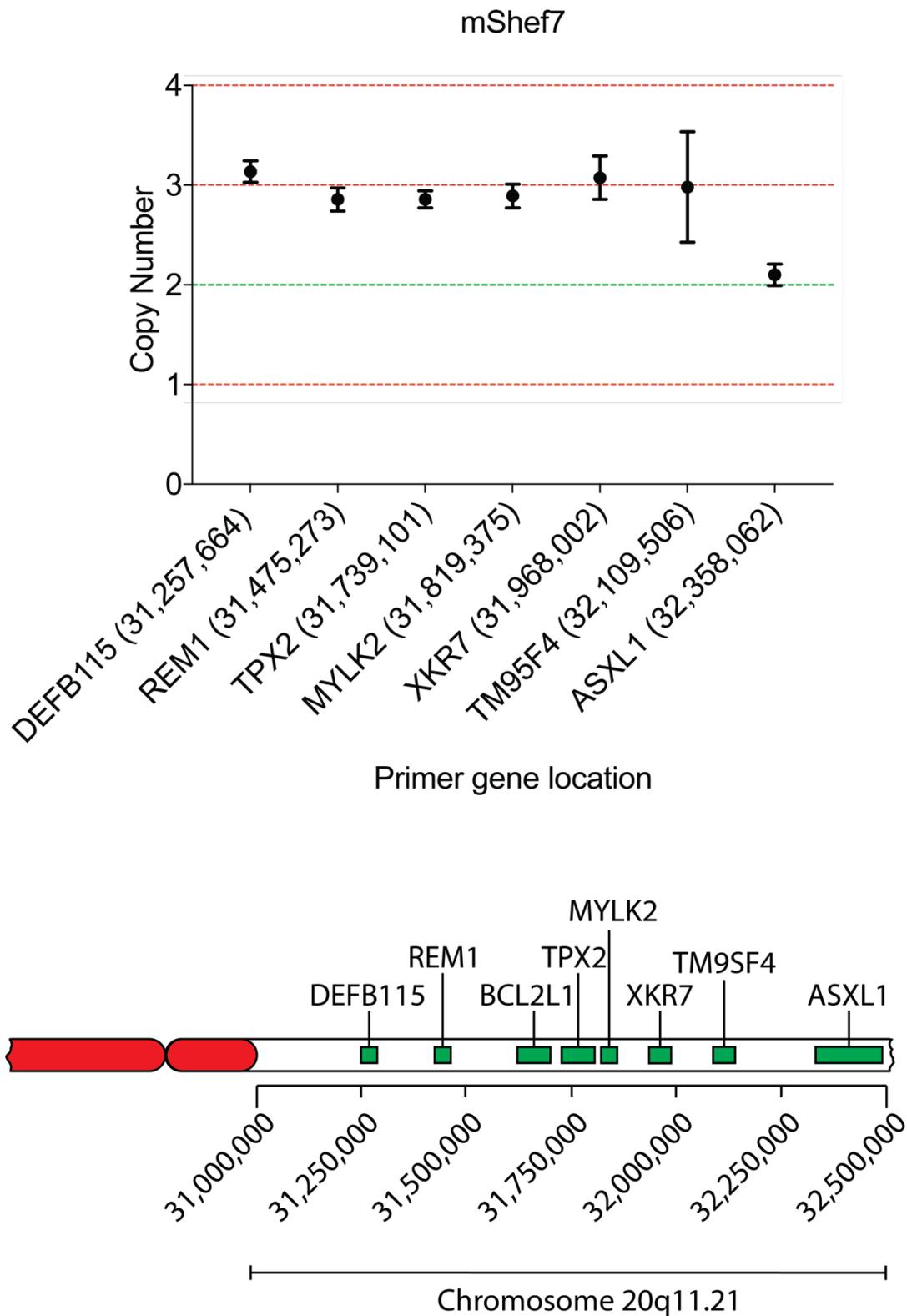
Figure 4.1 | The mShef7 chromosome 20 breakpoint was estimated by qPCR.
qPCR analysis of genes along the q11.21 region, result are the mean of three technical replicates ± s.d. (top). Below, a schematic of the chromosome 20 region detected above.

4.2.2 Protocol development for long read Oxford Nanopore sequencing detection of chromosome 20 breakpoints

Whole-genome long-read next-generation sequencing was performed using the MinION and GridION (Oxford Nanopore Technologies) sequencing systems. To acquire adequate sequencing coverage of chromosome 20, sequencing was performed 10 times on genomic DNA extracted from the clonal mShef7 cell line. Analysis of reads from one representative sequencing run showed the mean read length was 21kb, although some reads were over 100kb in size (**Figure 4.2a,b**). Data exported as FASTQ files were mapped to the chromosome 20 hg38 reference sequence using minimap2 sequence aligner (Li, 2018). File management, sorting and indexing was performed using Sambamba and Samtools (Li et al., 2009b, Tarasov et al., 2015). Finally, structural variants were identified using nanoSV (Cretu Stancu et al., 2017) (**Figure 4.3**).
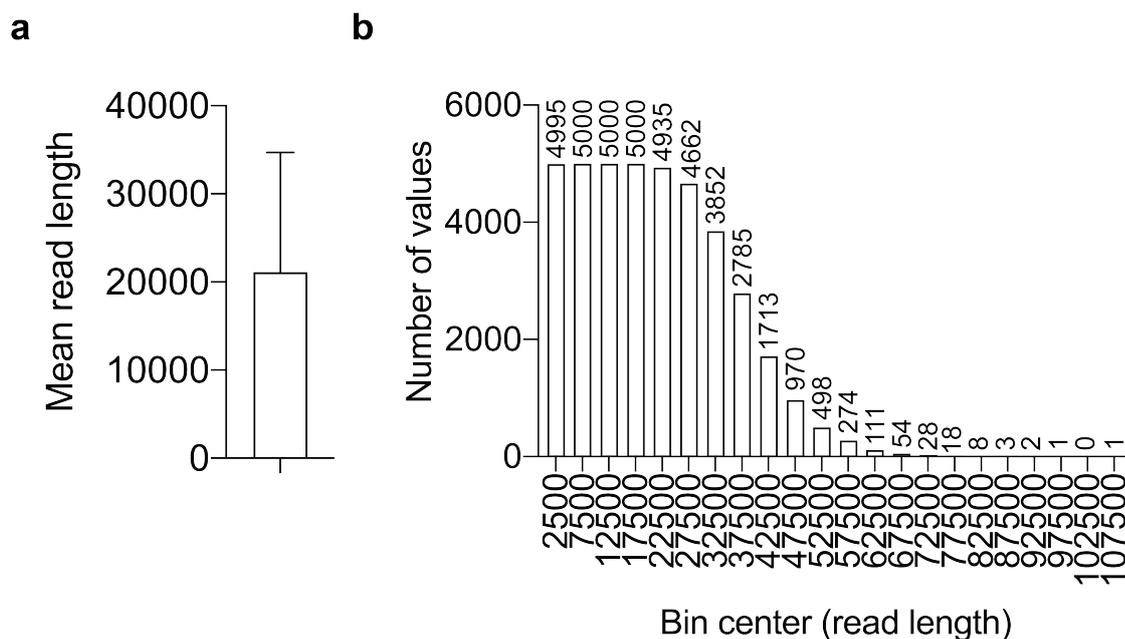
Figure 4.2 | The Oxford Nanopore Sequencing system generated long reads. a, Representative mean read length from a single run on the Oxford Nanopore GridION system. Displayed is the mean read length ± s.d. **b,** Histogram of read lengths generatyed from one sequencing run on the GridION using gDNA extracted from mShef7. Values above each bar correspond to the number of values with that read length bin center.

Long-read next-generation sequencing can facilitate breakpoint mapping in highly repetitive regions. The sequencing of mShef7 provided long reads in excess of conventional sequencing approaches.
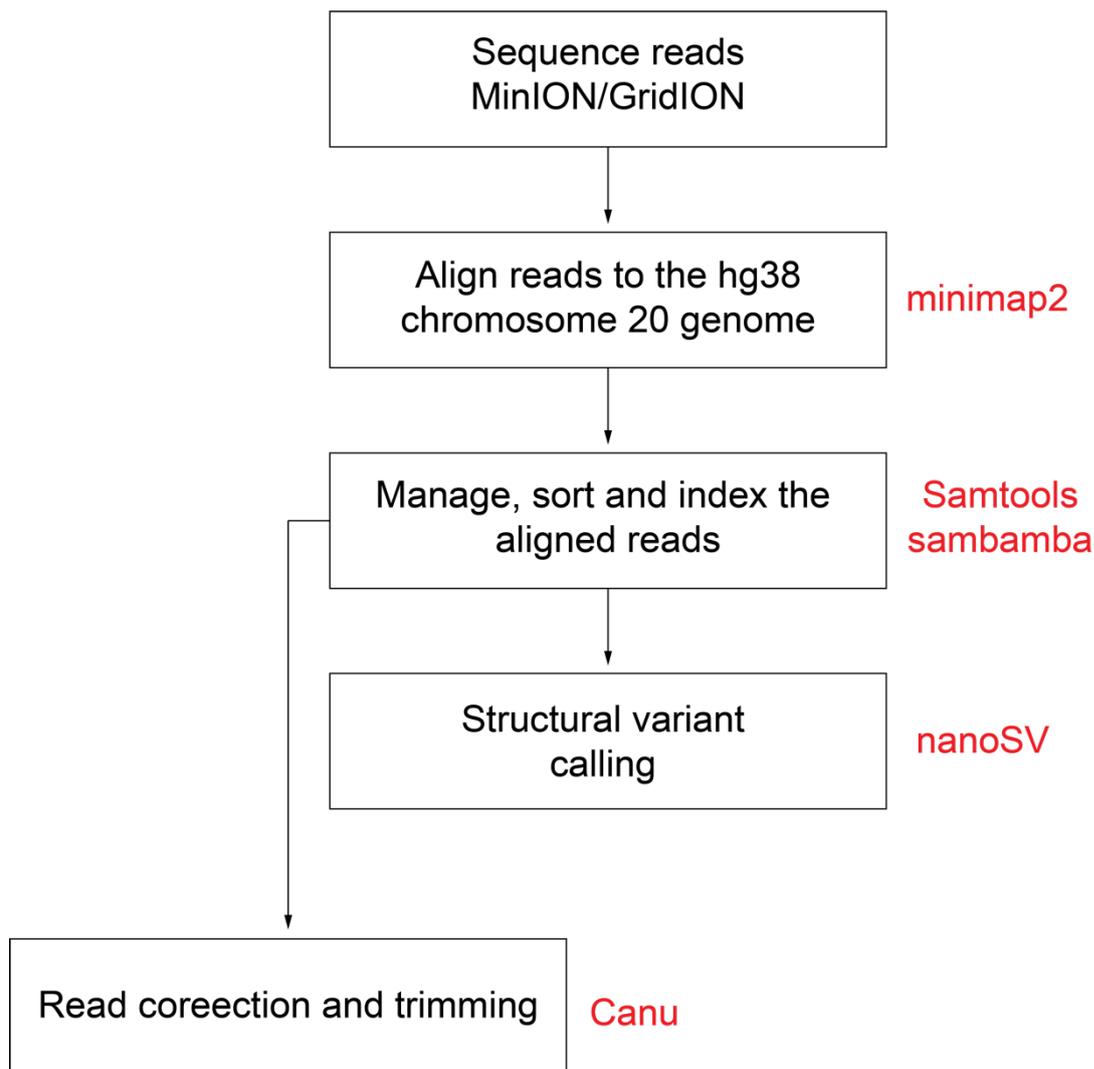
Figure 4.3 | Bioinformatic workflow for the identification of the chromosome 20 structural variant in mShef7. The schematic shows the bioinformatic stages in the analysis of the Oxford Nanopore sequencing data (black text). The bioinformatics tool used in each stage is listed with the step (red text).

4.2.3 Breakpoint detection in mShef7 human ES cell line

The distal breakpoint position was identified by scanning for enriched read depth between the TM9SF4 and ASXL1 genes using IGV viewer (Robinson et al., 2011). A region matching this was identified proximal to the gene KIF3B at locus 32,273,600kb of the hg38 genome (**Figure 4.4a**). It was observed that a number of the reads covering the breakpoint appeared to be largely soft clipped (**Figure 4.4a, inset**). Soft clipping allows the masking of portions of reads that do not map to the genome and are particularly useful in the detection of structural variants (Li et al., 2009b). A BLAT search of these read sequences showed they had been misaligned and matched to sequences elsewhere in the genome. The data in this region was corrected, trimmed and realigned using Canu which cleaned up the alignment significantly (Koren et al., 2017) (**Figure 4.4b**).

Figure 4.4 | Projected breakpoint region viewed in IGV genome viewer. a, IGV screenshot of aligned reads mapping to the chromosome 20q11.21 region between the TM9SF4 and ASXL1 genes. Grey reads (black arrow head) are aligned to the reference genome with the multi-coloured reads showing soft clipping alignment (red arrow head). Red box highlights the drop in sequencing read depth consistent with a copy number variant. Inset, a closer view of the soft clipped read sequences show homology and other mis-aligned reads. **b,** Screenshot of sequence alignment following canu correction and trimming to remove the mis-aligned reads. Note, all soft clipped reads align with the drop in sequencing depth.

a
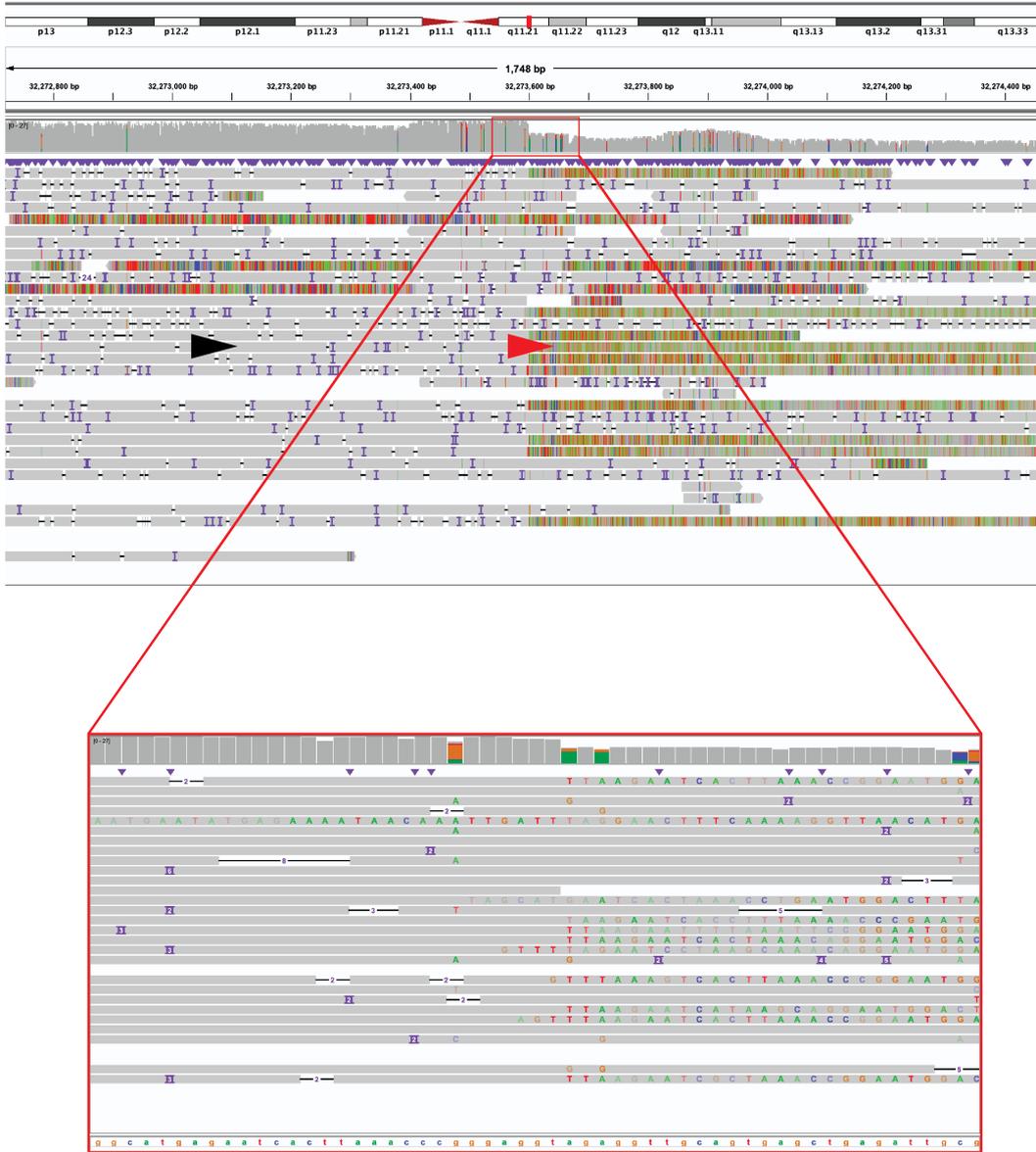
b

After the Canu correction, two types of reads remained. The first mapped to the reference sequence on both sides of the breakpoint and the second mapped proximal to the breakpoint but soft clipped after. The soft clipped portion of this read was assumed to be from the duplicated portion of the centromere. To test this theory, we copied the read sequence and divided it into two sections, the proximal breakpoint sequence and the distal breakpoint sequence. Each section was then queried using BLAT to identify its similarity to the genome. The proximal sequence matched to chromosome 20 with a high BLAT score of similarity (**Figure 4.5a**). The distal sequence matched with undefined sequence or sequence from other chromosomes with a much lower BLAT score of similarity (**Figure 4.5b**).

a

BLAT search result of read sequence proximal of the breakpoint

Human (hg38) BLAT Results

BLAT Search Results

Go back to chr9:111642287-111643900 on the Genome Browser.

Custom track name: Proximal of the breakpoint
Custom track description: Proximal of the breakpoint
Build a custom track with these results

ACTIONS		QUERY	SCORE	START	END	QSIZE	IDENTITY	CHROM	STRAND	START	END	SPAN
browser	details	Proximal	8358	33	10186	10186	95.5%	chr20	+	32263403	32273599	10197
browser	details	Proximal	1315	736	7826	10186	90.9%	chr1	+	246905251	247280488	375238
browser	details	Proximal	373	2915	5766	10186	90.2%	chr1	–	91926524	91960412	33889
browser	details	Proximal	331	5168	5862	10186	86.9%	chr18	+	1263124	1263726	603
browser	details	Proximal	304	5231	6222	10186	86.1%	chr2	+	71508892	71509911	1020
browser	details	Proximal	268	5231	6109	10186	85.1%	chr11	+	126259717	126260412	696
browser	details	Proximal	262	5231	5753	10186	83.5%	chr22	+	18123294	18123746	453
browser	details	Proximal	245	5366	6108	10186	91.0%	chr19	–	8245912	8247010	1099
browser	details	Proximal	238	5253	5786	10186	84.6%	chr10	+	78712946	78713389	444
browser	details	Proximal	226	4248	4948	10186	85.5%	chr14	+	73835118	73835857	740
browser	details	Proximal	221	5597	6283	10186	80.0%	chr19	+	9782932	9783426	495
browser	details	Proximal	221	5610	6361	10186	88.5%	chr14	+	64374755	64375763	1009
browser	details	Proximal	213	5607	6234	10186	86.9%	chr9	–	123955626	123956352	727
browser	details	Proximal	199	6094	6378	10186	93.2%	chr19	–	49726586	49726872	287
browser	details	Proximal	196	5246	5864	10186	84.2%	chr1	–	52544791	52545340	550
browser	details	Proximal	195	4738	5354	10186	84.0%	chr17	–	30756583	30757139	557
browser	details	Proximal	191	5207	5484	10186	90.2%	chr3	–	72880995	72881271	277

b

BLAT search result of read sequence distal of the breakpoint

Human (hg38) BLAT Results

BLAT Search Results

Go back to chr9:111642287-111643900 on the Genome Browser.

Custom track name: Distal of the breakpoint
Custom track description: Distal of the breakpoint
Build a custom track with these results

ACTIONS		QUERY	SCORE	START	END	QSIZE	IDENTITY	CHROM	STRAND	START	END	SPAN
browser	details	Distal	908	17	2832	2837	90.9%	chrUn_GL000216v2	–	10251	80280	70030
browser	details	Distal	794	45	2799	2837	92.2%	chr10	+	41861789	41914914	53126
browser	details	Distal	753	45	2768	2837	92.4%	chr4	–	49094681	49156293	61613
browser	details	Distal	652	46	2768	2837	92.4%	chr17_KI270730v1_random	+	36543	96328	59786
browser	details	Distal	639	51	2768	2837	92.7%	chr2	+	89837243	89840129	2887
browser	details	Distal	518	46	2276	2837	91.8%	chr10	–	42297336	42320109	22774
browser	details	Distal	501	792	2796	2837	91.4%	chrUn_KI270442v1	+	91266	93290	2025
browser	details	Distal	489	17	2796	2837	89.7%	chr10	+	41881757	41910897	29141
browser	details	Distal	476	81	2648	2837	89.2%	chr17_KI270729v1_random	–	3905	6057	2153
browser	details	Distal	475	963	2796	2837	92.3%	chr10	–	18565374	18573028	7655
browser	details	Distal	469	17	1531	2837	91.7%	chr21	–	10662938	10669475	6538
browser	details	Distal	466	59	2563	2837	91.3%	chrY	–	10759856	11008035	248180
browser	details	Distal	463	45	2729	2837	93.0%	chr17	+	21976563	21989791	13229
browser	details	Distal	453	45	1417	2837	92.8%	chrUn_KI270756v1	–	767	68433	67667
browser	details	Distal	431	97	2268	2837	91.0%	chr10	–	38786204	38816981	30778
browser	details	Distal	426	80	1369	2837	93.2%	chr4	+	49633982	49648711	14730
browser	details	Distal	415	17	2768	2837	91.2%	chr4	–	49101768	49155910	54143

Figure 4.5 | BLAT sequence alignment of a representative soft-slipped sequencing read proximal and distal of the breakpoint. a, The sequencing read proximal of the projected breakpoint aligns with high similarity to chromosome 20 region between the TM9SF4 and ASXL1 genes. **b,** The sequencing read sequence does not align with chromosome 20, the highest match is from an unmapped animal genome sequence

Currently, sequence information of centromeric regions for most chromosome has yet to be mapped. However, an unmapped draft sequence has been collated by the T2T consortium and contains the centromeric sequence from all chromosomes (Miga et al., 2019). The distal sequence mapped closely with the draft unmapped sequence (**Figure 4.6**).

These data present a candidate breakpoint region responsible for a tandem duplication in the mShef7 human ES cell line. Repetitive DNA sequence was successfully mapped and aligned to unmapped genomic locations.
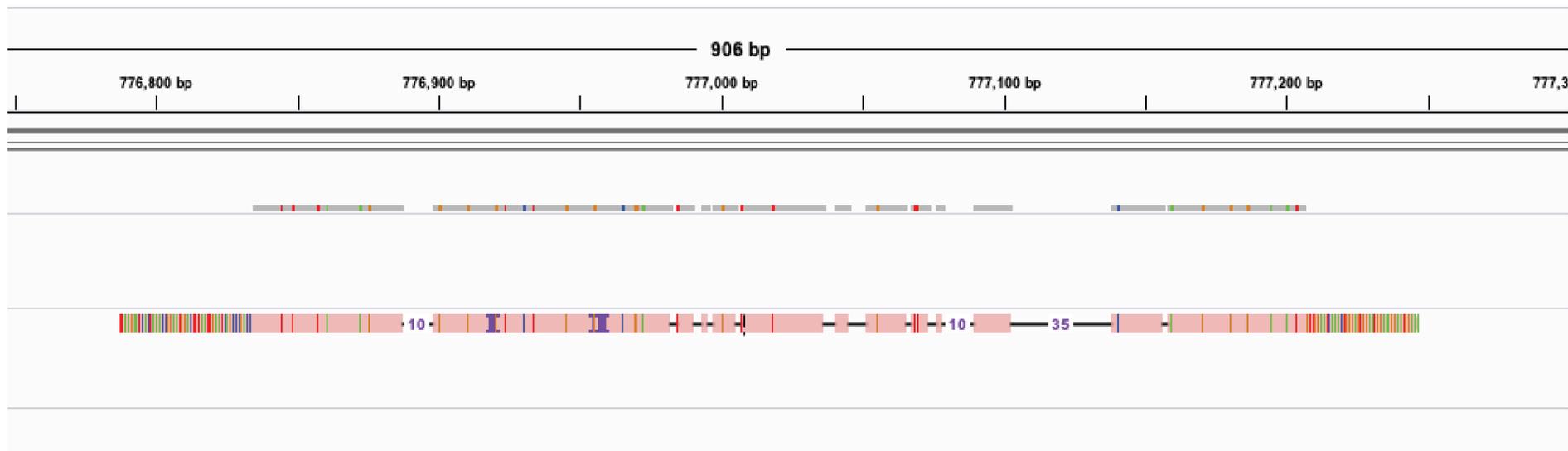
Figure 4.6 | The soft clipped sequence aligns to unmapped draft sequences provided by the T2T consortium. IGV genome viewer showing mapping of a consensus sequence to the draft unmapped T2T consortium sequence that contains centromeric sequencing data.

4.2.4 Breakpoint sequence maps to *Alu* retrotransposon sequence

Transposons are repetitive DNA elements that are known to result in genomic duplication (Ivics and Izsvák, 2010). The *Alu* family, *AluSz6* sequence was mapped to the human genome and presented as an ideogram heat map (**Figure 4.7**). Strikingly, the chromosome 20q11.21 region was enriched for *AluSz6* repetitive elements and on the whole, the q11.21 region has one of the highest abundancies of *AluSz6* sequences in the genome (**Figure 4.7**).

To identify whether the tandem duplication in the mShef7 line was caused by a transposable element, a 1000bp section of the reference sequence, consisting of 500bp sequence either side of the breakpoint was analysed using the Dfam DNA transposable element sequence alignment query tool (Hubley et al., 2015). An E-value is given to each alignment which is a statistical probability that the sequences have aligned by chance. An E-value score of generating a single match of a similar score by chance would be 1, the closer the value is to 0 the greater its significance. The Dfam database aligned and matched the breakpoint sequence to an *AluSz6* SINE repetitive element with a highly significant E-value of 6.1×10^{-95} (**Figure 4.8**).

Alu repetitive transposable elements are responsible for genomic instability. An *Alu* family member, *AluSz6* aligned to the breakpoint sequence mapped in these experiments with high similarity and provides a mechanism for tandem duplication observed on chromosome 20.

Figure 4.7 | Ideogram heat map of *AluSz6* sequences along the human genome. The chromosome number can be seen below each chromosome. The key displays the number of Dfam preferred hits for the *AluSz6* sequences and does not include all *AluSz6* sequences. The red box on chromosome 20 highlights the 20q11.21 region.

Job submitted: 2020-01-09T14:06:47.000Z
Duration: 8 seconds
Parameters: --species "Homo sapiens" --cut_ga

Query

Input ▼

| | None | | Other TE | | LINE | | SINE | | LTR | | DNA Transposon | | Unknown | | Satellite | | RNA |

5 matches

Download

	Sequence	Family	Bit Score	E-value	Model start	Model end	Alignment start	Alignment end	Strand
v	Input	AluSz6	316.6	6.1e-95	1	296	283	583	+

```
              .........*.........*.........*.........*.........*.........*.........*.........*
Model     1   ggccgggcgcgg....tggctcacgcctgtaatcccagcactttgggaggccgaggcggcggatcacttgaggccagga   76
              ggcc+gg+gc+g      tgg tca+gcctg aatcccagcactttgggaggc+gaggc+gg+ggatcacttgagg+cagga
Query   283   GGCCAGGTGCAGtgagTGGGTCATGCCTGCAATCCCAGCACTTTGGGAGGCTGAGGCAGGTGGATCACTTGAGGTCAGGA   362
PP            6888999988751111116899*********************************************************


              .........*.........*.........*.........*.........*.........*.........*.........*
Model    77   gttcgagaccagcctggccaacatggcaaaacccgtctctact.aaaaatacaaaaattagccggcgtggtggcgcgc   155
              ttcgagaccagcctgg caacatggc+aaaccc+g ctctact aaaaatacaaaaat agc+ggc+tggtggc+c++
Query   363   ATTCGAGACCAGCCTGGTCAACATGGCGAAACCCTGCCTCTACTgAAAAATACAAAAATGAGCCAGGCATGGTGGCACAT   442
PP            ******************************************9667789*****************************


              .........*.........*.........*.........*.........*.........*.........*.........*
Model   156   gcctgtaatcccagctactcgggaggctgaggcacgagaatcgcttgaacccgggaggcggaggttgcagtgagccgaga   235
              g ctgtaatcccagctactc ggaggctgaggca+gagaatc ctt aacccgggagg++gaggttgcagtgagc+gaga
Query   443   GTCTGTAATCCCAGCTACTCAGGAGGCTGAGGCATGAGAATCACTTAAACCCGGGAGGTAGAGGTTGCAGTGAGCTGAGA   522
PP            ********************************************************************************


              .........*.........*.........*.........*.........*.........*.....
Model   236   tcgcgccactgcactccagcctgggtgacagagtgagactctgtctcaaaaaaaaaaaaa   295
              t+gcgccac gcactccagcctggg+ acagag+ a a tctgtctcaaaaaaaaaaaaaa
Query   523   TTGCGCCACAGCACTCCAGCCTGGGCAACAGAGCAATATTCTGTCTCAAAAAAAAAAAAA   582
PP            ********************************************99999988776
```

Description

Alu elements derive from a pair of internally-deleted 7SL RNA monomers that are joined by an A-rich linker.

Classification and Taxa ⊘

Classification: Interspersed_Repeat; Transposable_Element; Retrotransposed_Element; LINE-dependent_Retroposon; SINE; 7SL-RNA_Promoter; No-core; L1-dependent ; Alu

Taxa: [...]; Primates ↗

132

Figure 4.8 | Dfam database alignment of transposable elements aligns with the mShef7 breakpoint. Screenshot of breakpoint alignment with the Dfam database. At the top is a schematic showing the different possible alignments with repetitive elements. The grey bar is the input sequence and the coloured bars indicate the different repetitive elements explained in the key. Details of the mapping and the model reference sequence can be seen below, statistics of mapping quality is also displayed.

4.3 Discussion

In this chapter, a next-generation sequencing approach was used to identify the breakpoint region of a chromosome 20 tandem duplication that arose as a result of prolonged *in vitro* culture. Using bioinformatic algorithms, the breakpoint sequence was aligned, with high certainty, to the *AluSz6* retrotransposon family member. The observation that the breakpoint lies within an *Alu* element strongly implies that the tandem duplication is associated with replication and potentially caused as a result of break-induced replication at the site of stalled forks. *Alu* elements are inverted DNA repeats and are symmetrical in their appearance. This symmetry promotes the formation of hairpin secondary DNA structure that are highly susceptible to replication fork collapse (Voineagu et al., 2008). The collapse of replication forks triggers repair by break-induced replication that is known to lead to amplifications when *Alu*-repeats are inserted into yeast (Narayanan et al., 2006).

To map the breakpoint position, a long-read next-generation sequencing approach was chosen. This has distinct advantages over conventional Illumina short-read sequencing when mapping breakpoint positions. With short-reads, it is possible to indirectly observe a structural variant using paired-end reads. The paired-end approach sequences both ends of a DNA fragment and the gap between them is then filled in by consensus from other sequencing reads. If these paired-end reads map in the wrong orientation or the distance between them is not what was expected, then it indicates the detection of a structural variant. However, this can be ambiguous and often becomes inaccurate when aligning sequences of repetitive regions, particularly when the repeat is longer than the read length (Cretu Stancu et al., 2017). Long read sequencing is able to overcome many of these issues as it can often directly sequence through breakpoint regions allowing for their direct observation and are more likely to span the entirety of a repetitive region (English et al., 2015).

The bioinformatics tool, nanoSV was tested on the detection of this breakpoint which relies on split read mapping as a basis for variant discovery. However, nanoSV failed to detect our variant as it was unable to align the section of the split read that mapped to the centromere. This is a technical issue, as there currently isn't a

complete reference sequence for the centromeric region on chromosome 20. The Telomere-to-Telomere consortium is currently generating the first complete assembly of the human genome. So far, they have successfully assembled chromosome X from one telomere to the other (Miga et al., 2019). However, in addition to this, they have released all sequence information that they have yet to map to their respective chromosomes. By generating a reference sequence of the Telomere-2 -Telomere draft sequence, it was possible to align the unmapped soft clipped sequence found at the breakpoint, demonstrating this came from a so far unmapped chromosomal region. Additionally, the sequence contains a highly repetitive (AATGG)n motif that is associated with the centromeres. However, at the time of writing this, it is not possible to confirm that this sequence is from the centromere of chromosome 20.

As the nanoSV tool was not suitable for the detection of our breakpoint, its position was instead inferred manually by combining analysis of the read depth with soft clipped read information. The approximate position of the breakpoint was known from qPCR analysis, which was explored to identify the projected breakpoint. A region of our sequencing was identified between the genes *TM9SF4* and *ASXL1* that had an approximate 0.5X drop in coverage, this was expected as by chance diploid regions should be sequenced half as much as duplicated regions. Further, three types of read were expected to map to the projected breakpoint. Read type 1 (**Figure 4.9a**), had a proportion of the read that mapped to the sequence as well as a portion that did not. These reads were soft clipped over the projected breakpoint (**Figure 4.9b**). Read type 2 (**Figure 4.9a,b**), mapped from end to end of the reference genome. However, it was also predicted that there would be a third read type. Read type 3 (**Figure 4.9a**), would result from sequencing directly through the breakpoint. I did not detect any read type 3 and would suggest that the size of centromere in the duplicated region was large and the reads were too short to sequence through this region.

a

Read type 1

Read type 2

Read type 3

Centromeric breakpoint

MYLK2

REM1

TPX2

TM9SF4

DEFB115

BCL2L1

XKR7

Distal breakpoint

MYLK2

REM1

TPX2

TM9SF4

DEFB115

BCL2L1

XKR7

ASXL1

Centromeric duplicated DNA

Tandem duplication

b

Consensus sequence

MYLK2

TPX2

REM1

DEFB115

BCL2L1

XKR7

TM9SF4

ASXL1

Soft-clipped read

Read type 1
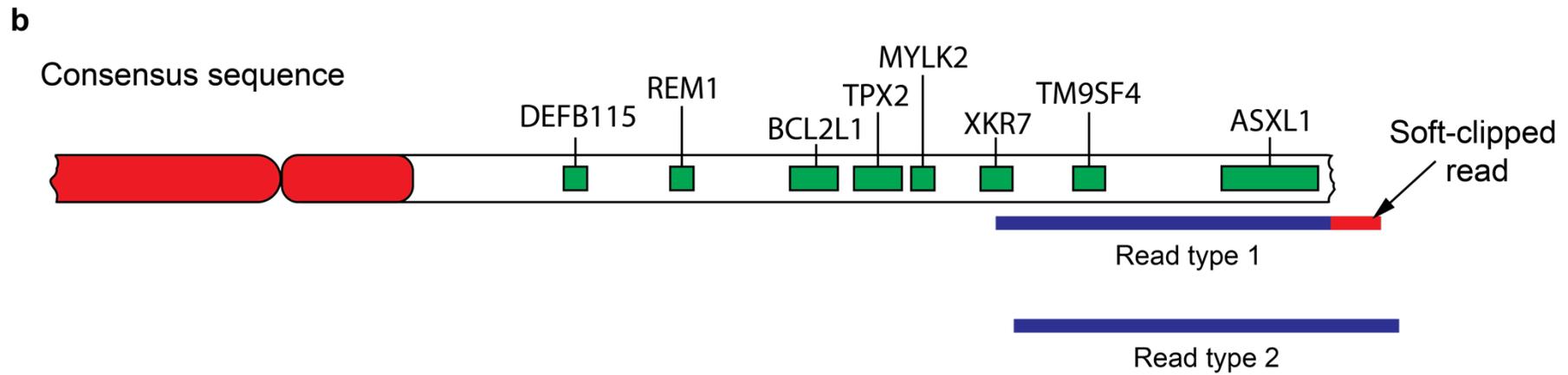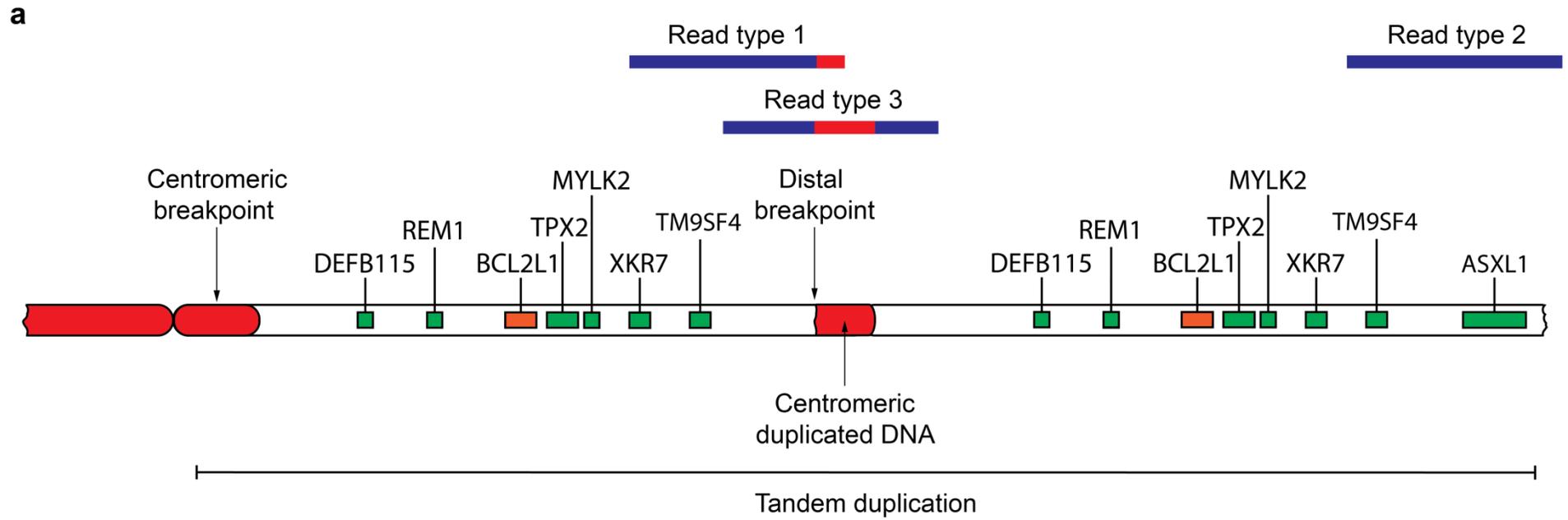
Read type 2

Figure 4.9 | Schematic of the read types spanning the projected breakpoint. a, Schematic of the tandem duplication in the mShef7 chromosome 20 variant cell line. Gene positions (green/orange boxes) are labelled. Schematic of the three expected read types mapping over the tandem duplication can be seen at the top of the schematic. **b,** Schematic of the reference sequence for chromosome 20q11.21, the alignment of read type 1 and 2 is displayed with the expected soft-clipping at the breakpoint region indicated.

Tandem duplications on chromosome 20 are observed with high frequency in human PSC and threaten the safe application of these cells in regenerative medicine. The data presented in this chapter identifies the repetitive *Alu* element as being the substrate in this genomic copy number variant. I propose, that collapse of replication forks within *Alu* sequences triggers repair by microhomology BIR. This HR mechanism requires strand invasion on a different template with the same microhomology. Should the strand invasion slip to a region microhomology elsewhere on the same chromosome it can result in the formation of tandem duplications (**Figure 4.10**). Errors in processing and repair of stalled forks that result from the replication of *Alu* elements are known to lead to genetic instability (Narayanan et al., 2006). With this novel understanding of the mechanisms leading to genetic instability in human PSC, the following chapters will investigate these cells susceptibility to replication stress and explore the potential to improve culture conditions to minimise its detrimental consequences.
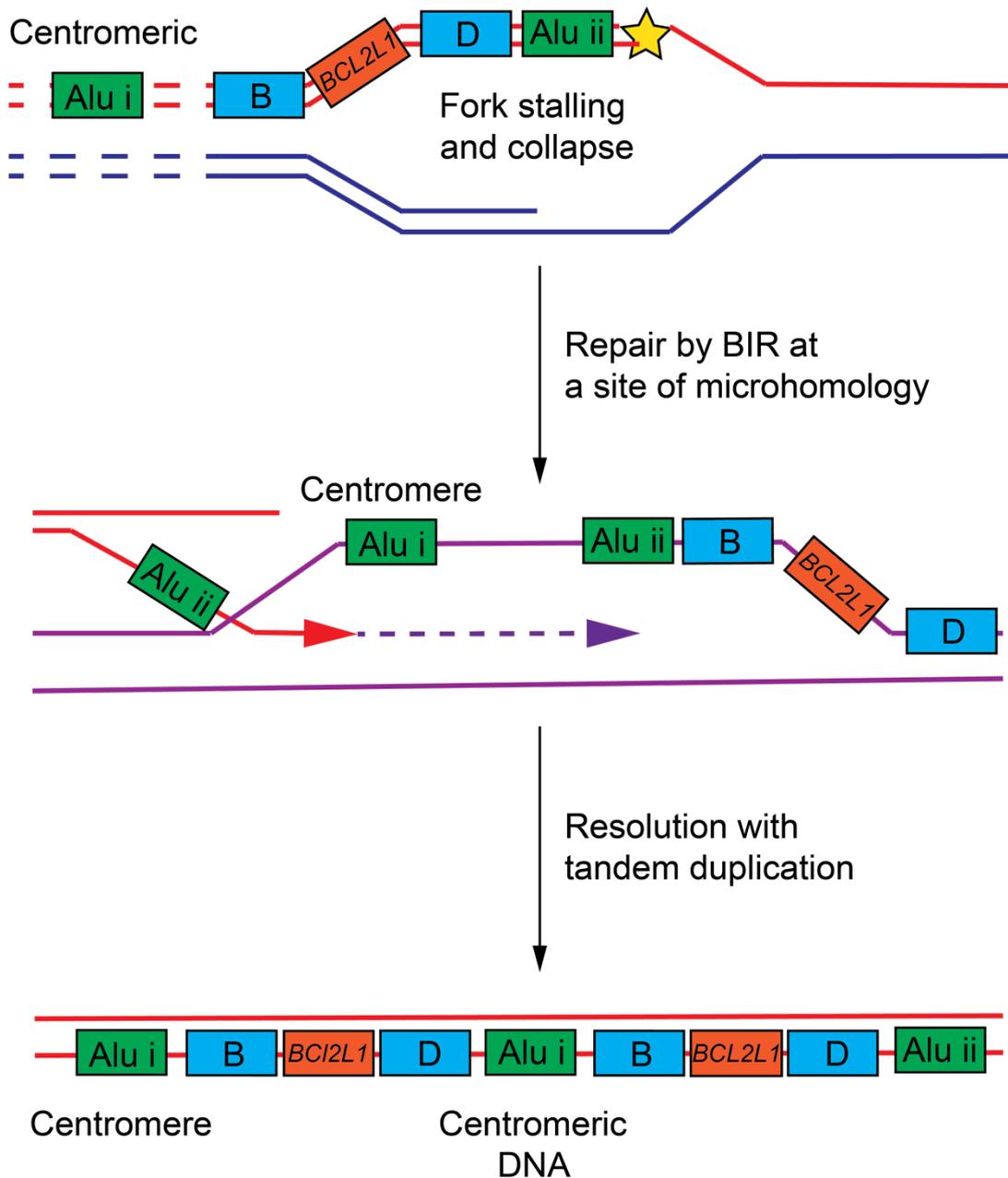
Figure 4.10 | Schematic showing the proposed mechanism of chromosome 20 tandem duplication in human PSC. Fork stalling and collapse in a region of microhomology, such as an *Alu* element occurs distal to the *BCL2L1* driver gene (top). Repair of the stalled fork by MMBIR proceeds by strand invasion at a region of microhomology located in the chromosome (middle). Re-replication results in tandem duplication that is selected for due to the *BCL2L1* driver gene amplification (top).

5 Human PSC are susceptible to replication stress and DNA damage

5.1 Introduction

The cell cycle of human PSC is considerably shorter than most untransformed somatic cell types and has been measured to last between 15 and 18 hours (Becker et al., 2006, Barbaric et al., 2014, Calder et al., 2013). The difference in cell cycle time is primarily due to a truncated G1 phase with S and G2/M phases comparable to somatic cells. The exact mechanism that leads to these cell cycle characteristics has yet to be fully elucidated. However, the rapid progression through the G1 phase is likely due to the relaxation of the retinoblastoma (Rb) restriction checkpoint. The Rb protein is normally found in one of three states, hypophosphorylated, monophosphorylated or hyperphosphorylated. The hypophosphorylated form suppresses the activity of the E2F transcription factor in G0 cells, whereas the monophosphorylated or hyperphosphorylated states release suppression on E2F and allows progression through the G1 phase of the cell cycle. The mono and hyperphosphorylation of Rb is facilitated by cyclin D/CDK4/6 and cyclin E/CDK2 and occurs in early and late G1 phase respectively, initiating the cellular processes required for proliferation to continue (Narasimha et al., 2014). Reports have shown that the CDK4/6 inhibitors, which include p16[INK4a] are silenced in human PSC (Zhang et al., 2009). Consistent with this loss of inhibition, human PSC have elevated activity of cyclin D-CDK4/6 (Becker et al., 2006) which would lead to the mono-phosphorylation of Rb. Further, it was also shown that these cells constitutively express cyclin E (Filipczyk et al., 2007) which would hyper-phosphorylate Rb, and shorten G1 phase. Therefore, it seems likely that the short G1 and rapid proliferation of human PSC are driven by atypical regulation of G1 and G1/S cell cycle cyclins that leads to the relaxation of the Rb restriction point.

Given the clear role of Rb-E2F in cell proliferation, it is not surprising that disruption of this pathway is often targeted by oncogenic mutations in a wide range of cancers (Sherr, 1996). One of the most prevalent mutations to affect this pathway results in the loss of the p16[INK4a] cyclin kinase inhibitor. Functionally, this leads to the activation of Rb through an unchecked high expression of cyclin D/CDK4 activity

(Lukas et al., 1995). In non-mutant cells, p16^{INK4a} normally acts to suppress CDK4 activity to promote cell cycle arrest in G1 phase (Lukas et al., 1995). This deregulation ultimately promotes cellular proliferation by allowing E2F activation of downstream transcriptional targets, including proteins involved in replication, nucleotide biosynthesis, DNA repair, DNA replication origins and cyclin E that further acts to inactivate Rb (Leone et al., 1998, Lundberg and Weinberg, 1998, Harbour et al., 1999). Deregulation of cell proliferation induces replication stress and the formation of under replicated regions and unresolved replication structures that if persist into mitosis can cause structural or numerical chromosomal instabilities (Lukas et al., 2011). Further, replication stress can induce double-strand breaks, which are precursors for genomic instability following erroneous repair (Costantino et al., 2014). These randomly formed mutations are then selected for when they increase tumour growth and enhance the cancers tumorigenicity.

It is striking then that human PSC share cell cycle characteristics similar to those observed in cancer and are susceptible to DNA damage (Simara et al., 2017) that has been associated with DNA replication (Vallabhaneni et al., 2018). Such perturbed replication in human PSC may explain the high frequency of mitotic errors that has been reported elsewhere (Zhang et al., 2019).

The following chapter is a report into an investigation into the dynamics of the cell cycle in human pluripotent stem cells. The expression of key regulators involved in the progression of the cell cycle are defined. Unlike previous studies, replication stress was investigated directly by looking at alterations in replication dynamics between pluripotent and somatic cell lines with single-molecule resolution (Simara et al., 2017, Vallabhaneni et al., 2018). Further, isogenic somatic and pluripotent cells were used to investigate the susceptibility of human PSC to DNA damage as a precursor to genetic instability that results from extended *in vitro* culture.

5.2 Results

5.2.1 Establishing a panel of isogenic pluripotent and differentiated cell lines

To investigate the differences between pluripotent and somatic cell states, irrespective of the cells genetic background, a panel of isogenic pluripotent and somatic cell lines were generated. First, the human iPSC line MIFF1, herein referred to as hiPSC1 was utilised (Desmarais et al., 2016). This human iPSC line was previously reprogrammed in Sheffield from a human foreskin fibroblast line (CRL2429), which was used as one somatic control line and will be referred to as fibroblast for simplicity. In addition to hiPSC1, two further cell lines were used in this study, a second human iPSC line, TC113 and a human ES cell line, mShef11 that will be referred to as hiPSC2 and hESC from here on. Further, all of these lines were differentiated by treating each of them with CHIR99021, an inhibitor of GSK3 beta, for 5 days.

To confirm that the 5-day treatment with CHIR99021 successfully differentiated the pluripotent cells, immunofluorescence staining for NANOG, a transcription factor associated with undifferentiated self-renewal and pluripotency was performed (**Figure 5.1a**). As I intended to investigate the cell cycle, DNA replication and susceptibility to replication stress and DNA damage, it was crucial that all these lines still maintained the ability to proliferate. The ability of the differentiated lines to continue proliferating was assessed by immunofluorescence staining for the Ki67 protein, which is associated with cell proliferation (**Figure 5.1b**), and by labelling the parent fibroblast line with EdU and analysing by flow cytometry (**Figure 5.1c**). Lastly, RT-qPCR analysis was used to confirm that the differentiated cells had acquired the expression of genes associated with differentiated lineages (**Figure 5.2a-c**). In some experiments, another human iPSC cell line, NCRM1 was included in the undifferentiated state and will be referred to as hiPSC3. The cell lines and their isogenic somatic control lines used in this chapter are listed with a description in **Table 5.1**.

a

	DAPI	NANOG	DAPI/NANOG
hiPSC1 (MIFF1)			
Diff			

b

	DAPI	Ki67	DAPI/Ki67
hiPSC1 (MIFF1)			
Diff			

c

EdU +ve
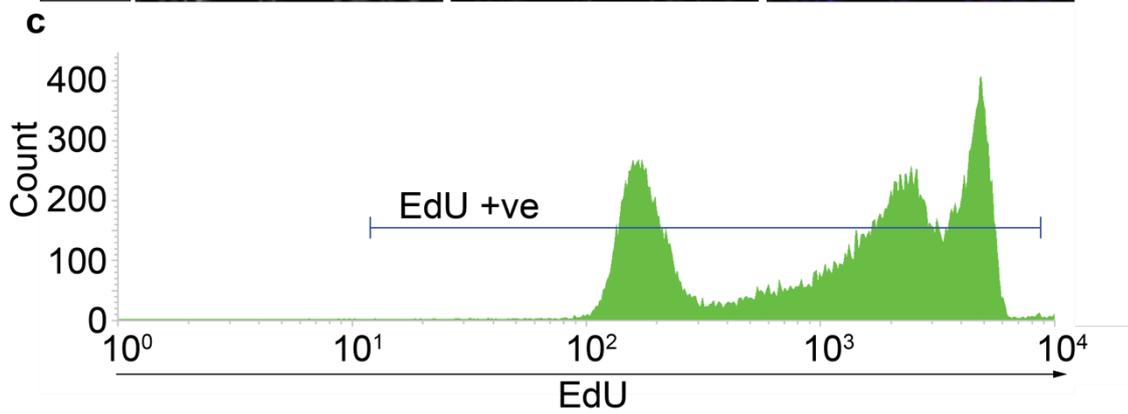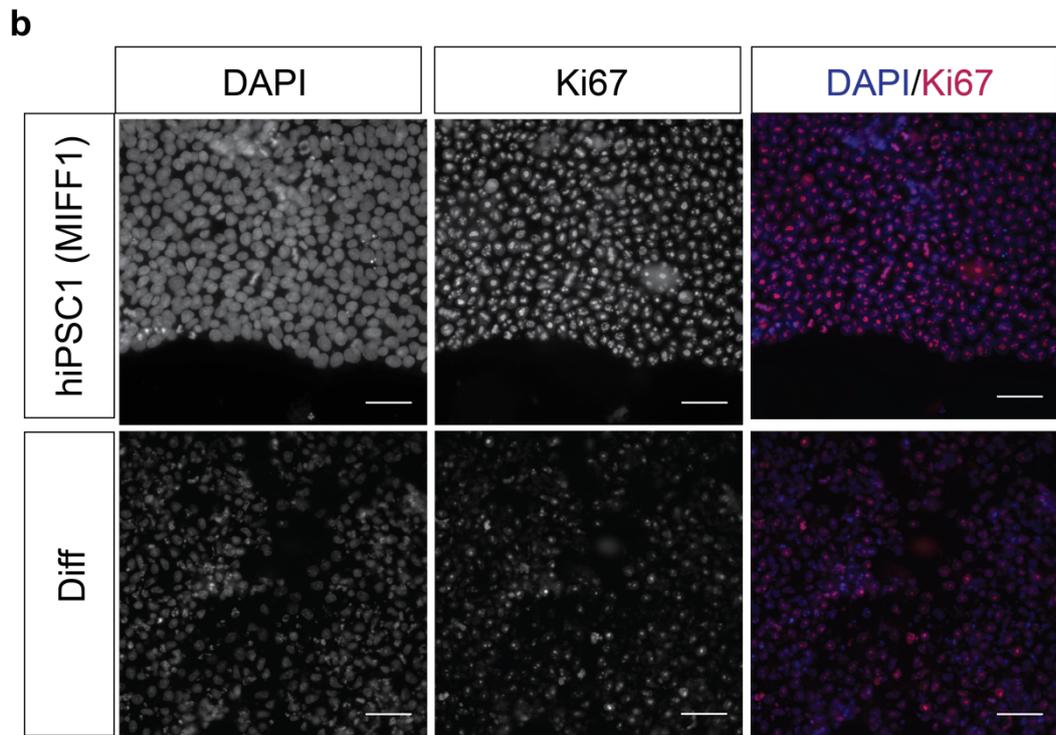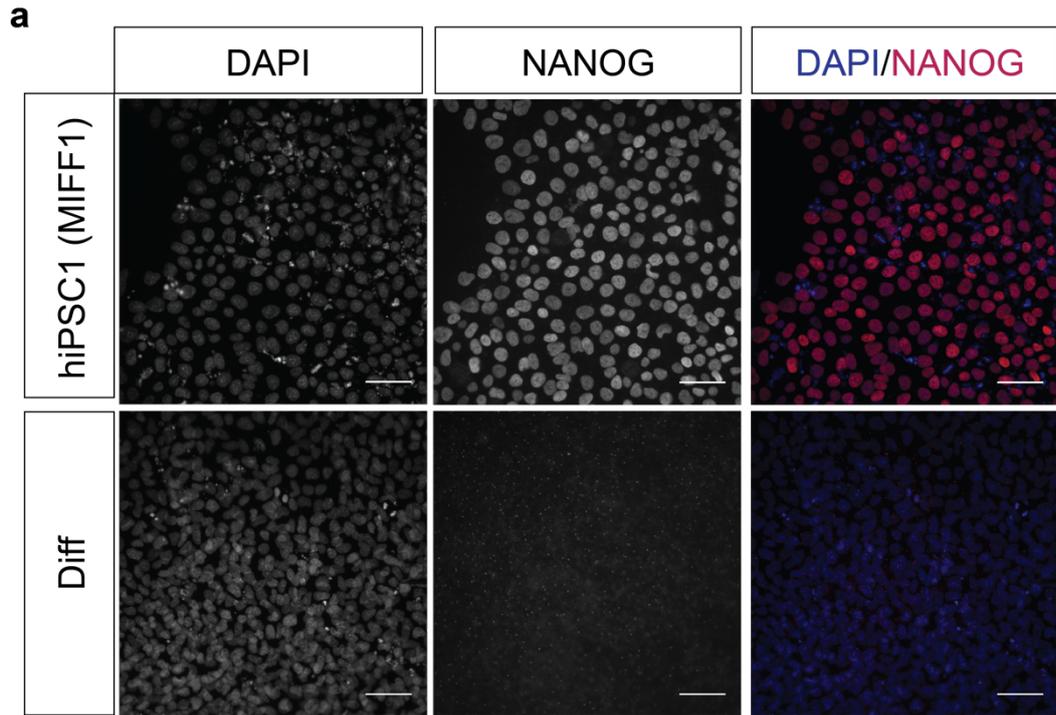
Figure 5.1 | Differentiation of human PSC lines for somatic isogenic controls.
a, Representative images of the hiPSC1 (MIFF1) line in the undifferentiated top and differentiated states (Diff). Images show nuclear DAPI staining, NANOG and the merged image, left to right. **b,** Representative images taken from the hiPSC1 line (MIFF1) and the differentiated derivatives (Diff) obtained from treatment with CHIR99021. Differentiated cell were deemed proliferative based on Ki67 staining (middle). Nuclear DAPI staining and a merged DAPI/Ki67 can be seen, left and right. Scale bars, 50μM. **c,** The parent fibroblast line to MIFF1 was monitored to ensure the line did not senesce by labelling cells with EdU and analysing by flow cytometry. The histogram plot shows a representative plot of EdU positive cells, cells were deemed 99.9% positive for EdU by this analysis.

Table 5.1 Table of pluripotent cell lines and isogenic somatic control lines used over the course of this chapter

Cell line	Description
Fibroblast	Parent human foreskin fibroblast cell line (CRL2429) that was reprogrammed into the MIFF1 human iPSC line.
hiPSC1 (MIFF1)	Daughter human iPSC line derived from fibroblast cell line CRL2429 (Desmarais et al., 2016) (hPSCreg: UOSi001-A). 46,XY.
Diff	hiPSC1 differentiated into somatic control line **(Figure 5.1 and 5.2)**
hiPSC2 (TC113)	Second human iPSC line TC113 (hPSCreg: RUCDRi002-A). 46,XY
Diff	hiPSC2 differentiated cells used as the corresponding somatic cell control.
hESC (mShef11)	Human ES cell line mShef11 (hPSCreg: UOSe015-A). 46,XY.
Diff	Differentiated derivative of the hESC1 cell line.
hiPSC3 (NCRM1)	An additional human iPSC that we did not differentiate (hPSCreg:CRMi003-A).

Figure 5.2 | Gene expression analysis of the panel of pluripotent cells and differentiated derivatives. RT-qPCR gene expression data of hiPSC1 (**a**) hiPSC2 (**b**) and hESC (**c**) compared to their differentiated derivatives. Genes associated with pluripotency, endoderm, ectoderm and mesoderm are displayed (left to right). Data are mean $\pm$s.d., two-tailed t-test, *P<0.05, **P<0.01, ***P<0.001, ****P<0.0001 (n = 3 experiments).
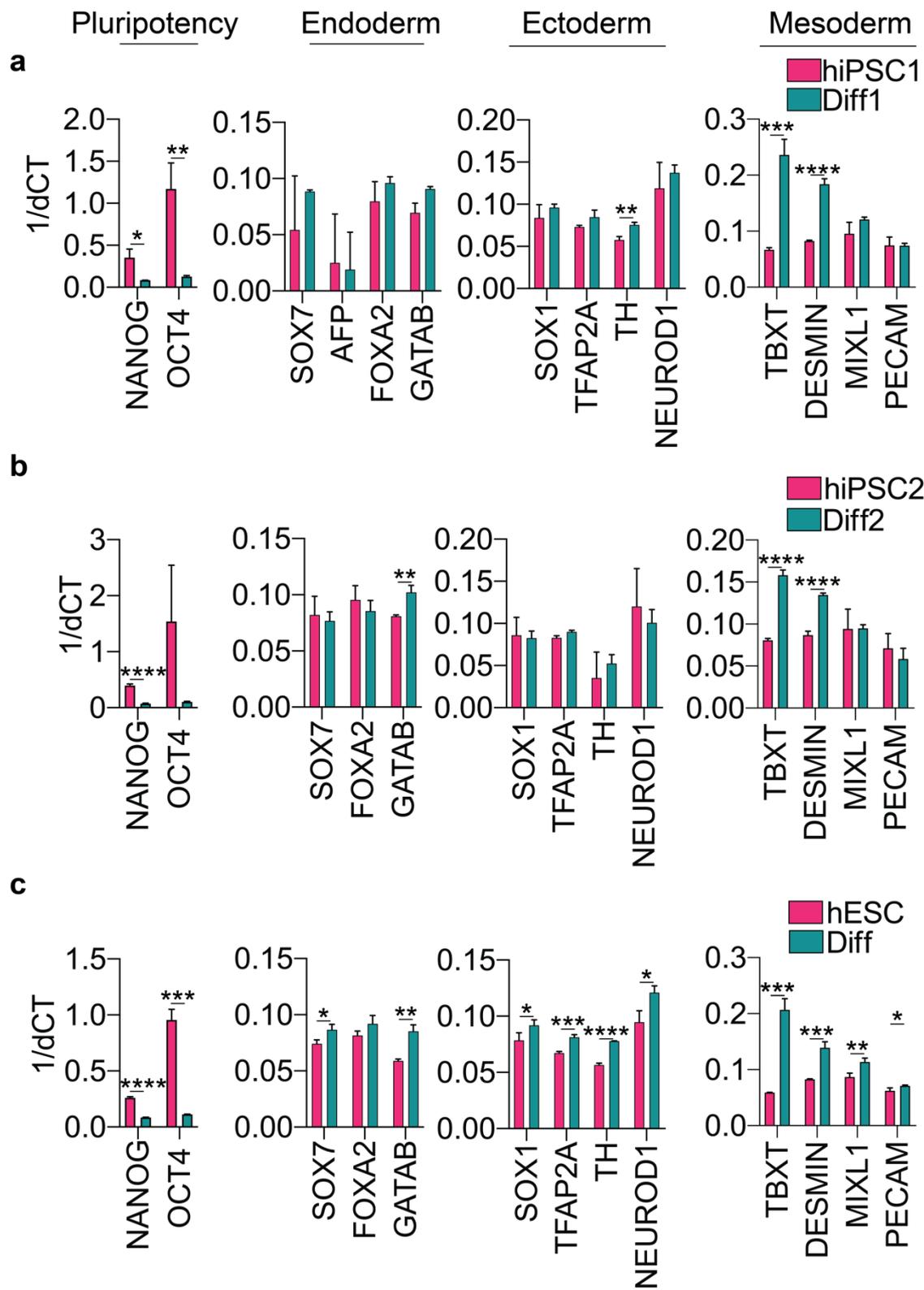
5.2.2 Comparative analysis of cell cycle time between pluripotent and somatic cell states

A common feature of cancers is genetic instability that arises from a dysregulation of the cell cycle. To investigate the potential of atypical cell cycle characteristic in human PSC contributing to mutation, I began by defining the cell cycle dynamics of our pluripotent cells in relation to the isogenic somatic control lines.

To accurately define the proportions of cells in each stage of the cell cycle, the fibroblast, hiPSC1 and hESC lines were grown to 50% confluency before pulse labelling with EdU and stained with DAPI. EdU can be added to cell culture media and will incorporate into newly synthesised DNA and therefore enable detection of those cells undergoing DNA synthesis in the S phase of the cell cycle. DAPI intercalates proportionally with the amount of DNA present in the nucleus allowing for separation of cells based on their DNA content. When analysed by flow cytometry, cells that only stained for DAPI made two distinct populations of 2N and 4N cells that represent the G1 and G2/M populations. A further population of co-positive EdU and DAPI cells marked those cells in the S phase of the cell cycle. When the proportions of fibroblast cells in each phase of the cell cycle was compared to its daughter hiPSC1 line, it was observed that there was a greater proportion of cells in G1 phase, 55% to 14% respectively (**Figure 5.3a,b**). This difference could be explained by a much larger proportion of cells in S phase in the pluripotent state, 68% compared to 21% in the fibroblast line (**Figure 5.3a,b**). Compared to G1 and S phase, the proportion of cells in G2/M phase was comparable between the cell states, 19% in the fibroblast line to 15% in the daughter hiPSC1 line (**Figure 5.3a,b**). To confirm this cell cycle distribution was not unique to human iPSC lines or specific to this cell line, the experiment was repeated using hESC. The hESC line also had a smaller proportion of its cells in G1 phase and a much larger S phase population when compared with the fibroblast line (**Figure 5.3c**). These results indicate that the cell cycle dynamics of pluripotent cells differ substantially to somatic cell lines.

a

S phase

10^4
10^3
10^2
10^1
10^0

G1 phase
G2&M phase

0   40K

Fibroblast
G1 = 55.4%
S = 21.3%
G2/M = 18.5%

b

S phase

10^4
10^3
10^2
10^1
10^0

G1 phase
G2&M phase

0   40K

hiPSC1
G1 = 14.6%
S = 67.6%
G2/M = 15.2%

c

S Phase

10^4
10^3
10^2
10^1
10^0

G1 phase
G2&M phase

0   40K

hESC
G1 = 27.5%
S = 59.0%
G2/M = 11.8%

EdU
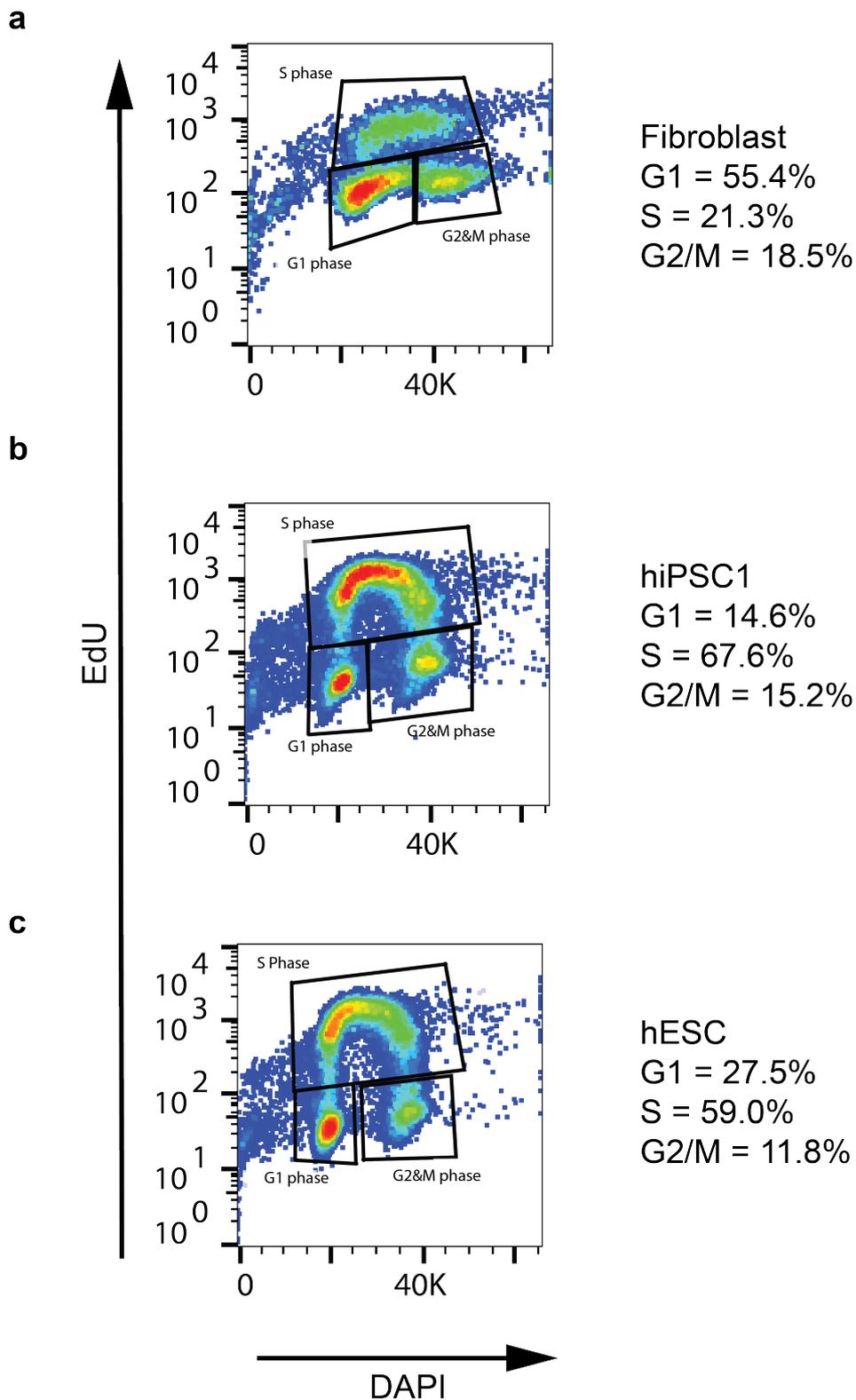
DAPI

Figure 5.3 | Pluripotent and somatic cells differ in their cell cycle distributions.
a-c, Flow cytometry density dot plots of dual labelled EdU and DAPI asynchronous cell populations (**a**) parent fibroblast (**b**) the daughter hiPSC1 and (**c**) an independent hESC line. Gates for G1, S and G2/M phase are displayed. Proportion of cells in each cell cycle phase is shown (right).

To measure the length of the cell cycle stages, EdU pulse chase analysis was performed, whereby cells plated in multi-well plates were pulse labelled with EdU for 45 minutes. One well was harvested every hour for approximately 30 hours and analysed by flow cytometry detecting both the EdU and DAPI stained cells. Representative plots from the pulse chase analysis using the hiPSC1 line can be seen in **Figure 5.4a**. Using the flow cytometry data, analyses of relative movement and the ratio of cells in the mid-S phase window were carried out as previously described (Begg et al., 1985). The S phase cells were labelled with EdU-488 and were identified by their green fluorescence. At the 0 hour time point the EdU labelled cells were evenly distributed between the unlabelled G1 and G2/M populations. During the subsequent time points the EdU labelled cells moved right in line with the G2/M population and later began to align with the G1 population (seen at 4Hr) before returning back into S phase (between 8Hr and 20Hr) (**Figure 5.4a**). A similar pattern, albeit slower, was observed in the fibroblast line.

The length of S phase was determined by measuring the movement of the EdU labelled cells relative to the G1 and G2/M populations using the following equation:

$$RM = \frac{\text{Mean DAPI fluorescence EdU cells - Mean DAPI fluorescence G1 Cells}}{\text{Mean DAPI fluorescence G2 Cells - Mean DAPI fluorescence G1 Cells}}$$

Briefly, at the 0 hour time point the DAPI fluorescence of the EdU labelled cells will be approximately half way between the G1 and G2/M cells and as such the RM = 0.5. Over the subsequent time points the EdU labelled cells will move over to the right hand side of the plot and will begin to equalise the DAPI fluorescence of the G2/M population. An RM of 1 is reached when all the EdU labelled cells have left S phase and entered G2, in hiPSC1 this was 6 hours (**Figure 5.4b**). The length of a cells G2/M phase will affect the shape of the relative movement curve, the degree to which it curves can be used to estimate the length of G2/M phase. The point at which the RM curve intersects with the line of best fit is proportional to the length of G2/M phase, in hiPSC1 this was 6.5 hours (**Figure 5.4b**). The total cell cycle time was deduced using the ratio of mid-S window calculation. Two narrow overlapping gates were drawn in the mid-S phase. The first gate spanned the EdU positive population only and the second larger gate spans both the EdU labelled and

unlabelled cells (Example gates can be seen in, **Figure 5.4a**). At each time point, the ratio of cells in each of these gates was determined and plotted graphically (**Figure 5.4c**). At the 0 hour time point all cells will be in both overlapping gates and the ratio will be 1. As time goes on the mid-S ratio decreased as the EdU labelled cells leave S phase and move through G2/M and G1 phase. Gradually, the ratio will increase again as the EdU labelled cells re-enter S phase. The time point where the ratio peaks is equal to the total cell cycle time, which was 18 hours in hiPSC1 cell line (**Figure 5.4c**). Finally, the length of G1 phase was deduced by subtracting the S phase time and the G2/M phase time from the total cell cycle time which equalled 5.5 hours. The same analysis was performed on the somatic parent fibroblast line to hiPSC1, where total cell cycle time was found to be considerably longer, lasting 23 hours. Interestingly, the length of S phase and G2/M phase was comparable, although G1 phase was 11 hours, double the length of the hiPSC1 line (**Figure 5.4d**). In hESC, a similar result was found with its cell cycle lasting 18 hours of which 4.5 hours was spent in G1 phase, 7 hours in S phase and 6.5 hours in G2/M phase (**Figure 5.4d**).

a

0Hr — Mid-S window 3.74, Total Cells 3.78
4Hr — Mid-S window 3.26, Total Cells 7.10
8Hr — Mid-S window 2.26, Total Cells 6.11
12Hr — Mid-S window 1.59, Total Cells 8.21
16Hr — Mid-S window 4.17, Total Cells 6.58
20Hr — Mid-S window 5.38, Total Cells 5.50
24Hr — Mid-S window 1.25, Total Cells 3.55
28Hr — Mid-S window 3.74, Total Cells 4.33
32Hr — Mid-S window 3.21, Total Cells 4.08

EdU (y-axis), DAPI (x-axis)

b Relative Movement vs Time after Labeling (Hours)

c Ratio Mid-S Window vs Time (Hours), $T_c = 18HR$

d Time (Hours) for G1, S, G2/M — Fibro, hiPSC1, hESC

Figure 5.4 | G1 phase is truncated in human PSC lines. a, Representative flow cytometry density plots from pulse chase analysis of hiPSC1 line. hiPSC1 cells were pulse labelled with EdU (y axis) for 45 minutes and harvested every hour and dual stained with DAPI (x axis). **b,** Relative movement of the EdU labelled S phase cells relative to the G1 and G2/M phase cells. G2/M length determined from the intersection with line of best fit and S phase length is equal to a relative movement of 1. **c,** Ratio of mid-S phase window of hiPSC1 cells over time. Total cell cycle time was determined from the elapsed time between peak ratios of mid S phase cells. **d,** Summary bar chart of cell cycle times as determined by pulse chase analysis. The daughter hiPSC1 (pink) line has a truncated G1 phase compared to the parent fibroblast (green, black stripe). Comparable cell cycle phase times were also found in the hESC line (pink, black stripe).

In addition to pulse chase analysis the cell cycle time of hiPSC1 and the parent fibroblast cell lines was measured using time-lapse microscopy. Single cells were plated at clonal density and images were acquired of the cells every 10 minutes using the Nikon Biostation CT incubated time-lapse microscope. The cell cycle time was determined by measuring the time elapsed between the first and second cellular divisions (**Figure 5.5a**). The fibroblast line had a cell cycle time of 23 hours which matched exactly the result attained by pulse chase analysis (**Figure 5.5b**). However, when the cell cycle time was measured in hiPSC1 cell line using time-lapse microscopy, the time elapsed was 20 hours, some 2 hours longer than the measurement made by pulse chase analysis (**Figure 5.5b**). Additionally, the hESC cell cycle time was also lengthened by 2 hours when measured by time-lapse microscopy compared to the pulse chase analysis (**Figure 5.5b**). Using time-lapse microscopy, the cell cycle time of an additional pluripotent line, hiPSC3 was measured and again the mean cell cycle time was 20 hours. Overall, these results indicate that cells in a pluripotent state have a shortened cell cycle time that results from a significantly shortened G1 phase.
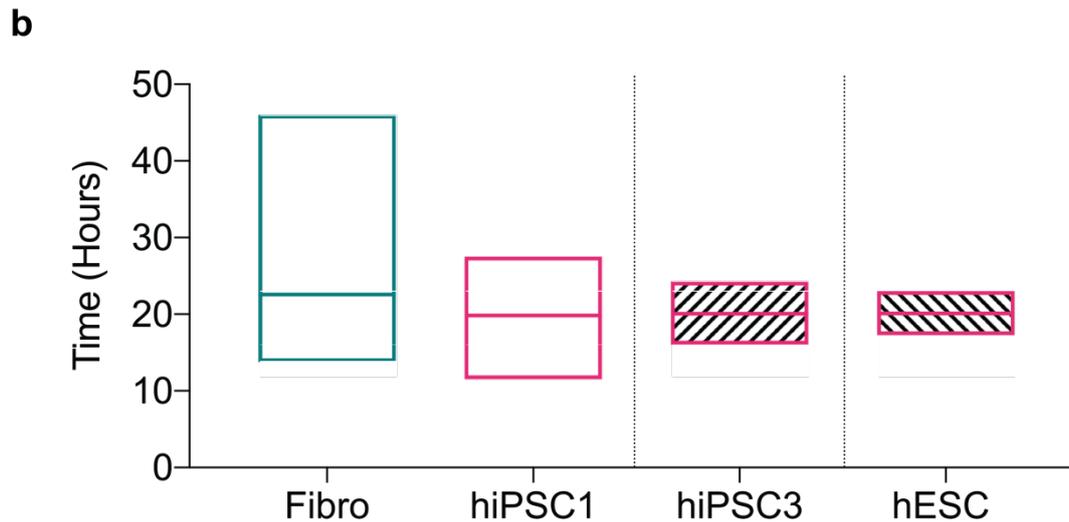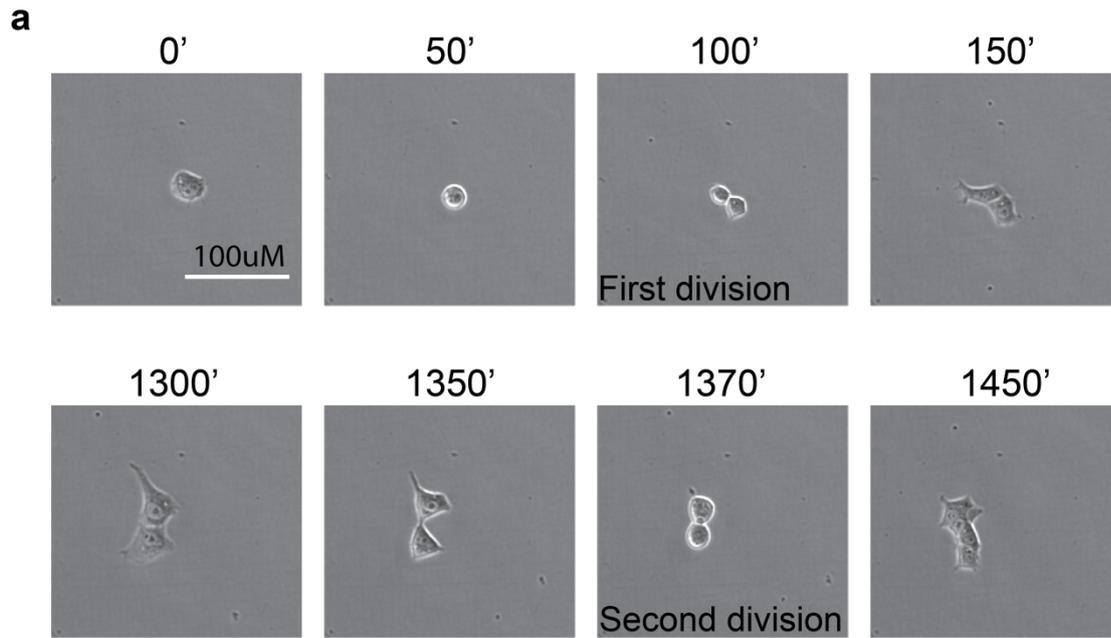
Figure 5.5 | Pluripotent cells proliferate faster than somatic cells. a, Representative images from time-lapse microscopy of hiPSC1 cells. Time elapsed in minutes is shown at the top of each image. The time point of the first and second division is labelled on the respective images. Sale bar, 100μM. **b,** Summary data of cell cycle time in the parent fibroblast (n=44), daughter hiPSC1 (n=76), and two independent pluripotent lines, hiPSC3 (n=31) and hESC (n=31). Box plots display line at mean and the minimum and maximum values.

5.2.3 Expression of components of the cell cycle control system in the pluripotent state

To better understand what drives the cell cycle dynamics in the pluripotent state, the hiPSC1 line was compared to its parent fibroblast line to identify atypical expression of cell cycle components. Using RT-qPCR analysis the major cyclins and CDK partners involved specifically in G1 and S phase progression were screened (**Figure 5.6**). The gene expression of *CDK4* and *CDK6* was found to be equivalent between hiPSC1 and the parent fibroblast line (**Figure 5.6**). However, cyclins, the regulatory proteins that phosphorylate and activate these CDK partners were atypically expressed in hIPSC1 compared to the parent fibroblast. What was particularly noteworthy was the relative expression of *cyclin D2*, *E1* and *E2* (*CCND2*, *CCNE1* and *CCNE2*) which were all highly expressed in the pluripotent state (**Figure 5.6**). The expression of *cyclin C* (*CCNC*) was also raised in hiPSC1 compared to its parent fibroblast but interestingly, *cyclin D1* and *D3* expression was found to be relatively decreased (**Figure 5.6**).
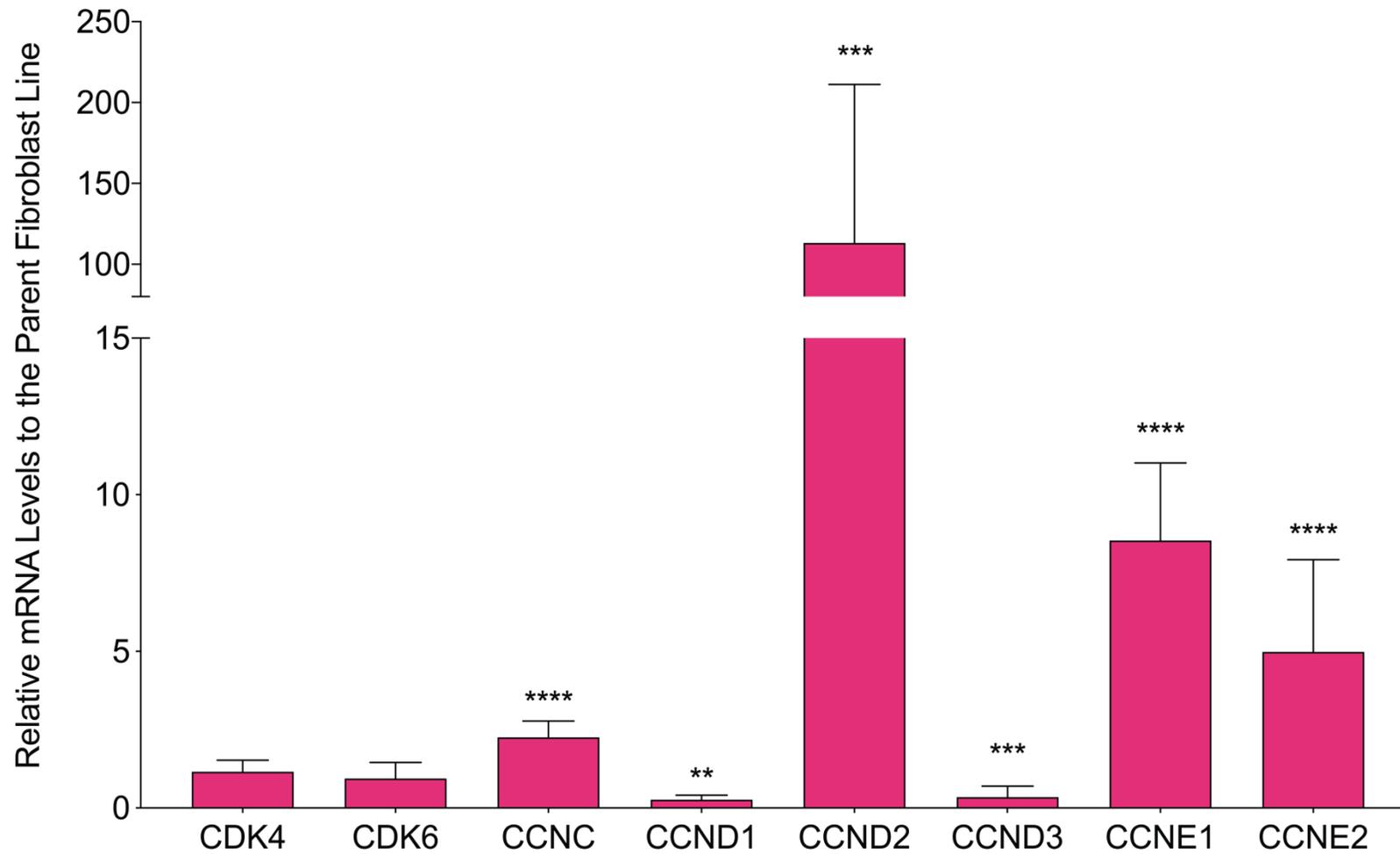
Figure 5.6 | Pluripotent cells have atypical G1 and S cell cycle component expression. RT-qPCR gene expression analysis for genes involved in G1 cell cycle progression. Data is relative gene expression of hiPSC1 line compared to the parent fibroblast line. Data are mean ±s.d., two-tailed *t*-test, *P<0.05, **P<0.01, ***P<0.001, ****P<0.0001 (*n* = 3 experiments).

Following this initial screen, analysis was focused on *cyclin D2*, *E1* and *E2*, which were identified as the most atypically expressed cyclins. Again, gene expression analysis by RT-qPCR showed *cyclin D2*, *E1* and *E2* to be all significantly raised in hiPSC1 compared to the parent fibroblast line (**Figure 5.7a**). These findings were confirmed in two further pluripotent lines, hiPSC2 and hESC and overall the gene expression of these cell cycle components was similarly expressed across all lines tested (**Figure 5.7a**). To ensure this high gene expression was being translated to protein, western blot analysis was performed for the *cyclin D2*, *E1* and *E2* protein in hiPSC1 and the parent fibroblast line. The protein expression of cyclin D2, E1 and E2 was also substantially elevated in hiPSC1 relative to its parent somatic fibroblast (**Figure 5.7b**). Consistent with the reduction in the length of G1 phase, cyclin D2, E1 and E2, which are known to phosphorylate the Rb-E2F restriction checkpoint and allow the rapid progression through G1 in cancer cells were also highly expressed in the pluripotent state.
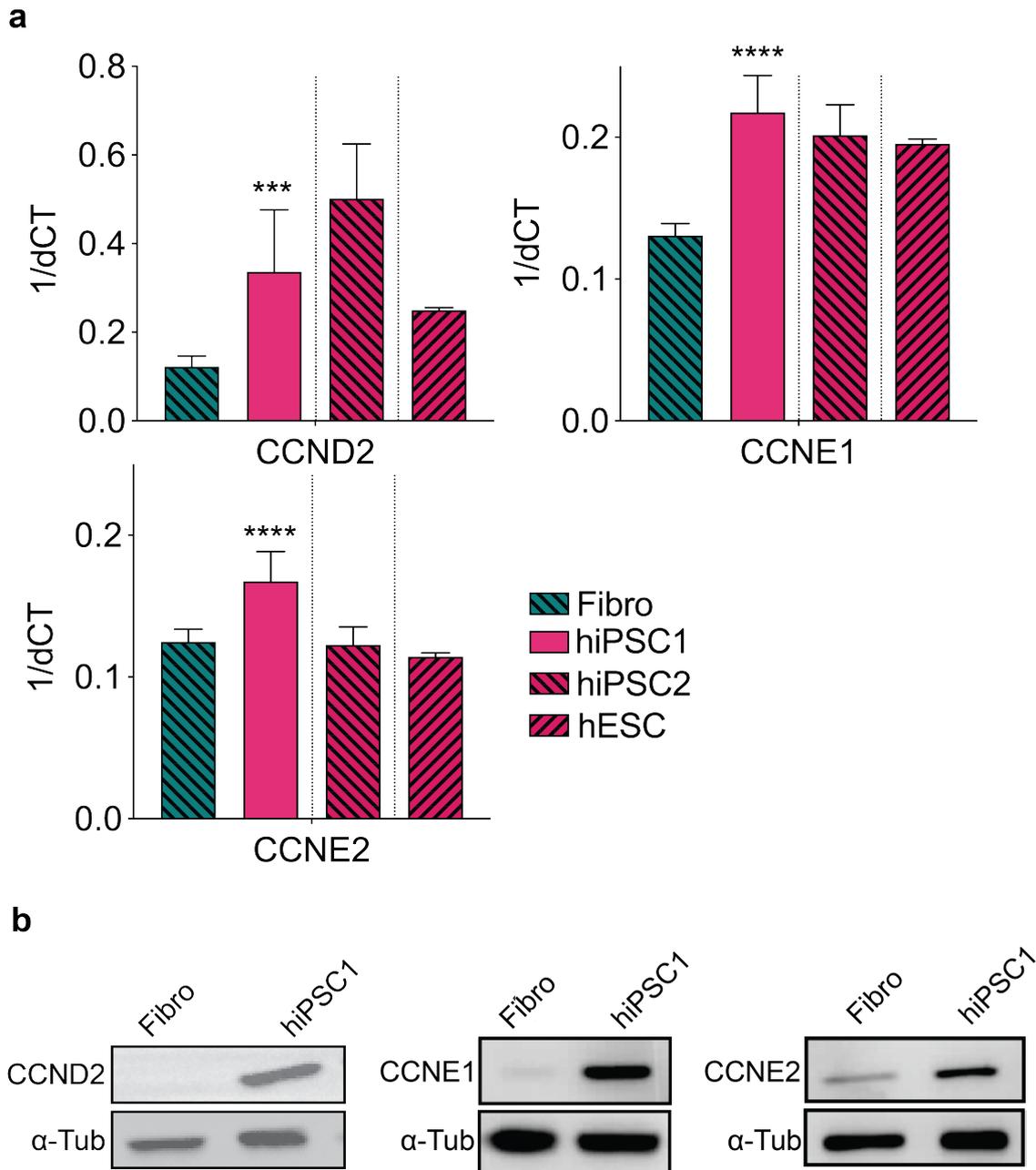
Figure 5.7 | Cyclin D2 and cyclin E are highly expressed in human PSC. a, RT-qPCR gene expression of *cyclin D2* (*CCND2*), *cyclin E1* (*CCNE1*) and *cyclin E2* (*CCNE2*) for the parent fibroblast line and the daughter hiPSC1. Additional pluripotent lines hiPSC2 and hESC are also displayed. Data are mean ±s.d., two-tailed *t*-test, *P<0.05, **P<0.01, ***P<0.001, ****P<0.0001 (*n* = 3 experiments). **b,** Western blot analysis showing relative protein expression of CCND2 (left), CCNE1 (middle) and CCNE2 (right) in the parent fibroblast and daughter hiPSC1 line. α-Tubulin loading control.

5.2.4 Cells in the pluripotent state display perturbed replication dynamics

Deregulation of the cell cycle can lead to replication stress during S phase and can be diagnosed in cultures of cells through the analysis of cells replication dynamics. The fibre assay allows newly synthesised DNA to be visualised by pulse labelling cells successively with the thymidine analogues chlorodeoxyuridine (CldU) and iododeoxyuridine (IdU) for 20 minutes each (**Figure 5.8**). The DNA fibres of pluripotent and somatic cells were analysed in a number of different ways. First, the total length of DNA fibres were measured, with the assumption that longer fibres were generated from unhindered replication forks (**Figure 5.8**). The average DNA replication fork rate for the CldU and IdU pulses was estimated using the previously defined conversion where $1\mu M$ corresponds to 2.59kb (**Figure 5.8**) (Jackson and Pombo, 1998). Finally, the abundance of replication origins was assessed by measuring the distance between replication origins. The distance between replication origins was assessed only on DNA fibres where it was possible to detect two forks on a single DNA fibre that were both moving bidirectionally, a representative image of such a DNA fibre can be seen in **Figure 5.8**.
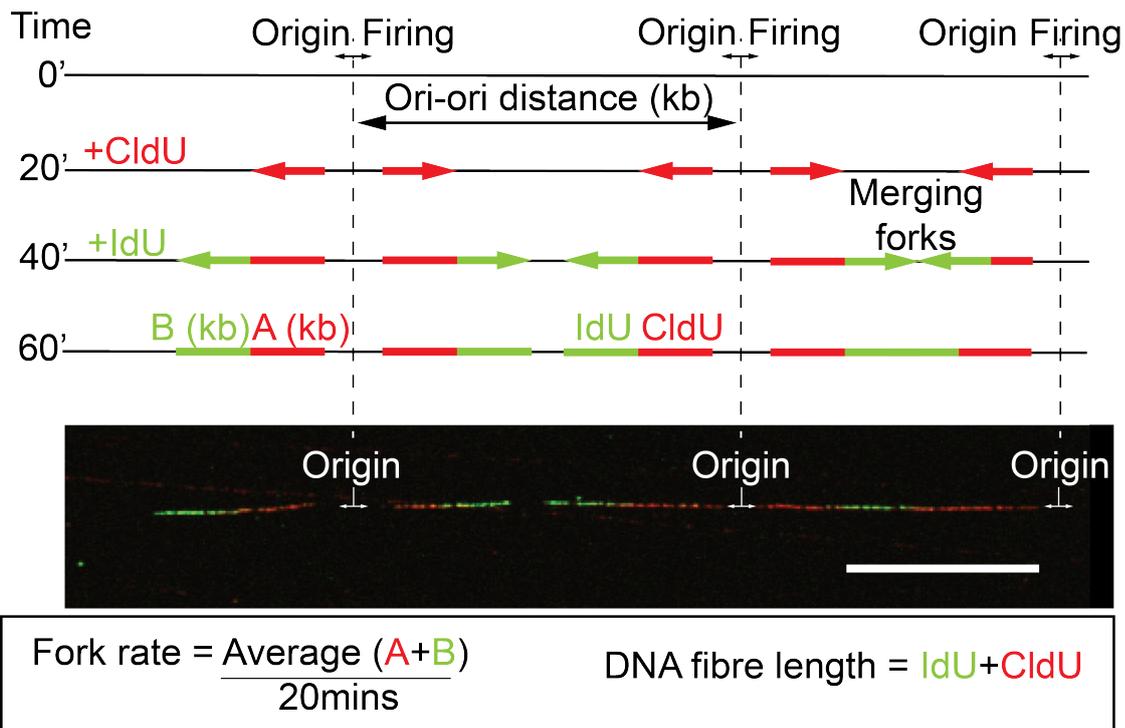
Figure 5.8 | DNA fibre assays monitor DNA replication dynamics. Schematic of DNA fibre assay. Sequential labelling of newly synthesised DNA fibres with first, chlorodeoxyuridine (CldU) and iododeoxyuridine (IdU) for 20 minutes each. Representative image of a single DNA fibre is displayed with multiple origins and bidirectional forks (scale bar = 10μM). We converted measurement to lengths in kb using the conversion factor 1μM = 2.59kb. Equations used in the determination of the replication dynamics are displayed in the box.

The replication dynamics of three pluripotent stem cell lines were compared to their differentiated somatic counterparts. A decreased DNA fibre length was observed in the undifferentiated hiPSC1 line compared with the parent fibroblast from which it was reprogrammed (**Figure 5.9a,b**). Interestingly, the length of the DNA fibres increased after differentiation of the hiPSC1 cells using CHIR99021 (**Figure 5.9a,b**). To confirm the pluripotent state was associated with decreased DNA fibre lengths, two further pluripotent lines: hiPSC2, hESC, and their differentiated derivatives were assessed (**Figure 5.9b**). Consistent with a decrease in fibre length, the rate of replication fork progression was also diminished in the pluripotent state. The pluripotent lines displayed a slower overall distribution of replication fork speeds compared to their somatic cellular counterparts (**Figure 5.9c**). Further, the mean fork rate was significantly slower in each of the pluripotent cell lines compared to the

differentiated lines (**Figure 5.9d**). When the progress of replication forks is slowed the cell can fire from dormant origins to ensure replication is completed on time (Taylor, 1977). By measuring the distance between replication origins, it is possible to determine the abundance of replication origins. A consistent decrease in the distance between origins of replication was observed in the three pluripotent cell lies that is consistent with an increase in origin abundance (**Figure 5.9e**). These novel data show a perturbation to replication dynamics, involving both reduced replication speed and increased numbers of replication origins in pluripotent cells compared with isogenic somatic counterparts.
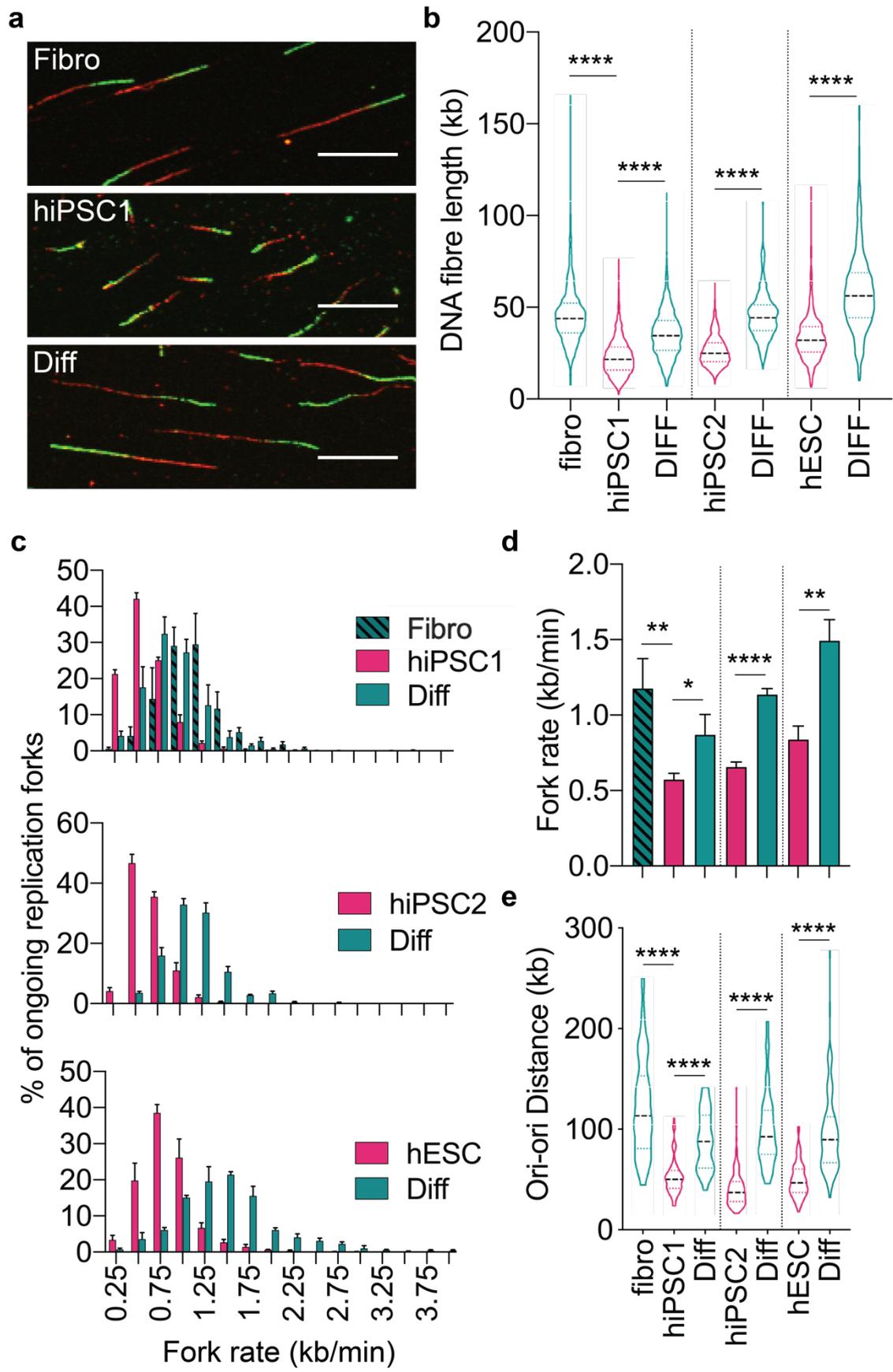
Figure 5.9 | Replication dynamics are perturbed in human PSC. Data presented is from our three pluripotent lines hiPSC1, hiPSC2 and hESC (pink) and their differentiated derivatives (green) **a,** Representative images of labelled DNA fibres from the parent fibroblast line (top), hiPSC1 line (middle) and the differentiated derivatives of hiPSC1 (bottom). Scale bars, 10μM. **b,** Combined length of CldU and IdU in individual fibres (n > 200 fibres per cell line per experiment, n = 3 experiments), Media length, 25th and 75th quartile are presented, two-tailed t-test, ****P<0.0001. **c,** Distribution of replication fork rates (n > 200 fibres per cell line per experiment, n = 3 experiments), data is the mean value from each experiment ± s.e.m **d,** Mean fork rates from (**d**) ±s.d., *P<0.05, **P<0.01, ****P<0.0001 (n = 3 experiments). **e,** Distribution of adjacent origins distance measurements (Ori-ori). Median distance, 25th and 75th quartile are presented, two-tailed t-test, ****P<0.0001. (n > 30 per cell line, per experiment. n = 3 experiments)

5.2.5 Pluripotent stem cells are susceptible to DNA damage

Replication stress, the slowing, stalling and collapse of replication forks, can lead to the formation of DNA lesions in the form of double-strand breaks (Bartkova et al., 2005). If these breaks are left unrepaired or repaired incorrectly, they can lead to genetic instability. To investigate the susceptibility of human PSC to DNA damage, two commonly used assays were employed to measure double-strand breaks *in vitro*. The primary assay used was the immunofluorescence staining for γH2AX, the findings from this assay were confirmed using the neutral comet assay, a single cell gel electrophoresis approach for monitoring double-strand breaks.

As DNA damage is more likely to be inflicted as a result of DNA replication in S phase it is important to be able to separate this population from the G1 phase cells particularly when comparing different cell states. Further, conventional analysis of γH2AX staining involves manually counting the number of foci in each cell which can be labour intensive and open to subjective analysis. Therefore, a CellProfiler pipeline was developed that is capable of determining a cell cycle phase based on DNA content whilst automatically counting the number γH2AX foci within that cell. This had the added benefit of reducing labour and substantially improving throughput. Briefly, images of nuclei stained with DAPI as well as images of the γH2AX foci were

acquired using the high content INCell microscope. The DAPI images were segmented by CellProfiler and at the same time, a measurement of the integrated intensity of the DAPI stain was made that would later be used to determine the cell cycle phase (**Figure 5.10a**). The integrated intensity is a measurement of pixel density per area and can be used to separate cells between 2N and 4N of DNA content. The segmented DAPI image was masked over the γH2AX image to eliminate any non-nuclear signal from non-specific binding or debris and improved the accuracy of analysis (**Figure 5.10b**). The pipeline then segmented the individual γH2AX foci and counted the numbers per cell (**Figure 5.10c**). The data was exported as a spreadsheet which indicated the number of γH2AX foci per cell and the integrated intensity of the DAPI stain. The cells in the G1 phase were discarded based on DNA content as the focus of this assay was to understand DNA damage induced during DNA replication (**Figure 5.10d**).

Figure 5.10 | Automated cell profiler analysis of γH2AX immuno-labelled cells.
a, Nuclei segmentation was performed based on DAPI stained nuclei. The original image (left), is segmented by CellProfiler using a global otsu thresholding strategy (middle) and accepted objects are then displayed (right). **b,** Segmented nuclear objects from (**a**) are used as a nuclear mask to remove non-specific binding and non-nuclear signals. **c,** Detection and segmentation of γH2AX signals that are related to the nuclei segmented in (**a**). **d,** Measurements of integrated intensity from segmented nuclei in (**a**) plotted to separate cells based on cell cycle stage.

a Nuclei segmentation from DAPI images

b Nuclear mask of H2AX signal

c Segmentation of H2AX signal

d DAPI integrated intensity measurement

Sub G1/G1

S/G2

% frequency

Integrated intensity
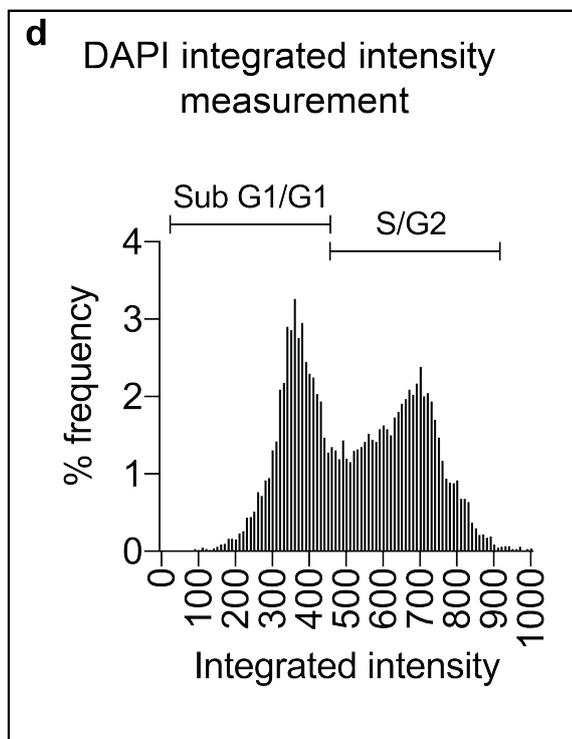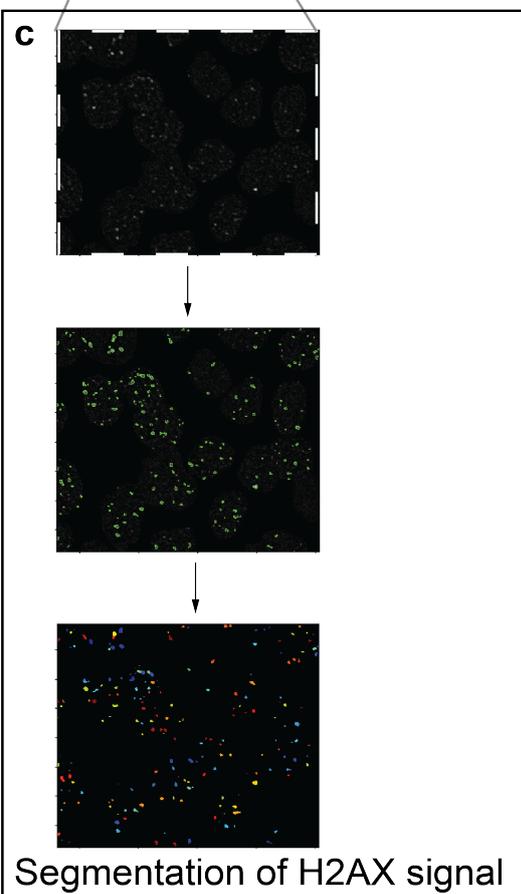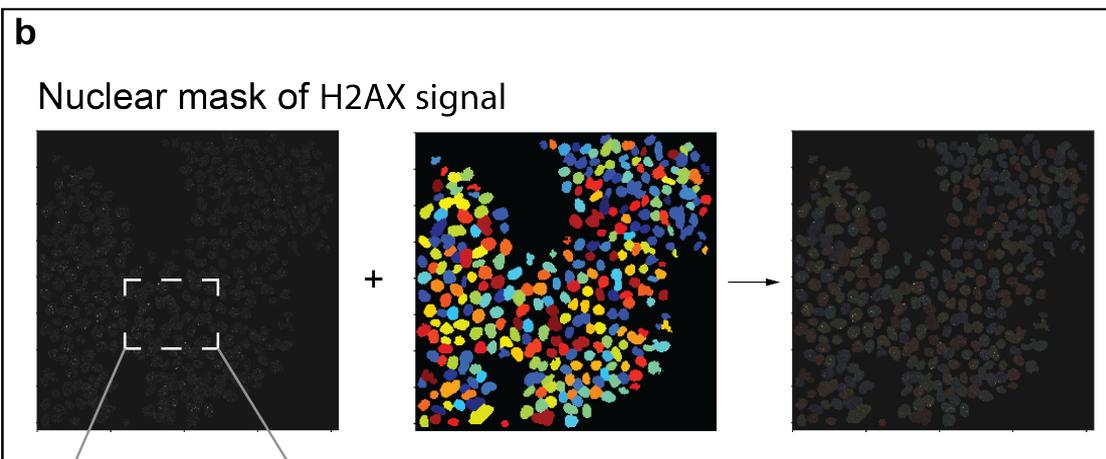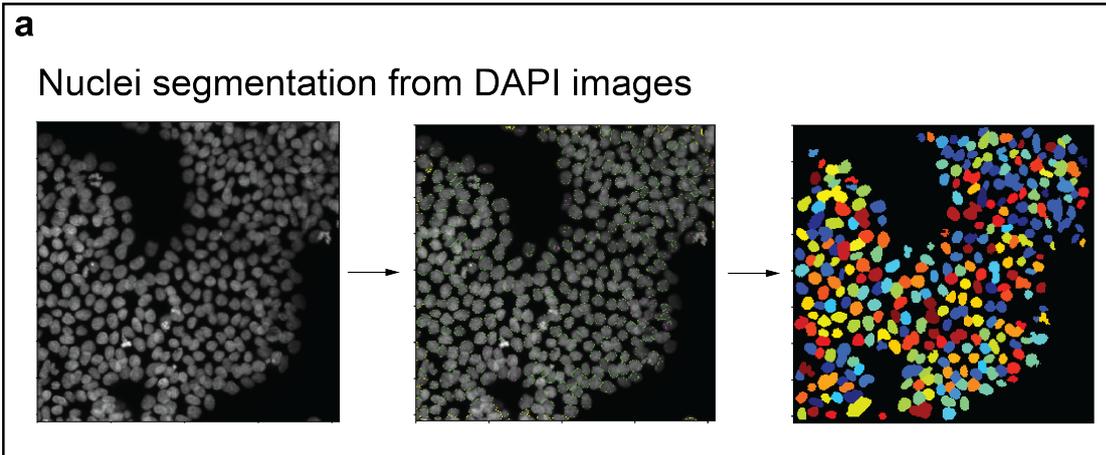
The γH2AX assay was used to determine the susceptibility of cells in the pluripotent state to double-stranded DNA breaks. Overall, greater numbers of γH2AX foci were observed in hiPSC1 compared to the parent fibroblast cell line and its differentiated derivatives (**Figure 5.11a**). Next, the frequency of cells in S/G2 phase with greater than eight γH2AX foci were measured. This measurement was chosen as it would represent the population of cells that have incurred significant numbers of double strand-breaks during S phase. In the parent fibroblast line, 19% of the S/G2 populations had more than eight γH2AX foci which significantly increased to 51% in the daughter pluripotent line hiPSC1 (**Figure 5.11b**). When hiPSC1 was differentiated, the frequency of cells with eight foci decreased again to 22% (**Figure 5.11b**). In our second pluripotent line, hiPSC2 compared to its differentiated derivatives, a similar trend of 52% to 29% respectively was observed (**Figure 5.11b**). Lastly, in the hESC line, 39% of the S/G2 phase cells had more than eight γH2AX foci which compared to only 22% of the differentiated hESC cells (**Figure 5.11b**).
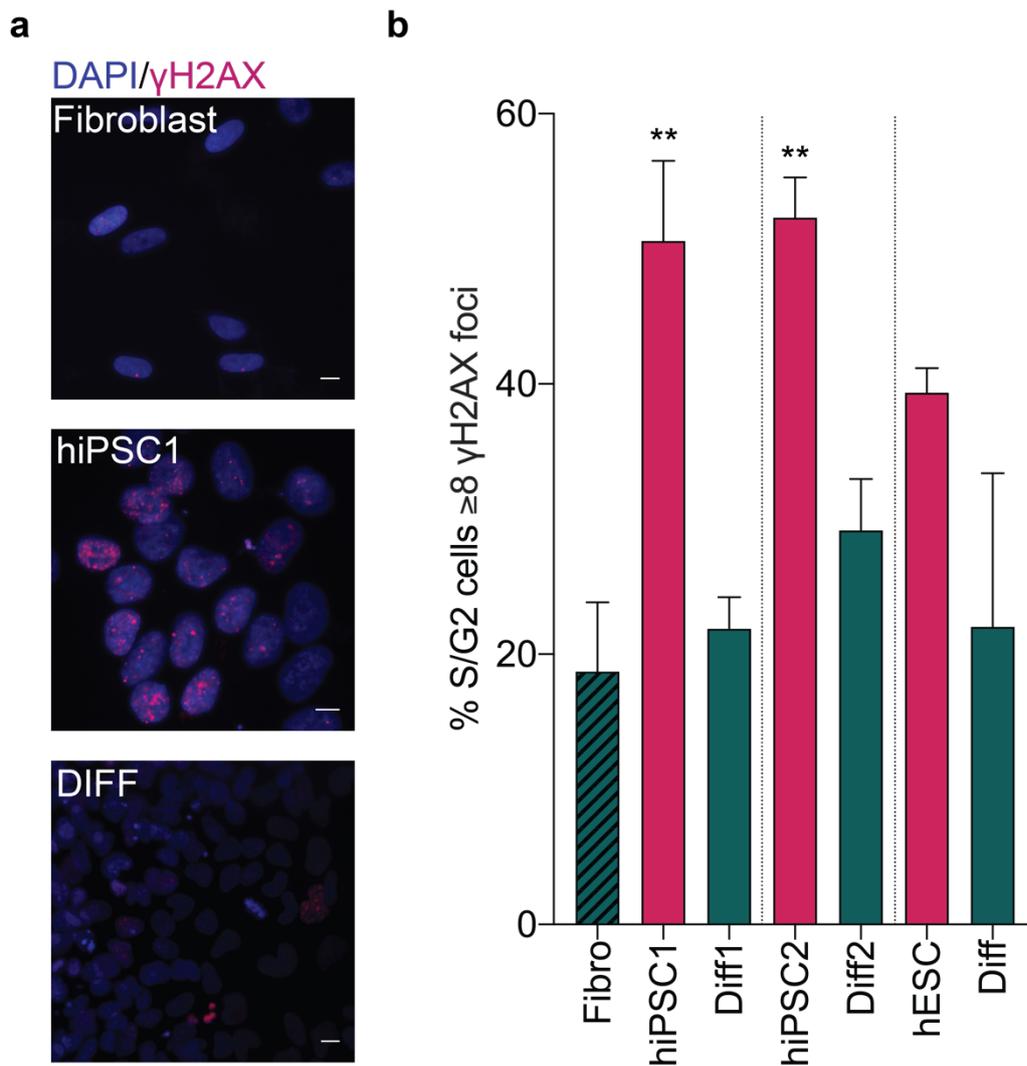
Figure 5.11 | Human PSC are susceptible to DNA damage. a, Representative merged images of DAPI and γH2AX images for the parent fibroblast line (top), hiPSC1 (middle) and the differentiated derivatives to hiPSC1 (bottom). scale bar, 10μM. **b,** Frequency of S/G2 cells with ≥8 γH2AX foci. The S/G2 phase was determined from nuclear DNA content. Data in **b** are mean values of 3 independent experiments ± s.d., for the isogenic somatic (fibroblast and differentiated derivatives of hiPSC1, hiPSC2 and hESC) (green) and pluripotent cell lines (hiPSC1, hiPSC2 and hESC) (pink), two-tailed *t*-test, **$P<0.01$, ($n > 100$ cells per cell line per experiment).

These results were confirmed using the neutral comet assay. Visually, the length of the tails in the pluripotent cells were longer than those in the somatic differentiated control lines (**Figure 5.12a**). To quantify this observation, tail moments were

measured using the Comet Assay IV data analysis software (Instem). The tail moment measured for hiPSC1, hiPSC2 and hESC ranged between 3.3 to 4.3 (**Figure 5.12b**). Comparatively, the fibroblast and differentiated cell lines had a tail moment of between 1.3 and 2.9 and was significantly lower than their pluripotent counterparts in each case (**Figure 5.12b**).

Altogether, the pluripotent cells possess a cell cycle that is consistent with many cancer cells that exploit a relaxed Rb-E2F to induce genetic instability and increase tumorigenic potential. Our results show that human PSC display perturbed replication dynamics and a high susceptibility to DNA damage.
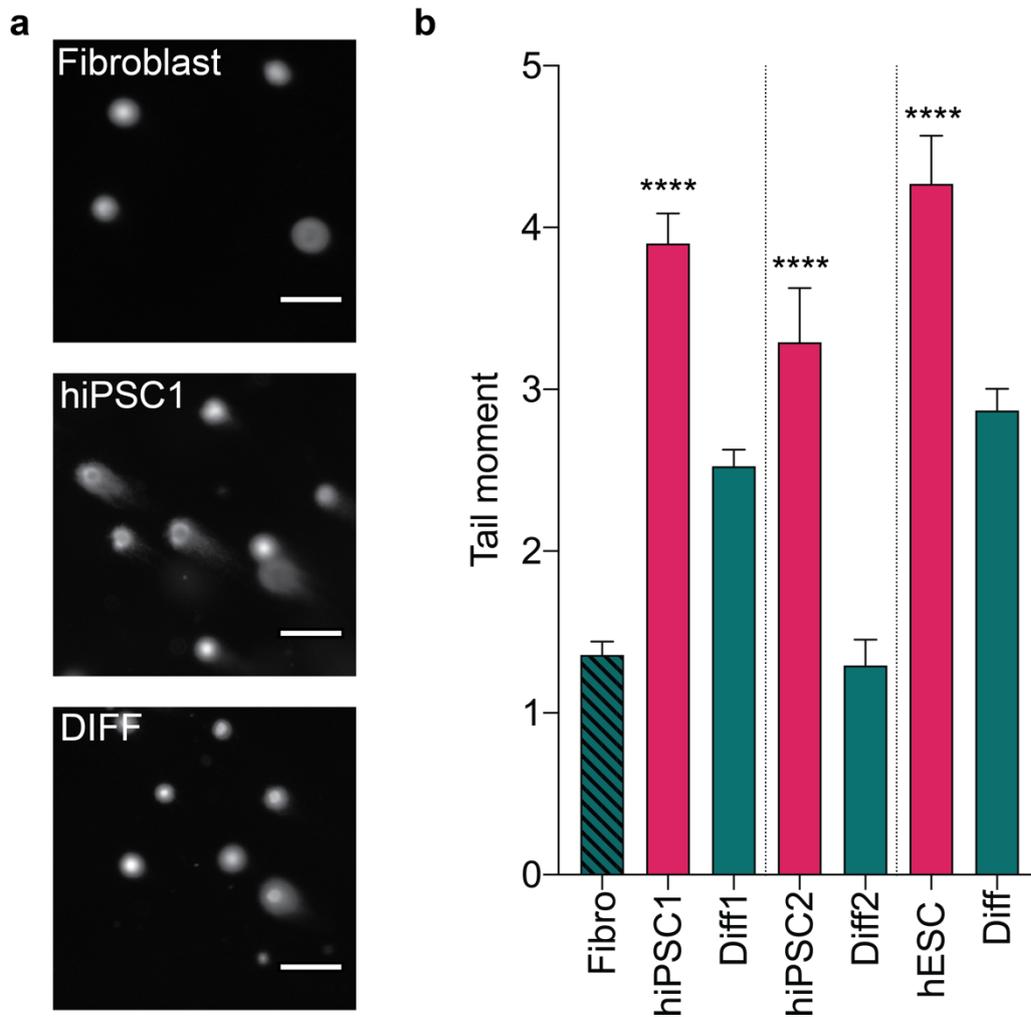
Figure 5.12 | Human PSC lines have increased numbers of double strand breaks. a, Representative comet images from one panel of isogenic somatic and pluripotent cells. Image acquired from the parent fibroblast (top) to the hiPSC1 (middle) and after the hiPSC1 was differentiated (bottom). **b,** Average tail moment from neutral comet assays on three independent human PSC lines (hiPSC1, hiPSC2 and hESC) (pink) and their differentiated derivatives (Fibro and Diff) (green). Data displayed is from 3 independent experiments $\pm$ s.e.m., two-tailed *t*-test, ****P<0.0001, ($n \geq 300$ cells per cell line per experiment).

5.3 Discussion

To investigate the mechanisms of mutation in human PSC that may be leading to the formation of structural and numerical instabilities, I examined unique features of cells in the pluripotent state. To do this it was important to establish isogenic cell lines that were both pluripotent and somatic in nature so that any comparisons made were irrespective of the cells genetic background. I differentiated each of our human PSC lines using CHIR99021, a small molecule inhibitor of the GSK3α enzyme that acts to downregulate β-catenin and C-Myc (Ye et al., 2012). GSK3α inhibition leads to WNT activation that differentiates cells to mesodermal and endodermal progenitors. The human PSC used in this study were robustly differentiated using this protocol as determined by immunofluorescence and RT-qPCR gene expression analysis.

During DNA replication a cell is subject to stresses that a senescent cell is not. One key feature of human PSC is their ability to self-renew endlessly. To ensure our differentiated cells had retained the capacity to proliferate, I differentiated the cells for a minimum period of 5 days and ensured a sample of the cells used in each experiment stained positively for Ki67, a marker of proliferation. Of all the cell lines used in this study hiPSC1 and its somatic counterparts provided the most comprehensive panel to investigate mechanistic features of pluripotent cells that are leading to genetic instability. The hiPSC1 line had been reprogrammed from a primary adult fibroblast line in Sheffield. By analysing both of these lines and its differentiated derivatives an assessment of the difference that occurs to a cell line after it acquires pluripotency and what happens following its differentiation back into a somatic state was made. One difference that was identified between these cell states is the rapid cell cycle dynamics that has been observed in human PSC. This is of particular interest as replication stress caused by rapid cell proliferation has been linked to DNA damage and cancer onset in adult stem cells (Tomasetti and Vogelstein, 2015, Walter et al., 2015)

Links between pluripotency and cancer have long been discussed. For instance, tumour progression has been linked to the reactivation of many genes that are required for embryonic development and are thought to provide cancer cells with the

ability to endlessly self-renew which is crucial for tumour formation (Bussard et al., 2010, Pezzolo et al., 2011, Illmensee and Mintz, 1976). One unique and innate property of pluripotent cells is their capacity to self-renew through unique cell cycle characteristics. Since the discovery of somatic cell reprogramming to human iPSC our ability to understand the potential mechanisms cancers use to unlock the capacity to endlessly self-renewal have been expanded. Of note, it has been shown that the silencing of p16^{INK4a} is a common feature of human iPS cells and ES cells and its suppression during reprogramming is capable of increasing the efficiency of generating human iPSC lines (Li et al., 2009a). The p16^{INK4a} binds to and inhibits the cyclin D dependant CDK4/6 kinases that are important in relieving the inhibitory action of the Rb tumour suppressor. Silencing of p16^{INK4a} leads to elevated activity of cyclin D/CDK4 (Lukas et al., 1995) and in turn, upregulation of E2F transcription factor targets that include cyclin E (Leone et al., 1998). In this study, I have shown an elevation in cyclin D2 and cyclin E that is associated with the pluripotent state. This is consistent with the cell cycle profiles that I observed using flow cytometry of dual EdU and DAPI labelled cells. The pluripotent lines had a far greater proportion of cells that resided in the S phase and a smaller population in G1 phase. Further, I revealed that pluripotent cells display a truncated G1 phase that can be explained by the atypical expression of the cell cycle cyclins. I used two independent assays to measure the cell cycle time in both somatic and human PSC lines. Both assays determined the cell cycle time in the fibroblast line to be 23 hours. Interestingly, when I determined the cell cycle time of three independent human PSC lines by time-lapse microscopy, we found it to be consistently longer by 2 hours than when its length was determined by pulse-chase analysis. An explanation for this could be the difference in experimental conditions, unlike in the pulse-chase analysis, the time-lapse assay required measurements taken from single cells so that the cell divisions were accurately determined. The dissociation of human PSC into single cells has previously been associated with extensive stress leading to cell death (Chen et al., 2010, Ohgushi et al., 2010, Watanabe et al., 2007) and it may be that cell to cell contact mediates the survival and proliferation of these cells (Fox et al., 2008, Andrews et al., 1982). Regardless, we observed a faster rate of proliferation in human PSC than isogenic somatic counterparts that are consistent with atypical G1 cyclin expression.

With features of some cancer cells thought to emanate from their ability to unlock pluripotent characteristics, I explored the association of cancer cells genetic instability and cell cycle characteristics further. Genetic instability is a key feature of cancer development with many different models suggested as potential causes. Current research has focused on oncogene activation that induces DNA replication stress and double-strand breaks. Oncogenic activation of the Rb-E2F pathway, a master regulator of cell proliferation has been shown to lead to perturbed progression of replication forks (Bester et al., 2011) and replication stress through the uncoordinated activation of the gene targets for E2F that include, amongst others, the nucleotide biosynthesis pathway (Leone et al., 1998, Lundberg and Weinberg, 1998, Harbour et al., 1999). This uncoordinated nucleotide synthesis is thought to be inadequate for the needs of the now rapidly proliferating cells (Bester et al., 2011). Previous studies have identified that depletion of nucleotide pools leads to slow DNA replication rates (Anglana et al., 2003), increased DNA replication origin abundance (Ge et al., 2007) and the formation of numerous double-strand breaks (Saintigny et al., 2001). Attempts to investigate similar symptoms of replication stress in human PSC have relied on the expression of proteins associated with replication fork stalling and repair (Vallabhaneni et al., 2018). However, the differences that they observed with these assays could be explained by pluripotent cells atypical DNA damage response and repair pathways, such as the reliance on HR over NHEJ. To overcome these limitations, I established several techniques to monitor the replication dynamics and DNA damage in human PSC relative the isogenic somatic control lines. The fibre assay provides a direct approach to understand differences in replication dynamics between cell states across the entire genome and at a single-molecule level. Strikingly, the fibre analysis revealed perturbed replication dynamics in the pluripotent cells compared to the somatic control lines. The speed of replication fork progression was decreased across all our lines compared to the somatic counterpart lines and we noted a decrease in the distance between origins of replication which suggested that as a consequence of slower replication fork speed, the cells were firing from dormant origins. To monitor the susceptibility of our pluripotent lines to DNA damage, I used two assays routinely employed for the *in vitro* detection of double-strand breaks. By using phopho-specific antibodies for serine 139 residue, it is possible to image the phosphorylated form of this protein that appears as foci within the nuclei of cells.

H2AX becomes phosphorylated to generate γH2AX in the vicinity local to a double-strand break (Rogakou et al., 1998). As such, counting the number of foci in a cell can be used to sensitively determine the number of double-strand breaks. However, γH2AX does not allow the user to monitor the exact fate of a double-strand break and may not decay following its resolution (Kinner et al., 2008). For these reasons, we performed additional double-strand break analysis using a single cell gel electrophoresis approach known as the comet assay. Measurements of a cells DNA migration through agarose gel as well as the relative amount of DNA present is known as the tail moment and can be used as an index of DNA damage. Consistent with previous findings, I found the pluripotent cells to have extensive γH2AX that is particularly associated with the S/G2 phase of the cell cycle (Vallabhaneni et al., 2018, Simara et al., 2017). I confirmed that the extensive γH2AX in these cells was associated with numerous double-strand breaks by measuring tail moments by neutral comet assays.

Altogether, these data suggest that the pluripotent cells possess cell cycle characteristic that are similar to those observed in early-stage cancer development. Silencing of p16^{INK4a}, that is observed in all human PSC would account for the high expression of Cyclin D2 and Cyclin E and the rapid progression through G1 phase that is reported here. Further, I found pluripotent cells possessed perturbed replication dynamics and a susceptibility to DNA damage that was strikingly similar to cancer cells that display a relaxed Rb-E2F and aberrant activation of the nucleotide biosynthesis pathways (Bester et al., 2011).

In summary, the aberrant activation of the Rb-E2F pathway in human PSC provides a likely cause of replication stress and DNA damage in human PSC. Using the findings uncovered in this chapter the following section will explore the optimisation of human PSC cell culture conditions that aim to reduce replication stress and minimise genetic instability in human PSC cultures.

6 Replication stress in human PSC is rescued with exogenous nucleosides

6.1 Introduction

In the previous chapter, the resemblance of cancer cells to human PSC was discussed. Silencing of p16^{INK4a} and overexpression Cyclin E and D2 are common features of human PSC (Li et al., 2009a), which enforces proliferation through the activation of E2F, which is also a common feature of certain cancers (Sekido et al., 1998). It has been proposed that the oncogenic activation of E2F promoted G1/S transition and enforces proliferation (Becker et al., 2006). This improper control of the cell cycle can lead to dysregulation of replication initiation, origin firing and increased replication/transcription collisions leading to replication stress directly or depletion of nucleotide pools that can initiate replication fork stalling and collapse (Bester et al., 2011, Jones et al., 2013, Halazonetis et al., 2008, Ekholm-Reed et al., 2004).

Localised DNA damage can form from replication stress after DNA forks stall and collapse, but evidence has also shown that a consequence of replication stress is then formation structural and numerical chromosomal instabilities (Burrell et al., 2013). Copy number variants may arise due to erroneous repair of breaks that emanate from collapsed forks. A form of homologous recombination DNA repair, known as break-induced replication is employed to repair DNA damage left by collapsed forks when homology is present at only one side of the break. Erroneous repair by break-induced replication generates amplifications with microhomology at the breakpoint (Costantino et al., 2014). Further, when the analysis of a panel of colon adenocarcinoma cell lines was performed, those lines that were chromosomally unstable presented replicative stress characteristics, whereas those that were stable did not. Replication stress in the unstable lines led to defects in chromosome segregation (Burrell et al., 2013) as a result of under-replicated regions and unresolved replication structures that proceeded into mitosis (Lukas et al., 2011).

Regardless of the mechanism surrounding replication stress and its consequences, it was shown that the addition of exogenous nucleosides *in vitro* can help to reduce it. It was identified that following overexpression of cyclin E, rNTP pools were reduced, and replication stress proceeded, although this was rescued with the addition of exogenous nucleosides *in vitro* (Bester et al., 2011). Further, the supplementation of exogenous nucleoside in chromosomally unstable colon adenocarcinomas decreased both replication stress-induced DNA damage and chromosome segregation errors that occurred during mitosis (Burrell et al., 2013). However, for the cells to make use of the exogenous nucleosides they must first convert them to dNTPs using the nucleotide biosynthesis pathway (**Figure 6.1**). The rate-limiting step in the formation of dNTPs is the reduction of NDPs to dNDPs by the enzyme ribonucleotide reductase (RNR). In response to DNA damage and replication stress, RNR subunit activity is upregulated through checkpoint controls (Zegerman and Diffley, 2009). Under normal conditions, E2F6 begins to accumulate through S phase to suppress the activity of E2F gene targets. Cells activate CHK1 checkpoint in response to DNA damage and replication stress to inhibit E2F6 and prolong E2F gene activation of downstream targets. One such target is the RNR subunit RRM2 that buffers the detrimental effects that can result from sustained replication stress and thereby prevent DNA damage and cell death (Bertoli et al., 2013, Herlihy and de Bruin, 2017). CHK1 also acts to stall the cell cycle (Sørensen et al., 2003, Sørensen and Syljuåsen, 2012) and coordinates the firing from dormant origins in the presence of under replicated regions (Ge and Blow, 2010). This is crucial to prevent cells progressing into mitosis before the completion of DNA synthesis, preventing lagging chromosomes, chromosome breaks, aneuploidy or mitotic catastrophe (Ishida et al., 2001, Saavedra et al., 2003, Manning and Dyson, 2012). The function of CHK1 in this way is essential with its loss leading to cells that are highly sensitive to apoptosis in response to DNA replication stress and DNA damage (Takai et al., 2000).

In human PSC, initiation of replication stress by treating cells with thymidine or cisplatin resulted in these cells committing to apoptosis whereas somatic cells favoured arrest of DNA synthesis, which restarted at a more favourable time (Desmarais et al., 2012). In response to replication stress, human PSC did not generate ssDNA and did not recruit RPA. Without RPA, human PSC failed to activate CHK1 in response to replication stress, although CHK1 was still activated

from other forms of DNA damage (Desmarais et al., 2012). Furthermore, markers of DNA repair by homologous recombination and fork restart were only present at low levels and instead of repair and fork restart, these cells favoured apoptosis to eliminate cells with a high degree of replication stress (Desmarais et al., 2012). This likely reflects the specific demands of the early embryo where rapid proliferation is required, yet genetic instability could be catastrophic for the whole embryo. Despite this, mutations appear during in vitro culture that favour anti-apoptosis and provide a selective advantage to the mutant cells by overcoming replication stress-induced apoptosis.

Current culture media contain only the basic factors that maintain human PSC in the self-renewal state, with little consideration for components that may ensure the cells genetic stability. With the observations that human PSC are susceptible to replication stress and DNA damage, I explored the potential to improve cell culture media to minimise mutations that arise from replication stress whilst also reducing apoptosis that, during prolonged culture, encourages the selection of mutant cells that evade apoptosis.

Figure 6.1 | Schematic of purine and pyrimidine metabolism. The schematic illustrates the nucleotide salvage and de-novo synthesis. The metabolite unit is written in black with the enzyme used to catalyse the reaction written in red. The schematic is separated into purines (top) and pyrimidines (bottom). In particular the schematic illustrates the metabolism of dNTP from nucleosides.

Purine metabolism

RNA   **DNA**   **DNA**   **RNA**

GTP   dGTP   dATP   ATP

Nucleoside diphosphate kinase (NDPK)

AMP-activated kinase (AMPK)

Nucleoside diphosphate kinase (NDPK)

GDP —*Ribonucleotide reductase (RNR)*→ dGDP   dADP ←*Ribonucleotide reductase (RNR)*— ADP

GMP Kinase

AMP-activated kinase (AMPK)

GMP reductase

AMP deaminase

GMP ←→   IMP   ←→ AMP

GMP synthase/ IMP dehydrogenase

AS synthase/lysase

5'-nucleotidase (5'-NT)

Adenosine kinase (AK)

5'-nucleotidase (5'-NT)

Guanosine   Adensoine

Hypoxanthine-guanine phosphoribosyltransferase (HGPRT)

Adenosine phosphoribotransferase (APRT)

Purine nucleoside phosphorylase (PNP)

Purine nucleoside phosphorylase (PNP)

Guanine   Adenine

Pyrimidine metabolism

Thymidine synthase (TS)

RNA　　　　　　　　**RNA**　　**DNA**　　　　　　**DNA**

CTP synthase

UTP　　　　　dUMP　　　CTP　　　　dCTP　　　　　　dTTP

Nucleoside triphosphate phosphatase　　Nucleoside diphosphate kinase　　Nucleoside triphosphate phosphatase　　Nucleoside diphosphate kinase　　Nucleoside diphosphate kinase

UDP　　　dUDP　　　CDP　　　dCDP　　　　　　dTDP

Ribonucleotide reductase (RNR)　　Ribonucleotide reductase (RNR)

Nucleoside diphosphatase phospatase　　Uridylate kinase　　Nucleoside diphosphatase phospatase　　Cytidylate kinase　　Thymidylate kinase

UMP　　　　　　CMP　　　　　　　　　　dTMP

5'-nucleotidase (5'-NT)　　Uridine Kinase　　5'-nucleotidase (5'-NT)　　Cytidine Kinase　　Thymidine kinase

Uracil phosphoribosyl-transferase

Uridine　　　　Cytidine　　　　　　　Thymidine

Uracil phosphoribosyl-transferase　　Uridine phosphorylase

Uracil

178

6.2 Results

6.2.1 Optimisation of nucleoside concentration for human PSC culture

To determine whether exogenous nucleosides can minimise the effect of replication stress on human PSC it was first important to determine the concentration that would provide the greatest benefit. A range of nucleoside concentrations were tested for the ability to reduce numbers of γH2AX foci. Cytidine, guanosine, uridine and adenosine were mixed at equimolar concentrations into the commercially available cell culture medium, mTeSR. A low level of thymidine was already included in mTeSR formula and so was titered at a lesser amount. Following 72 hours of culture in nucleoside (mTeSR + AUCTG) or control (mTeSR) conditions, the cells were fixed and the expression of γH2AX was quantified by immunofluorescence using the InCell high content microscope and analysis performed using the image analysis pipeline that I previously developed (section 5.2.5). The mean number of γH2AX foci per hiPSC1 or hiPSC2 cell in S/G2 phase was plotted and compared to the negative control (**Figure 6.2a,b**). A dose-dependent reduction of γH2AX foci was observed across multiple experiments in both cell lines. The concentration that produced the greatest response of cytidine, guanosine, uridine and adenosine equalled 15μM and thymidine was at 6μM (**Figure 6.2a,b**).

a
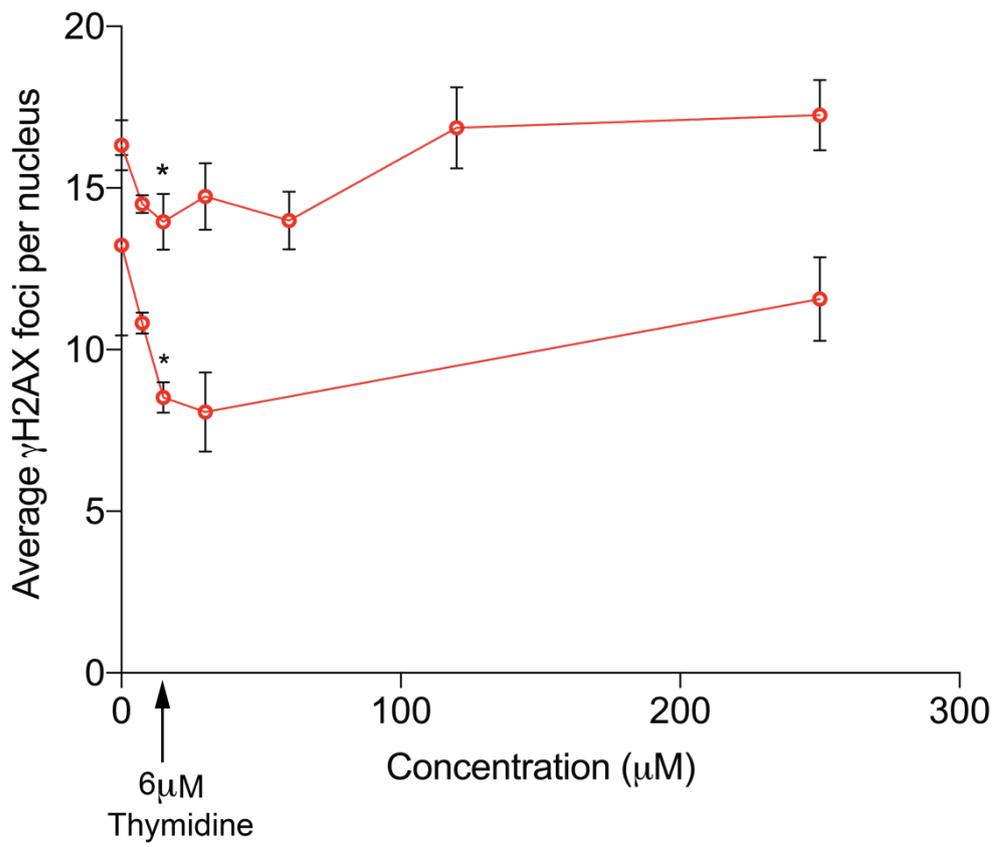
b
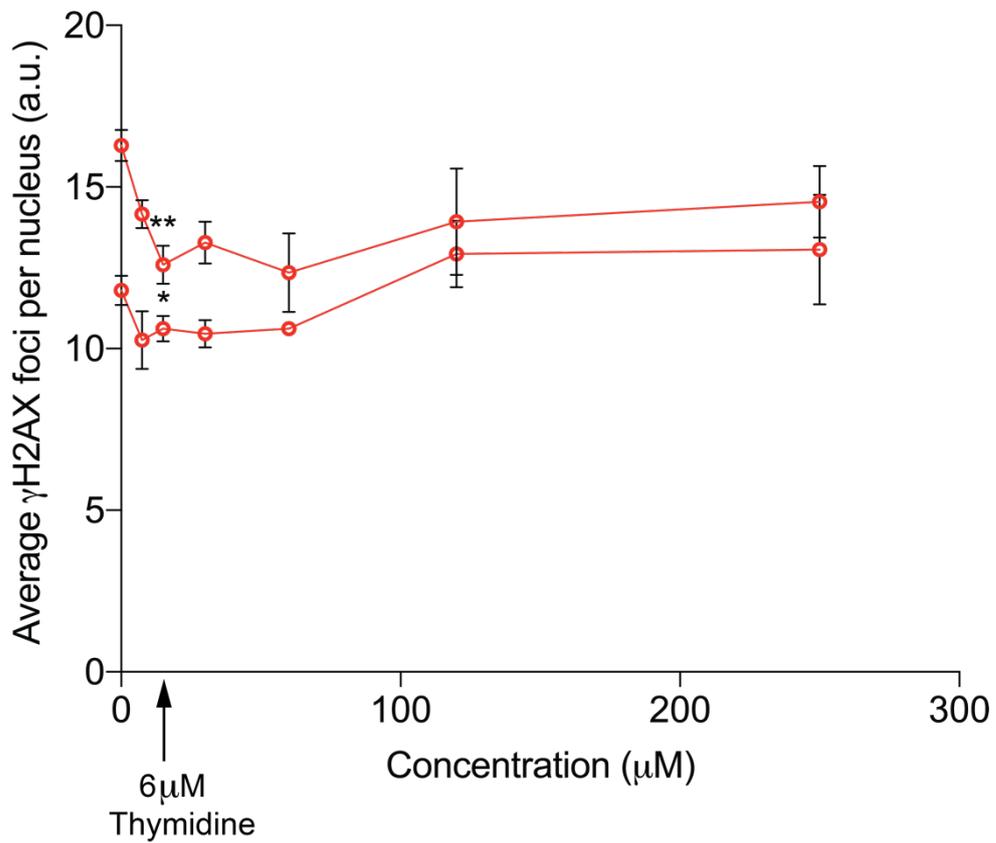
Figure 6.2 | Exogenous nucleosides reduced γH2AX expression in a dose dependant manner. a, hiPSC1 grown in mTeSR that was titered with exogenous nucleosides. Data presented is the average number of γH2AX foci per cell in each condition. Each line represents an individual experiment. **b,** Results from two independent experiments using a second cell line, hiPSC2. Dose dependant response was observed in both cell lines. Data are mean ±s.d. of three technical repeats, two-tailed *t*-test, *P<0.05, **P<0.01, ***P<0.001, ****P<0.0001.

6.2.2 Exogenous nucleosides maintain human PSC in a self-renewal state

Exogenous nucleosides can rescue replication stress of cells *in vitro* (Bester et al., 2011, Burrell et al., 2013). To determine if exogenous nucleosides could be included in cell culture media and reduce replication stress in human PSC, it was first crucial to ensure that cells maintained with the addition of exogenous nucleosides retained pluripotency. When hiPSC1 cells were grown in the presence of exogenous nucleosides they retained the expression of pluripotency-associated cell surface antigens, SSEA3, TRA-160, TRA-185 and TRA-181, over ten passages (**Figure 6.3a-d**). Overall, cells grown in mTeSR with the addition of exogenous nucleosides (Nuc) presented a greater proportion of cells that were positive for the respective cell surface antigen than when grown in mTeSR alone (control) (**Figure 6.3a-d**).

Figure 6.3 | Exogenous nucleosides maintain human PSC in a self-renewal state. The Human PSC line, hiPSC1 was grown for 10 passages in exogenous nucleosides (red) or control conditions (black). A sample of cells were taken every second passage and immuno-stained based on the expression of four pluripotency associated cell surface antigens, SSEA3 (**a**), TRA-160 (**b**), TRA-185 (**c**) and TRA-181 (**d**). P3X was used as a negative control in each case (blue) and the proportion of positive cells in each case was gated based on P3X.

c TRA-185

% Positive cells
NUC/Cntrl

Passage 2
100%
100%

Passage 4
100%
100%

Passage 6
100%
100%

Passage 8
100%
100%

Passage 10
100%
100%

d TRA-181

% Positive cells
NUC/Cntrl

Passage 2
98.7%
90.0%

Passage 4
93.8%
91.9%

Passage 6
89.3%
80.1%

Passage 8
97.3%
96.9%

Passage 10
98.9%
81.2%

Frequency (%)

Antigen expression

183
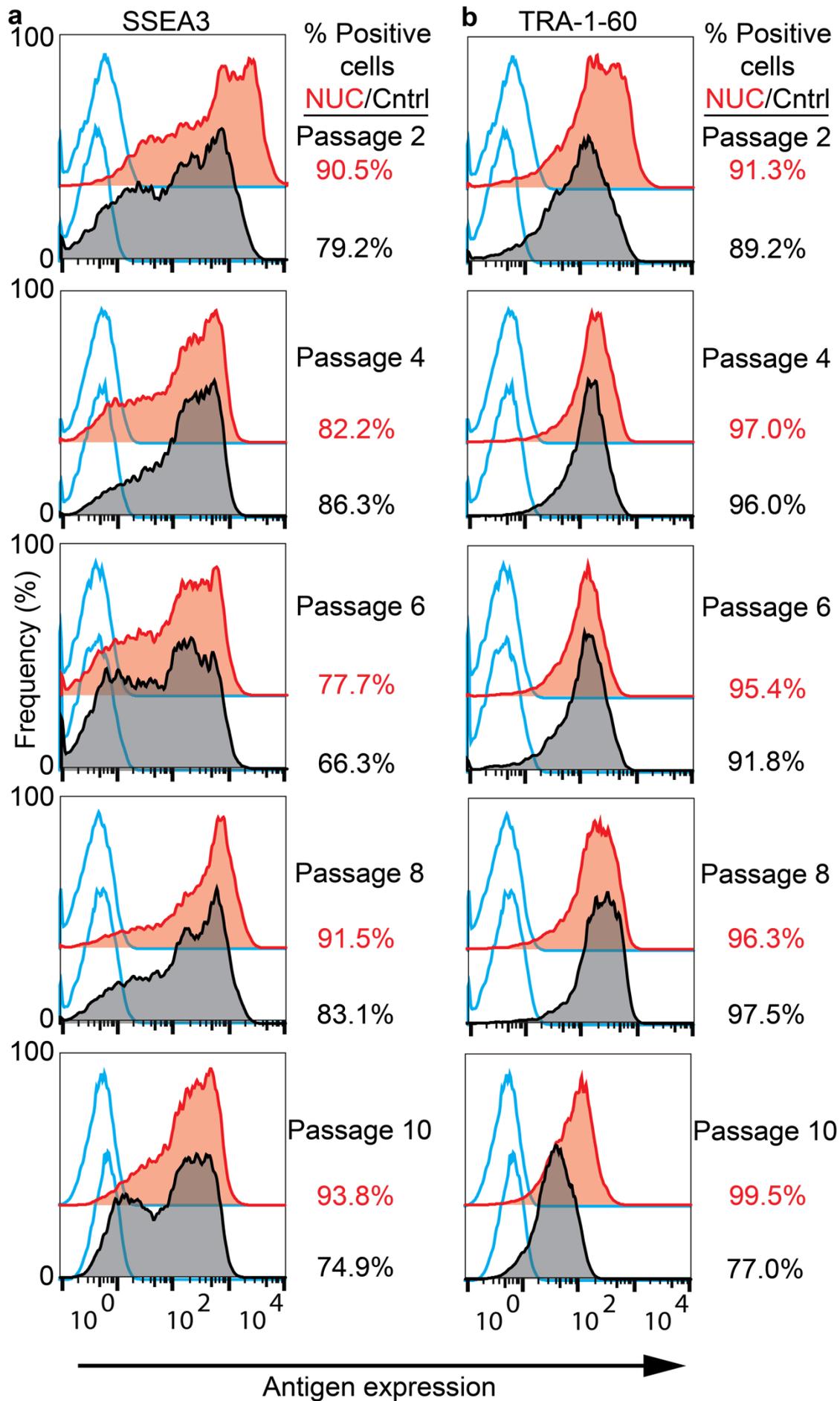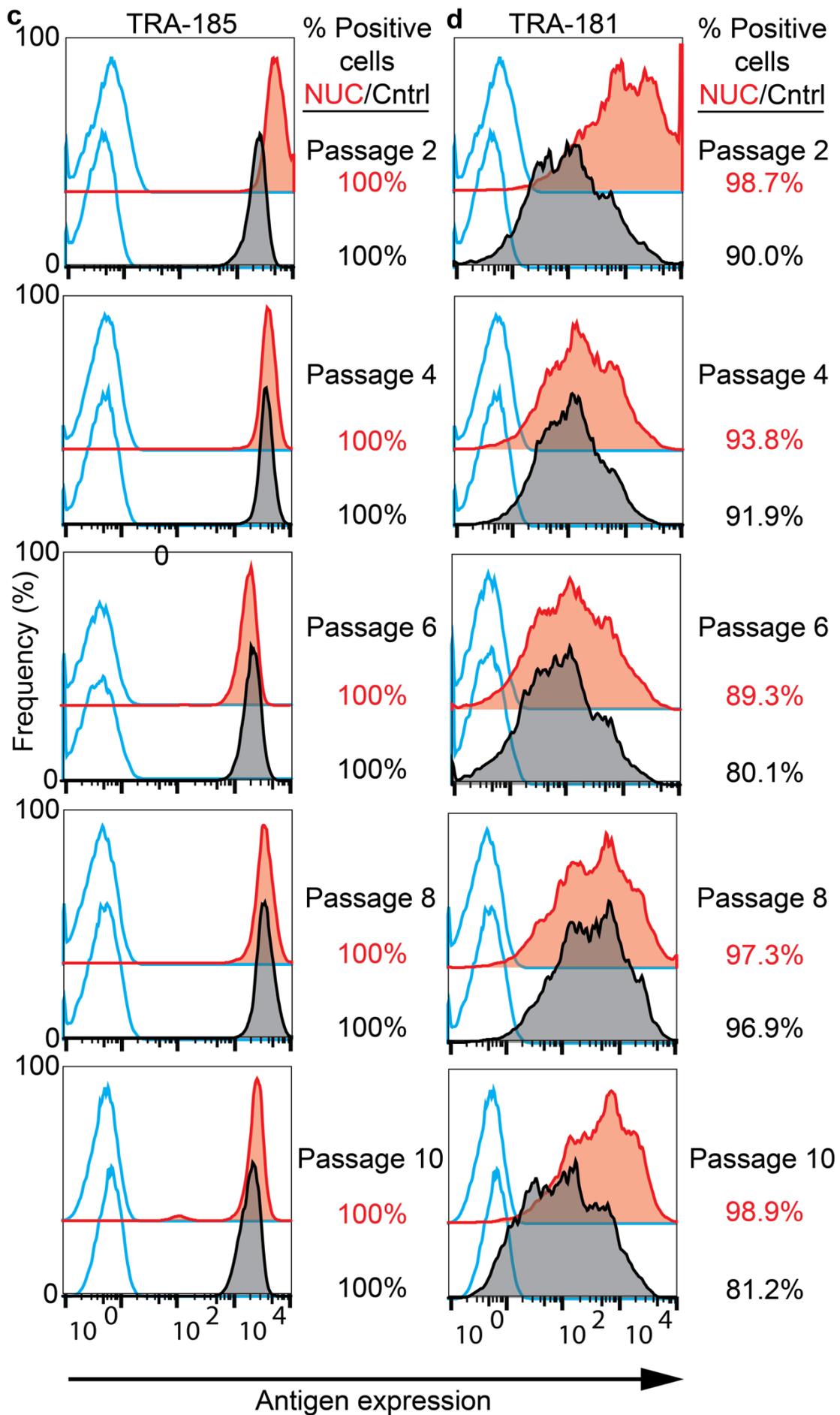
Following culture for 10 passages in the presence of exogenous nucleosides, a flask of cells was taken to assess their capacity to differentiate into the three embryonic germ layers. Human PSC grown in the control condition or in the presence of exogenous nucleosides were spontaneously differentiated into embryoid bodies under neutral conditions that are devoid of small molecule inhibitors that would otherwise direct differentiation to one of the three germ layers (**Figure 6.4a,b**). Interestingly, the embryoid bodies formed from the nucleoside condition were consistently larger and more irregular in form than those grown in the absence of exogenous nucleosides (**Figure 6.4a,b**). Gene expression analysis was performed on the 10-day old embryoid bodies using RT-qPCR for genes associated with pluripotency, mesoderm, ectoderm and endoderm genes. Human PSC grown in the presence of exogenous nucleosides spontaneously differentiated as observed by a loss of the pluripotent transcription factor OCT4 (**Figure 6.4c**). In comparison to cells maintained in self-renewal conditions, the differentiated cells presented an upregulation of gene expression associated with the three germ layers (**Figure 6.4d-f**). Together, these results show that the addition of exogenous nucleosides to human PSC culture media can maintain human PSC in a pluripotent state.

Figure 6.4 | Human PSC maintained in exogenous nucleosides spontaneously differentiate under neutral embryoid body conditions. Representative images of neutral embryoid bodies that were formed from hiPSC1 grown in control conditions (**a**) or in the presence of exogenous nucleosides (**b**). Gene expression analysis by RT-qPCR of the neutral embryoid bodies following 10 days spontaneous differentiation under neutral conditions. Gene expression for a panel of genes associated with pluripotency (**c**), endoderm (**d**), ectoderm (**e**) and mesoderm (**f**). Data are mean ±s.d., two-tailed *t*-test, *$P<0.05$, **$P<0.01$, ***$P<0.001$, ****$P<0.0001$ (*n* = 3 experiments).
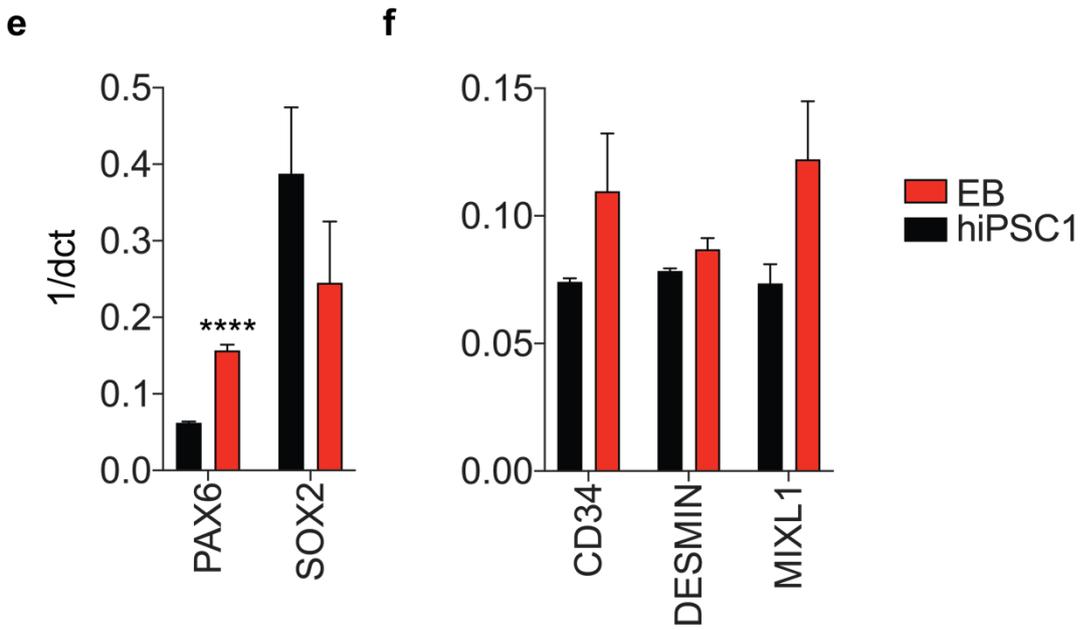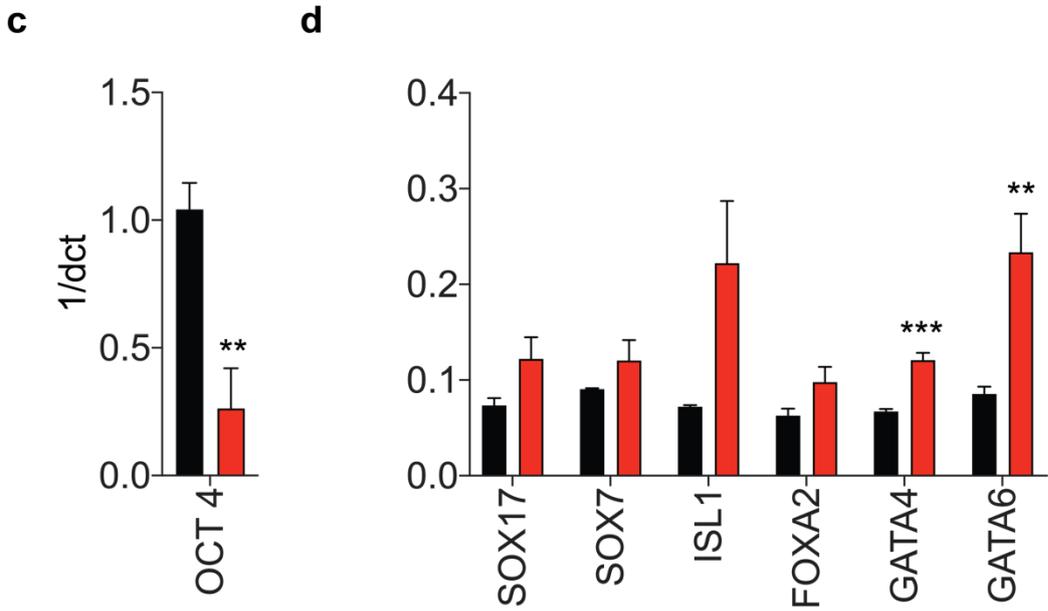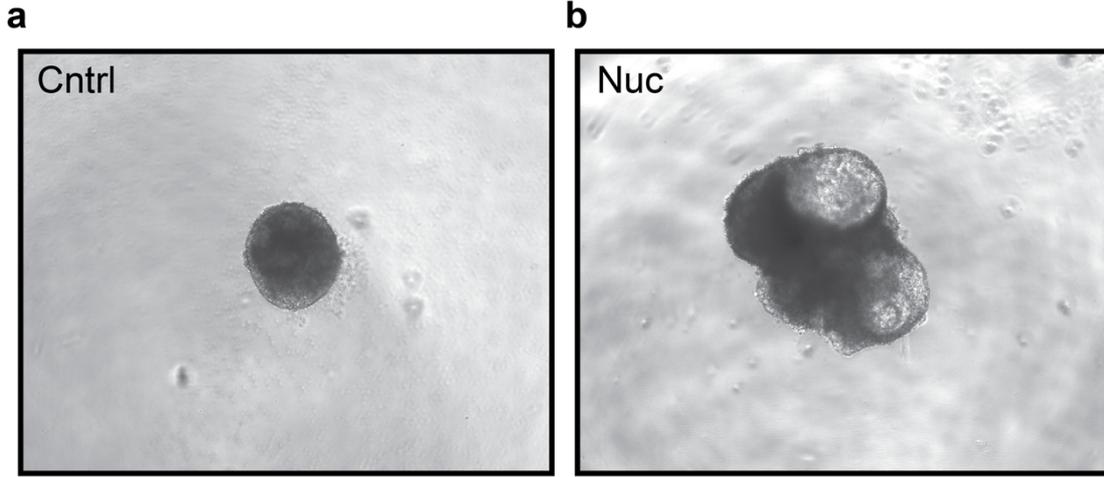
6.2.3 Exogenous nucleosides reduce DNA damage in cultures of human PSC

In the previous chapter, it was shown that the pluripotent state is susceptible to DNA damage during S/G2 phase of the cell cycle. With the concentration of nucleosides optimised for pluripotent cell culture, the frequency of cells with high levels of damage induced during S/G2 phase of the cell cycle was examined. After 72 hours in culture with exogenous nucleosides, the frequency of cells with more than 8 γH2AX foci significantly decreased in hiPSC1, hiPSC2 and hESC lines (**Figure 6.5a,b**). To confirm these findings, the neutral comet assay was used to directly measure DNA damage of individual cells. The addition of exogenous nucleosides reduced the average tail moment as determined by the neutral come assay (**Figure 6.5c,d**). These findings show that the addition of exogenous nucleosides to human PSC media has the benefit of reducing the susceptibility of human PSC to DNA damage.

Figure 6.5 | Susceptibility to DNA damage is reduced in cultures of human PSC with the addition of exogenous nucleosides. a, Representative images of γH2AX stained hiPSC1 cells grown in control (top) and nucleoside (bottom) conditions for 72 hours. γH2AX (red – 647) and nuclear counterstain (blue – DAPI). White arrows indicate each nuclei that contains 8 or more γH2AX nuclei. Data are mean $\pm$s.d., two-tailed t-test, *P<0.05, **P<0.01, ***P<0.001, ****P<0.0001 (n = 3 experiments). **b,** Frequency of cells with greater than or equal to 8 γH2AX foci in hiPSC1, hiPSC2 and hESC grown in control (black) or nucleoside conditions (red). **c,** Representative images from comet assays performed on hiPSC1 when grown in control (top) or nucleoside (bottom) conditions. **d,** Quantification of tail moments in hiPSC1 grown in control (black) and nucleoside (red). Data from 3 independent experiments are displayed. Mean $\pm$s.d., two-tailed t-test, *P<0.05, **P<0.01, ***P<0.001, ****P<0.0001.

a

Cntrl

Nuc

b

S/G2 Cells with >8 γH2AX foci

Cntrl-hiPSC1
Nuc-hiPSC1
Cntrl-hiPSC2
Nuc-hiPSC2
Cntrl-hESC
Nuc-hESC

c

Cntrl

Nuc

d

Tail Moment

Cntrl-hiPSC1
Nuc-hiPSC1
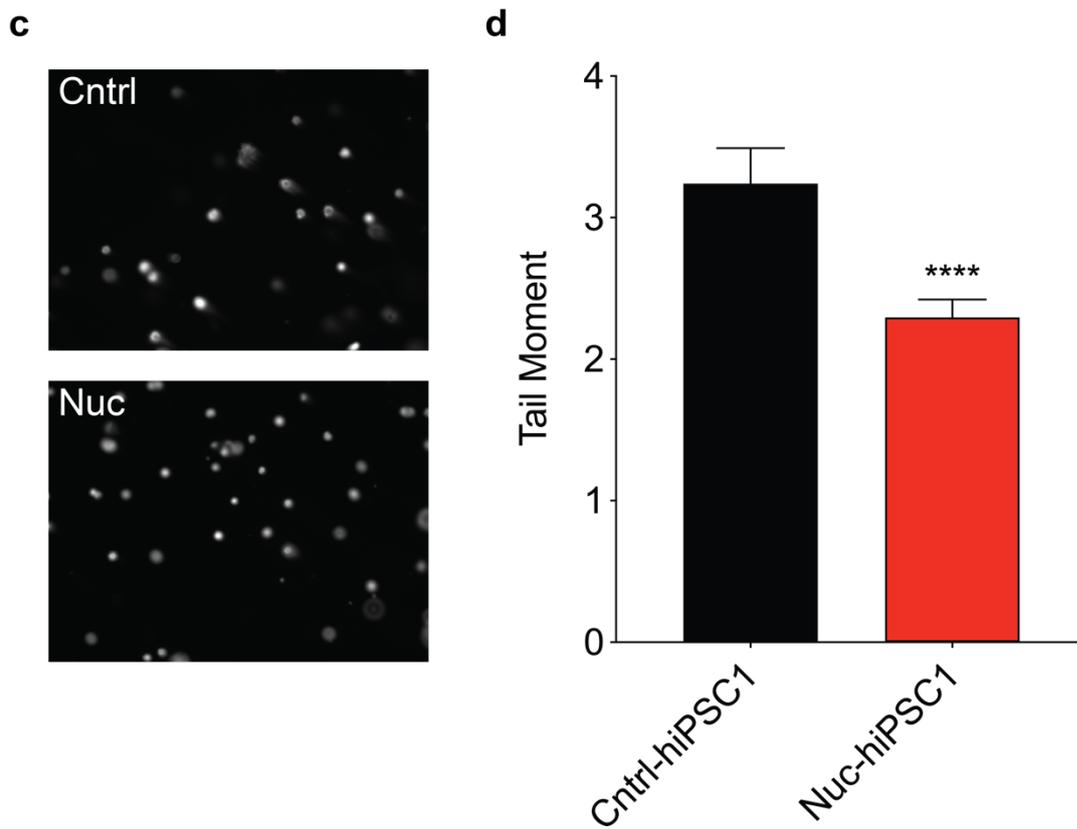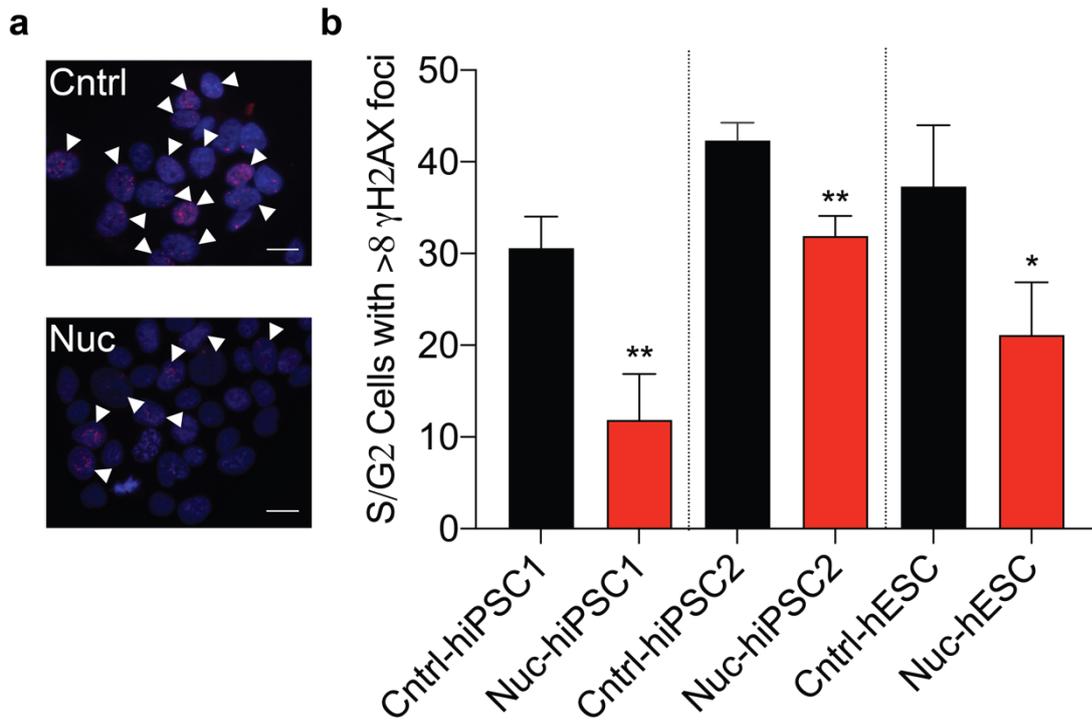
6.2.4 Exogenous nucleosides alleviate perturbed replication dynamics in human PSC

A common source of DNA damage during S phase of the cell cycle is replication stress that is caused by the slowing, stalling or collapse of replication forks (Ichijima et al., 2010). As exogenous nucleosides reduced the susceptibility of human PSC to DNA damage, it was next investigated whether this was a result of decreased replication stress by performing fibre assays to monitor replication dynamics.

The DNA fibre assay was performed on the hiPSC1 line in control conditions. The majority of fibres observed were small and their length was consistent with what we found during our experimentation in section 5.2.4 (**Figure 6.6a,b**). However, when the same cells were grown with the addition of exogenous nucleosides, the length of the DNA fibres increased (**Figure 6.6a,b**). Interestingly, upon the addition of exogenous nucleosides, the average fibre length was 47.9kb which was comparable to the fibres measured in the isogenic somatic cell lines which measured 45.5 and 35.5kb for the parent fibroblast and the hiPSC1 differentiated lines respectively (see section 5.2.4).

The fibre lengths now matched more closely the lengths measured in the parent fibroblast. When cells were grown in cell culture media supplemented with exogenous nucleosides the distribution of replication fork rates increased (**Figure 6.6c**). Further, the mean speed of replication fork rate was significantly higher in the presence of exogenous nucleosides measuring 1.2kb/min. Again, this was now akin to the fork rates observed in the isogenic parent fibroblast and hiPSC1 differentiated cell lines, which measured 1.2 and 0.9kb/min respectively (**Figure 6.6d**).
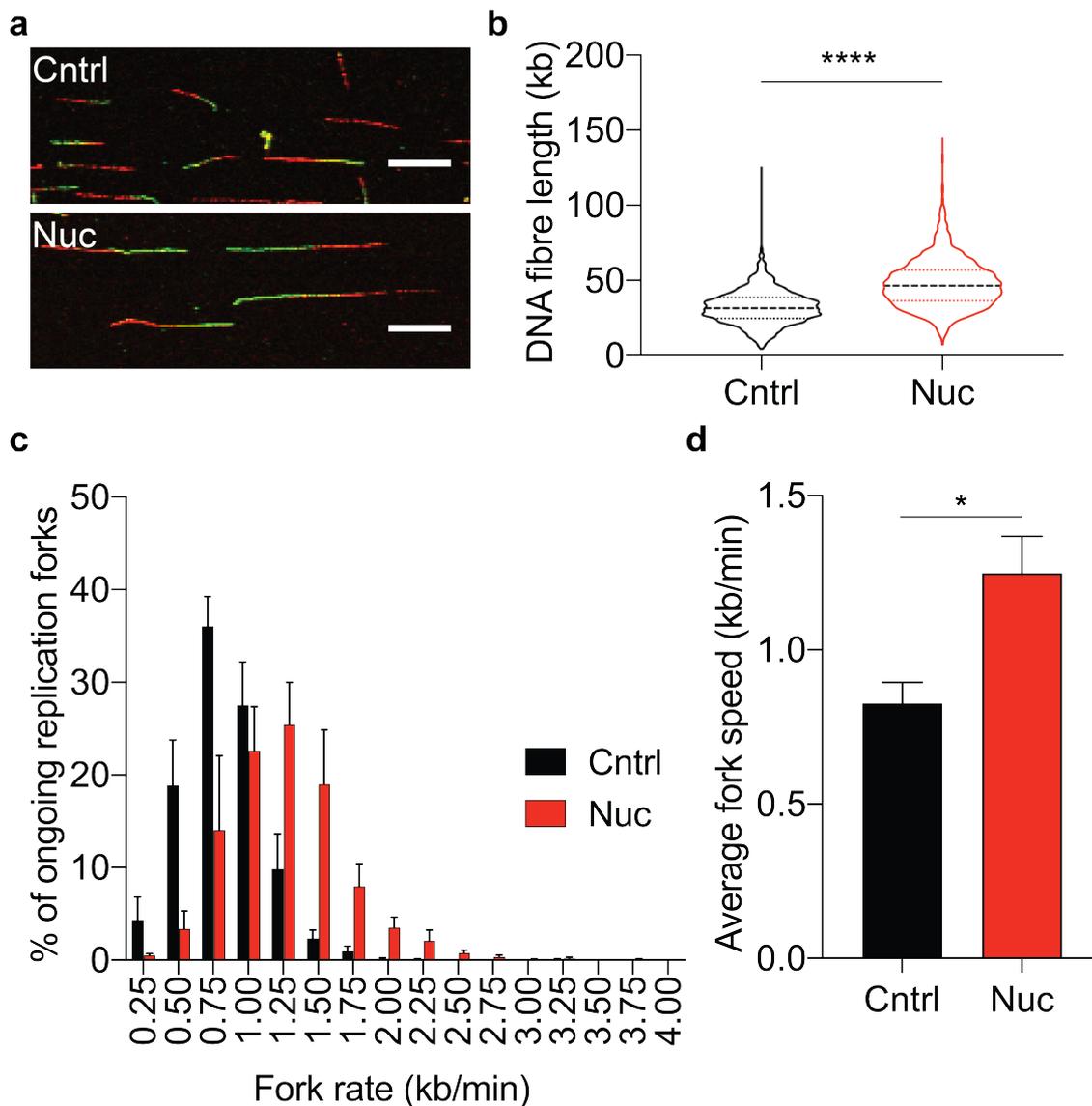
Figure 6.6 | Exogenously supplied nucleosides minimise replication stress in human PSC. a, Representative images of labelled DNA fibres from the hiPSC1 cell line grown in control (top) and nucleoside (bottom) conditions. Scale bar, 10μM. **b,** Quantification of the combined CldU and IdU DNA fibre lengths in hiPSC1 grown in control (black) or nucleoside conditions (red) (n > 200 fibres per experiment, n = 3 experiments). Violin plots of the distribution of fibre lengths, median distance and the 25th and 75th percentile are presented. **c,** Distribution of replication fork rates in control (black) and nucleoside (red) conditions (n > 200 fibres per experiment, n = 3 experiments). Data is mean ±s.e.m. **d,** Average fork rate of data from (**c**) mean ±s.d. (n = 3 experiments). two-tailed t-test, *P<0.05, **P<0.01, ***P<0.001, ****P<0.0001.

Dormant replication origins are fired in response to slow replication fork speeds to complete replication in under replicated regions (Anglana et al., 2003, Ge et al., 2007). Consistent with a reduction in replication stress, the distance between replication origins increased which is constant with a decrease in the replication origin density (**Figure 6.7a**). In line with this finding, the frequency of the red only CldU tracts that arise from stalled or collapsed forks before pulse labelling with the green IdU label also decreased (**Figure 6.7b**).
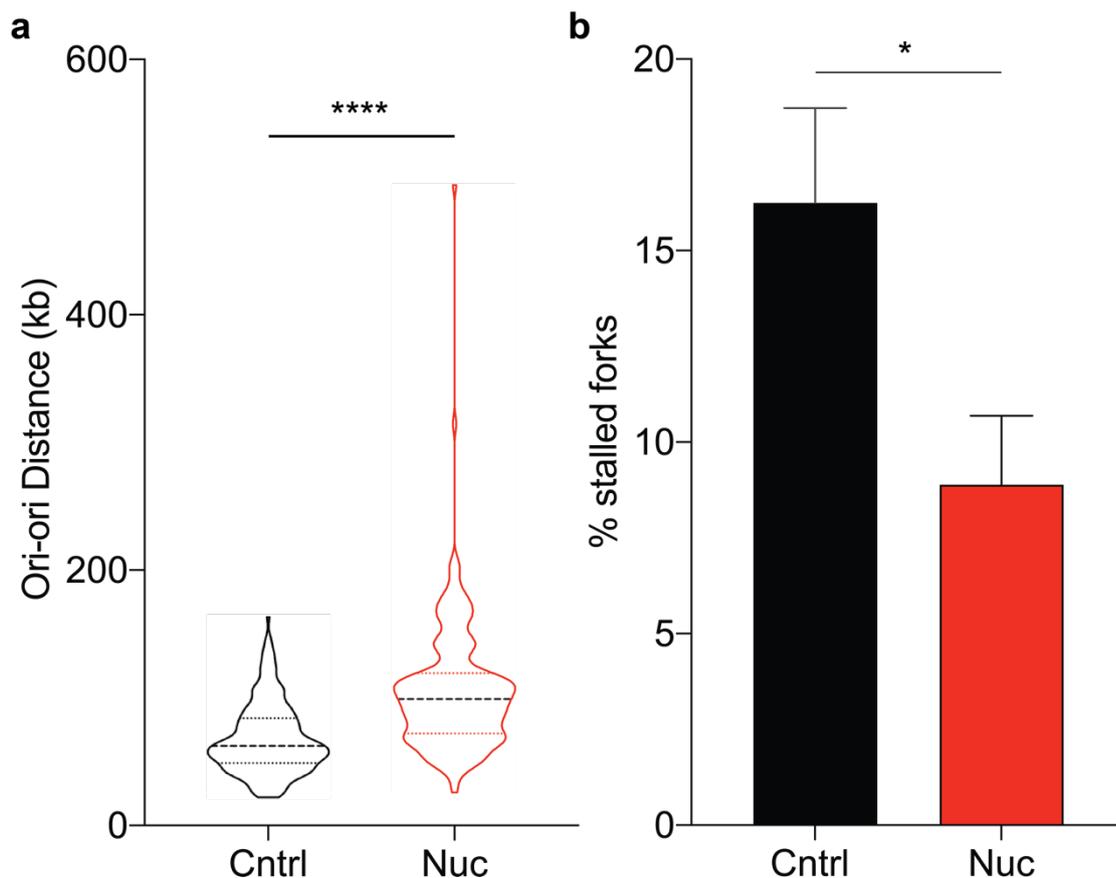
Figure 6.7 | Exogenous nucleosides decrease origin density and reduces frequency of stalled forks. a, Origin density determined from measurements between replication origins (ori-ori) in hiPSC1 cell line. Median distance and the 25th and 75th percentile are presented ($n > 150$ measurements per cell line, $n = 3$ experiments). **e,** Frequency of CldU only tracts in hiPSC1 that denotes stalled forks mean $\pm$s.d. ($n > 700$ per condition per experiment, $n = 3$ experiments). two-tailed t-test, *P<0.05, **P<0.01, ***P<0.001, ****P<0.0001.

Further, analysis using the hiPSC2 and hESC cell lines substantiated these findings. The analysis showed that hiPSC2 and hESC grown with exogenous nucleosides had a distribution of replication fork rates with greater speeds and an overall increase in the mean replication fork rate (**Figure 6.8a-d**). In addition, both hiPSC2 and hESC presented an increased distance between origins of replication in the presence of exogenous nucleosides compared to when the same cells were grown in control conditions (**Figure 6.8e,f**).

The addition of exogenous nucleosides to the cell culture media improved the replication dynamics of human PSC. The improvements made with exogenous nucleosides enhanced replication dynamics to levels that were previously observed in the isogenic somatic cell states.
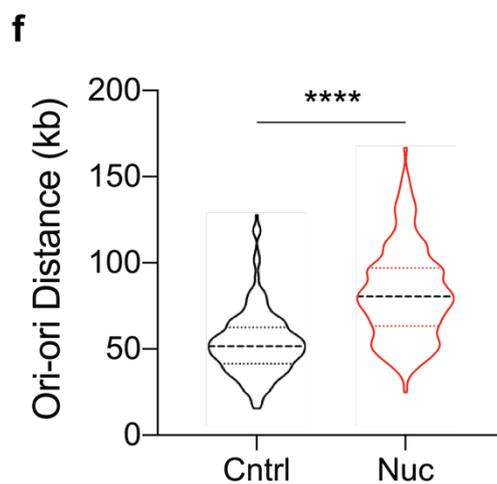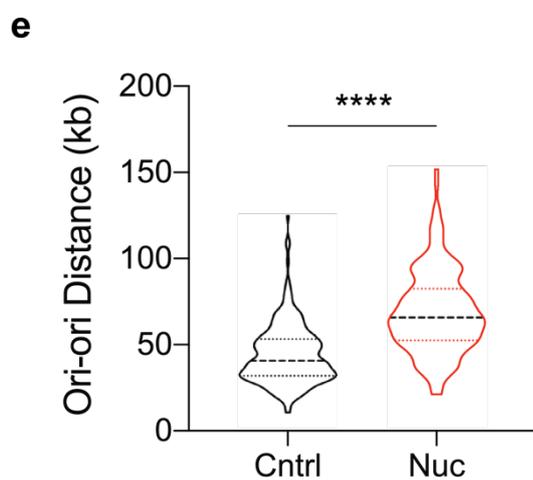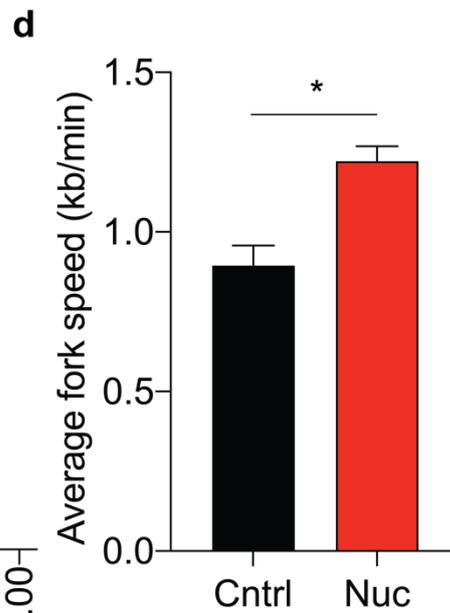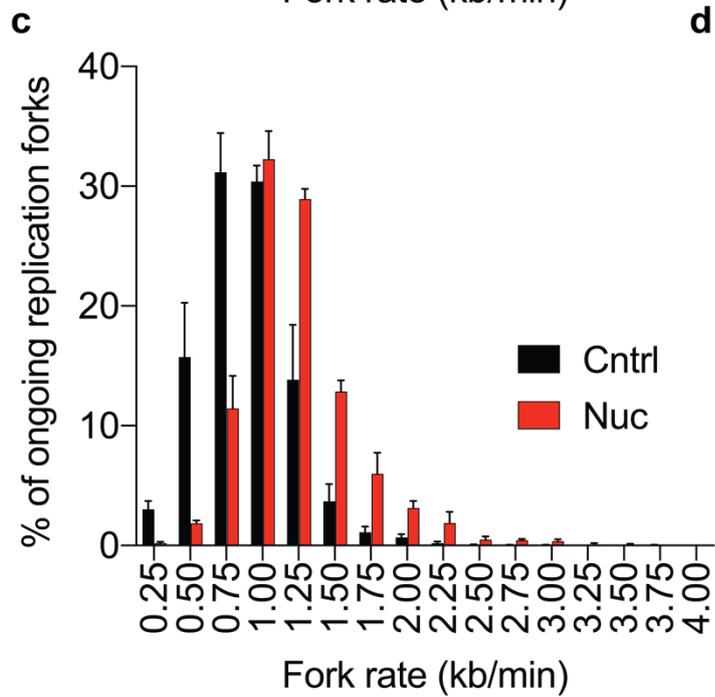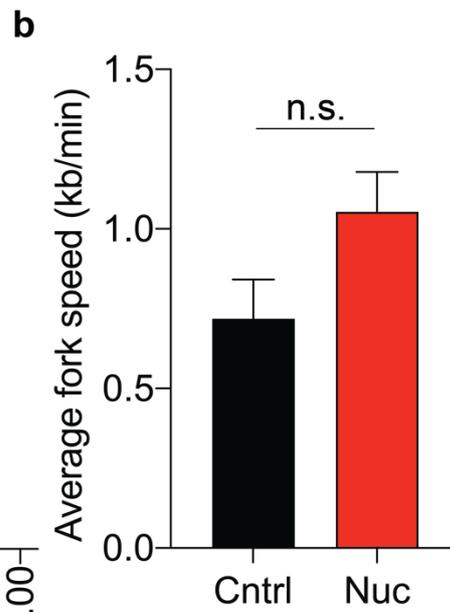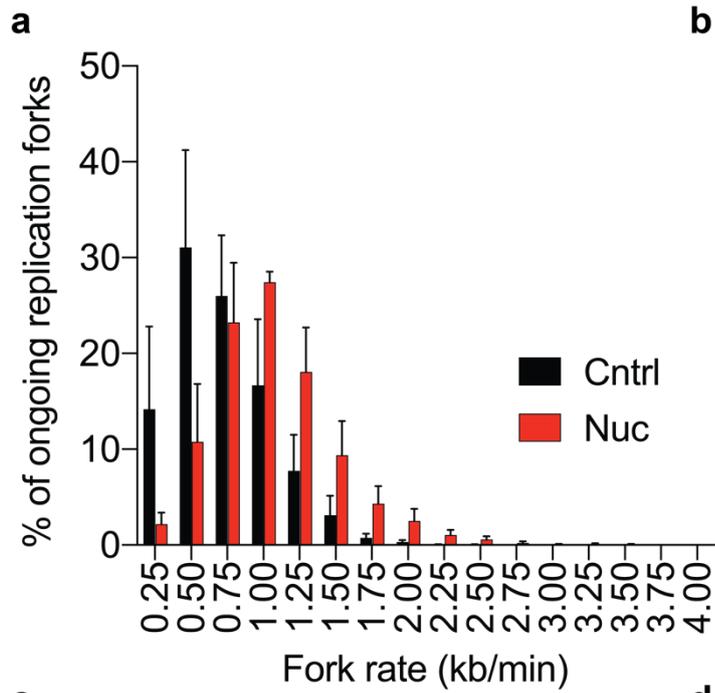
Figure 6.8 | Nucleosides effect on replication dynamics was confirmed in two further human PSC cell lines. a,c, Distribution of replication fork rates in control (black) and nucleoside (red) conditions in the hiPSC2 (**a**) and hESC (**c**) cell line (n > 200 fibres per experiment, n = 3 experiments) Data is mean ±s.e.m. **b,d,** Average fork rate of the data from (**a**) and (**c**) respectively. Data from 3 independent experiments are displayed. Mean ±s.d. **e,f,** Origin density determined from measurements between replication origins (ori-ori) in hiPSC1 cell line. Median distance and the 25th and 75th percentile are presented (n > 160 measurements per cell line, n = 3 experiments). two-tailed t-test, *P<0.05, **P<0.01, ***P<0.001, ****P<0.0001.

6.2.5 Exogenous nucleosides reduce errors during mitosis

The persistence of under replicated regions into mitosis can result in mitotic errors. The impairment of chromosome segregation can result in lagging chromosomes and chromosome bridges (Ichijima et al., 2010, Gisselsson, 2008). If a chromosome or part of a chromosome is not separated efficiently it can become compartmentalised into a micronucleus (Ford et al., 1988).

To understand the effect exogenous nucleosides have on mitosis in human PSC, the hiPSC1 cell line was stably transfected with H2B-RFP to fluorescently label the cell's chromatin and allows the tracking of chromosome dynamics in live cells. Images were acquired every minute for 2 hours using time-lapse microscopy and the length of prometaphase, metaphase and anaphase were measured when hiPSC1-H2B-RFP was grown in exogenous nucleosides or the control condition. An extension of the time spent in these phases may indicate mitotic checkpoint activation that slows or stalls transition from prometaphase to anaphase. Overall, the time spent in each of these phases was similar in both conditions although the time in anaphase was significantly shorter in hiPSC1 grown in the presence of exogenous nucleosides (**Figure 6.9a**).

Using the time-lapse analysis, the frequency of lagging chromosomes, chromosome bridges and micronuclei were determined (**Figure 6.9b-d**). Consistent with previous findings, a high frequency of mitotic errors in human PSC was observed under

standard culture conditions (Zhang et al., 2019) (**Figure 6.9e**). Strikingly, a significant decrease in the frequency of mitotic errors was observed with the addition of exogenous nucleosides over four independent experiments (**Figure 6.9e**). The addition of exogenous nucleosides to human PSC not only improves the localised impact of replication stress but also minimises its detrimental consequences on mitosis.

Figure 6.9 | Exogenous nucleosides reduce the frequency of mitotic errors in human PSC. The hiPSC1 cell line was fluorescently labelled with H2B-RFP to monitor mitosis. **a,** Violin plots of the distribution of prometaphase, metaphase and anaphase mitotic phase times. hiPSC1 cell line was grown in control (black) and nucleoside (red) conditions. **b-d,** Representative images of mitotic errors observed in hiPSC1 line labelled with H2B-RFP. White arrow head shows the mitotic error. Scale bar, 10μM. **e,** Average frequency of mitotic errors observed in control (black) and nucleoside (red) conditions. Data are mean ±s.d. two-tailed *t*-test, *P<0.05, **P<0.01, ***P<0.001, ****P<0.0001.

a

b Multipolar mitosis

c Lagging chromosome

d Chromosome bridge

e
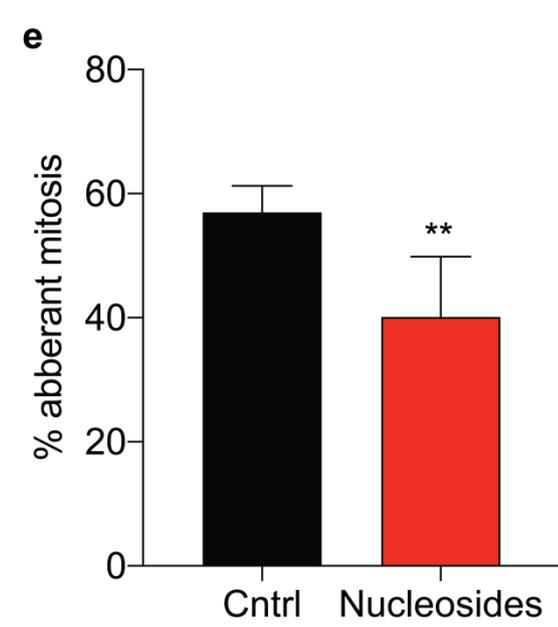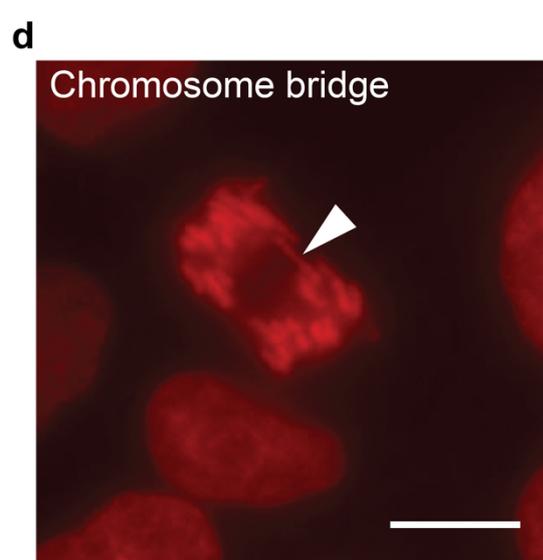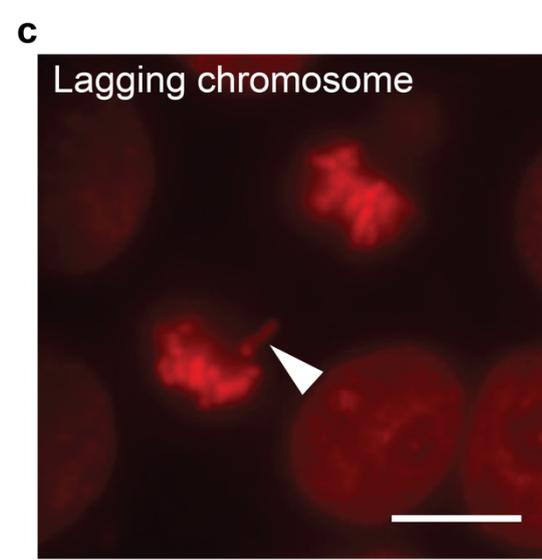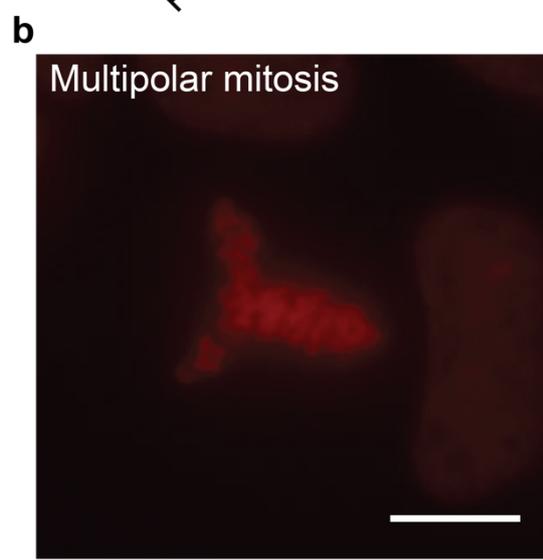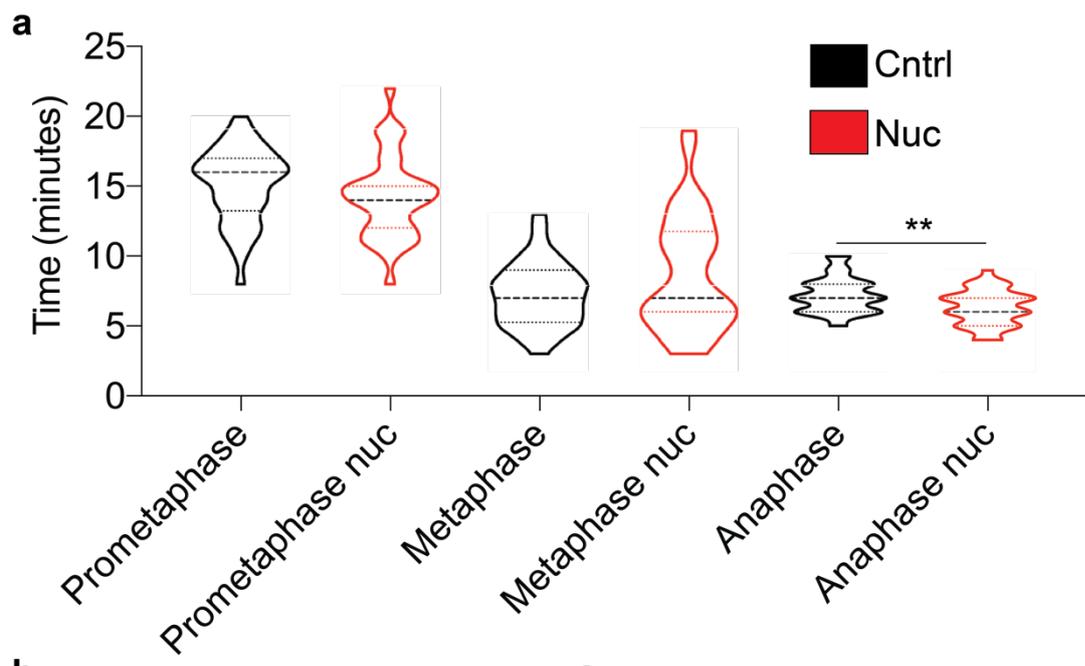
6.2.6 Exogenous nucleosides improve the growth dynamics of human PSC cultures

Unlike somatic cells, human PSC do not activate CHK1 in response to replication stress (Desmarais et al., 2012). In the absence of this checkpoint, the cells respond to replication stress by activating apoptosis instead of repairing stalled and collapsed forks.

To understand whether the beneficial effects of exogenous nucleosides on DNA damage, replication dynamics and mitosis also improved cell growth, I performed experiments to analyse growth rate. Cells were seeded at equal density into multiple 96 well plates and grown in control or nucleoside conditions for 96 hours. After 24-48 hours the colonies in the nucleoside condition appeared more condensed and rounded when compared to the spread out and spikey morphology of the control condition (**Figure 6.10a,b**). Each day a cell count was performed by fixing one plate from each condition and counting cell numbers based on nuclear DAPI stain. Nucleosides improved the growth rate of human PSC with a particular advantage during initial seeding and survival over the first 24 hours (**Figure 6.10c**).
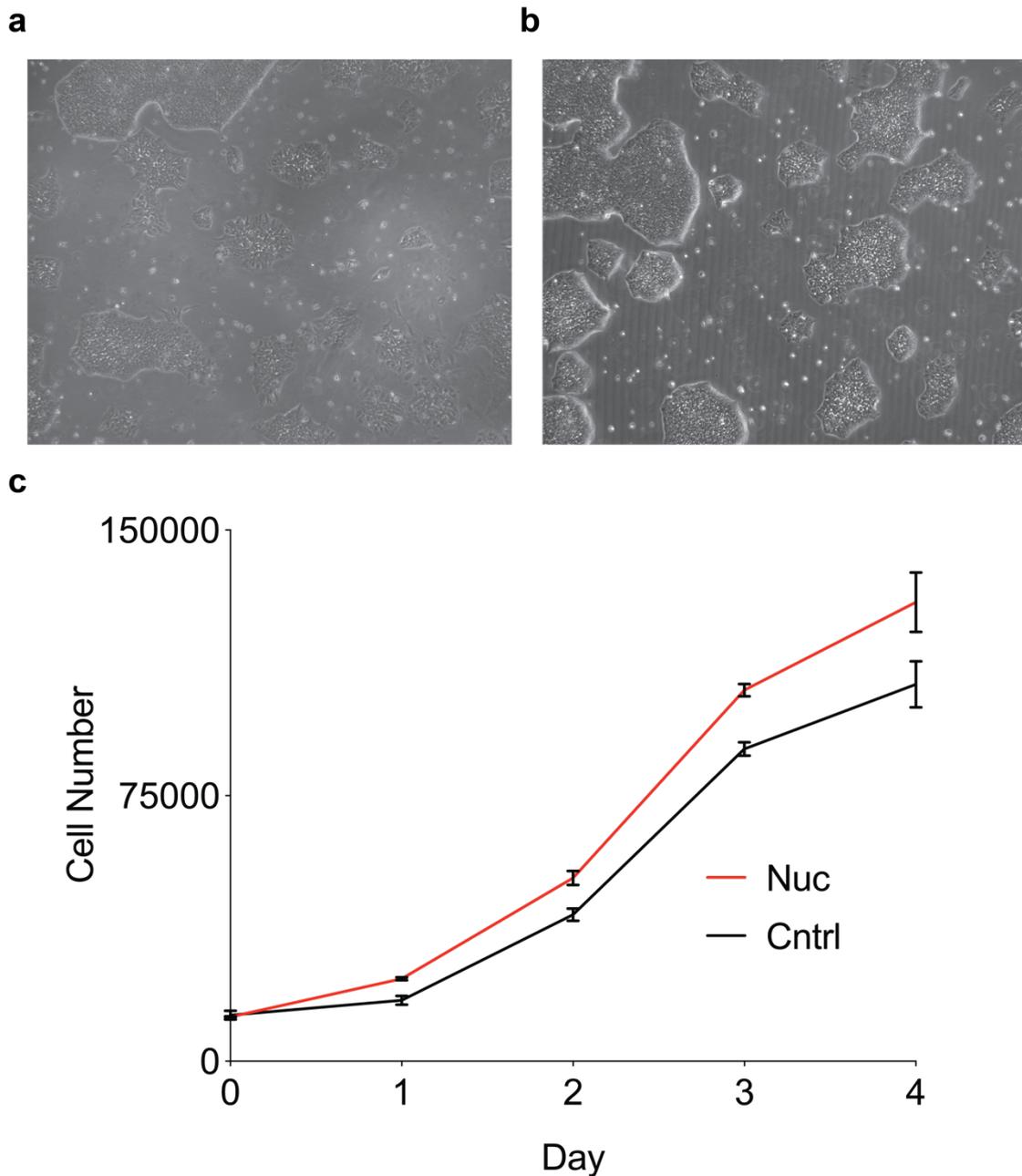
Figure 6.10 | Exogenous nucleosides improve the growth kinetics of human PSC. a,b, Representative brightfield images of hiPSC1 cell line grown in control (**a**) and nucleoside conditions (**b**). Images taken at 4X magnification. **c,** Assessment of growth rate over 96 hours determined from cell counts based on DAPI stained nuclei. Data is mean ±s.d. of three technical replicates.

To determine if the initial improvement in seeding was due to a reduction in apoptosis, hiPSC1 cells were seeded in a multi-well plate in control or nucleoside conditions. The cells were harvested after 3, 6 or 24 hours and immunostained for cleaved caspase 3, a measure of early apoptosis. After 3 hours, 27% of the cells seeded into the nucleoside condition were positive for cleaved caspase 3, which was less than the control condition which had a 39% cleaved caspase 3 positive population (**Figure 6.11a,d**). At the subsequent time points, there was no difference in the proportions of cleaved caspase 3 positive cells (**Figure 6.11b-d**). These data indicate that exogenous nucleosides improved human PSC growth, with a particular survival advantage achieved through improved survival post-plating.

Figure 6.11 | Exogenous nucleosides reduces apoptosis during initial cell seeding. a-c, Flow cytometry histograms showing the frequency of cleaved caspase 3 positive hiPSC1 cells grown in control (black) or nucleoside (red) conditions, 3 (**a**), 6 (**b**) or 24 (**c**) hours after seeding. **d,** Summary data of **a-d** of the percentage of cleaved caspase 3 positive cells in each condition.
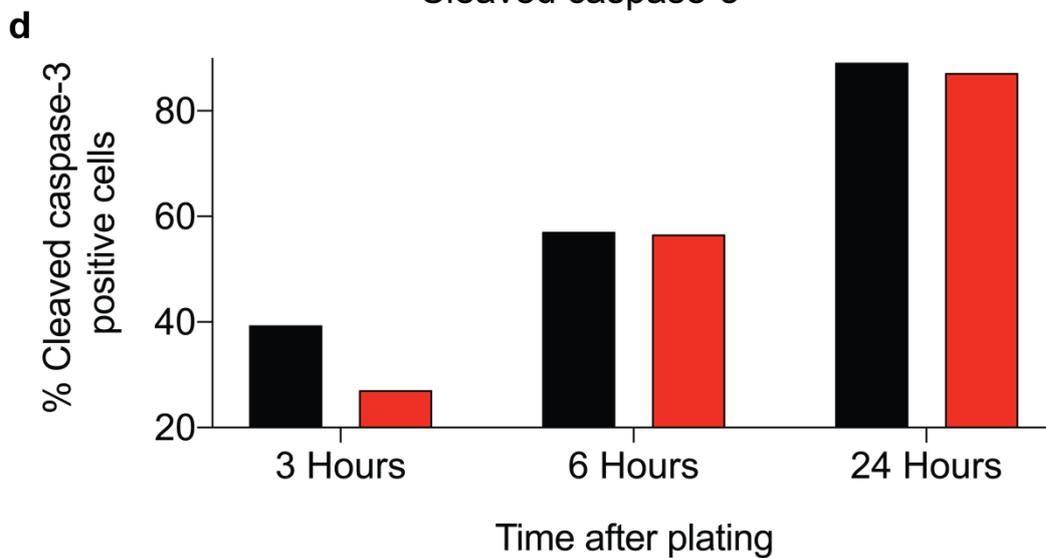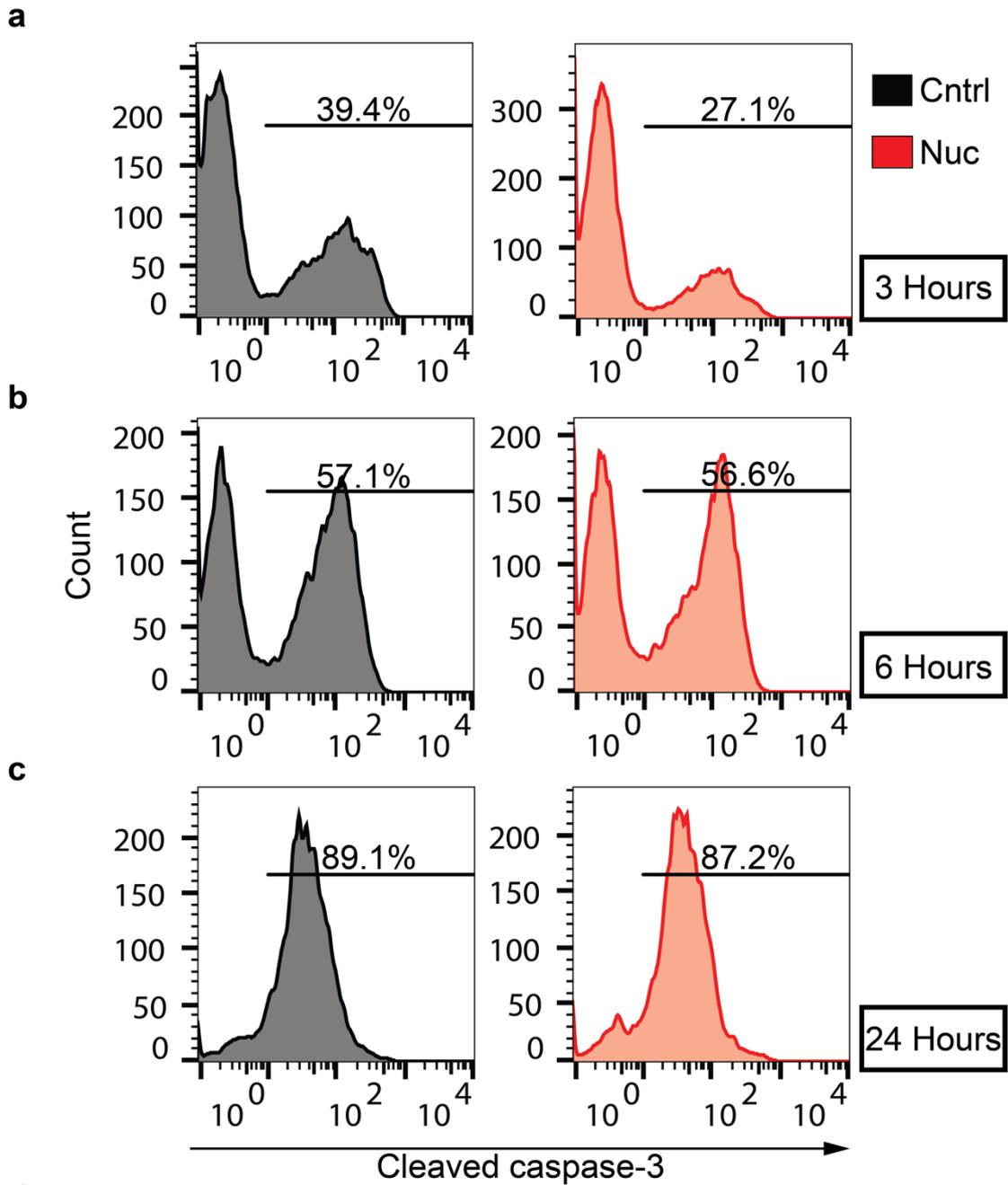
6.2.7 Exogenous nucleosides improved the cloning efficiency of human PSC

Plating human PSC at higher densities improves cell viability. The passaging of single cells at low density is associated with poor survival and thought to play a role in the selection of genetically variant cells that possess a survival advantage (Barbaric et al., 2014).

To investigate the effect of exogenous nucleosides on the survival of human PSC at low density, clonogenic assays were performed. To ensure that any colonies that arose at the end of the experiment were derived from a single cell, the hiPSC1 line was dissociated and plated at 500 cells/cm^2, a density at which cell-cell distance is larger than the maximum distance human PSC can migrate (Barbaric et al., 2014). The cells were grown for 5 days and the resulting colonies were stained for the pluripotency transcription factor OCT4. The cloning efficiency was calculated as the percentage of OCT4 positive colonies as a proportion of the initial seeding cell number. The cloning efficiency improved approximately two-fold across three independent experiments in the hiPSC1 cell line grown in the presence of exogenous nucleosides (**Figure 6.12a**). The experiment was repeated using the hiPSC2 cell line and again, exogenous nucleosides improved the cloning efficiency by approximately two-fold (**Figure 6.12b**).

Figure 6.12 | Exogenous nucleosides increase the cloning efficiency of two human PSC lines. a,b, hiPSC1 and hiPSC2 respectively show increased cloning efficiency when plated in exogenous nucleosides (red) when compared to the control conditions (black). Data are mean ±s.d. of three technical replicates, two-tailed *t*-test, *P<0.05, **P<0.01, ***P<0.001, ****P<0.0001.

a

b
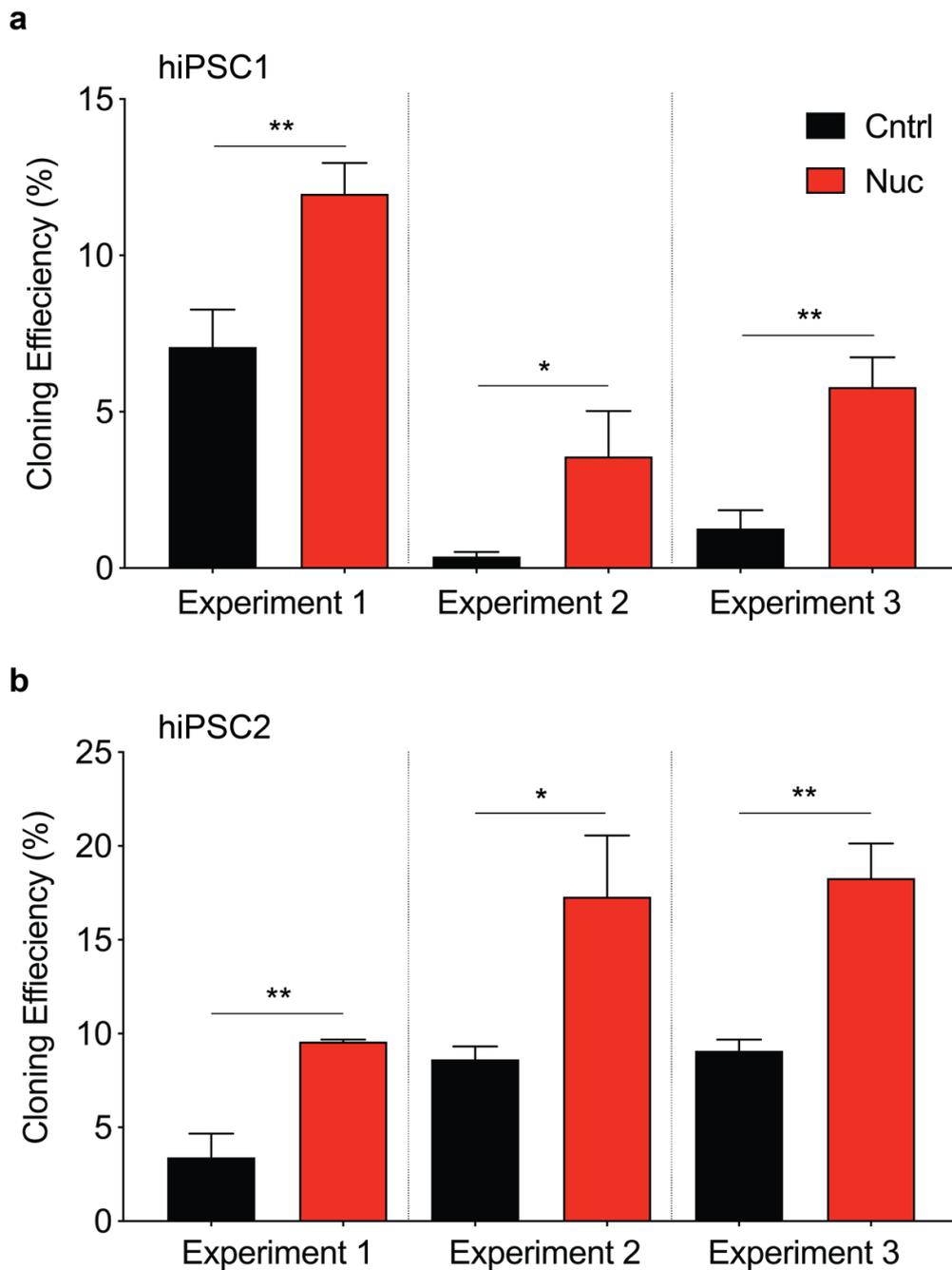
6.2.8 Exogenous nucleosides improved survival is consistent with decreased replication stress and mitotic catastrophe

To better understand the mechanism through which exogenous nucleosides improves cell survival at low density we performed a time-lapse microscopy experiment to track the growth of single cells through successive cell divisions. A random selection of 75 hiPSC1 cells from control or nucleosides conditions were chosen if they attached and performed a first division. The growth of these single cells was then tracked and plotted as lineage trees (**Figure 6.13a,b**).

a

Cntrl

Time (Hours)
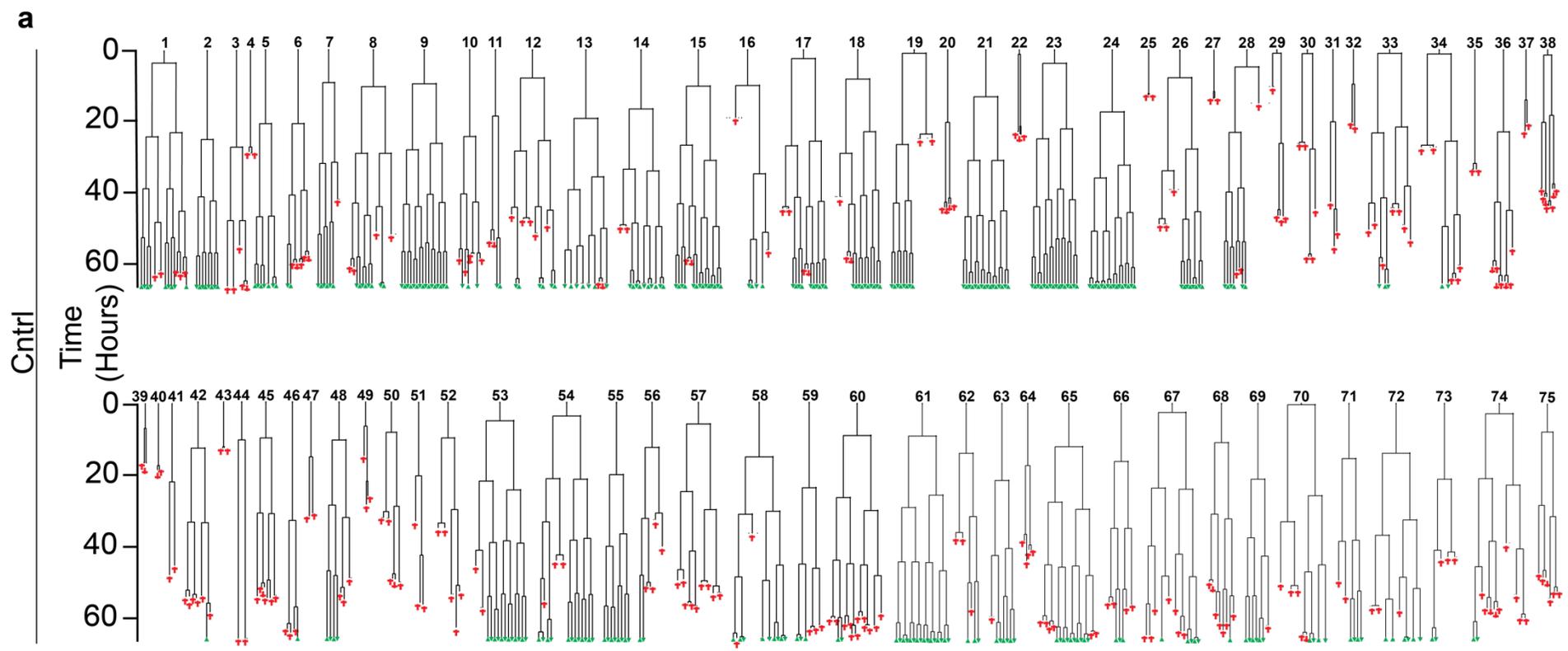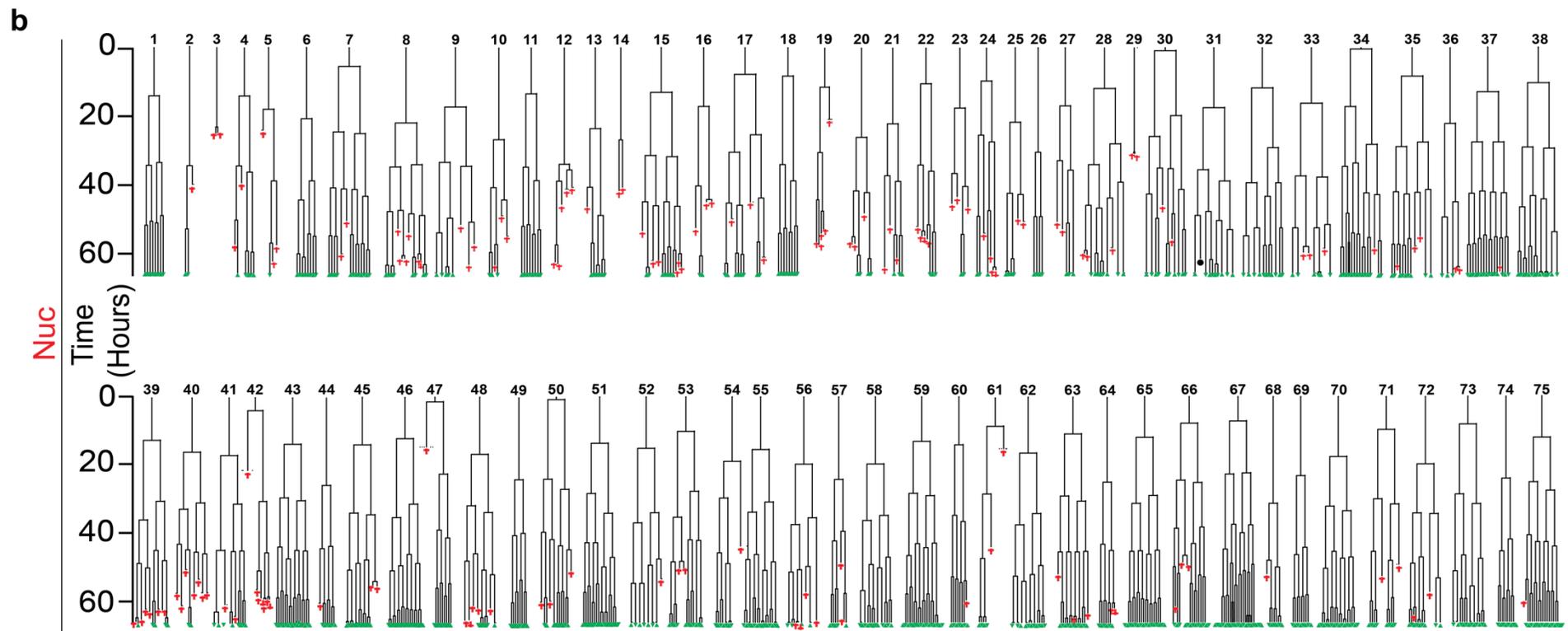
Figure 6.13 | Exogenous nucleosides improve survival post plating and following successive divisions. a, Lineage trees tracing 75 randomly selected cells that reach the first division in control conditions **b,** Lineage trees of 75 single hiPSC1 cells grown in nucleoside conditions that reach the first division. Time in hours is shown down the y-axis, forks in the lineage tree indicate a cell division. Green triangles show the surviving cell at the end of the timelapse experiment and the red crosses indicate where a cell has died.

Of the cells that seeded and attached, 68% survived beyond the first division in the control condition. When the cells were grown with exogenous nucleosides, 79% attached and entered mitosis post plating (**Figure 6.14a**). It has been previously shown that post plating, the majority of surviving cells that reach the first division are those from the G2 fraction of the cell cycle (Barbaric et al., 2014). The time from plating to the first division increased from 13 hours to 16 hours when nucleosides were added to the cell culture media, indicating improved survival of cells from earlier cell cycle phases (**Figure 6.14b**). This is consistent with previous observations that human PSC activate apoptosis in response to replication stress (Desmarais et al., 2012). When the time from the first to second division was measured, the time decreased from 22 hours to 19 hours when exogenous nucleosides were added to the cell culture media (**Figure 6.14c**). Next, the lineage tree analysis was used to examine the capacity of the cells to form colonies. When the cells were grown in exogenous nucleosides they were more likely to form a colony (**Figure 6.13a,b and 6.14d**). The overall number of cells in each colony was also greater (**Figure 6.13a,b and 6.14e**). Far fewer of those in the control condition consisted of greater than 8 cells and the greatest proportion colonies contained between 1 and 4 cells when compared to the nucleoside condition (**Figure 6.13a,b and 6.14f**). Finally, the survival of the daughter cells from each cell division was determined. The outcome of each division had three possible scenarios, survival of both daughter cells (SS), survival of one daughter cells and the death of other (SD) or the death of both daughter cells (DD) (**Figure 6.14g**). In the absence of nucleosides, there was a consistently higher number of cell divisions that resulted in the death of both daughter cells, this result was anticipated and indicated that mitotic errors caused by DNA replication stress are catastrophic for both daughter cells (**Figure 6.14h**). Together these results indicate that nucleosides reduce replication stress in human PSC and this improves the growth dynamics of cells which is manifested through a pronounced survival of cells following seeding and cell division.

Figure 6.14 | Time-lapse microscopy of human PSC cultures supplemented with exogenous nucleosides improve survival of cells following plating and after successive divisions. a-h, Summary data of the lineage tress analysis shown in figure 6.12. **a,** Percentage of cells that attached and survived to the first division in control (black) and nucleoside (red) conditions. The growth dynamics of these surviving cells were plotted as lineage trees in figure 6.12. **b,** Box plot of the distribution of time in hours from attachment to first division in control (black) and nucleoside (red) condition. Box and whiskers denote the mean and the 5th and 95th percentiles. Symbols are those values that fall outside of the 5th and 95th percentile. **c,** Box plot of the distribution of time in hours from the first cellular division to the second in control (black) and nucleoside (red) condition. Box and whiskers denote the mean and the 5th and 95th percentiles. Symbols are those values that fall outside of the 5th and 95th percentile. **d,** The number of cells that formed a colony out of the randomly selected 75 cells traced in the time-lapse analysis. **e,** Box plot of final colony sizes after the 72 hour time-lapse experiment. A colony was classified as having 2 more cells. Box and whiskers denote the mean and the 5th and 95th percentiles. **f,** Histogram showing the distribution of colony sizes formed after the 72 hour time-lapse experiment in control (black) and nucleoside (red) conditions. **g,** Schematic illustrating the scoring method used in (**h**). Survival of both cells after cell division was scored as SS, the death of one daughter cell and the survival of the other was denoted as SD and the death of both daughter cells was scored as DD. **h,** The frequency of the first (left), second (middle) and third (right) cell division that resulted in the daughter cells SS, SD or DD. two-tailed *t*-test, *P<0.05, **P<0.01, ***P<0.001, ****P<0.0001.
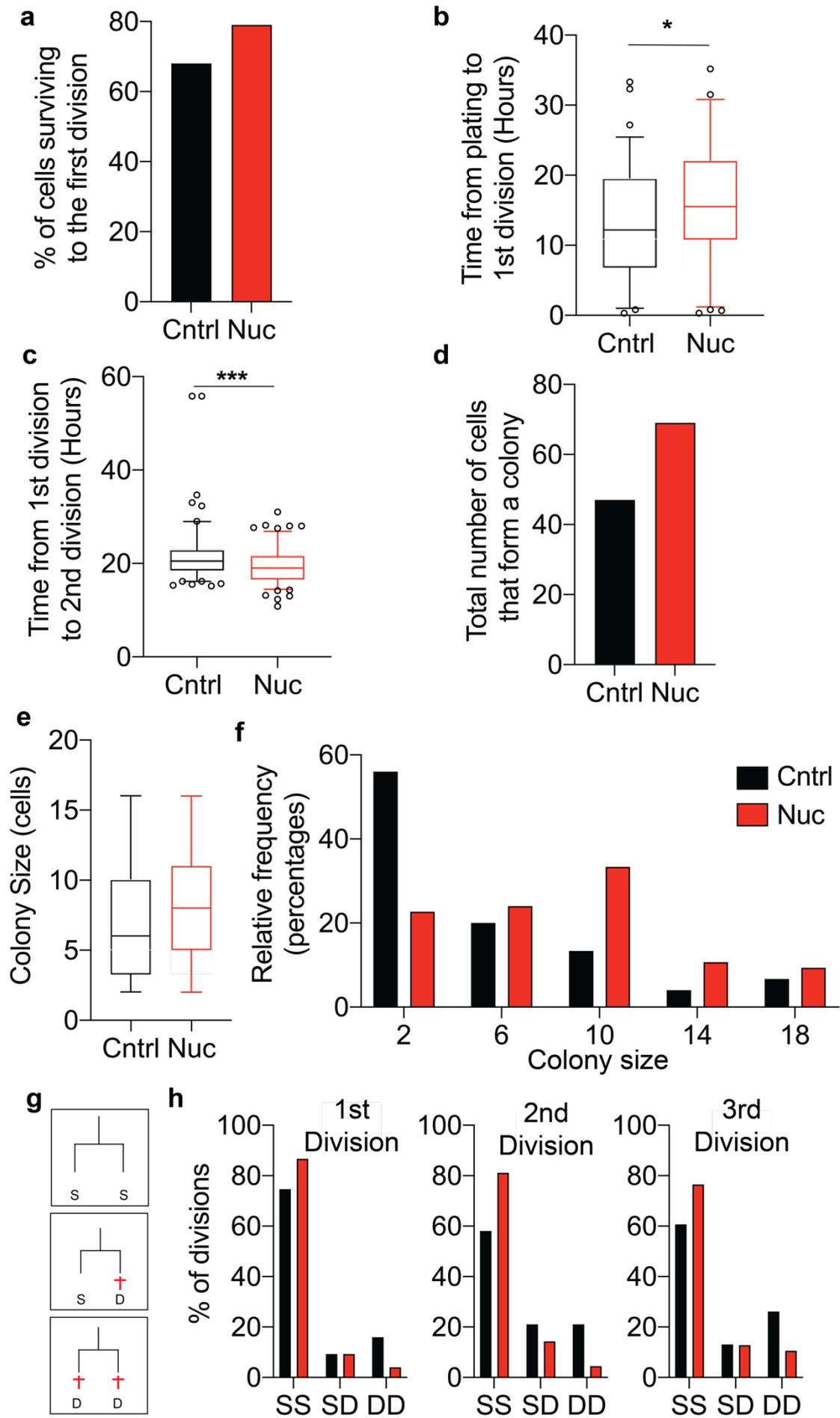
6.3 Discussion

The maintenance of genetic stability in human PSC is imperative for the safe future application of these cells in regenerative medicine. However, previous publications have observed a high frequency of DNA damage in these cells (Vallabhaneni et al., 2018, Simara et al., 2017). Errors in DNA repair and the errors caused by DNA damage during replication can lead to chromosomal instability, like those reported to afflict human PSC during prolonged culture (Amps et al., 2011, Seth et al., 2011). In the previous chapter, I confirmed a susceptibility to DNA damage in the pluripotent state and characterised replication stress as a probable cause that is likely due to uncoordinated entry into S phase. It has previously been shown that an uncoordinated entry into S phase can cause depletion of dNTP pools and subsequent DNA replication stress, DNA damage and genome instability (Bester et al., 2011). Reports elsewhere have shown that DNA replication stress can be alleviated with the addition of exogenous nucleosides which can also reduce mitotic errors that lead to chromosomal instability (Bester et al., 2011, Burrell et al., 2013).

The addition of exogenous nucleosides to the cell culture media of human PSC substantially improved replication dynamics and with this, decreased the detrimental consequences of replication stress including DNA damage, mitotic error and apoptosis. I found nucleosides decreased DNA damage in human PSC when measured by two statistically independent assays. In the previous chapter, distinct differences in the replication dynamics of cells in pluripotent and somatic states were observed. Using fibre assay measurements, I found exogenous nucleosides completely restored DNA fibre lengths, fork speeds and origin density to measurements made in the somatic cell lines. Of these results, a decrease in origin density was most interesting and has led to a potential mechanism for dNTP pool depletion in human PSC. These results would suggest increased origin density in human PSC is a response to depleted dNTP pools, with dormant origins fired to replicate regions around collapsed and stalled forks (Ge et al., 2007). Instead, I propose that replication stress in human PSC fits with the model where uncoordinated entry into S phase forces cell proliferation without the adequate dNTP pools required for DNA replication (Bester et al., 2011).

Supplementing chromosomally unstable cancer cell lines with nucleosides reduced both DNA damage and segregation errors during mitosis that supported a role for replication stress in driving the formation of structural and numerical chromosome instabilities (Burrell et al., 2013). Conventionally, studies have used metaphase spread analysis to measure the frequency of chromosomal instabilities that arise during in vitro experimentation. As particular genetic variants are selected for and enriched during the culture of human PSC, the frequency of chromosome changes may not reflect the mutation rate. For this reason, I chose to examine mitosis using fluorescently labelled histone H2B-RFP cell line and found exogenous nucleosides reduced the frequency of mitotic errors. However, the frequency of error was still high and would suggest that replication stress is only a contributing factor in the events leading to chromosomal instability. Overall, the frequency of mitotic errors under normal conditions was high and confirmed previous reports that human PSC are susceptible to mitotic errors (Zhang et al., 2019). By using the same approach, Zhang et al found that 30% of mitosis resulted in errors, although their analysis only included chromosome bridges and lagging chromosome, whereas I also included counts of multipolar divisions and micronuclei (Zhang et al., 2019). When time spent in each phase of mitosis was measured, the time to complete anaphase was significantly lengthened in the absence of nucleosides. This observation is consistent with the activation of the spindle assembly checkpoint mitotic arrest prior to cell death and highlighted further the ability of nucleosides to reduces mitotic error during human PSC culture (Masamsetti et al., 2019).

Unsurprisingly, the addition of exogenous nucleosides to cell culture media was able to maintain human PSC in the pluripotent state. However, when pluripotency-associated cell surface antigens were analysed, the proportion of positive cells increased with the addition of nucleosides, signifying an improvement in the survival of human PSC. Further, the embryoid bodies formed from the nucleoside condition were much larger and although this is outside of the scope of this project, it would suggest that nucleosides benefited differentiation of human PSC by reducing cell death. This is critical as regenerative medicine requires huge numbers of differentiated cells. Some application such as cardiac regeneration will require one billion or more cardiomyocytes to provide effective remuscularization of the heart (Chong et al., 2014).

Observations of colony morphology and examination of cells grown at low density provided context to how nucleosides improved cell survival. In particular, I found that the time from cell seeding to the first division increased. In a study by Barbaric et al, it was shown that genetically normal cells plated at low density had a greater chance of survival if they were in the G2 phase of the cell cycle (Barbaric et al., 2014). An extension in the time to the first division like I observed, indicates that human PSC in G1 and S phase were now more likely to survive. This is consistent with a reduction in replication stress-induced apoptosis, that is the default response of human PSC to replication stress as they do not activate CHK1 to stall the cell cycle (Desmarais et al., 2012). In cultures devoid of exogenous nucleosides, lineage tree analysis showed that cell death soon after cell division was prevalent. This may be explained by the high frequency of mitotic errors leading to mitotic catastrophe and cell death of both daughter cells in a TP53 dependant manner (Zhang et al., 2019). Overall, this led to an increase in colony numbers and greater colony size at the end of the experiment.

These data suggest that the susceptibility of human PSC to replication stress and DNA damage is entirely due to shortages in dNTP pools which can be rectified with the addition of exogenous nucleosides. To test this notion, a collaboration was initiated with Professor Baek Kim at (Emory University, Atlanta), who has a system by which individual dNTP concentrations from cell extracts can be measured by primer extension DNA synthesis using HIV-1 reverse transcriptase (Diamond et al., 2004). Surprisingly, the initial results showed that the dNTP pools were larger in the hiPSC1 cell line compared with the parent fibroblast from which it was reprogrammed. A likely explanation for this would be the difference in proportions of S phase cells in each of these cell states. In the previous chapter, I found that 68% of the hiPSC1 line was in the S phase of the cell cycle compared to only 21% in the parent fibroblast (see section 5.22). Cellular dNTP pools are accumulated late in G1 phase and throughout S phase, controlled by the expression of RRM2, a subunit of the rate-limiting RNR enzyme in the nucleotide biosynthesis reaction (Engström et al., 1985). Consistent with this hypothesis, I found that RRM2 gene and protein expression was higher in the hiPSC1 line than the parent fibroblast.

However, a further surprising result was that an increase in dNTP pools was not detected in cultures of human PSC grown in the presence of nucleosides. This might indicate a more complex relationship between endogenous and exogenous dNTP pools than first anticipated. However, unfortunately, because of a lack of time and opportunity for further collaboration with the Atlanta group, I was unable to explore this further and rule out the possibility that the unexpected result represented a technical problem with the assay or the preparation of samples. Considering my other results this could highlight an issue with the preparation of samples or the assay itself.

Collectively, these data suggest that dNTP pools are depleted from rapid proliferation and uncoordinated activation of nucleotide biosynthesis in human PSC. It seems likely that this is a feature of *in vitro* culture as endangering cells of the inner cell mass could be catastrophic. There is some evidence to suggest that *in vivo* these cells maintain rapid rates of proliferation by generating dNTPs on the go. A study using data from *Drosophila, Xenopus* and sea urchins showed that maternally supplied dNTPs are not sufficient to supply DNA replication up until the maternal to zygotic transition, that normally occurs around the mid-blastula transition and marks the period where development comes under zygotic control (Song et al., 2017). Instead, it's been proposed that leading up to the maternal to zygotic transition the depletion of maternal dNTPs allosterically upregulates the activity of RNR in the zygote, coinciding with the breakdown of the maternally loaded mRNA that provides the metabolites needed for nucleotide biosynthesis (Vastenhouw et al., 2019).

In summary, the culture of human PSC with exogenous nucleosides may better reflect the conditions of the early embryo. The observation presented here show that nucleosides reduced DNA replication stress and alleviated the stress-induced damage and mitotic errors and will provide a means to reduce the appearance of recurrent genetic changes that will otherwise threaten the safe exploitation of human PSC in regenerative medicine.

7 Discussion

Ever since the discovery that human PSC could be maintained under *in vitro* conditions, their promise in the field of cell-based regenerative medicine has been touted. However, human PSC acquire genetic changes upon prolonged culture which may preclude their use in regenerative medicine applications based on safety (Amps et al., 2011, Seth et al., 2011). The acquisition of genetic changes affects chromosomes in a recurrent and non-random nature (Draper et al., 2004, Olariu et al., 2010). The recurrence of these genetic changes suggests that acquisition requires two steps, first a random mutation and second, selection of that mutation should it provide the variant cell with a growth advantage. It is now understood that selection provides the mechanism responsible for the recurrent nature of these variants, yet the origins and mechanisms that are responsible for these mutations remain elusive. The successful translation of human PSC derived therapy will require strategies for minimising the appearance of mutations, so far this has been hindered by the absence of a comprehensive understanding of the origins of mutation in cultures of human PSC.

In this body of work, significant evidence as to the origins of mutation in cultures of human PSC has been provided. The data presented here is the first conclusive evidence that human PSC are susceptible to replication stress and DNA damage during *in vitro* culture. In addition to the mechanistic insights to the origins of mutation, data has also been presented on the development of improved culture conditions that better replicates the early embryonic environment and will facilitate the culture of genetically stable human PSC for future regenerative medicine applications.

7.1 The origins of mutation in human PSC

To determine the origins of mutation in human PSC, it was first necessary to focus the study on one of the recurrent karyotypic changes. Of all the recurrent changes, the chromosome 20 variant is the best characterised. Additionally, it frequently presents as tandem duplication that always includes the 20q11.21 region, which suggests a common mechanism of mutation.

To elucidate the potential origins of mutation, breakpoint sequencing was performed on a chromosome 20 variant cell line which was used to infer the mechanism of mutation. However, it was first necessary to accurately detect candidate cell lines for sequencing. A battery of assays that are frequently employed during the detection of genetically variant human PSC were tested to interpret their limits of detection and in addition, the novel C-FISH assay was developed that improved the sensitivity of detecting chromosome 20 variant cell lines. Using the approaches that were tested and developed here, the mShef7 cell line was identified that had a mosaic population of chromosome 20 variant cells. The variant cells were cloned and subsequently sequenced using Oxford Nanopore long-read next-generation sequencing system. Bioinformatic analysis of this data elucidated the breakpoint sequence, which mapped to an *AluSz6* retrotransposon element. This discovery provided the first evidence that this tandem duplication could be arising as a result of problems with replication.

Early replicating fragile site locations are enriched for SINE repetitive elements that include the *Alu* retrotransposons (Barlow et al., 2013). These regions have been shown to cause recurrent genetic instability in B cells leading to B lymphoma that is potentially a consequence of B cells rapid proliferation upon their activation (Barlow et al., 2013). Interestingly, replication stress predisposes early replicating fragile sites to DNA double-strand breaks as these regions are vulnerable to fork collapse (Barlow et al., 2013). Depletion of dNTP pools, inhibition of ATR and atypical regulation of DNA repair by HR, have all been found to increase the fragility of these regions (Barlow et al., 2013). The *Alu* retrotransposons are classified as inverted DNA repeats. It's been suggested that that double-strand breaks from collapsed replication forks can be repaired by a subclass of HR mediated repair, known as break-induced replication. In yeast studies, the repair lesions by break-induced replication at sites of *Alu* elements resulted in a higher frequency of translocations and amplifications (Narayanan et al., 2006). Armed with this novel breakpoint data and its association with replication, human PSC susceptibility to replication stress was investigated as a mechanism for mutation during *in vitro* culture.

Like activated B cells and many cancer cell types, human PSC rapidly divide which may be exposing them to high endogenous levels of replicative stress. By performing comparative experiments on human PSC and isogenic somatic control lines, the cell cycle time in human PSC was found to be significantly shorter as a result of a truncated G1 phase. Silencing of the $p16^{INK4a}$ CDK4/6 inhibitor is necessary to reprogram cells to human iPSC and also during the maintenance of cells in the pluripotent state (Li et al., 2009a). The lack of $p16^{INK4a}$ suppression on CDK4/6 would explain the relaxed Rb-E2F pathway that exists in human PSC (Filipczyk et al., 2007) and as *cyclin E* is a transcriptional target of E2F, this would explain the elevated *cyclin E* expression reported in this study. Cyclin E is an oncogene and has been studied extensively to understand how it induces replication stress. Bester et al have shown that uncoordinated entry into S phase may allow replication to begin without the cell first building the necessary dNTP pools (Bester et al., 2011). Nucleotide pools are critical for normal DNA replication, depletion of dNTP pools has been shown to slow DNA replication, increase origin density and cause DNA damage, particularly at common fragile sites (Anglana et al., 2003, Saintigny et al., 2001, Ge et al., 2007, Yan et al., 1987). To investigate replication stress, DNA fibre assays were performed. It was discovered that the replication rate was slower and the density of replication origins increased in the pluripotent state compared with isogenic somatic lines. Additionally, these pluripotent cells also exhibited a susceptibility to DNA damage that was associated with S phase.

In conclusion, these data have shown that the chromosome 20 variant breakpoint mapped to an *Alu* retrotransposon sequence. *Alu* elements are highly enriched in early replicating fragile sites, which are prone to breakage as a result of replication stress and are frequently found at regions of recurrent chromosomal instabilities. Replication stress, leading to fork collapse is a common feature of rapidly proliferating cells. Human PSC display a rapid cell cycle, punctuated by a truncated G1 phase and symptoms of replication stress and DNA damage. These observations, I propose are the origin of mutation in human PSC that, in particular, lead to the recurrent tandem duplications observed to arise during prolonged culture of human PSC.

7.2 Alleviating the stress posed by DNA replication

The overlap between pluripotency and cancer cells is startling. Oncogenic activation of the Rb-E2F pathway can result in the low nucleotide pools from uncoordinated activation of the nucleotide biosynthesis pathway (Bester et al., 2011). The addition of nucleosides to cultures of cells overexpressing oncogenes decreased replication stress, DNA damage, chromosome segregation errors and tumorigenicity (Bester et al., 2011, Burrell et al., 2013). Likewise, in cultures of human PSC, the exogenous supply of nucleosides decreased replication stress, DNA damage and reduced the number of mitotic errors. In addition to these findings, the behaviour of human PSC grown in nucleosides was examined in comparison to standard culture conditions. It's been previously shown that the major bottlenecks that exert selective pressure on human PSC exist during passaging and restrict the survival of cells post-plating (Barbaric et al., 2014). Behavioural analysis showed that human PSC were less likely to die following plating as a result of replication stress or mitotic catastrophe when grown with the addition of exogenous nucleosides.

It has previously been shown that human PSC uniquely respond to replication stress. Unlike somatic cells, human PSC lack the CHK1 checkpoint which leads to apoptosis rather than cell cycle stalling and repair that could otherwise risk error and potential mutation (Desmarais et al., 2012). This is likely a characteristic of the early embryo to minimise the risk of mutations that could be catastrophic and result in pregnancy loss or congenital disease and surely explains why human PSC have a low mutation rate (Thompson et al., 2019). However, both this study and a previous study have shown that human PSC are frequently exposed to errors during mitosis (Zhang et al., 2019). Despite this, the actual frequency of chromosomal structural and numerical instabilities that arise from these events is surprisingly low (Amps et al., 2011, Merkle et al., 2017). These independent sets of data suggest that chromosomal instabilities that provide a selective advantage to the variant cell are rare and there are distinct mechanisms in play that eliminate cells with mitotic errors during mitosis. Interestingly, CHK1 has been demonstrated to play a role in unperturbed mitosis. Loss of CHK1 leads to chromosome misalignment and lagging chromosomes in human cells (Tang et al., 2006). Loss of CHK1 likely causes errors in mitosis due to its role in the localisation of Aurora B to the kinetochore, which is

required for accurate chromosome segregation (Peddibhotla et al., 2009, Zachos et al., 2007). Therefore, if the loss of CHK1 does not lead to apoptosis during S phase it may destabilize segregation during mitosis and lead an increase in mitotic errors. How then, would the human PSC selectively die in response to mitotic errors? This may be linked to the interplay between CHK1 and the spindle assembly checkpoint. It has been demonstrated that CHK1 depletion prevents the recruitment of the spindle assembly checkpoint proteins to aneuploid cells but not to diploid cells. As aneuploid cells are more likely have abnormal mitoses they are reliant on an intact spindle assembly checkpoint, which, if not present, can lead to the activation of p53 and cell death of the aneuploid cells in a spindle assembly checkpoint dependant manner (Vitale et al., 2007). As human PSC during culture acquire anti-apoptotic mutations, it may be these that desensitize the cell to both replication and mitotic stresses, enabling their survival and selective advantage in cultures (Zhang et al., 2019).

7.3 Future research direction

The amazing capacity for human PSC to endlessly proliferate and yet, seemingly be resistant to mutation has important implications in the study of a diverse number of other fields. The strength of using human PSC as a model in cancer biology and fertility should be considered. For example, the use of CHK1 inhibitors to selectively kill cancer cells has been reported on extensively, and yet, this mechanism is inherent to human PSC and likely maintains their stability. Further elucidation of the mechanisms that human PSC use to ensure that mutations rarely arise, despite being able to endlessly renew, may uncover new drug targets for the treatment of cancer.

Although it seems counterintuitive, insight now suggests that chromosome segregation errors regulate fertility in women of different ages (Gruhn et al., 2019). In women who are of reproductive age, 20-30% of oocytes are aneuploid, whereas up to 70% of oocytes are aneuploid in women of advanced maternal age which correlates with fertility in these age groups (Wang et al., 2017, Capalbo et al., 2017). Following fertilisation of the egg, the early embryo implants and the cells expand rapidly to produce the numbers necessary for gastrulation. Studies using mouse

embryos have shown that these cells are under extensive replication stress and are hypersensitive to DNA damage, undergoing apoptosis without cell cycle stalling (Laurent and Blasi, 2015, Heyer et al., 2000). It has been proposed that access to nutrients at this stage may regulate the expansion of these cells. Studies using Drosophila, Xenopus and sea urchins have shown that the early embryo does not have sufficient stockpiles of dNTPs to supply the cells of the early embryo during these rapid stages of proliferation (Song et al., 2017). Instead, it has been proposed that the embryo overcomes these shortages by benefitting from maternally supplied dNTPs and a de-novo capacity to synthesise new dNTPs (Liu et al., 2019). However, for the early embryo to synthesise new dNTPs they require the precursors. Around the same time, maternally loaded mRNA transcripts are broken down in a process known as the maternal to zygotic transition, releasing free nucleotides that can be recycled by the embryo (Vastenhouw et al., 2019). The consequences of not having adequate dNTPs can be catastrophic. Exhaustion of dNTP pools in the embryos of mice, Xenopus, Zebrafish and Drosophila led to severe cell cycle defects and arrest in mitosis which was followed by mitotic catastrophe (Liu et al., 2019, Song et al., 2017, Vastag et al., 2011, Newport and Dasso, 1989, Scott et al., 1971, Zhang et al., 2014, Zhang et al., 2008). These results show a profound overlap with the data presented in this study. The findings that culture media containing exogenous nucleosides benefited the growth of human PSC may reflect that these conditions better replicate the conditions of the early embryo. Future collaborations are required between the fields of embryogenesis and stem cell biology to capitalise on findings to not only improve the safe expansion of human PSC but also improve fertility with reproductive ageing that will better match the increasing human lifespan.

8 Conclusion

Recurrent genetic changes in cultures of human PSC may compromise their use in research and regenerative medicine. This study aimed to seek the origins of mutation in cultures of human PSC. By performing long-read next-generation sequencing on the chromosome tandem duplication of a variant human PSC line, the breakpoint region was mapped and localised to an *Alu* repetitive element. This is a novel discovery and up until now, breakpoint sequencing of variant human PSC has not been published. From this finding, it was possible to infer that these sites are particularly susceptible to breaks during replication stress. Early replication fragile sites that include repetitive elements such as *Alu* sequences are particularly vulnerable to the double-strand that results from dNTP pool starvation causing replication fork collapse (Barlow et al., 2013). From the analysis of replication dynamics in human PSC in comparison to somatic cell lines, it was discovered that not only was the breakpoint in a region susceptible to replication stress but that these cells were exposed to replication stress under routine culture and had extensive DNA damage that was associated with the S phase of the cells cycle.

Rapid proliferation in cancer cells can lead to replication stress from uncoordinated entry into S phase with inadequate dNTP pools. Like cancer, human PSC display a rapid cell cycle that is driven by mechanisms that relax the Rb-E2F pathway. In studies involving cancer, the addition of exogenous nucleosides alleviated replication stress and genomic instability (Burrell et al., 2013, Bester et al., 2011). Exogenous nucleosides reduced replication stress, DNA damage and the frequency of mitotic errors in cultures of human PSC. Further, the addition of exogenous nucleosides enabled human PSC to better overcome selective bottlenecks associated with re-plating. These findings suggest that exogenous nucleosides not only reduce the mechanisms that lead to mutation but will also reduce the ability of variant cells to overtake a culture by improving the survival of normal human PSC and limiting the anti-apoptotic selective advantage of the variant human PSC. Finally, the data presented here will greatly improve the efficiency of growing human PSC *in vitro,* whilst reducing the acquisition of genomic damage, a critical finding to ensure the safety of human PSC derived therapeutics.

9 References

ADAMS, B. R., GOLDING, S. E., RAO, R. R. & VALERIE, K. 2010. Dynamic dependence on ATR and ATM for double-strand break repair in human embryonic stem cells and neural descendants. *PLoS One,* 5**,** e10001.

AHUJA, A. K., JODKOWSKA, K., TELONI, F., BIZARD, A. H., ZELLWEGER, R., HERRADOR, R., ORTEGA, S., HICKSON, I. D., ALTMEYER, M., MENDEZ, J. & LOPES, M. 2016. A short G1 phase imposes constitutive replication stress and fork remodelling in mouse embryonic stem cells. *Nat Commun,* 7**,** 10660.

ALLERS, T. & LICHTEN, M. 2001. Differential timing and control of noncrossover and crossover recombination during meiosis. *Cell,* 106**,** 47-57.

AMIR, H., TOUBOUL, T., SABATINI, K., CHHABRA, D., GARITAONANDIA, I., LORING, J. F., MOREY, R. & LAURENT, L. C. 2017. Spontaneous Single-Copy Loss of TP53 in Human Embryonic Stem Cells Markedly Increases Cell Proliferation and Survival. *Stem Cells,* 35**,** 872-885.

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., CHIAO, E., CHOI, Y. M., CHOO, A. B., COLLINS, D., COLMAN, A., CROOK, J. M., DALEY, G. Q., DALTON, A., DE SOUSA, P. A., DENNING, C., DOWNIE, J., DVORAK, P., MONTGOMERY, K. D., FEKI, A., FORD, A., FOX, V., FRAGA, A. M., FRUMKIN, T., GE, L., GOKHALE, P. J., GOLAN-LEV, T., GOURABI, H., GROPP, M., LU, G., HAMPL, A., HARRON, K., HEALY, L., HERATH, W., HOLM, F., HOVATTA, O., HYLLNER, J., INAMDAR, M. S., IRWANTO, A. K., ISHII, T., JACONI, M., JIN, Y., KIMBER, S., KISELEV, S., KNOWLES, B. B., KOPPER, O., KUKHARENKO, V., KULIEV, A., LAGARKOVA, M. A., LAIRD, P. W., LAKO, M., LASLETT, A. L., LAVON, N., LEE, D. R., LEE, J. E., LI, C., LIM, L. S., LUDWIG, T. E., MA, Y., MALTBY, E., MATEIZEL, I., MAYSHAR, Y., MILEIKOVSKY, M., MINGER, S. L., MIYAZAKI, T., MOON, S. Y., MOORE, H., MUMMERY, C., NAGY, A., NAKATSUJI, N., NARWANI, K., OH, S. K., OLSON, C., OTONKOSKI, T., PAN, F., PARK, I. H., PELLS, S., PERA, M. F., PEREIRA, L. V., QI, O., RAJ,

G. S., REUBINOFF, B., ROBINS, A., ROBSON, P., ROSSANT, J., SALEKDEH, G. H., SCHULZ, T. C., et al. 2011. Screening ethnically diverse human embryonic stem cells identifies a chromosome 20 minimal amplicon conferring growth advantage. *Nat Biotechnol,* 29**,** 1132-44.

ANDREWS, P. W., BRONSON, D. L., BENHAM, F., STRICKLAND, S. & KNOWLES, B. B. 1980. A comparative study of eight cell lines derived from human testicular teratocarcinoma. *Int J Cancer,* 26**,** 269-80.

ANDREWS, P. W., CASPER, J., DAMJANOV, I., DUGGAN-KEEN, M., GIWERCMAN, A., HATA, J., VON KEITZ, A., LOOIJENGA, L. H., MILLÁN, J. L., OOSTERHUIS, J. W., PERA, M., SAWADA, M., SCHMOLL, H. J., SKAKKEBAEK, N. E., VAN PUTTEN, W. & STERN, P. 1996. Comparative analysis of cell surface antigens expressed by cell lines derived from human germ cell tumours. *Int J Cancer,* 66**,** 806-16.

ANDREWS, P. W., GOODFELLOW, P. N., SHEVINSKY, L. H., BRONSON, D. L. & KNOWLES, B. B. 1982. Cell-surface antigens of a clonal human embryonal carcinoma cell line: morphological and antigenic differentiation in culture. *Int J Cancer,* 29**,** 523-31.

ANDREWS, P. W., MATIN, M. M., BAHRAMI, A. R., DAMJANOV, I., GOKHALE, P. & DRAPER, J. S. 2005. Embryonic stem (ES) cells and embryonal carcinoma (EC) cells: opposite sides of the same coin. *Biochem Soc Trans,* 33**,** 1526-30.

ANGLANA, M., APIOU, F., BENSIMON, A. & DEBATISSE, M. 2003. Dynamics of DNA replication in mammalian somatic cells: nucleotide pool modulates origin choice and interorigin spacing. *Cell,* 114**,** 385-94.

ARLT, M. F., MULLE, J. G., SCHAIBLEY, V. M., RAGLAND, R. L., DURKIN, S. G., WARREN, S. T. & GLOVER, T. W. 2009. Replication stress induces genome-wide copy number changes in human cells that resemble polymorphic and pathogenic variants. *Am J Hum Genet,* 84**,** 339-50.

ATKIN, N. B. & BAKER, M. C. 1982. Specific chromosome change, i(12p), in testicular tumours? *Lancet,* 2**,** 1349.

AUDEBERT, M., SALLES, B. & CALSOU, P. 2004. Involvement of poly(ADP-ribose) polymerase-1 and XRCC1/DNA ligase III in an alternative route for DNA double-strand breaks rejoining. *J Biol Chem,* 279**,** 55117-26.

AVERY, S., HIRST, A. J., BAKER, D., LIM, C. Y., ALAGARATNAM, S., SKOTHEIM, R. I., LOTHE, R. A., PERA, M. F., COLMAN, A., ROBSON, P., ANDREWS, P.

W. & KNOWLES, B. B. 2013. BCL-XL mediates the strong selective advantage of a 20q11.21 amplification commonly found in human embryonic stem cell cultures. *Stem Cell Reports,* 1**,** 379-86.

BAKER, D., HIRST, A. J., GOKHALE, P. J., JUAREZ, M. A., WILLIAMS, S., WHEELER, M., BEAN, K., ALLISON, T. F., MOORE, H. D., ANDREWS, P. W. & BARBARIC, I. 2016. Detecting Genetic Mosaicism in Cultures of Human Pluripotent Stem Cells. *Stem Cell Reports,* 7**,** 998-1012.

BAKER, D. E., 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. *Nat Biotechnol,* 25**,** 207-15.

BARBARIC, I., BIGA, V., GOKHALE, P. J., JONES, M., STAVISH, D., GLEN, A., COCA, D. & ANDREWS, P. W. 2014. Time-lapse analysis of human embryonic stem cells reveals multiple bottlenecks restricting colony formation and their relief upon culture adaptation. *Stem Cell Reports,* 3**,** 142-55.

BARKER, R. A., PARMAR, M., STUDER, L. & TAKAHASHI, J. 2017. Human Trials of Stem Cell-Derived Dopamine Neurons for Parkinson's Disease: Dawn of a New Era. *Cell Stem Cell,* 21**,** 569-573.

BARLOW, J. H., FARYABI, R. B., CALLÉN, E., WONG, N., MALHOWSKI, A., CHEN, H. T., GUTIERREZ-CRUZ, G., SUN, H. W., MCKINNON, P., WRIGHT, G., CASELLAS, R., ROBBIANI, D. F., STAUDT, L., FERNANDEZ-CAPETILLO, O. & NUSSENZWEIG, A. 2013. Identification of early replicating fragile sites that contribute to genome instability. *Cell,* 152**,** 620-32.

BARNUM, K. J. & O'CONNELL, M. J. 2014. Cell cycle regulation by checkpoints. *Methods Mol Biol,* 1170**,** 29-40.

BARTEK, J., LUKAS, C. & LUKAS, J. 2004. Checking on DNA damage in S phase. *Nat Rev Mol Cell Biol,* 5**,** 792-804.

BARTEK, J. & LUKAS, J. 2003. Chk1 and Chk2 kinases in checkpoint control and cancer. *Cancer Cell,* 3**,** 421-9.

BARTKOVA, J., HOREJSÍ, Z., KOED, K., KRÄMER, A., TORT, F., ZIEGER, K., GULDBERG, P., SEHESTED, M., NESLAND, J. M., LUKAS, C., ØRNTOFT, T., LUKAS, J. & BARTEK, J. 2005. DNA damage response as a candidate anti-cancer barrier in early human tumorigenesis. *Nature,* 434**,** 864-70.

BARTKOVA, J., REZAEI, N., LIONTOS, M., KARAKAIDOS, P., KLETSAS, D., ISSAEVA, N., VASSILIOU, L. V., KOLETTAS, E., NIFOROU, K., ZOUMPOURLIS, V. C., TAKAOKA, M., NAKAGAWA, H., TORT, F., FUGGER, K., JOHANSSON, F., SEHESTED, M., ANDERSEN, C. L., DYRSKJOT, L., ØRNTOFT, T., LUKAS, J., KITTAS, C., HELLEDAY, T., HALAZONETIS, T. D., BARTEK, J. & GORGOULIS, V. G. 2006. Oncogene-induced senescence is part of the tumorigenesis barrier imposed by DNA damage checkpoints. *Nature,* 444**,** 633-7.

BECKER, K. A., GHULE, P. N., LIAN, J. B., STEIN, J. L., VAN WIJNEN, A. J. & STEIN, G. S. 2010. Cyclin D2 and the CDK substrate p220(NPAT) are required for self-renewal of human embryonic stem cells. *J Cell Physiol,* 222**,** 456-64.

BECKER, K. A., GHULE, P. N., THERRIEN, J. A., LIAN, J. B., STEIN, J. L., VAN WIJNEN, A. J. & STEIN, G. S. 2006. Self-renewal of human embryonic stem cells is supported by a shortened G1 cell cycle phase. *J Cell Physiol,* 209**,** 883-93.

BEGG, A. C., MCNALLY, N. J., SHRIEVE, D. C. & KÄRCHER, H. 1985. A method to measure the duration of DNA synthesis and the potential doubling time from a single sample. *Cytometry,* 6**,** 620-6.

BEN-PORATH, I., THOMSON, M. W., CAREY, V. J., GE, R., BELL, G. W., REGEV, A. & WEINBERG, R. A. 2008. An embryonic stem cell-like gene expression signature in poorly differentiated aggressive human tumors. *Nature genetics,* 40**,** 499-507.

BERTOLI, C., KLIER, S., MCGOWAN, C., WITTENBERG, C. & DE BRUIN, R. A. 2013. Chk1 inhibits E2F6 repressor function in response to replication stress to maintain cell-cycle transcription. *Curr Biol,* 23**,** 1629-37.

BESTER, A. C., RONIGER, M., OREN, Y. S., IM, M. M., SARNI, D., CHAOAT, M., BENSIMON, A., ZAMIR, G., SHEWACH, D. S. & KEREM, B. 2011. Nucleotide deficiency promotes genomic instability in early stages of cancer development. *Cell,* 145**,** 435-46.

BHUTANI, K., NAZOR, K. L., WILLIAMS, R., TRAN, H., DAI, H., DŽAKULA, Ž., CHO, E. H., PANG, A. W. C., RAO, M., CAO, H., SCHORK, N. J. & LORING, J. F. 2016. Whole-genome mutational burden analysis of three pluripotency induction methods. *Nat Commun,* 7**,** 10536.

BLUM, B., BAR-NUR, O., GOLAN-LEV, T. & BENVENISTY, N. 2009. The anti-apoptotic gene survivin contributes to teratoma formation by human embryonic stem cells. *Nat Biotechnol,* 27**,** 281-7.

BRUGAROLAS, J., CHANDRASEKARAN, C., GORDON, J. I., BEACH, D., JACKS, T. & HANNON, G. J. 1995. Radiation-induced cell cycle arrest compromised by p21 deficiency. *Nature,* 377**,** 552-7.

BRYANT, H. E., SCHULTZ, N., THOMAS, H. D., PARKER, K. M., FLOWER, D., LOPEZ, E., KYLE, S., MEUTH, M., CURTIN, N. J. & HELLEDAY, T. 2005. Specific killing of BRCA2-deficient tumours with inhibitors of poly(ADP-ribose) polymerase. *Nature,* 434**,** 913-7.

BUISSON, R., BOISVERT, J. L., BENES, C. H. & ZOU, L. 2015. Distinct but Concerted Roles of ATR, DNA-PK, and Chk1 in Countering Replication Stress during S Phase. *Mol Cell,* 59**,** 1011-24.

BURRELL, R. A., MCCLELLAND, S. E., ENDESFELDER, D., GROTH, P., WELLER, M. C., SHAIKH, N., DOMINGO, E., KANU, N., DEWHURST, S. M., GRONROOS, E., CHEW, S. K., ROWAN, A. J., SCHENK, A., SHEFFER, M., HOWELL, M., KSCHISCHO, M., BEHRENS, A., HELLEDAY, T., BARTEK, J., TOMLINSON, I. P. & SWANTON, C. 2013. Replication stress links structural and numerical cancer chromosomal instability. *Nature,* 494**,** 492-496.

BUSSARD, K. M., BOULANGER, C. A., BOOTH, B. W., BRUNO, R. D. & SMITH, G. H. 2010. Reprogramming human cancer cells in the mouse mammary gland. *Cancer Res,* 70**,** 6336-43.

BYUN, T. S., PACEK, M., YEE, M. C., WALTER, J. C. & CIMPRICH, K. A. 2005. Functional uncoupling of MCM helicase and DNA polymerase activities activates the ATR-dependent checkpoint. *Genes Dev,* 19**,** 1040-52.

CALDER, A., ROTH-ALBIN, I., BHATIA, S., PILQUIL, C., LEE, J. H., BHATIA, M., LEVADOUX-MARTIN, M., MCNICOL, J., RUSSELL, J., COLLINS, T. & DRAPER, J. S. 2013. Lengthened G1 phase indicates differentiation status in human embryonic stem cells. *Stem Cells Dev,* 22**,** 279-95.

CALSOU, P., DELTEIL, C., FRIT, P., DROUET, J. & SALLES, B. 2003. Coordinated assembly of Ku and p460 subunits of the DNA-dependent protein kinase on DNA ends is necessary for XRCC4-ligase IV recruitment. *J Mol Biol,* 326**,** 93-103.

CAPALBO, A., HOFFMANN, E. R., CIMADOMO, D., UBALDI, F. M. & RIENZI, L. 2017. Human female meiosis revised: new insights into the mechanisms of chromosome segregation and aneuploidies from advanced genomics and time-lapse imaging. *Hum Reprod Update,* 23**,** 706-722.

CAREY, B. W., MARKOULAKI, S., HANNA, J., SAHA, K., GAO, Q., MITALIPOVA, M. & JAENISCH, R. 2009. Reprogramming of murine and human somatic cells using a single polycistronic vector. *Proc Natl Acad Sci U S A,* 106**,** 157-62.

CARPENTER, A. E., JONES, T. R., LAMPRECHT, M. R., CLARKE, C., KANG, I. H., FRIMAN, O., GUERTIN, D. A., CHANG, J. H., LINDQUIST, R. A., MOFFAT, J., GOLLAND, P. & SABATINI, D. M. 2006. CellProfiler: image analysis software for identifying and quantifying cell phenotypes. *Genome Biol,* 7**,** R100.

CAYROU, C., COULOMBE, P., VIGNERON, A., STANOJCIC, S., GANIER, O., PEIFFER, I., RIVALS, E., PUY, A., LAURENT-CHABALIER, S., DESPRAT, R. & MÉCHALI, M. 2011. Genome-scale analysis of metazoan replication origins reveals their organization in specific but flexible sites defined by conserved features. *Genome Res,* 21**,** 1438-49.

CHAKRADHAR, S. 2016. An eye to the future: Researchers debate best path for stem cell–derived therapies. *Nature Medicine,* 22**,** 116-119.

CHEN, G., GULBRANSON, D. R., HOU, Z., BOLIN, J. M., RUOTTI, V., PROBASCO, M. D., SMUGA-OTTO, K., HOWDEN, S. E., DIOL, N. R., PROPSON, N. E., WAGNER, R., LEE, G. O., ANTOSIEWICZ-BOURGET, J., TENG, J. M. & THOMSON, J. A. 2011. Chemically defined conditions for human iPSC derivation and culture. *Nat Methods,* 8**,** 424-9.

CHEN, G., HOU, Z., GULBRANSON, D. R. & THOMSON, J. A. 2010. Actin-myosin contractility is responsible for the reduced viability of dissociated human embryonic stem cells. *Cell Stem Cell,* 7**,** 240-8.

CHENG, L., HANSEN, NANCY F., ZHAO, L., DU, Y., ZOU, C., DONOVAN, FRANK X., CHOU, B.-K., ZHOU, G., LI, S., DOWEY, SARAH N., YE, Z., CHANDRASEKHARAPPA, SETTARA C., YANG, H., MULLIKIN, JAMES C. & LIU, P. P. 2012a. Low Incidence of DNA Sequence Variation in Human Induced Pluripotent Stem Cells Generated by Nonintegrating Plasmid Expression. *Cell Stem Cell,* 10**,** 337-344.

CHENG, L., HANSEN, N. F., ZHAO, L., DU, Y., ZOU, C., DONOVAN, F. X., CHOU, B. K., ZHOU, G., LI, S., DOWEY, S. N., YE, Z., CHANDRASEKHARAPPA, S. C., YANG, H., MULLIKIN, J. C., LIU, P. P. & PROGRAM, N. C. S. 2012b. Low incidence of DNA sequence variation in human induced pluripotent stem cells generated by nonintegrating plasmid expression. *Cell Stem Cell,* 10**,** 337-44.

CHOI, S., TOLEDO, L. I., FERNANDEZ-CAPETILLO, O. & BAKKENIST, C. J. 2011. CGK733 does not inhibit ATM or ATR kinase activity in H460 human lung cancer cells. *DNA Repair (Amst),* 10**,** 1000-1; author reply 1002.

CHONG, J. J., YANG, X., DON, C. W., MINAMI, E., LIU, Y. W., WEYERS, J. J., MAHONEY, W. M., VAN BIBER, B., COOK, S. M., PALPANT, N. J., GANTZ, J. A., FUGATE, J. A., MUSKHELI, V., GOUGH, G. M., VOGEL, K. W., ASTLEY, C. A., HOTCHKISS, C. E., BALDESSARI, A., PABON, L., REINECKE, H., GILL, E. A., NELSON, V., KIEM, H. P., LAFLAMME, M. A. & MURRY, C. E. 2014. Human embryonic-stem-cell-derived cardiomyocytes regenerate non-human primate hearts. *Nature,* 510**,** 273-7.

CLEAVER, J. E. 1968. Defective repair replication of DNA in xeroderma pigmentosum. *Nature,* 218**,** 652-6.

COSTANTINO, L., SOTIRIOU, S. K., RANTALA, J. K., MAGIN, S., MLADENOV, E., HELLEDAY, T., HABER, J. E., ILIAKIS, G., KALLIONIEMI, O. P. & HALAZONETIS, T. D. 2014. Break-induced replication repair of damaged forks induces genomic duplications in human cells. *Science,* 343**,** 88-91.

COSTANZO, V., SHECHTER, D., LUPARDUS, P. J., CIMPRICH, K. A., GOTTESMAN, M. & GAUTIER, J. 2003. An ATR- and Cdc7-dependent DNA damage checkpoint that inhibits initiation of DNA replication. *Mol Cell,* 11**,** 203-13.

COUCH, F. B., BANSBACH, C. E., DRISCOLL, R., LUZWICK, J. W., GLICK, G. G., BÉTOUS, R., CARROLL, C. M., JUNG, S. Y., QIN, J., CIMPRICH, K. A. & CORTEZ, D. 2013. ATR phosphorylates SMARCAL1 to prevent replication fork collapse. *Genes Dev,* 27**,** 1610-23.

COWAN, C. A., KLIMANSKAYA, I., MCMAHON, J., ATIENZA, J., WITMYER, J., ZUCKER, J. P., WANG, S., MORTON, C. C., MCMAHON, A. P., POWERS, D. & MELTON, D. A. 2004. Derivation of embryonic stem-cell lines from human blastocysts. *N Engl J Med,* 350**,** 1353-6.

CRETU STANCU, M., VAN ROOSMALEN, M. J., RENKENS, I., NIEBOER, M. M., MIDDELKAMP, S., DE LIGT, J., PREGNO, G., GIACHINO, D., MANDRILE, G., ESPEJO VALLE-INCLAN, J., KORZELIUS, J., DE BRUIJN, E., CUPPEN, E., TALKOWSKI, M. E., MARSCHALL, T., DE RIDDER, J. & KLOOSTERMAN, W. P. 2017. Mapping and phasing of structural variation in patient genomes using nanopore sequencing. *Nat Commun,* 8**,** 1326.

CROSS, J., PETERS, G., WU, Z., BROHEDE, J. & HANNAN, G. N. 2007. Resolution of trisomic mosaicism in prenatal diagnosis: estimated performance of a 50K SNP microarray. *Prenat Diagn,* 27**,** 1197-204.

DANOVI, D., MEULMEESTER, E., PASINI, D., MIGLIORINI, D., CAPRA, M., FRENK, R., DE GRAAF, P., FRANCOZ, S., GASPARINI, P., GOBBI, A., HELIN, K., PELICCI, P. G., JOCHEMSEN, A. G. & MARINE, J. C. 2004. Amplification of Mdmx (or Mdm4) directly contributes to tumor formation by inhibiting p53 tumor suppressor activity. *Mol Cell Biol,* 24**,** 5835-43.

DARR, H., MAYSHAR, Y. & BENVENISTY, N. 2006. Overexpression of NANOG in human ES cells enables feeder-free growth while inducing primitive ectoderm features. *Development,* 133**,** 1193-201.

DE JAGER, M., VAN NOORT, J., VAN GENT, D. C., DEKKER, C., KANAAR, R. & WYMAN, C. 2001. Human Rad50/Mre11 is a flexible complex that can tether DNA ends. *Mol Cell,* 8**,** 1129-35.

DEAN, F. B., NELSON, J. R., GIESLER, T. L. & LASKEN, R. S. 2001. Rapid amplification of plasmid and phage DNA using Phi 29 DNA polymerase and multiply-primed rolling circle amplification. *Genome Res,* 11**,** 1095-9.

DEININGER, P. L. & BATZER, M. A. 1999. Alu repeats and human disease. *Mol Genet Metab,* 67**,** 183-93.

DEININGER, P. L., MORAN, J. V., BATZER, M. A. & KAZAZIAN, H. H. 2003. Mobile elements and mammalian genome evolution. *Curr Opin Genet Dev,* 13**,** 651-8.

DENG, C., ZHANG, P., HARPER, J. W., ELLEDGE, S. J. & LEDER, P. 1995. Mice lacking p21CIP1/WAF1 undergo normal development, but are defective in G1 checkpoint control. *Cell,* 82**,** 675-84.

DESCIPIO, C., CONLIN, L., ROSENFELD, J., TEPPERBERG, J., PASION, R., PATEL, A., MCDONALD, M. T., ARADHYA, S., HO, D., GOLDSTEIN, J., MCGUIRE, M., MULCHANDANI, S., MEDNE, L., RUPPS, R., SERRANO, A.

H., THORLAND, E. C., TSAI, A. C., HILHORST-HOFSTEE, Y., RUIVENKAMP, C. A., VAN ESCH, H., ADDOR, M. C., MARTINET, D., MASON, T. B., CLARK, D., SPINNER, N. B. & KRANTZ, I. D. 2012. Subtelomeric deletion of chromosome 10p15.3: clinical findings and molecular cytogenetic characterization. *Am J Med Genet A,* 158A**,** 2152-61.

DESMARAIS, J. A., HOFFMANN, M. J., BINGHAM, G., GAGOU, M. E., MEUTH, M. & ANDREWS, P. W. 2012. Human embryonic stem cells fail to activate CHK1 and commit to apoptosis in response to DNA replication stress. *Stem Cells,* 30**,** 1385-93.

DESMARAIS, J. A., UNGER, C., DAMJANOV, I., MEUTH, M. & ANDREWS, P. 2016. Apoptosis and failure of checkpoint kinase 1 activation in human induced pluripotent stem cells under replication stress. *Stem Cell Res Ther,* 7**,** 17.

DI MICCO, R., FUMAGALLI, M., CICALESE, A., PICCININ, S., GASPARINI, P., LUISE, C., SCHURRA, C., GARRE', M., NUCIFORO, P. G., BENSIMON, A., MAESTRO, R., PELICCI, P. G. & D'ADDA DI FAGAGNA, F. 2006. Oncogene-induced senescence is a DNA damage response triggered by DNA hyper-replication. *Nature,* 444**,** 638-42.

DIAMOND, T. L., ROSHAL, M., JAMBURUTHUGODA, V. K., REYNOLDS, H. M., MERRIAM, A. R., LEE, K. Y., BALAKRISHNAN, M., BAMBARA, R. A., PLANELLES, V., DEWHURST, S. & KIM, B. 2004. Macrophage tropism of HIV-1 depends on efficient cellular dNTP utilization by reverse transcriptase. *J Biol Chem,* 279**,** 51545-53.

DIANOV, G. L., SLEETH, K. M., DIANOVA, I. I. & ALLINSON, S. L. 2003. Repair of abasic sites in DNA. *Mutat Res,* 531**,** 157-63.

DOLEZALOVA, D., MRAZ, M., BARTA, T., PLEVOVA, K., VINARSKY, V., HOLUBCOVA, Z., JAROS, J., DVORAK, P., POSPISILOVA, S. & HAMPL, A. 2012. MicroRNAs regulate p21(Waf1/Cip1) protein expression and the DNA damage response in human embryonic stem cells. *Stem Cells,* 30**,** 1362-72.

DOS SANTOS SILVA, I. & SWERDLOW, A. J. 1991. Ovarian germ cell malignancies in England: epidemiological parallels with testicular cancer. *British Journal of Cancer,* 63**,** 814-818.

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. *Nat Biotechnol,* 22**,** 53-4.

DULIĆ, V., KAUFMANN, W. K., WILSON, S. J., TLSTY, T. D., LEES, E., HARPER, J. W., ELLEDGE, S. J. & REED, S. I. 1994. p53-dependent inhibition of cyclin-dependent kinase activities in human fibroblasts during radiation-induced G1 arrest. *Cell,* 76**,** 1013-23.

EKHOLM-REED, S., MÉNDEZ, J., TEDESCO, D., ZETTERBERG, A., STILLMAN, B. & REED, S. I. 2004. Deregulation of cyclin E in human cells interferes with prereplication complex assembly. *J Cell Biol,* 165**,** 789-800.

EL-DEIRY, W. S., TOKINO, T., VELCULESCU, V. E., LEVY, D. B., PARSONS, R., TRENT, J. M., LIN, D., MERCER, W. E., KINZLER, K. W. & VOGELSTEIN, B. 1993. WAF1, a potential mediator of p53 tumor suppression. *Cell,* 75**,** 817-25.

ELLIOTT, B., RICHARDSON, C. & JASIN, M. 2005. Chromosomal translocation mechanisms at intronic alu elements in mammalian cells. *Mol Cell,* 17**,** 885-94.

ENGLISH, A. C., SALERNO, W. J., HAMPTON, O. A., GONZAGA-JAUREGUI, C., AMBRETH, S., RITTER, D. I., BECK, C. R., DAVIS, C. F., DAHDOULI, M., MA, S., CARROLL, A., VEERARAGHAVAN, N., BRUESTLE, J., DREES, B., HASTIE, A., LAM, E. T., WHITE, S., MISHRA, P., WANG, M., HAN, Y., ZHANG, F., STANKIEWICZ, P., WHEELER, D. A., REID, J. G., MUZNY, D. M., ROGERS, J., SABO, A., WORLEY, K. C., LUPSKI, J. R., BOERWINKLE, E. & GIBBS, R. A. 2015. Assessing structural variation in a personal genome-towards a human reference diploid genome. *BMC Genomics,* 16**,** 286.

ENGSTRÖM, Y., ERIKSSON, S., JILDEVIK, I., SKOG, S., THELANDER, L. & TRIBUKAIT, B. 1985. Cell cycle-dependent expression of mammalian ribonucleotide reductase. Differential regulation of the two subunits. *J Biol Chem,* 260**,** 9114-6.

EVANS, M. J. & KAUFMAN, M. H. 1981. Establishment in culture of pluripotential cells from mouse embryos. *Nature,* 292**,** 154-6.

FAAST, R., WHITE, J., CARTWRIGHT, P., CROCKER, L., SARCEVIC, B. & DALTON, S. 2004. Cdk6-cyclin D3 activity in murine ES cells is resistant to inhibition by p16(INK4a). *Oncogene,* 23**,** 491-502.

FALCK, J., COATES, J. & JACKSON, S. P. 2005. Conserved modes of recruitment of ATM, ATR and DNA-PKcs to sites of DNA damage. *Nature,* 434**,** 605-11.

FALCK, J., MAILAND, N., SYLJUÅSEN, R. G., BARTEK, J. & LUKAS, J. 2001. The ATM-Chk2-Cdc25A checkpoint pathway guards against radioresistant DNA synthesis. *Nature,* 410**,** 842-7.

FAZELI, A., LIEW, C. G., MATIN, M. M., ELLIOTT, S., JEANMEURE, L. F., WRIGHT, P. C., MOORE, H. & ANDREWS, P. W. 2011. Altered patterns of differentiation in karyotypically abnormal human embryonic stem cells. *Int J Dev Biol,* 55**,** 175-80.

FEIJOO, C., HALL-JACKSON, C., WU, R., JENKINS, D., LEITCH, J., GILBERT, D. M. & SMYTHE, C. 2001. Activation of mammalian Chk1 during DNA replication arrest: a role for Chk1 in the intra-S phase checkpoint monitoring replication origin firing. *J Cell Biol,* 154**,** 913-23.

FILIPCZYK, A. A., LASLETT, A. L., MUMMERY, C. & PERA, M. F. 2007. Differentiation is coupled to changes in the cell cycle regulatory apparatus of human embryonic stem cells. *Stem Cell Res,* 1**,** 45-60.

FINCH, B. W. & EPHRUSSI, B. 1967. RETENTION OF MULTIPLE DEVELOPMENTAL POTENTIALITIES BY CELLS OF A MOUSE TESTICULAR TERATOCARCINOMA DURING PROLONGED CULTURE in vitro AND THEIR EXTINCTION UPON HYBRIDIZATION WITH CELLS OF PERMANENT LINES. *Proc Natl Acad Sci U S A,* 57**,** 615-21.

FINCH, R. A., DONOVIEL, D. B., POTTER, D., SHI, M., FAN, A., FREED, D. D., WANG, C. Y., ZAMBROWICZ, B. P., RAMIREZ-SOLIS, R., SANDS, A. T. & ZHANG, N. 2002. mdmx is a negative regulator of p53 activity in vivo. *Cancer Res,* 62**,** 3221-5.

FORD, J. H., SCHULTZ, C. J. & CORRELL, A. T. 1988. Chromosome elimination in micronuclei: a common cause of hypoploidy. *Am J Hum Genet,* 43**,** 733-40.

FOX, V., GOKHALE, P. J., WALSH, J. R., MATIN, M., JONES, M. & ANDREWS, P. W. 2008. Cell-cell signaling through NOTCH regulates human embryonic stem cell proliferation. *Stem Cells,* 26**,** 715-23.

FUSAKI, N., BAN, H., NISHIYAMA, A., SAEKI, K. & HASEGAWA, M. 2009. Efficient induction of transgene-free human pluripotent stem cells using a vector based on Sendai virus, an RNA virus that does not integrate into the host genome. *Proceedings of the Japan Academy. Series B, Physical and biological sciences,* 85**,** 348-362.

GADHIKAR, M. A., ZHANG, J., SHEN, L., RAO, X., WANG, J., ZHAO, M., KALU, N. N., JOHNSON, F. M., BYERS, L. A., HEYMACH, J., HITTELMAN, W. N., UDAYAKUMAR, D., PANDITA, R. K., PANDITA, T. K., PICKERING, C. R., REDWOOD, A. B., PIWNICA-WORMS, H., SCHLACHER, K., FREDERICK, M. J. & MYERS, J. N. 2018. Deletion in Head and Neck Cancer Cells Is Associated with CDK2 Activation, Replication Stress, and Vulnerability to CHK1 Inhibition. *Cancer Res,* 78**,** 781-797.

GE, X. Q. & BLOW, J. J. 2010. Chk1 inhibits replication factory activation but allows dormant origin firing in existing factories. *J Cell Biol,* 191**,** 1285-97.

GE, X. Q., JACKSON, D. A. & BLOW, J. J. 2007. Dormant origins licensed by excess Mcm2-7 are required for human cells to survive replicative stress. *Genes Dev,* 21**,** 3331-41.

GISSELSSON, D. 2008. Classification of chromosome segregation errors in cancer. *Chromosoma,* 117**,** 511-9.

GONZALEZ, M. A., TACHIBANA, K. E., LASKEY, R. A. & COLEMAN, N. 2005. Control of DNA replication and its potential clinical exploitation. *Nat Rev Cancer,* 5**,** 135-41.

GONZALEZ, S. & SERRANO, M. 2006. A new mechanism of inactivation of the INK4/ARF locus. *Cell Cycle,* 5**,** 1382-4.

GORGOULIS, V. G., VASSILIOU, L. V., KARAKAIDOS, P., ZACHARATOS, P., KOTSINAS, A., LILOGLOU, T., VENERE, M., DITULLIO, R. A., KASTRINAKIS, N. G., LEVY, B., KLETSAS, D., YONETA, A., HERLYN, M., KITTAS, C. & HALAZONETIS, T. D. 2005. Activation of the DNA damage checkpoint and genomic instability in human precancerous lesions. *Nature,* 434**,** 907-13.

GOZZETTI, A. & LE BEAU, M. M. 2000. Fluorescence in situ hybridization: uses and limitations. *Semin Hematol,* 37**,** 320-33.

GRANDELA, C., PERA, M. F., GRIMMOND, S. M., KOLLE, G. & WOLVETANG, E. J. 2007. p53 is required for etoposide-induced apoptosis of human embryonic stem cells. *Stem Cell Res,* 1**,** 116-28.

GROTH, P., AUSLÄNDER, S., MAJUMDER, M. M., SCHULTZ, N., JOHANSSON, F., PETERMANN, E. & HELLEDAY, T. 2010. Methylated DNA causes a physical block to replication forks independently of damage signalling, O(6)-

methylguanine or DNA single-strand breaks and results in DNA damage. *J Mol Biol,* 402**,** 70-82.

GRUHN, J. R., ZIELINSKA, A. P., SHUKLA, V., BLANSHARD, R., CAPALBO, A., CIMADOMO, D., NIKIFOROV, D., CHAN, A. C., NEWNHAM, L. J., VOGEL, I., SCARICA, C., KRAPCHEV, M., TAYLOR, D., KRISTENSEN, S. G., CHENG, J., ERNST, E., BJØRN, A. B., COLMORN, L. B., BLAYNEY, M., ELDER, K., LISS, J., HARTSHORNE, G., GRØNDAHL, M. L., RIENZI, L., UBALDI, F., MCCOY, R., LUKASZUK, K., ANDERSEN, C. Y., SCHUH, M. & HOFFMANN, E. R. 2019. Chromosome errors in human eggs shape natural fertility over reproductive life span. *Science,* 365**,** 1466-1469.

GUAN, X. Y., XU, J., ANZICK, S. L., ZHANG, H., TRENT, J. M. & MELTZER, P. S. 1996. Hybrid selection of transcribed sequences from microdissected DNA: isolation of genes within amplified region at 20q11-q13.2 in breast cancer. *Cancer Res,* 56**,** 3446-50.

GUBBAY, J., COLLIGNON, J., KOOPMAN, P., CAPEL, B., ECONOMOU, A., MÜNSTERBERG, A., VIVIAN, N., GOODFELLOW, P. & LOVELL-BADGE, R. 1990. A gene mapping to the sex-determining region of the mouse Y chromosome is a member of a novel family of embryonically expressed genes. *Nature,* 346**,** 245-50.

GUO, Z., KUMAGAI, A., WANG, S. X. & DUNPHY, W. G. 2000. Requirement for Atr in phosphorylation of Chk1 and cell cycle regulation in response to DNA replication blocks and UV-damaged DNA in Xenopus egg extracts. *Genes Dev,* 14**,** 2745-56.

HALAZONETIS, T. D., GORGOULIS, V. G. & BARTEK, J. 2008. An oncogene-induced DNA damage model for cancer development. *Science,* 319**,** 1352-5.

HARBOUR, J. W., LUO, R. X., DEI SANTI, A., POSTIGO, A. A. & DEAN, D. C. 1999. Cdk phosphorylation triggers sequential intramolecular interactions that progressively block Rb functions as cells move through G1. *Cell,* 98**,** 859-69.

HARDY, K. 1997. Cell death in the mammalian blastocyst. *Mol Hum Reprod,* 3**,** 919-25.

HARDY, K., STARK, J. & WINSTON, R. M. 2003. Maintenance of the inner cell mass in human blastocysts from fragmented embryos. *Biol Reprod,* 68**,** 1165-9.

HAUPT, S., MEJÍA-HERNÁNDEZ, J. O., VIJAYAKUMARAN, R., KEAM, S. P. & HAUPT, Y. 2019. The long and the short of it: the MDM4 tail so far. *Journal of Molecular Cell Biology,* 11**,** 231-244.

HERLIHY, A. E. & DE BRUIN, R. A. 2017. The Role of the Transcriptional Response to DNA Replication Stress. *Genes (Basel),* 8.

HEYER, B. S., MACAULEY, A., BEHRENDTSEN, O. & WERB, Z. 2000. Hypersensitivity to DNA damage leads to increased apoptosis during early mouse development. *Genes Dev,* 14**,** 2072-84.

HILLS, S. A. & DIFFLEY, J. F. 2014. DNA replication and oncogene-induced replicative stress. *Curr Biol,* 24**,** R435-44.

HIRAO, A., KONG, Y. Y., MATSUOKA, S., WAKEHAM, A., RULAND, J., YOSHIDA, H., LIU, D., ELLEDGE, S. J. & MAK, T. W. 2000. DNA damage-induced activation of p53 by the checkpoint kinase Chk2. *Science,* 287**,** 1824-7.

HOGAN, B., FELLOUS, M., JACOB, F. & AVNER, P. 1977. Isolation of a human teratoma cell line which expresses F9 antigen. *Nature,* 270**,** 515-518.

HUANG, X., TRAN, T., ZHANG, L., HATCHER, R. & ZHANG, P. 2005. DNA damage-induced mitotic catastrophe is mediated by the Chk1-dependent mitotic exit DNA damage checkpoint. *Proc Natl Acad Sci U S A,* 102**,** 1065-70.

HUBLEY, R., FINN, R. D., CLEMENTS, J., EDDY, S. R., JONES, T. A., BAO, W., SMIT, A. F. A. & WHEELER, T. J. 2015. The Dfam database of repetitive DNA families. *Nucleic Acids Research,* 44**,** D81-D89.

HYKA-NOUSPIKEL, N., DESMARAIS, J., GOKHALE, P. J., JONES, M., MEUTH, M., ANDREWS, P. W. & NOUSPIKEL, T. 2012. Deficient DNA damage response and cell cycle checkpoints lead to accumulation of point mutations in human embryonic stem cells. *Stem Cells,* 30**,** 1901-10.

ICHIJIMA, Y., YOSHIOKA, K., YOSHIOKA, Y., SHINOHE, K., FUJIMORI, H., UNNO, J., TAKAGI, M., GOTO, H., INAGAKI, M., MIZUTANI, S. & TERAOKA, H. 2010. DNA lesions induced by replication stress trigger mitotic aberration and tetraploidy development. *PLoS One,* 5**,** e8821.

ILLMENSEE, K. & MINTZ, B. 1976. Totipotency and normal differentiation of single teratocarcinoma cells cloned by injection into blastocysts. *Proc Natl Acad Sci U S A,* 73**,** 549-53.

ISHIDA, S., HUANG, E., ZUZAN, H., SPANG, R., LEONE, G., WEST, M. & NEVINS, J. R. 2001. Role for E2F in control of both DNA replication and mitotic functions as revealed from DNA microarray analysis. *Mol Cell Biol,* 21**,** 4684-99.

IVICS, Z. & IZSVÁK, Z. 2010. Repetitive elements and genome instability. *Semin Cancer Biol,* 20**,** 197-9.

JACKSON, D. A. & POMBO, A. 1998. Replicon clusters are stable units of chromosome structure: evidence that nuclear organization contributes to the efficient activation and propagation of S phase in human cells. *J Cell Biol,* 140**,** 1285-95.

JACOB, F. 1978. The Leeuwenhoek Lecture, 1977. Mouse teratocarcinoma and mouse embryo. *Proc R Soc Lond B Biol Sci,* 201**,** 249-70.

JIANG, H., REINHARDT, H. C., BARTKOVA, J., TOMMISKA, J., BLOMQVIST, C., NEVANLINNA, H., BARTEK, J., YAFFE, M. B. & HEMANN, M. T. 2009. The combined status of ATM and p53 link tumor development with therapeutic response. *Genes Dev,* 23**,** 1895-909.

JIAO, X., WOOD, L. D., LINDMAN, M., JONES, S., BUCKHAULTS, P., POLYAK, K., SUKUMAR, S., CARTER, H., KIM, D., KARCHIN, R. & SJÖBLOM, T. 2012. Somatic mutations in the Notch, NF-KB, PIK3CA, and Hedgehog pathways in human breast cancers. *Genes Chromosomes Cancer,* 51**,** 480-9.

JIN, D. H., PARK, S. E., LEE, J., KIM, K. M., KIM, S., KIM, D. H. & PARK, J. 2015. Copy Number Gains at 8q24 and 20q11-q13 in Gastric Cancer Are More Common in Intestinal-Type than Diffuse-Type. *PLoS One,* 10**,** e0137657.

JONES, R. M., MORTUSEWICZ, O., AFZAL, I., LORVELLEC, M., GARCÍA, P., HELLEDAY, T. & PETERMANN, E. 2013. Increased replication initiation and conflicts with transcription underlie Cyclin E-induced replication stress. *Oncogene,* 32**,** 3744-53.

JUDSON, R. L., BABIARZ, J. E., VENERE, M. & BLELLOCH, R. 2009. Embryonic stem cell-specific microRNAs promote induced pluripotency. *Nature biotechnology,* 27**,** 459-461.

KANNAGI, R., COCHRAN, N. A., ISHIGAMI, F., HAKOMORI, S., ANDREWS, P. W., KNOWLES, B. B. & SOLTER, D. 1983. Stage-specific embryonic antigens (SSEA-3 and -4) are epitopes of a unique globo-series ganglioside isolated from human teratocarcinoma cells. *EMBO J,* 2**,** 2355-61.

KATSUNO, Y., SUZUKI, A., SUGIMURA, K., OKUMURA, K., ZINELDEEN, D. H., SHIMADA, M., NIIDA, H., MIZUNO, T., HANAOKA, F. & NAKANISHI, M. 2009. Cyclin A-Cdk1 regulates the origin firing program in mammalian cells. *Proc Natl Acad Sci U S A,* 106**,** 3184-9.

KEARNEY, L. 2001. Molecular cytogenetics. *Best Pract Res Clin Haematol,* 14**,** 645-69.

KIDD, J. M., COOPER, G. M., DONAHUE, W. F., HAYDEN, H. S., SAMPAS, N., GRAVES, T., HANSEN, N., TEAGUE, B., ALKAN, C., ANTONACCI, F., HAUGEN, E., ZERR, T., YAMADA, N. A., TSANG, P., NEWMAN, T. L., TÜZÜN, E., CHENG, Z., EBLING, H. M., TUSNEEM, N., DAVID, R., GILLETT, W., PHELPS, K. A., WEAVER, M., SARANGA, D., BRAND, A., TAO, W., GUSTAFSON, E., MCKERNAN, K., CHEN, L., MALIG, M., SMITH, J. D., KORN, J. M., MCCARROLL, S. A., ALTSHULER, D. A., PEIFFER, D. A., DORSCHNER, M., STAMATOYANNOPOULOS, J., SCHWARTZ, D., NICKERSON, D. A., MULLIKIN, J. C., WILSON, R. K., BRUHN, L., OLSON, M. V., KAUL, R., SMITH, D. R. & EICHLER, E. E. 2008. Mapping and sequencing of structural variation from eight human genomes. *Nature,* 453**,** 56-64.

KIESSLING, A. A., ANDERSON, S. & ANDERSON, S. C. 2003. *Human embryonic stem cells: an introduction to the science and therapeutic potential*, Jones & Bartlett Learning.

KINNER, A., WU, W., STAUDT, C. & ILIAKIS, G. 2008. Gamma-H2AX in recognition and signaling of DNA double-strand breaks in the context of chromatin. *Nucleic Acids Res,* 36**,** 5678-94.

KOLOMIETZ, E., MEYN, M. S., PANDITA, A. & SQUIRE, J. A. 2002. The role of Alu repeat clusters as mediators of recurrent chromosomal aberrations in tumors. *Genes Chromosomes Cancer,* 35**,** 97-112.

KOREN, S., WALENZ, B. P., BERLIN, K., MILLER, J. R., BERGMAN, N. H. & PHILLIPPY, A. M. 2017. Canu: scalable and accurate long-read assembly via adaptive. *Genome Res,* 27**,** 722-736.

KUMAR, S., BLANGERO, J. & CURRAN, J. E. 2018. Induced Pluripotent Stem Cells in Disease Modeling and Gene Identification. *In:* DISTEFANO, J. K. (ed.) *Disease Gene Identification: Methods and Protocols.* New York, NY: Springer New York.

KÖHLER, G. & MILSTEIN, C. 1975. Continuous cultures of fused cells secreting antibody of predefined specificity. *Nature,* 256**,** 495-7.

LAING, O., HALLIWELL, J. & BARBARIC, I. 2019. Rapid PCR Assay for Detecting Common Genetic Variants Arising in Human Pluripotent Stem Cell Cultures. *Curr Protoc Stem Cell Biol,* 49**,** e83.

LAMM, N., BEN-DAVID, U., GOLAN-LEV, T., STORCHOVÁ, Z., BENVENISTY, N. & KEREM, B. 2016. Genomic Instability in Human Pluripotent Stem Cells Arises from Replicative Stress and Chromosome Condensation Defects. *Cell Stem Cell,* 18**,** 253-61.

LANDER, E. S., LINTON, L. M., BIRREN, B., NUSBAUM, C., ZODY, M. C., BALDWIN, J., DEVON, K., DEWAR, K., DOYLE, M., FITZHUGH, W., FUNKE, R., GAGE, D., HARRIS, K., HEAFORD, A., HOWLAND, J., KANN, L., LEHOCZKY, J., LEVINE, R., MCEWAN, P., MCKERNAN, K., MELDRIM, J., MESIROV, J. P., MIRANDA, C., MORRIS, W., NAYLOR, J., RAYMOND, C., ROSETTI, M., SANTOS, R., SHERIDAN, A., SOUGNEZ, C., STANGE-THOMANN, Y., STOJANOVIC, N., SUBRAMANIAN, A., WYMAN, D., ROGERS, J., SULSTON, J., AINSCOUGH, R., BECK, S., BENTLEY, D., BURTON, J., CLEE, C., CARTER, N., COULSON, A., DEADMAN, R., DELOUKAS, P., DUNHAM, A., DUNHAM, I., DURBIN, R., FRENCH, L., GRAFHAM, D., GREGORY, S., HUBBARD, T., HUMPHRAY, S., HUNT, A., JONES, M., LLOYD, C., MCMURRAY, A., MATTHEWS, L., MERCER, S., MILNE, S., MULLIKIN, J. C., MUNGALL, A., PLUMB, R., ROSS, M., SHOWNKEEN, R., SIMS, S., WATERSTON, R. H., WILSON, R. K., HILLIER, L. W., MCPHERSON, J. D., MARRA, M. A., MARDIS, E. R., FULTON, L. A., CHINWALLA, A. T., PEPIN, K. H., GISH, W. R., CHISSOE, S. L., WENDL, M. C., DELEHAUNTY, K. D., MINER, T. L., DELEHAUNTY, A., KRAMER, J. B., COOK, L. L., FULTON, R. S., JOHNSON, D. L., MINX, P. J., CLIFTON, S. W., HAWKINS, T., BRANSCOMB, E., PREDKI, P., RICHARDSON, P., WENNING, S., SLEZAK, T., DOGGETT, N., CHENG, J. F., OLSEN, A., LUCAS, S., ELKIN, C., UBERBACHER, E., FRAZIER, M., et al. 2001. Initial sequencing and analysis of the human genome. *Nature,* 409**,** 860-921.

LARSSON, C., ALI, M. A., PANDZIC, T., LINDROTH, A. M., HE, L. & SJÖBLOM, T. 2017. Loss of DIP2C in RKO cells stimulates changes in DNA methylation and epithelial-mesenchymal transition. *BMC Cancer,* 17**,** 487.

LASKEN, R. S. 2013. Single-cell sequencing in its prime. *Nat Biotechnol,* 31**,** 211-2.

LAURENT, A. & BLASI, F. 2015. Differential DNA damage signalling and apoptotic threshold correlate with mouse epiblast-specific hypersensitivity to radiation. *Development,* 142**,** 3675-85.

LAURIE, N. A., DONOVAN, S. L., SHIH, C. S., ZHANG, J., MILLS, N., FULLER, C., TEUNISSE, A., LAM, S., RAMOS, Y., MOHAN, A., JOHNSON, D., WILSON, M., RODRIGUEZ-GALINDO, C., QUARTO, M., FRANCOZ, S., MENDRYSA, S. M., GUY, R. K., MARINE, J. C., JOCHEMSEN, A. G. & DYER, M. A. 2006. Inactivation of the p53 pathway in retinoblastoma. *Nature,* 444**,** 61-6.

LAVIN, M. F. 2007. ATM and the Mre11 complex combine to recognize and signal DNA double-strand breaks. *Oncogene,* 26**,** 7749-58.

LEBER, R., WISE, T. W., MIZUTA, R. & MEEK, K. 1998. The XRCC4 gene product is a target for and interacts with the DNA-dependent protein kinase. *J Biol Chem,* 273**,** 1794-801.

LEE, J. H. & PAULL, T. T. 2005. ATM activation by DNA double-strand breaks through the Mre11-Rad50-Nbs1 complex. *Science,* 308**,** 551-4.

LEONE, G., DEGREGORI, J., YAN, Z., JAKOI, L., ISHIDA, S., WILLIAMS, R. S. & NEVINS, J. R. 1998. E2F3 activity is regulated during the cell cycle and is required for the induction of S phase. *Genes Dev,* 12**,** 2120-30.

LEY, T. J., MARDIS, E. R., DING, L., FULTON, B., MCLELLAN, M. D., CHEN, K., DOOLING, D., DUNFORD-SHORE, B. H., MCGRATH, S., HICKENBOTHAM, M., COOK, L., ABBOTT, R., LARSON, D. E., KOBOLDT, D. C., POHL, C., SMITH, S., HAWKINS, A., ABBOTT, S., LOCKE, D., HILLIER, L. W., MINER, T., FULTON, L., MAGRINI, V., WYLIE, T., GLASSCOCK, J., CONYERS, J., SANDER, N., SHI, X., OSBORNE, J. R., MINX, P., GORDON, D., CHINWALLA, A., ZHAO, Y., RIES, R. E., PAYTON, J. E., WESTERVELT, P., TOMASSON, M. H., WATSON, M., BATY, J., IVANOVICH, J., HEATH, S., SHANNON, W. D., NAGARAJAN, R., WALTER, M. J., LINK, D. C., GRAUBERT, T. A., DIPERSIO, J. F. & WILSON, R. K. 2008. DNA sequencing of a cytogenetically normal acute myeloid leukaemia genome. *Nature,* 456**,** 66-72.

LI, H. 2018. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics,* 34**,** 3094-3100.

LI, H., COLLADO, M., VILLASANTE, A., STRATI, K., ORTEGA, S., CAÑAMERO, M., BLASCO, M. A. & SERRANO, M. 2009a. The Ink4/Arf locus is a barrier for iPS cell reprogramming. *Nature,* 460**,** 1136-9.

LI, H., HANDSAKER, B., WYSOKER, A., FENNELL, T., RUAN, J., HOMER, N., MARTH, G., ABECASIS, G., DURBIN, R. & SUBGROUP, G. P. D. P. 2009b. The Sequence Alignment/Map format and SAMtools. *Bioinformatics,* 25**,** 2078-9.

LIEW, C. G., DRAPER, J. S., WALSH, J., MOORE, H. & ANDREWS, P. W. 2007. Transient and stable transgene expression in human embryonic stem cells. *Stem Cells,* 25**,** 1521-8.

LINDAHL, T. & BARNES, D. E. 2000. Repair of endogenous DNA damage. *Cold Spring Harb Symp Quant Biol,* 65**,** 127-33.

LIU, B., WINKLER, F., HERDE, M., WITTE, C. P. & GROßHANS, J. 2019. A Link between Deoxyribonucleotide Metabolites and Embryonic Cell-Cycle Control. *Curr Biol,* 29**,** 1187-1192.e3.

LIU, Q., GUNTUKU, S., CUI, X. S., MATSUOKA, S., CORTEZ, D., TAMAI, K., LUO, G., CARATTINI-RIVERA, S., DEMAYO, F., BRADLEY, A., DONEHOWER, L. A. & ELLEDGE, S. J. 2000. Chk1 is an essential kinase that is regulated by Atr and required for the G(2)/M DNA damage checkpoint. *Genes Dev,* 14**,** 1448-59.

LLORENTE, B., SMITH, C. E. & SYMINGTON, L. S. 2008. Break-induced replication: what is it and what is it for? *Cell Cycle,* 7**,** 859-64.

LOPES, M., COTTA-RAMUSINO, C., PELLICIOLI, A., LIBERI, G., PLEVANI, P., MUZI-FALCONI, M., NEWLON, C. S. & FOIANI, M. 2001. The DNA replication checkpoint response stabilizes stalled replication forks. *Nature,* 412**,** 557-61.

LUKAS, C., FALCK, J., BARTKOVA, J., BARTEK, J. & LUKAS, J. 2003. Distinct spatiotemporal dynamics of mammalian checkpoint regulators induced by DNA damage. *Nat Cell Biol,* 5**,** 255-60.

LUKAS, C., SAVIC, V., BEKKER-JENSEN, S., DOIL, C., NEUMANN, B., PEDERSEN, R. S., GRØFTE, M., CHAN, K. L., HICKSON, I. D., BARTEK, J. & LUKAS, J. 2011. 53BP1 nuclear bodies form around DNA lesions generated by mitotic transmission of chromosomes under replication stress. *Nat Cell Biol,* 13**,** 243-53.

LUKAS, J., PARRY, D., AAGAARD, L., MANN, D. J., BARTKOVA, J., STRAUSS, M., PETERS, G. & BARTEK, J. 1995. Retinoblastoma-protein-dependent cell-cycle inhibition by the tumour suppressor p16. *Nature,* 375**,** 503-6.

LUNDBERG, A. S. & WEINBERG, R. A. 1998. Functional inactivation of the retinoblastoma protein requires sequential modification by at least two distinct cyclin-cdk complexes. *Mol Cell Biol,* 18**,** 753-61.

LUO, L. Z., GOPALAKRISHNA-PILLAI, S., NAY, S. L., PARK, S. W., BATES, S. E., ZENG, X., IVERSON, L. E. & O'CONNOR, T. R. 2012. DNA repair in human pluripotent stem cells is distinct from that in non-pluripotent human cells. *PLoS One,* 7**,** e30541.

LUPSKI, J. R. 1998. Genomic disorders: structural features of the genome can lead to DNA rearrangements and human disease traits. *Trends Genet,* 14**,** 417-22.

MA, X., ZHENG, W., WEI, D., MA, Y., WANG, T., WANG, J., LIU, Q. & YANG, S. 2006. High-level expression, purification and pro-apoptosis activity of HIV-TAT-survivin (T34A) mutant to cancer cells in vitro. *J Biotechnol,* 123**,** 367-78.

MA, Y., PANNICKE, U., SCHWARZ, K. & LIEBER, M. R. 2002. Hairpin opening and overhang processing by an Artemis/DNA-dependent protein kinase complex in nonhomologous end joining and V(D)J recombination. *Cell,* 108**,** 781-94.

MADDEN, D. T., DAVILA-KRUGER, D., MELOV, S. & BREDESEN, D. E. 2011. Human embryonic stem cells express elevated levels of multiple pro-apoptotic BCL-2 family members. *PLoS One,* 6**,** e28530.

MAILAND, N., FALCK, J., LUKAS, C., SYLJUÂSEN, R. G., WELCKER, M., BARTEK, J. & LUKAS, J. 2000. Rapid destruction of human Cdc25A in response to DNA damage. *Science,* 288**,** 1425-9.

MANNING, A. L. & DYSON, N. J. 2012. RB: mitotic implications of a tumour suppressor. *Nat Rev Cancer,* 12**,** 220-6.

MARTIN, G. R. 1981. Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. *Proc Natl Acad Sci U S A,* 78**,** 7634-8.

MASAMSETTI, V. P., LOW, R. R. J., MAK, K. S., O'CONNOR, A., RIFFKIN, C. D., LAMM, N., CRABBE, L., KARLSEDER, J., HUANG, D. C. S., HAYASHI, M. T. & CESARE, A. J. 2019. Replication stress induces mitotic death through parallel pathways regulated by WAPL and telomere deprotection. *Nat Commun,* 10**,** 4224.

MASSELINK, H. & BERNARDS, R. 2000. The adenovirus E1A binding protein BS69 is a corepressor of transcription through recruitment of N-CoR. *Oncogene,* 19**,** 1538-46.

MATSUMOTO, Y., SUZUKI, N., NAMBA, N., UMEDA, N., MA, X. J., MORITA, A., TOMITA, M., ENOMOTO, A., SERIZAWA, S., HIRANO, K., SAKAIA, K., YASUDA, H. & HOSOI, Y. 2000. Cleavage and phosphorylation of XRCC4 protein induced by X-irradiation. *FEBS Lett,* 478**,** 67-71.

MATSUOKA, S., HUANG, M. & ELLEDGE, S. J. 1998. Linkage of ATM to cell cycle regulation by the Chk2 protein kinase. *Science,* 282**,** 1893-7.

MAYNARD, S., SWISTOWSKA, A. M., LEE, J. W., LIU, Y., LIU, S. T., DA CRUZ, A. B., RAO, M., DE SOUZA-PINTO, N. C., ZENG, X. & BOHR, V. A. 2008. Human embryonic stem cells have enhanced repair of multiple forms of DNA damage. *Stem Cells,* 26**,** 2266-74.

MERKLE, F. T., GHOSH, S., KAMITAKI, N., MITCHELL, J., AVIOR, Y., MELLO, C., KASHIN, S., MEKHOUBAD, S., ILIC, D., CHARLTON, M., SAPHIER, G., HANDSAKER, R. E., GENOVESE, G., BAR, S., BENVENISTY, N., MCCARROLL, S. A. & EGGAN, K. 2017. Human pluripotent stem cells recurrently acquire and expand dominant negative P53 mutations. *Nature,* 545**,** 229-233.

MESRI, M., WALL, N. R., LI, J., KIM, R. W. & ALTIERI, D. C. 2001. Cancer gene therapy using a survivin mutant adenovirus. *J Clin Invest,* 108**,** 981-90.

MEYER, N., JACONI, M., LANDOPOULOU, A., FORT, P. & PUCÉAT, M. 2000. A fluorescent reporter gene as a marker for ventricular specification in ES-derived cardiac cells. *FEBS Lett,* 478**,** 151-8.

MICHL, J., ZIMMER, J. & TARSOUNAS, M. 2016. Interplay between Fanconi anemia and homologous recombination pathways in genome integrity. *EMBO J,* 35**,** 909-23.

MIGA, K. H., KOREN, S., RHIE, A., VOLLGER, M. R., GERSHMAN, A., BZIKADZE, A., BROOKS, S., HOWE, E., PORUBSKY, D., LOGSDON, G. A., SCHNEIDER, V. A., POTAPOVA, T., WOOD, J., CHOW, W., ARMSTRONG, J., FREDRICKSON, J., PAK, E., TIGYI, K., KREMITZKI, M., MARKOVIC, C., MADURO, V., DUTRA, A., BOUFFARD, G. G., CHANG, A. M., HANSEN, N. F., THIBAUD-NISSEN, F., SCHMITT, A. D., BELTON, J.-M., SELVARAJ, S., DENNIS, M. Y., SOTO, D. C., SAHASRABUDHE, R., KAYA, G., QUICK, J.,

LOMAN, N. J., HOLMES, N., LOOSE, M., SURTI, U., RISQUES, R. A., GRAVES LINDSAY, T. A., FULTON, R., HALL, I., PATEN, B., HOWE, K., TIMP, W., YOUNG, A., MULLIKIN, J. C., PEVZNER, P. A., GERTON, J. L., SULLIVAN, B. A., EICHLER, E. E. & PHILLIPPY, A. M. 2019. Telomere-to-telomere assembly of a complete human X chromosome. *bioRxiv*, 735928.

MIGLIORINI, D., LAZZERINI DENCHI, E., DANOVI, D., JOCHEMSEN, A., CAPILLO, M., GOBBI, A., HELIN, K., PELICCI, P. G. & MARINE, J. C. 2002. Mdm4 (Mdmx) regulates p53-induced growth arrest and neuronal cell death during early embryonic mouse development. *Mol Cell Biol,* 22**,** 5527-38.

MILHOLLAND, B., DONG, X., ZHANG, L., HAO, X., SUH, Y. & VIJG, J. 2017. Differences between germline and somatic mutation rates in humans and mice. *Nat Commun,* 8**,** 15183.

MIMITOU, E. P. & SYMINGTON, L. S. 2008. Sae2, Exo1 and Sgs1 collaborate in DNA double-strand break processing. *Nature,* 455**,** 770-4.

MIMITOU, E. P. & SYMINGTON, L. S. 2009. Nucleases and helicases take center stage in homologous recombination. *Trends Biochem Sci,* 34**,** 264-72.

MINOCHERHOMJI, S., YING, S., BJERREGAARD, V. A., BURSOMANNO, S., ALELIUNAITE, A., WU, W., MANKOURI, H. W., SHEN, H., LIU, Y. & HICKSON, I. D. 2015. Replication stress activates DNA repair synthesis in mitosis. *Nature,* 528**,** 286-90.

MISCEO, D., CARDONE, M. F., CARBONE, L., D'ADDABBO, P., DE JONG, P. J., ROCCHI, M. & ARCHIDIACONO, N. 2004. Evolutionary History of Chromosome 20. *Molecular Biology and Evolution,* 22**,** 360-366.

MITSUI, K., TOKUZAWA, Y., ITOH, H., SEGAWA, K., MURAKAMI, M., TAKAHASHI, K., MARUYAMA, M., MAEDA, M. & YAMANAKA, S. 2003. The homeoprotein Nanog is required for maintenance of pluripotency in mouse epiblast and ES cells. *Cell,* 113**,** 631-42.

MOMCILOVIC, O., KNOBLOCH, L., FORNSAGLIO, J., VARUM, S., EASLEY, C. & SCHATTEN, G. 2010. DNA damage responses in human induced pluripotent stem cells and embryonic stem cells. *PLoS One,* 5**,** e13410.

MOMCILOVIĆ, O., CHOI, S., VARUM, S., BAKKENIST, C., SCHATTEN, G. & NAVARA, C. 2009. Ionizing radiation induces ataxia telangiectasia mutated-dependent checkpoint signaling and G(2) but not G(1) cell cycle arrest in pluripotent human embryonic stem cells. *Stem Cells,* 27**,** 1822-35.

MURRY, C. E. & KELLER, G. 2008. Differentiation of embryonic stem cells to clinically relevant populations: lessons from embryonic development. *Cell,* 132**,** 661-80.

NARASIMHA, A. M., KAULICH, M., SHAPIRO, G. S., CHOI, Y. J., SICINSKI, P. & DOWDY, S. F. 2014. Cyclin D activates the Rb tumor suppressor by mono-phosphorylation. *Elife,* 3.

NARAYANAN, V., MIECZKOWSKI, P. A., KIM, H. M., PETES, T. D. & LOBACHEV, K. S. 2006. The pattern of gene amplification is determined by the chromosomal location of hairpin-capped breaks. *Cell,* 125**,** 1283-96.

NAVIN, N., KENDALL, J., TROGE, J., ANDREWS, P., RODGERS, L., MCINDOO, J., COOK, K., STEPANSKY, A., LEVY, D., ESPOSITO, D., MUTHUSWAMY, L., KRASNITZ, A., MCCOMBIE, W. R., HICKS, J. & WIGLER, M. 2011. Tumour evolution inferred by single-cell sequencing. *Nature,* 472**,** 90-4.

NEELSEN, K. J., ZANINI, I. M., MIJIC, S., HERRADOR, R., ZELLWEGER, R., RAY CHAUDHURI, A., CREAVIN, K. D., BLOW, J. J. & LOPES, M. 2013. Deregulated origin licensing leads to chromosomal breaks by rereplication of a gapped DNA template. *Genes Dev,* 27**,** 2537-42.

NEGANOVA, I. & LAKO, M. 2008. G1 to S phase cell cycle transition in somatic and embryonic stem cells. *J Anat,* 213**,** 30-44.

NEWPORT, J. & DASSO, M. 1989. On the coupling between DNA replication and mitosis. *J Cell Sci Suppl,* 12**,** 149-60.

NODA, A., NING, Y., VENABLE, S. F., PEREIRA-SMITH, O. M. & SMITH, J. R. 1994. Cloning of senescent cell-derived inhibitors of DNA synthesis using an expression screen. *Exp Cell Res,* 211**,** 90-8.

NOUSPIKEL, T. 2009. DNA repair in mammalian cells : Nucleotide excision repair: variations on versatility. *Cell Mol Life Sci,* 66**,** 994-1009.

NURSE, P. 1997. Regulation of the eukaryotic cell cycle. *European Journal of Cancer,* 33**,** 1002-1004.

NUSSENZWEIG, A. & NUSSENZWEIG, M. C. 2007. A backup DNA repair pathway moves to the forefront. *Cell,* 131**,** 223-5.

OHGUSHI, M., MATSUMURA, M., EIRAKU, M., MURAKAMI, K., ARAMAKI, T., NISHIYAMA, A., MUGURUMA, K., NAKANO, T., SUGA, H., UENO, M., ISHIZAKI, T., SUEMORI, H., NARUMIYA, S., NIWA, H. & SASAI, Y. 2010.

Molecular pathway and cell state responsible for dissociation-induced apoptosis in human pluripotent stem cells. *Cell Stem Cell,* 7**,** 225-39.

OKITA, K., ICHISAKA, T. & YAMANAKA, S. 2007. Generation of germline-competent induced pluripotent stem cells. *Nature,* 448**,** 313-317.

OKITA, K., NAKAGAWA, M., HYENJONG, H., ICHISAKA, T. & YAMANAKA, S. 2008. Generation of mouse induced pluripotent stem cells without viral vectors. *Science,* 322**,** 949-53.

OLARIU, V., HARRISON, N. J., COCA, D., GOKHALE, P. J., BAKER, D., BILLINGS, S., KADIRKAMANATHAN, V. & ANDREWS, P. W. 2010. Modeling the evolution of culture-adapted human embryonic stem cells. *Stem Cell Res,* 4**,** 50-6.

OOSTERHUIS, J. W. & LOOIJENGA, L. H. 2005. Testicular germ-cell tumours in a broader perspective. *Nat Rev Cancer,* 5**,** 210-22.

PARANT, J., CHAVEZ-REYES, A., LITTLE, N. A., YAN, W., REINKE, V., JOCHEMSEN, A. G. & LOZANO, G. 2001. Rescue of embryonic lethality in Mdm4-null mice by loss of Trp53 suggests a nonoverlapping pathway with MDM2 to regulate p53. *Nat Genet,* 29**,** 92-5.

PEDDIBHOTLA, S., LAM, M. H., GONZALEZ-RIMBAU, M. & ROSEN, J. M. 2009. The DNA-damage effector checkpoint kinase 1 is essential for chromosome segregation and cytokinesis. *Proc Natl Acad Sci U S A,* 106**,** 5159-64.

PERA, M. F. & TROUNSON, A. O. 2004. Human embryonic stem cells: prospects for development. *Development,* 131**,** 5515-25.

PERRIER, A. L., TABAR, V., BARBERI, T., RUBIO, M. E., BRUSES, J., TOPF, N., HARRISON, N. L. & STUDER, L. 2004. Derivation of midbrain dopamine neurons from human embryonic stem cells. *Proc Natl Acad Sci U S A,* 101**,** 12543-8.

PERRY, G. H., BEN-DOR, A., TSALENKO, A., SAMPAS, N., RODRIGUEZ-REVENGA, L., TRAN, C. W., SCHEFFER, A., STEINFELD, I., TSANG, P., YAMADA, N. A., PARK, H. S., KIM, J. I., SEO, J. S., YAKHINI, Z., LADERMAN, S., BRUHN, L. & LEE, C. 2008. The fine-scale and complex architecture of human copy-number variation. *Am J Hum Genet,* 82**,** 685-95.

PEZZOLO, A., PARODI, F., MARIMPIETRI, D., RAFFAGHELLO, L., COCCO, C., PISTORIO, A., MOSCONI, M., GAMBINI, C., CILLI, M., DEAGLIO, S., MALAVASI, F. & PISTOIA, V. 2011. Oct-4+/Tenascin C+ neuroblastoma cells

serve as progenitors of tumor-derived endothelial cells. *Cell Res,* 21**,** 1470-86.

PFISTER, S. X., MARKKANEN, E., JIANG, Y., SARKAR, S., WOODCOCK, M., ORLANDO, G., MAVROMMATI, I., PAI, C. C., ZALMAS, L. P., DROBNITZKY, N., DIANOV, G. L., VERRILL, C., MACAULAY, V. M., YING, S., LA THANGUE, N. B., D'ANGIOLELLA, V., RYAN, A. J. & HUMPHREY, T. C. 2015. Inhibiting WEE1 Selectively Kills Histone H3K36me3-Deficient Cancers by dNTP Starvation. *Cancer Cell,* 28**,** 557-568.

QIN, H., YU, T., QING, T., LIU, Y., ZHAO, Y., CAI, J., LI, J., SONG, Z., QU, X., ZHOU, P., WU, J., DING, M. & DENG, H. 2007. Regulation of apoptosis and differentiation by p53 in human embryonic stem cells. *J Biol Chem,* 282**,** 5842-52.

RAGLAND, R. L., PATEL, S., RIVARD, R. S., SMITH, K., PETERS, A. A., BIELINSKY, A. K. & BROWN, E. J. 2013. RNF4 and PLK1 are required for replication fork collapse in ATR-deficient cells. *Genes Dev,* 27**,** 2259-73.

RASSEKH, S. R., CHAN, S., HARVARD, C., DIX, D., QIAO, Y. & RAJCAN-SEPAROVIC, E. 2008. Screening for submicroscopic chromosomal rearrangements in Wilms tumor using whole-genome microarrays. *Cancer Genet Cytogenet,* 182**,** 84-94.

ROBINSON, J. T., THORVALDSDÓTTIR, H., WINCKLER, W., GUTTMAN, M., LANDER, E. S., GETZ, G. & MESIROV, J. P. 2011. Integrative genomics viewer. *Nat Biotechnol,* 29**,** 24-6.

RODRIGUEZ, E., HOULDSWORTH, J., REUTER, V. E., MELTZER, P., ZHANG, J., TRENT, J. M., BOSL, G. J. & CHAGANTI, R. S. 1993. Molecular cytogenetic analysis of i(12p)-negative human male germ cell tumors. *Genes Chromosomes Cancer,* 8**,** 230-6.

ROGAKOU, E. P., PILCH, D. R., ORR, A. H., IVANOVA, V. S. & BONNER, W. M. 1998. DNA double-stranded breaks induce histone H2AX phosphorylation on serine 139. *J Biol Chem,* 273**,** 5858-68.

ROSLER, E. S., FISK, G. J., ARES, X., IRVING, J., MIURA, T., RAO, M. S. & CARPENTER, M. K. 2004. Long-term culture of human embryonic stem cells in feeder-free conditions. *Dev Dyn,* 229**,** 259-74.

ROUHANI, F. J., NIK-ZAINAL, S., WUSTER, A., LI, Y., CONTE, N., KOIKE-YUSA, H., KUMASAKA, N., VALLIER, L., YUSA, K. & BRADLEY, A. 2016.

Mutational History of a Human Cell Lineage from Somatic to Induced Pluripotent Stem Cells. *PLoS Genet,* 12**,** e1005932.

RUBIN, C. M. 1998. The Genetic Basis of Human Cancer. *Annals of Internal Medicine,* 129**,** 759-759.

RUIZ, S., LOPEZ-CONTRERAS, A. J., GABUT, M., MARION, R. M., GUTIERREZ-MARTINEZ, P., BUA, S., RAMIREZ, O., OLALDE, I., RODRIGO-PEREZ, S., LI, H., MARQUES-BONET, T., SERRANO, M., BLASCO, M. A., BATADA, N. N. & FERNANDEZ-CAPETILLO, O. 2015. Limiting replication stress during somatic cell reprogramming reduces genomic instability in induced pluripotent stem cells. *Nat Commun,* 6**,** 8036.

SAAVEDRA, H. I., MAITI, B., TIMMERS, C., ALTURA, R., TOKUYAMA, Y., FUKASAWA, K. & LEONE, G. 2003. Inactivation of E2F3 results in centrosome amplification. *Cancer Cell,* 3**,** 333-46.

SAINI, N., RAMAKRISHNAN, S., ELANGO, R., AYYAR, S., ZHANG, Y., DEEM, A., IRA, G., HABER, J. E., LOBACHEV, K. S. & MALKOVA, A. 2013. Migrating bubble during break-induced replication drives conservative DNA synthesis. *Nature,* 502**,** 389-92.

SAINTIGNY, Y., DELACÔTE, F., VARÈS, G., PETITOT, F., LAMBERT, S., AVERBECK, D. & LOPEZ, B. S. 2001. Characterization of homologous recombination induced by replication inhibition in mammalian cells. *EMBO J,* 20**,** 3861-70.

SALK, J. J., SCHMITT, M. W. & LOEB, L. A. 2018. Enhancing the accuracy of next-generation sequencing for detecting rare and subclonal mutations. *Nat Rev Genet,* 19**,** 269-285.

SAN FILIPPO, J., SUNG, P. & KLEIN, H. 2008. Mechanism of eukaryotic homologous recombination. *Annu Rev Biochem,* 77**,** 229-57.

SANJIV, K., HAGENKORT, A., CALDERÓN-MONTAÑO, J. M., KOOLMEISTER, T., REAPER, P. M., MORTUSEWICZ, O., JACQUES, S. A., KUIPER, R. V., SCHULTZ, N., SCOBIE, M., CHARLTON, P. A., POLLARD, J. R., BERGLUND, U. W., ALTUN, M. & HELLEDAY, T. 2016. Cancer-Specific Synthetic Lethality between ATR and CHK1 Kinase Activities. *Cell Rep,* 17**,** 3407-3416.

SCHLAEGER, T. M., DAHERON, L., BRICKLER, T. R., ENTWISLE, S., CHAN, K., CIANCI, A., DEVINE, A., ETTENGER, A., FITZGERALD, K., GODFREY, M.,

GUPTA, D., MCPHERSON, J., MALWADKAR, P., GUPTA, M., BELL, B., DOI, A., JUNG, N., LI, X., LYNES, M. S., BROOKES, E., CHERRY, A. B. C., DEMIRBAS, D., TSANKOV, A. M., ZON, L. I., RUBIN, L. L., FEINBERG, A. P., MEISSNER, A., COWAN, C. A. & DALEY, G. Q. 2014. A comparison of non-integrating reprogramming methods. *Nature Biotechnology,* 33**,** 58.

SCHUBBERT, S., SHANNON, K. & BOLLAG, G. 2007. Hyperactive Ras in developmental disorders and cancer. *Nat Rev Cancer,* 7**,** 295-308.

SCHÖLER, H. R., HATZOPOULOS, A. K., BALLING, R., SUZUKI, N. & GRUSS, P. 1989. A family of octamer-specific proteins present during mouse embryogenesis: evidence for germline-specific expression of an Oct factor. *EMBO J,* 8**,** 2543-50.

SCOTT, W. J., RITTER, E. J. & WILSON, J. G. 1971. DNA synthesis inhibition and cell death associated with hydroxyurea teratogenesis in rat embryos. *Dev Biol,* 26**,** 306-15.

SEKIDO, Y., FONG, K. M. & MINNA, J. D. 1998. Progress in understanding the molecular pathogenesis of human lung cancer. *Biochim Biophys Acta,* 1378**,** F21-59.

SETH, T., BENJAMIN, N., MICHAEL, N., TORI, S.-B., KIMBERLEY, L., ERIK, M. & KAREN, M. 2011. Karyotypic abnormalities in human induced pluripotent stem cells and embryonic stem cells. *Nature Biotechnology,* 29**,** 313-314.

SHERR, C. J. 1996. Cancer cell cycles. *Science,* 274**,** 1672-7.

SHERR, C. J. & MCCORMICK, F. 2002. The RB and p53 pathways in cancer. *Cancer Cell,* 2**,** 103-12.

SHEVINSKY, L. H., KNOWLES, B. B., DAMJANOV, I. & SOLTER, D. 1982. Monoclonal antibody to murine embryos defines a stage-specific embryonic antigen expressed on mouse embryos and human teratocarcinoma cells. *Cell,* 30**,** 697-705.

SHINOHARA, A., OGAWA, H., MATSUDA, Y., USHIO, N., IKEO, K. & OGAWA, T. 1993. Cloning of human, mouse and fission yeast recombination genes homologous to RAD51 and recA. *Nat Genet,* 4**,** 239-43.

SHINOHARA, A., OGAWA, H. & OGAWA, T. 1992. Rad51 protein involved in repair and recombination in S. cerevisiae is a RecA-like protein. *Cell,* 69**,** 457-70.

SHIOTANI, B. & ZOU, L. 2009. Single-stranded DNA orchestrates an ATM-to-ATR switch at DNA breaks. *Mol Cell,* 33**,** 547-58.

SIMARA, P., TESAROVA, L., REHAKOVA, D., MATULA, P., STEJSKAL, S., HAMPL, A. & KOUTNA, I. 2017. DNA double-strand breaks in human induced pluripotent stem cell reprogramming and long-term in vitro culturing. *Stem Cell Res Ther,* 8**,** 73.

SIMONS, A., SHAFFER, L. G. & HASTINGS, R. J. 2013. Cytogenetic Nomenclature: Changes in the ISCN 2013 Compared to the 2009 Edition. *Cytogenet Genome Res,* 141**,** 1-6.

SIMPSON, N. E. 1988. The map of chromosome 20. *J Med Genet,* 25**,** 794-804.

SKOTHEIM, R. I., MONNI, O., MOUSSES, S., FOSSÅ, S. D., KALLIONIEMI, O. P., LOTHE, R. A. & KALLIONIEMI, A. 2002. New insights into testicular germ cell tumorigenesis from gene expression profiling. *Cancer Res,* 62**,** 2359-64.

SMITH, C. E., LLORENTE, B. & SYMINGTON, L. S. 2007. Template switching during break-induced replication. *Nature,* 447**,** 102-5.

SMITH, G. C. & JACKSON, S. P. 1999. The DNA-dependent protein kinase. *Genes Dev,* 13**,** 916-34.

SONG, Y., MARMION, R. A., PARK, J. O., BISWAS, D., RABINOWITZ, J. D. & SHVARTSMAN, S. Y. 2017. Dynamic Control of dNTP Synthesis in Early Embryos. *Dev Cell,* 42**,** 301-308.e3.

STEAD, E., WHITE, J., FAAST, R., CONN, S., GOLDSTONE, S., RATHJEN, J., DHINGRA, U., RATHJEN, P., WALKER, D. & DALTON, S. 2002. Pluripotent cell division cycles are driven by ectopic Cdk2, cyclin A/E and E2F activities. *Oncogene,* 21**,** 8320-33.

STEINEMANN, D., GÖHRING, G. & SCHLEGELBERGER, B. 2013. Genetic instability of modified stem cells - a first step towards malignant transformation? *Am J Stem Cells,* 2**,** 39-51.

STEVENS, C., SMITH, L. & LA THANGUE, N. B. 2003. Chk2 activates E2F-1 in response to DNA damage. *Nat Cell Biol,* 5**,** 401-9.

STEVENS, L. C. & LITTLE, C. C. 1954. Spontaneous Testicular Teratomas in an Inbred Strain of Mice. *Proc Natl Acad Sci U S A,* 40**,** 1080-7.

SUGIURA, M., KASAMA, Y., ARAKI, R., HOKI, Y., SUNAYAMA, M., UDA, M., NAKAMURA, M., ANDO, S. & ABE, M. 2014. Induced Pluripotent Stem Cell Generation-Associated Point Mutations Arise during the Initial Stages of the Conversion of These Cells. *Stem Cell Reports,* 2**,** 52-63.

SØRENSEN, C. S. & SYLJUÅSEN, R. G. 2012. Safeguarding genome integrity: the checkpoint kinases ATR, CHK1 and WEE1 restrain CDK activity during normal DNA replication. *Nucleic Acids Res,* 40**,** 477-86.

SØRENSEN, C. S., SYLJUÅSEN, R. G., FALCK, J., SCHROEDER, T., RÖNNSTRAND, L., KHANNA, K. K., ZHOU, B. B., BARTEK, J. & LUKAS, J. 2003. Chk1 regulates the S phase checkpoint by coupling the physiological turnover and ionizing radiation-induced accelerated proteolysis of Cdc25A. *Cancer Cell,* 3**,** 247-58.

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**,** 861-72.

TAKAHASHI, K. & YAMANAKA, S. 2006. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell,* 126**,** 663-76.

TAKAI, H., NAKA, K., OKADA, Y., WATANABE, M., HARADA, N., SAITO, S., ANDERSON, C. W., APPELLA, E., NAKANISHI, M., SUZUKI, H., NAGASHIMA, K., SAWA, H., IKEDA, K. & MOTOYAMA, N. 2002. Chk2-deficient mice exhibit radioresistance and defective p53-mediated transcription. *EMBO J,* 21**,** 5195-205.

TAKAI, H., TOMINAGA, K., MOTOYAMA, N., MINAMISHIMA, Y. A., NAGAHAMA, H., TSUKIYAMA, T., IKEDA, K., NAKAYAMA, K. & NAKANISHI, M. 2000. Aberrant cell cycle checkpoint function and early embryonic death in Chk1(-/-) mice. *Genes Dev,* 14**,** 1439-47.

TANG, J., ERIKSON, R. L. & LIU, X. 2006. Checkpoint kinase 1 (Chk1) is required for mitotic progression through negative regulation of polo-like kinase 1 (Plk1). *Proc Natl Acad Sci U S A,* 103**,** 11964-9.

TANIGUCHI, T., GARCIA-HIGUERA, I., XU, B., ANDREASSEN, P. R., GREGORY, R. C., KIM, S. T., LANE, W. S., KASTAN, M. B. & D'ANDREA, A. D. 2002. Convergence of the fanconi anemia and ataxia telangiectasia signaling pathways. *Cell,* 109**,** 459-72.

TANNER, M. M., TIRKKONEN, M., KALLIONIEMI, A., ISOLA, J., KUUKASJÄRVI, T., COLLINS, C., KOWBEL, D., GUAN, X.-Y., TRENT, J., GRAY, J. W., MELTZER, P. & KALLIONIEMI, O.-P. 1996. Independent Amplification and

Frequent Co-Amplification of Three Nonsyntenic Regions on the Long Arm of Chromosome 20 in Human Breast Cancer. *Cancer Research,* 56**,** 3441.

TARASOV, A., VILELLA, A. J., CUPPEN, E., NIJMAN, I. J. & PRINS, P. 2015. Sambamba: fast processing of NGS alignment formats. *Bioinformatics,* 31**,** 2032-4.

TAYLOR, J. H. 1977. Increase in DNA replication sites in cells held at the beginning of S phase. *Chromosoma,* 62**,** 291-300.

THOMPSON, O., VON MEYENN, F., HEWITT, Z., ALEXANDER, J., WOOD, A., WEIGHTMAN, R., GREGORY, S., KRUEGER, F., ANDREWS, S., BARBARIC, I., GOKHALE, P. J., MOORE, H. D., REIK, W., MILO, M., NIK-ZAINAL, S., YUSA, K. & ANDREWS, P. W. 2019. Low rates of acquisition of de novo mutations in human pluripotent stem cells under different culture conditions.

THOMPSON, O., VON MEYENN, F., HEWITT, Z., ALEXANDER, J., WOOD, A., WEIGHTMAN, R., GREGORY, S., KRUEGER, F., ANDREWS, S., BARBARIC, I., GOKHALE, P. J., MOORE, H. D., REIK, W., MILO, M., NIK-ZAINAL, S., YUSA, K. & ANDREWS, P. W. 2020. Low rates of acquisition of de novo mutations in human pluripotent stem cells under different culture conditions.

THOMSON, J. A., ITSKOVITZ-ELDOR, J., SHAPIRO, S. S., WAKNITZ, M. A., SWIERGIEL, J. J., MARSHALL, V. S. & JONES, J. M. 1998. Embryonic stem cell lines derived from human blastocysts. *Science,* 282**,** 1145-7.

TOLEDO, L. I., ALTMEYER, M., RASK, M. B., LUKAS, C., LARSEN, D. H., POVLSEN, L. K., BEKKER-JENSEN, S., MAILAND, N., BARTEK, J. & LUKAS, J. 2013. ATR prohibits replication catastrophe by preventing global exhaustion of RPA. *Cell,* 155**,** 1088-103.

TOMASETTI, C. & VOGELSTEIN, B. 2015. Cancer etiology. Variation in cancer risk among tissues can be explained by the number of stem cell divisions. *Science,* 347**,** 78-81.

TONON, G., WONG, K.-K., MAULIK, G., BRENNAN, C., FENG, B., ZHANG, Y., KHATRY, D. B., PROTOPOPOV, A., YOU, M. J., AGUIRRE, A. J., MARTIN, E. S., YANG, Z., JI, H., CHIN, L. & DEPINHO, R. A. 2005. High-resolution genomic profiles of human lung cancer. *Proceedings of the National Academy of Sciences of the United States of America,* 102**,** 9625.

TRASK, B. J. 2002. Human cytogenetics: 46 chromosomes, 46 years and counting. *Nat Rev Genet,* 3**,** 769-78.

UMEZU, K., HIRAOKA, M., MORI, M. & MAKI, H. 2002. Structural analysis of aberrant chromosomes that occur spontaneously in diploid Saccharomyces cerevisiae: retrotransposon Ty1 plays a crucial role in chromosomal rearrangements. *Genetics,* 160**,** 97-110.

VALLABHANENI, H., LYNCH, P. J., CHEN, G., PARK, K., LIU, Y., GOEHE, R., MALLON, B. S., BOEHM, M. & HURSH, D. A. 2018. High Basal Levels of γH2AX in Human Induced Pluripotent Stem Cells Are Linked to Replication-Associated DNA Damage and Repair. *Stem Cells,* 36**,** 1501-1513.

VALLI, R., MARLETTA, C., PRESSATO, B., MONTALBANO, G., LO CURTO, F., PASQUALI, F. & MASERATI, E. 2011. Comparative genomic hybridization on microarray (a-CGH) in constitutional and acquired mosaicism may detect as low as 8% abnormal cells. *Mol Cytogenet,* 4**,** 13.

VASTAG, L., JORGENSEN, P., PESHKIN, L., WEI, R., RABINOWITZ, J. D. & KIRSCHNER, M. W. 2011. Remodeling of the metabolome during early frog development. *PLoS One,* 6**,** e16881.

VASTENHOUW, N. L., CAO, W. X. & LIPSHITZ, H. D. 2019. The maternal-to-zygotic transition revisited. *Development,* 146**,** dev161471.

VELASCO, G., GRKOVIC, S. & ANSIEAU, S. 2006. New insights into BS69 functions. *J Biol Chem,* 281**,** 16546-50.

VITALE, I., GALLUZZI, L., VIVET, S., NANTY, L., DESSEN, P., SENOVILLA, L., OLAUSSEN, K. A., LAZAR, V., PRUDHOMME, M., GOLSTEYN, R. M., CASTEDO, M. & KROEMER, G. 2007. Inhibition of Chk1 kills tetraploid tumor cells through a p53-dependent pathway. *PLoS One,* 2**,** e1337.

VOET, T., KUMAR, P., VAN LOO, P., COOKE, S. L., MARSHALL, J., LIN, M. L., ZAMANI ESTEKI, M., VAN DER AA, N., MATEIU, L., MCBRIDE, D. J., BIGNELL, G. R., MCLAREN, S., TEAGUE, J., BUTLER, A., RAINE, K., STEBBINGS, L. A., QUAIL, M. A., D'HOOGHE, T., MOREAU, Y., FUTREAL, P. A., STRATTON, M. R., VERMEESCH, J. R. & CAMPBELL, P. J. 2013. Single-cell paired-end genome sequencing reveals structural variation per cell cycle. *Nucleic Acids Res,* 41**,** 6119-38.

VOINEAGU, I., NARAYANAN, V., LOBACHEV, K. S. & MIRKIN, S. M. 2008. Replication stalling at unstable inverted repeats: Interplay between DNA

hairpins and fork stabilizing proteins. *Proceedings of the National Academy of Sciences,* 105**,** 9936.

WALKER, J. R., CORPINA, R. A. & GOLDBERG, J. 2001. Structure of the Ku heterodimer bound to DNA and its implications for double-strand break repair. *Nature,* 412**,** 607-14.

WALTER, D., LIER, A., GEISELHART, A., THALHEIMER, F. B., HUNTSCHA, S., SOBOTTA, M. C., MOEHRLE, B., BROCKS, D., BAYINDIR, I., KASCHUTNIG, P., MUEDDER, K., KLEIN, C., JAUCH, A., SCHROEDER, T., GEIGER, H., DICK, T. P., HOLLAND-LETZ, T., SCHMEZER, P., LANE, S. W., RIEGER, M. A., ESSERS, M. A., WILLIAMS, D. A., TRUMPP, A. & MILSOM, M. D. 2015. Exit from dormancy provokes DNA-damage-induced attrition in haematopoietic stem cells. *Nature,* 520**,** 549-52.

WANG, J., QIN, S., LI, F., LI, S., ZHANG, W., PENG, J., ZHANG, Z., GONG, Q., WU, J. & SHI, Y. 2014. Crystal structure of human BS69 Bromo-ZnF-PWWP reveals its role in H3K36me3 nucleosome binding. *Cell Res,* 24**,** 890-3.

WANG, K., GUZMAN, A. K., YAN, Z., ZHANG, S., HU, M. Y., HAMANEH, M. B., YU, Y. K., TOLU, S., ZHANG, J., KANAVY, H. E., YE, K., BARTHOLDY, B. & BOUHASSIRA, E. E. 2019. Ultra-High-Frequency Reprogramming of Individual Long-Term Hematopoietic Stem Cells Yields Low Somatic Variant Induced Pluripotent Stem Cells. *Cell Rep,* 26**,** 2580-2592.e7.

WANG, S., HASSOLD, T., HUNT, P., WHITE, M. A., ZICKLER, D., KLECKNER, N. & ZHANG, L. 2017. Inefficient Crossover Maturation Underlies Elevated Aneuploidy in Human Female Meiosis. *Cell,* 168**,** 977-989.e17.

WARREN, L., MANOS, P. D., AHFELDT, T., LOH, Y.-H., LI, H., LAU, F., EBINA, W., MANDAL, P. K., SMITH, Z. D., MEISSNER, A., DALEY, G. Q., BRACK, A. S., COLLINS, J. J., COWAN, C., SCHLAEGER, T. M. & ROSSI, D. J. 2010. Highly efficient reprogramming to pluripotency and directed differentiation of human cells with synthetic modified mRNA. *Cell stem cell,* 7**,** 618-630.

WATANABE, K., UENO, M., KAMIYA, D., NISHIYAMA, A., MATSUMURA, M., WATAYA, T., TAKAHASHI, J. B., NISHIKAWA, S., MUGURUMA, K. & SASAI, Y. 2007. A ROCK inhibitor permits survival of dissociated human embryonic stem cells. *Nat Biotechnol,* 25**,** 681-6.

WERBOWETSKI-OGILVIE, T. E., BOSSÉ, M., STEWART, M., SCHNERCH, A., RAMOS-MEJIA, V., ROULEAU, A., WYNDER, T., SMITH, M. J., DINGWALL,

S., CARTER, T., WILLIAMS, C., HARRIS, C., DOLLING, J., WYNDER, C., BOREHAM, D. & BHATIA, M. 2009. Characterization of human embryonic stem cells with features of neoplastic progression. *Nat Biotechnol,* 27**,** 91-7.

WHITE, J. & DALTON, S. 2005. Cell cycle control of embryonic stem cells. *Stem Cell Rev,* 1**,** 131-8.

WHITE, J., STEAD, E., FAAST, R., CONN, S., CARTWRIGHT, P. & DALTON, S. 2005. Developmental activation of the Rb-E2F pathway and establishment of cell cycle-regulated cyclin-dependent kinase activity during embryonic stem cell differentiation. *Mol Biol Cell,* 16**,** 2018-27.

WILES, M. V. & JOHANSSON, B. M. 1999. Embryonic stem cell development in a chemically defined medium. *Exp Cell Res,* 247**,** 241-8.

WILLIAMS, B. P., DANIELS, G. L., PYM, B., SHEER, D., POVEY, S., OKUBO, Y., ANDREWS, P. W. & GOODFELLOW, P. N. 1988. Biochemical and genetic analysis of the OKa blood group antigen. *Immunogenetics,* 27**,** 322-9.

WILLIAMS, R. S., MONCALIAN, G., WILLIAMS, J. S., YAMADA, Y., LIMBO, O., SHIN, D. S., GROOCOCK, L. M., CAHILL, D., HITOMI, C., GUENTHER, G., MOIANI, D., CARNEY, J. P., RUSSELL, P. & TAINER, J. A. 2008. Mre11 dimers coordinate DNA end bridging and nuclease processing in double-strand-break repair. *Cell,* 135**,** 97-109.

WILLIS, A., JUNG, E. J., WAKEFIELD, T. & CHEN, X. 2004. Mutant p53 exerts a dominant negative effect by preventing wild-type p53 from binding to the promoter of its target genes. *Oncogene,* 23**,** 2330-8.

WINDHOFER, F., WU, W., WANG, M., SINGH, S. K., SAHA, J., ROSIDI, B. & ILIAKIS, G. 2007. Marked dependence on growth state of backup pathways of NHEJ. *Int J Radiat Oncol Biol Phys,* 68**,** 1462-70.

WONG, C. C., LOEWKE, K. E., BOSSERT, N. L., BEHR, B., DE JONGE, C. J., BAER, T. M. & REIJO PERA, R. A. 2010. Non-invasive imaging of human embryos before embryonic genome activation predicts development to the blastocyst stage. *Nat Biotechnol,* 28**,** 1115-21.

WU, S. M. & HOCHEDLINGER, K. 2011. Harnessing the potential of induced pluripotent stem cells for regenerative medicine. *Nat Cell Biol,* 13**,** 497-505.

WU, W., WANG, M., SINGH, S. K., MUSSFELDT, T. & ILIAKIS, G. 2008. Repair of radiation induced DNA double strand breaks by backup NHEJ is enhanced in G2. *DNA Repair (Amst),* 7**,** 329-38.

WYMAN, C., RISTIC, D. & KANAAR, R. 2004. Homologous recombination-mediated double-strand break repair. *DNA Repair (Amst),* 3**,** 827-33.

YAN, Z. A., LI, X. Z. & ZHOU, X. T. 1987. The effect of hydroxyurea on the expression of the common fragile site at 3p14. *J Med Genet,* 24**,** 593-6.

YANG, D., WELM, A. & BISHOP, J. M. 2004. Cell division and cell survival in the absence of survivin. *Proc Natl Acad Sci U S A,* 101**,** 15100-5.

YE, S., TAN, L., YANG, R., FANG, B., QU, S., SCHULZE, E. N., SONG, H., YING, Q. & LI, P. 2012. Pleiotropy of glycogen synthase kinase-3 inhibition by CHIR99021 promotes self-renewal of embryonic stem cells from refractory mouse strains. *PLoS One,* 7**,** e35892.

ZACHOS, G., BLACK, E. J., WALKER, M., SCOTT, M. T., VAGNARELLI, P., EARNSHAW, W. C. & GILLESPIE, D. A. 2007. Chk1 is required for spindle checkpoint function. *Dev Cell,* 12**,** 247-60.

ZAGORDI, O., KLEIN, R., DÄUMER, M. & BEERENWINKEL, N. 2010. Error correction of next-generation sequencing data and reliable estimation of HIV quasispecies. *Nucleic Acids Res,* 38**,** 7400-9.

ZEGERMAN, P. & DIFFLEY, J. F. 2009. DNA replication as a target of the DNA damage checkpoint. *DNA Repair (Amst),* 8**,** 1077-88.

ZEMAN, M. K. & CIMPRICH, K. A. 2014. Causes and consequences of replication stress. *Nat Cell Biol,* 16**,** 2-9.

ZHANG, J., HIRST, A. J., DUAN, F., QIU, H., HUANG, R., JI, Y., BAI, L., ZHANG, F., ROBINSON, D., JONES, M., LI, L., WANG, P., JIANG, P., ANDREWS, P. W., BARBARIC, I. & NA, J. 2019. Anti-apoptotic Mutations Desensitize Human Pluripotent Stem Cells to Mitotic Stress and Enable Aneuploid Cell Survival. *Stem Cell Reports,* 12**,** 557-571.

ZHANG, J., WILLERS, H., FENG, Z., GHOSH, J. C., KIM, S., WEAVER, D. T., CHUNG, J. H., POWELL, S. N. & XIA, F. 2004. Chk2 phosphorylation of BRCA1 regulates DNA double-strand break repair. *Mol Cell Biol,* 24**,** 708-18.

ZHANG, L., KENDRICK, C., JÜLICH, D. & HOLLEY, S. A. 2008. Cell cycle progression is required for zebrafish somite morphogenesis but not segmentation clock function. *Development,* 135**,** 2065-70.

ZHANG, M., KOTHARI, P., MULLINS, M. & LAMPSON, M. A. 2014. Regulation of zygotic genome activation and DNA damage checkpoint acquisition at the mid-blastula transition. *Cell Cycle,* 13**,** 3828-38.

ZHANG, X., NEGANOVA, I., PRZYBORSKI, S., YANG, C., COOKE, M., ATKINSON, S. P., ANYFANTIS, G., FENYK, S., KEITH, W. N., HOARE, S. F., HUGHES, O., STRACHAN, T., STOJKOVIC, M., HINDS, P. W., ARMSTRONG, L. & LAKO, M. 2009. A role for NANOG in G1 to S transition in human embryonic stem cells through direct binding of CDK6 and CDC25A. *J Cell Biol,* 184**,** 67-82.

ZHANG, Y. & HUNTER, T. 2014. Roles of Chk1 in cell biology and cancer therapy. *Int J Cancer,* 134**,** 1013-23.

ZHAO, H., WATKINS, J. L. & PIWNICA-WORMS, H. 2002. Disruption of the checkpoint kinase 1/cell division cycle 25A pathway abrogates ionizing radiation-induced S and G2 checkpoints. *Proc Natl Acad Sci U S A,* 99**,** 14795-800.

ZWEIG, M. H. & CAMPBELL, G. 1993. Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. *Clin Chem,* 39**,** 561-77.