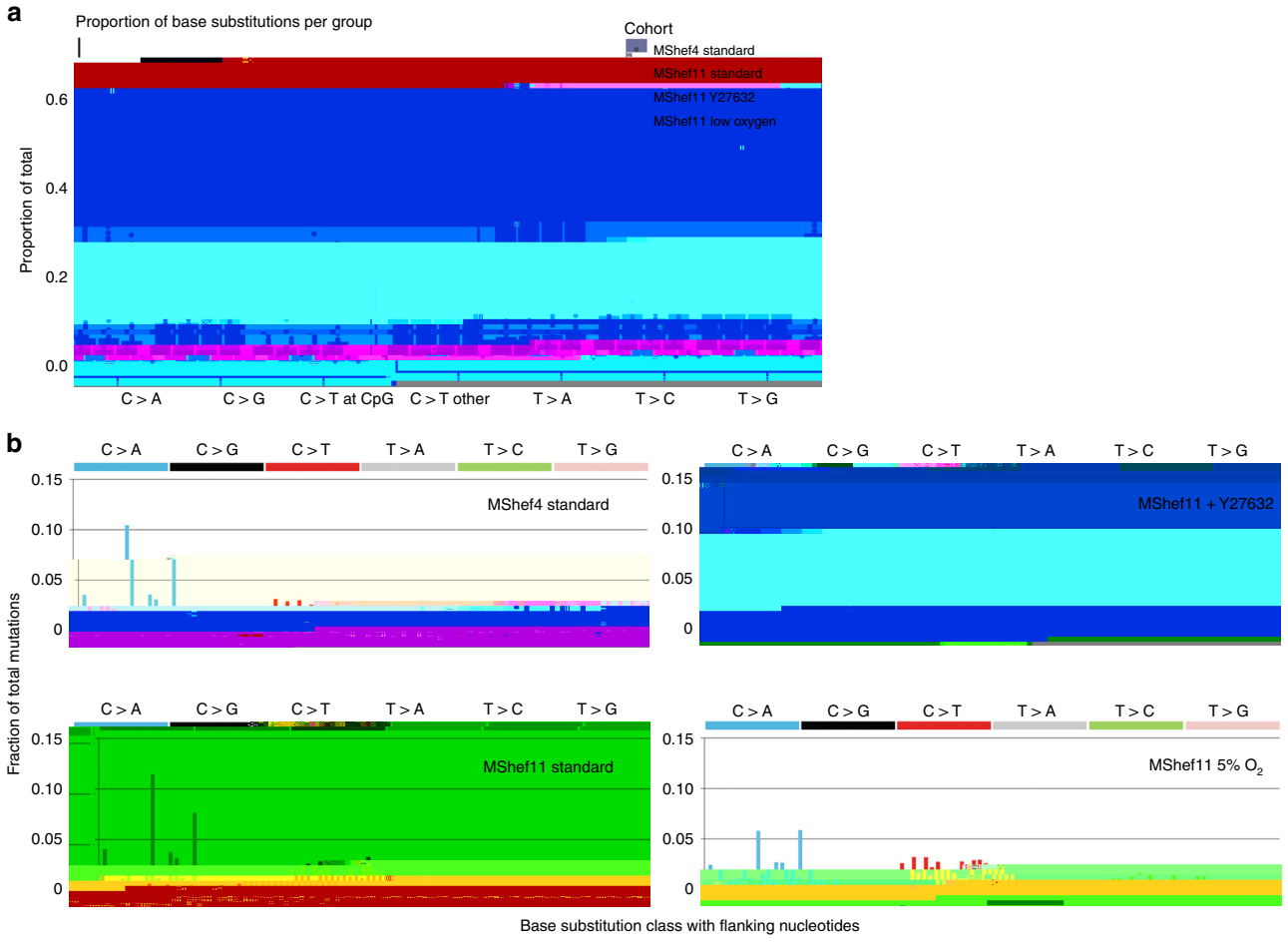
The presence of mutations in human pluripotent stem cells (PSC), whether embryonic stem (ES) cells or induced pluripotent stem (iPS) cells, is a concern for their safe use in therapeutic applications. Indeed, in one case, a potential trial of retinal pigment cells from an autologous iPS cell line was abandoned because the cells carried a mutation of unknown significance[1]. Certainly some such variants are likely to have been present in the embryos or somatic cells from which particular PSC were derived and can be classed as 'variants of origin'[2,3]. However, the propensity of PSC to acquire genetic variants on prolonged passage poses additional concerns, not only because of the diffi

over a variety of histone marks were also calculated (Supplementary Fig. 5b). In the case of three types of histone mark (H3K4me1, H3K4me3, and H3K36me3), MShef4 had slightly



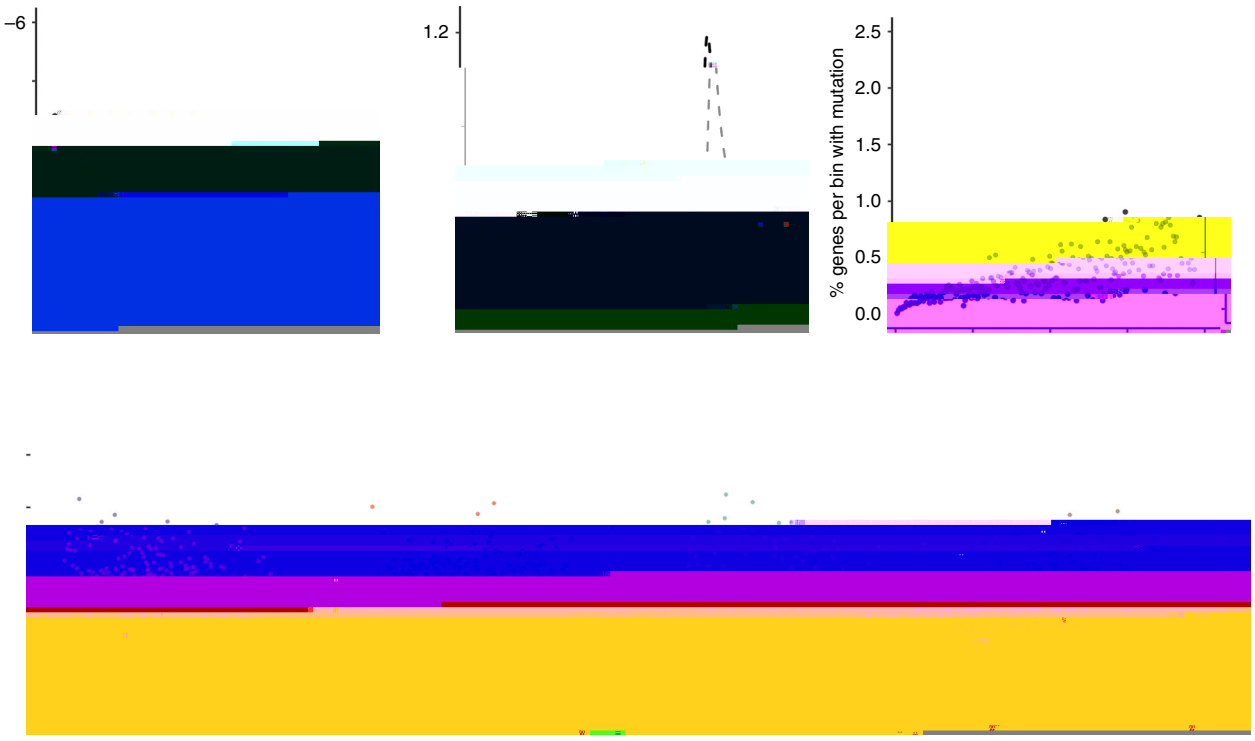
. Ba ra car -r l a al cara r ac f v a cal c S 4 a ar (N = 20), S 11 a ar (N = 1), S 11 + Y27632 (N = 20) a S 11 xy (N = 21) cl Ba r ca a r ac cal f a rac c ll / r e r ca v al cl ac r . A ar v r a a ar rval aa . A a l t- S c l r xy av a fia r c A ar v r c ar a arc (P = 0.037) l Wd x al a a ll cra ar (P = 0.045). -r l a al car w f c a aaa ll cl r ac r e . ac f x l a v 16cal a f 5'a 3' cl fil a , r l 6 l cal . A a A a ar rval aa . ca ca c rr al 30 a a r a a aa l S ac S ac r (S) aaa [53] x l r a l y f a . r ar dry ll r e r a r a a l r fil . S a S c l r a ar c x dr a a l r fil ra xy c dr a r filc ar a ll r r S c aaa r rv S l ay aa 2.

cell lines and growth conditions, with a predominance of C > A transversions (particularly in the C*A* and C context) and C > T transitions (Fig. 3b). When we compared the mutational profile of each condition with 30 annotated in the Catalogue of Somatic Mutations in Cancer (COSMIC) database, our data correlated most closely with COSMIC signature 18 (cosine correlation = 0.873), which is associated with oxidative stress and is a hallmark of in vitro cell culture[25,26]. The reduced C > A component observed in MShef11 under low oxygen fits with a model of reduced oxidative stress, as oxidative species predominantly affect guanine (which is captured in the C > A class). Similarity-clustering showed that MShef4 and MShef11 grown in standard conditions acquired the most similar mutation profiles, followed by MShef11 cultured with Y27632, and the low oxygen condition as an outgroup (Fig. 3c) indicating that low oxygen culture induced the largest difference in mutational profile of all conditions tested.

. To detect INDELs we used the *D* INDEL-calling algorithm (github.com/genome/pindel). Following quality control and excluding calls with length greater than 100 bp, as well as 291 calls that were recurrent in subclones of the same cohort, a total of 1171 de novo INDELs remained for analysis including single-base INDELs [572 insertions; 573 deletions; 26 complex] (Supplementary Figures 1–4; Supplementary Data 3). Out of 1171 INDELs, 578 are single-base INDELs of which 350 are insertions and 228 are deletions. Taken together, the median INDEL mutation rate was ~10-fold lower than that of base-pair substitutions (Fig. 4a). As with SNVs, we observed a lower mutation rate in MShef11 grown in low oxygen compared to MShef11 grown under standard conditions (0.15×10^{-10} vs 0.26×10^{-10} INDELs per day, per base-pair; = 0.02) and we observed a significant difference in median deletion mutation rates between MShef11 grown under low oxygen vs. MShef11 grown in standard conditions (0.05×10^{-10} vs 0.16×10^{-10} deletions per day, per base-pair; = 0.0008) (Fig. 4a, b). We detected no systematic deviation in INDEL mutation rates per chromosome (Supplementary Fig. 6a), and we did not find any enrichment of INDELs at regions associated with common recurrent change in human PSC (for example chromosomes 1q, 12p, 17q, 20q). Due to the overall low number of INDEL mutations, it was not possible to make further meaningful analysis of rates within different regions of the genome.

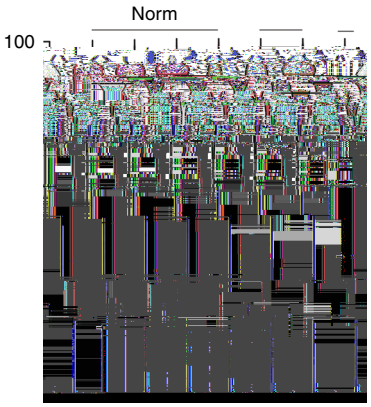
All growth-condition groups showed some evidence of larger structural rearrangements. Using the BRASS structural rearrangement-calling algorithm (www.github.com/cancerit/BRASS

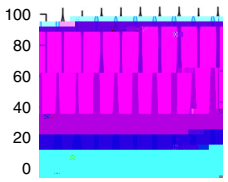
assembly (including those unannotated) (Supplementary Data 1) 4694 mutations mapped to introns and 225 SNVs occuured in exons resulting in 90 missense, 7 nonsense and 27 synonymous ammino accid substitutions. A further 985 SNVs mapped to regions within 1 kb up- or downstream of genes, 114 to 5' and 3' untranslated regions (UTRs), 27 to genes encoding non-coding

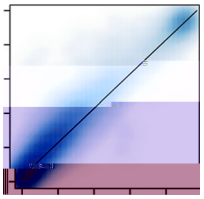


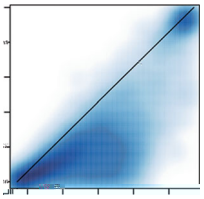
both parental clones grown in low oxygen (G8 and G2) had a significant decrease in methylation compared to those grown in standard conditions (both $= < 0.001$), with a particularly large decrease in G8 and a mean reduction of 3.6% (Fig. 6a). The reduction in methylation seen under low oxygen appeared reversible, as the subclones derived from G2 and G8 in low oxygen but expanded under standard conditions prior to WGBS showed reversion back to 91% methylation, equivalent to their counterparts from standard conditions. (Fig. 6a). Overall we found a weak relationship between promoter methylation and gene expression.

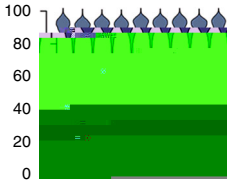
MShef4 subclones (J1-20) as a group showed striking hypermethylation of CpG island-containing (CGI) promoters compared to MShef11 subclones (Supplementary Fig. 8). To test if this effect was due to a single aberrant subclone skewing the grouped data, or if the hypermethylation had occurred in many independent subclones, we used higher-depth WGBS data from a subset of seven independent MShef4 subclones (Supplementary Data 10, 11) and compared the methylation levels of each to their parental clone, B8. Each subclone showed substantial hypermethylation of a subset of CpG island-containing promoters (Fig. 6b, top panels; Supplementary Fig. 9, left panels). This showed that the effect was not restricted to a single aberrant subclone and had likely occurred in all subclones in the cohort following single-cell deposition and expansion. Of the 11,677 CGI promoters measured, 1905 were hypermethylated by at least 20% across all seven subclones (Supplementary Data 12) although the levels were highly variable. However, only two CGI-containing promoters were hypermethylated by an equivalent level (±5%) across all subclones (these were directly upstream of the long non-coding RNA *236383.12* and the ribosomal gene *A5-8 5*, located on chromosomes 21 and 22, respectively) indicating no consistency in the patterns of hypermethylation and no common predecessor with the epigenetic alterations. In contrast, methylation of non-CGI promoters was similar between each subclone and its parental clone, B8 (Fig. 6b, bottom panels; Supplementary Fig. 9, right panels). We measured the expression of de novo methyltransferase genes (DNMTs) in MShef4 subclones (Supplementary Data 13) and found significantly elevated expression of *D 3B* and *D 3* ($= 7.2^{-8}$; $= 2^{-8}$) compared to MShef11 subclones (Fig. 6c). Both genes are involved in de novo methylation, and their relative overexpression in MShef4 subclones may account for the observed hypermethylation of CpG island-containing promoters.

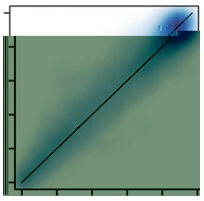

Norm
100

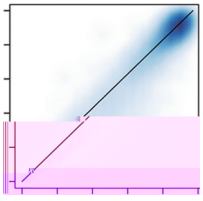

100
80
60
40
20
0

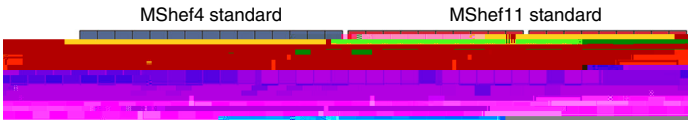


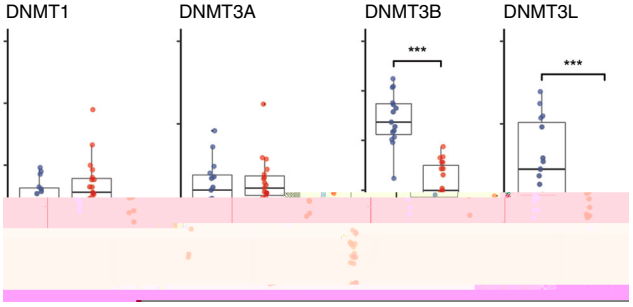



100
80
60
40
20
0






MShef4 standard
MShef11 standard


DNMT1
DNMT3A
DNMT3B
DNMT3L
***
***

. Given the

between MShef11 subclones, which showed greater variation within, and between growth-condition groups (Supplementary Fig. 12). In a set of subclones that showed significant variation in imprint expression, and for which sufficient WGBS depth was

other, making it difficult to accurately assess the effect of this condition on DNA methylation. We also saw significant hypermethylation of CpG island-containing promoters following subcloning of MShef4. This hypermethylation had occurred in multiple independent subclones, suggesting a global effect across the entire cohort. This event might be a stress-induced response to single-cell cloning and is worth considering in the context of PSC maintenance, given the dynamic nature of epigenetic change in PSCs. Finally, we saw a significant variation in the expression of imprinted genes and the methylation status of imprint control regions (ICRs). Overall, we detected variability in ICR methylation in subclones, compared to their parental clones or the starting cultures from which they were derived. MShef4 subclones exhibited a different pattern of imprint expression to MShef11 subclones, with little variation between subclones. For some imprints, MShef11 subclones showed significant subclonal variation, and we were able to detect specific instances in which changes in ICR methylation correlated with changes in accompanying imprint expression. These findings suggest that DNA methylation is highly dynamic, responsive to growth-conditions and culture practices, and can vary in a cell-line-specific manner. However, as we did not find a correlation between epigenetic change and mutation, the two may be unrelated forms of variation.

Overall, the striking conclusion from this study is the low mutation rate in human PSC, whether affecting SNV or INDELS, despite the frequent reports of common genetic variants in the literature. Most likely, the latter reflects an ascertainment bias. In the ISCI study of 120 pairs of human PSC in early and late passage, 79 lines remained karyotypically normal[2] while in a sequencing study of 140 human ES cell lines[12], only six acquired mutations in TP53, all results consistent with an underlying low mutation rate in human PSC. Of course, one unknown is whether PSC lines that have acquired growth advantages through long periods in culture may have an altered mutation rate, perhaps a mutator phenotype. Evidently, the mutation burden in human PSC can be reduced by culture conditions, such in a low oxygen environment, but it seems that the appearance of common variants is largely a consequence of selection rather than underlying mutation. Minimising the appearance of such variants will, then, depend primarily upon identification and moderating of the mechanisms by which they exhibit a growth advantage.

. Derivation and maintenance of the MShef4 and MShef11 cell lines was performed in the Sheffield Centre for Stem Cell Biology under HFEA licence R0115-8-A (Centre 0191) and HTA licence 22510, in a clean room setting, following strict standard operating procedures. The embryos used to derive MShef4 (frozen embryo) and MShef11 (fresh embryo) were donated from different Assisted Conception Units, and so likely from different donors, following fully informed consent, with no financial benefit to the donors, and were surplus or unsuitable for their IVF treatment. Briefly, the embryos, were cultured using standard IVF culture media (Medicult), to the blastocyst stage. Following removal of the trophectoderm using a dissection laser the embryos were explanted whole onto either mitotically inactivated human neonatal fibroblasts (human feeders) in standard KOSR/KODMEM (Life Technologies) medium in the case of MShef4 or onto Laminin-511 (Biolamina) and Nutristem medium (Biological Industries) in the case of MShef11. Both cell lines were initially maintained at 37ºC under 5% O_2/5% CO_2, until the lines were established, after which maintenance switched 5% CO_2 in air at 37 °C. Cultures were passaged using a manual technique, cutting selected colonies under a dissection microscope at an average split ratio of 1:2 every 7 days. Both lines have been deposited at the UK Stem Cell bank.

Parental material for these experiments was taken from fully characterised research bank frozen stocks known to contain high quality undifferentiated human ES cells, with a normal 46, XY karyotype frozen at passage P36 (MShef4) and P15 (MShef11). Cells were thawed onto mitomycin C-inactivated human fibroblast feeders, maintained in Nutristem and passaged using the manual cutting technique. After expansion for seven at pasiing