Article Human induced pluripotent stem cells display a similar mutation burden as embryonic pluripotent cells in vivo
Haaart, K.A.L.; Manders, F.; Ubels, J.; Verheul, M.; Roosmalen, M.J. van; Groenen, N.M.; ...; Boxtel, R. van

Citation
Haaart, K. A. L., Manders, F., Ubels, J., Verheul, M., Roosmalen, M. J. van, Groenen, N. M., ... Boxtel, R. van. (2022). Article Human induced pluripotent stem cells display a similar mutation burden as embryonic pluripotent cells in vivo. Iscience, 25(2).
doi:10.1016/j.isci.2022.103736

Version: Publisher's Version
License: Creative Commons CC BY-NC-ND 4.0 license
Downloaded from: https://hdl.handle.net/1887/3563160

Note: To cite this publication please use the final published version (if applicable).
Human induced pluripotent stem cells display a similar mutation burden as embryonic pluripotent cells in vivo

Highlights

- Shared mutations in fetal cells allow pre-gastrulation mutation rate estimation
- Human pluripotent cells have an in vivo mutation rate of 1.65 mutations per division
- iPSCs generated from fetal cells show a similar mutation rate in hypoxic conditions
- Similar mutational processes are active in iPSCs and pre-gastrulation embryonic cells
Human induced pluripotent stem cells display a similar mutation burden as embryonic pluripotent cells in vivo

Karlijn A.L. Hasaart, 1,2 Freek Manders, 1,2 Joske Ubels, 1,2 Mark Verheul, 1,2 Markus J. van Roosmalen, 1,2 Niels M. Groenen, 1,2 Rurika Oka, 1,2 Ewart Kuijk, 2,3 Susana M. Chuva de Sousa Lopes, 4 and Ruben van Boxtel 1,2,5,*

SUMMARY
Induced pluripotent stem cells (iPSCs) hold great promise for regenerative medicine, but genetic instability is a major concern. Embryonic pluripotent cells also accumulate mutations during early development, but how this relates to the mutation burden in iPSCs remains unknown. Here, we directly compared the mutation burden of cultured iPSCs with their isogenic embryonic cells during human embryogenesis. We generated developmental lineage trees of human fetuses by phylogenetic inference from somatic mutations in the genomes of multiple stem cells, which were derived from different germ layers. Using this approach, we characterized the mutations acquired pre-gastrulation and found a rate of 1.65 mutations per cell division. When cultured in hypoxic conditions, iPSCs generated from fetal stem cells of the assessed fetuses displayed a similar mutation rate and spectrum. Our results show that iPSCs maintain a genomic integrity during culture at a similar degree as their pluripotent counterparts do in vivo.

INTRODUCTION
Human pluripotent stem cells resemble the pluripotent epiblast of the early post-implantation embryo in the capacity to differentiate toward all three germ layers and any somatic cell type (Courtot et al., 2014; Lau et al., 2020; Yamanaka, 2020). Human pluripotent stem cell lines can be generated from the inner cell mass of blastocyst stage embryos or generated from more differentiated somatic cells using exogenous introduction of the four Yamanaka transcription factors (POUSF1, SOX2, KLF4, MYC) (Takahashi and Yamanaka, 2006).

In the last decades, induced pluripotent stem cells (iPSCs) have extensively been used for disease modeling and drug discovery (Rowe and Daley, 2019). Moreover, iPSCs are an attractive source for regenerative medicine, as they can be generated from the patient’s own cell, and the first clinical trials for stem cell therapies have recently been launched (Yamanaka, 2020). However, previous research has shown that stem cells in culture, such as iPSCs, acquire mutations, which can cause genetic safety issues and has hampered the use of stem cell therapies in clinic (Kucab et al., 2019; Kuijk et al., 2020; Rouhani et al., 2016; Thompson et al., 2020). For example, TPS3 mutations, base substitutions, and duplications of chromosomes 1, 12, 17, and 20 are often identified in iPSCs (Kuijk et al., 2020; Laurent et al., 2011; Martins-Taylor et al., 2011; Merkle et al., 2017). These findings suggest that iPSCs in culture may suffer from enhanced genetic instability. However, stem cells also accumulate mutations during normal life (Blokzijl et al., 2016; Hasaart et al., 2020; Lee-Six et al., 2018; Mitchell et al., 2021; Moore et al., 2021; Osorio et al., 2018). To which extent the in vitro mutation accumulation in iPSCs varies from the natural genetic instability of pluripotent and totipotent embryonic cells in vivo remains unknown.

Here, we directly compared the in vitro mutation accumulation in iPSCs with the mutation burden of their isogenic embryonic cells during human embryogenesis. For this, we cataloged the somatic mutations acquired pre-gastrulation in four human fetuses using a phylogenetic approach and generated iPSCs of the same stem cells to determine culture-associated mutagenesis. We show that iPSCs, when cultured under hypoxic conditions to reduce culture-associated mutations (Kuijk et al., 2020; Li and Marbán, 2010), maintain a similar genetic integrity as their embryonic counterparts do in vivo.
RESULTS

Mutation accumulation during early embryogenesis

Somatic mutations acquired by a cell are propagated to their future progeny. Therefore, mutations acquired during early embryogenesis will still be present in the genomes of adult cells, unless these are deleterious (Behjati et al., 2014; Ju et al., 2017; Lee-Six et al., 2018; Moore et al., 2021; Osorio et al., 2018). In addition, mutations shared by multiple cells derived from different germ layers of the same individual were acquired by an ancestral pluripotent or totipotent progenitor (Ju et al., 2017). We used this principle to determine the mutation rate pre-gastrulation during early human embryonic development. For this, we performed whole-genome sequencing (WGS) of multiple endoderm-derived intestinal stem cells (ISCs), mesoderm-derived hematopoietic stem and progenitor cells (HSPCs), and ectoderm/mesoderm-derived bulk skin of the same fetuses and constructed developmental lineage trees (Figure S1). To obtain sufficient DNA of a single cell for WGS, we clonally expanded the fetal stem cells in vitro (Hasaart et al., 2020). This approach allowed us to assess all mutations in the parental cell. We used DNA of bulk skin, of the same fetuses to control for germline mutations. Mutations that were acquired during the in vitro clonal expansion step could be excluded based on their low variant allele frequency (VAF) (Huber et al., 2019; Jager et al., 2017). Of note, we focused on mutations that were shared by multiple fetal cells of different germ layers and therefore accumulated in vivo in a shared ancestral cell, which was present during early embryogenesis pre-gastrulation.

In total, we sequenced 23 ISCs and 38 HSPCs of fetuses that were aborted because of the presence of a constitutive trisomy of chromosome 21 (T21; n = 2) or without medical indication (n = 2). Of note, the two latter fetuses were karyotypically normal (D21). In each fetus we identified 10–40 somatic mutations that were acquired before gastrulation (Figures 1A and 2A–2D). Of these, two mutations were located in protein coding sequences but were not identified as potential cancer driver mutations (Figure 1B and Table 1). Mutations that are acquired pre-gastrulation are present in the upper branches of a developmental lineage tree and can therefore be used to determine the early somatic mutation rate (Figures 2A–2D). Furthermore, mutations acquired during the same cell division will have a comparable VAF in bulk skin. In contrast, mutations acquired in subsequent cell divisions will have a lower VAF in bulk skin, assuming that no selection or genetic drift of embryonic clones took place (Ju et al., 2017). Some branch points are converted to polytopies, which suggests that during embryonic development, cell divisions have occurred without the

Figure 1. Identified mutations in embryonic cells pre-gastrulation and culture-associated mutations in iPSCs and fetal ISCs

(A) Pie charts representing the relative number of identified single base substitutions and indels.
(B) Pie charts representing the relative number of identified mutations affecting protein-coding regions of the genome (left) and relative number of protein-coding mutations causing missense, synonymous, stop gained, splice region and non-coding transcript exon, and splice acceptor and intron mutations (right) (Tables S1 and S2).
Figure 2. Developmental lineage trees of fetal stem cells derived from different germ layers

(A–D) Developmental lineage trees of four human fetuses (T21; n = 2, D21; n = 2) to show the clonal relationship of single intestinal stem cells (ISC) and single hematopoietic stem and progenitor cells (HSPC) (23 ISCs and 38 HSPCs). Each tip represents a single expanded cell. The green lines represent somatic mutations acquired pre-gastrulation, whereas blue lines represent somatic mutations present in the endoderm and red lines somatic mutations present in the mesoderm. Cell divisions in the branches indicated with a number are used to determine the somatic mutation rate pre-gastrulation. The pie charts represent the median contribution of the contributing mutations to bulk skin. Only the pie charts of the mutations acquired in cell divisions used to determine the somatic mutation rate pre-gastrulation are shown. The red parts indicate the contribution of the mutations to bulk skin (Figures S2B–S2E).
acquisition of any somatic mutation (Figures 2A–2D). Importantly, mutations that were acquired in lineages, which were lost during development, cannot be annotated. We used the mutations in the different branches of the constructed phylogenies to determine the pre-gastrulation somatic mutation rate during human embryogenesis (see STAR Methods). In total, we could use 17 independent cell divisions, which all took place before gastrulation to estimate the somatic mutation rate pre-gastrulation (see STAR Methods).

Often, more cell divisions took place before gastrulation; however, not all can be used for the estimation (see STAR Methods). To determine the pre-gastrulation somatic mutation rate as accurate as possible, we only included cell divisions in which the VAF of the mutations in the daughter cells was equal to the VAF of the mutations in the mother cell. We estimated a pre-gastrulation somatic mutation rate of 1.65 \( G^{0.93} \) (SD) mutations per cell division (Figure 3A). This mutation rate is in line with previous studies, in which the early embryonic mutation rate was determined by sequencing of clonal expansions of adult individuals (Coorens et al., 2021; Lee-Six et al., 2018; Park et al., 2021). Previously, we have shown an increased somatic mutation load in multipotent stem and progenitor cells in T21 fetuses (Hasaart et al., 2020). However, the mutation rate pre-gastrulation is similar between T21 and karyotypically normal human fetuses (Figure S2A). To validate our finding, we estimated the somatic mutation rate using a complementary approach. We performed deep targeted sequencing for the mutations, which we identified in the fetal clones, on bulk tissues derived from different germ layers of the assessed fetuses (Figure S1). Using this approach, we found a similar somatic mutation rate pre-gastrulation (Figure S2F).

### Mutation accumulation in iPSCs

We used the ISC cultures of the fetuses, which we analyzed by WGS, to generate isogenic iPSC lines (Figure S1). Using RNA sequencing analysis, we demonstrated that the iPSC lines show a different transcriptional profile compared to the ISC cultures. Moreover, we confirmed a similar expression profile of the

| Variant | Position   | Gene   | Predicted effect | Mutation type | Cell type | Fetus  |
|---------|------------|--------|------------------|---------------|-----------|--------|
| G>C     | 1:203800025| ZBED6  | Moderate         | Missense      | iPSC      | NR2 (D21) |
| G>A     | 1:203800166| ZBED6  | Moderate         | Missense      | iPSC      | NR2 (D21) |
| G>A     | 1:203800309| ZBED6  | Low              | Synonymous    | iPSC      | NR2 (D21) |
| G>A     | 3:132475016| DNAJC13| Moderate         | Missense      | iPSC      | NR2 (D21) |
| G>A     | 1:53277029 | LRP8   | Low              | Synonymous    | iPSC      | NR2 (D21) |
| G>A     | 2:109501510| SH3RF3 | Moderate         | Missense      | iPSC      | NR2 (D21) |
| G>A     | 19:15110164| SYDE1  | Low              | Synonymous    | iPSC      | OS1 (T21) |
| G>A     | 1:16992069 | ATP13A2| Moderate         | Missense      | iPSC      | OS1 (T21) |
| G>A     | 12:49049850| KMTD2  | Low              | Synonymous    | iPSC      | OS1 (T21) |
| G>A     | 8:127740738| MYC    | Moderate         | Missense      | iPSC      | NR1 (D21) |
| C>T     | 13:85794862| SLITK6 | Low              | Synonymous    | Pre-gastrulation | NR1 (D21) |
| C>T     | 3:126072855| KLF15  | Low              | Splice region & non-coding transcript exon | Pre-gastrulation | NR1 (D21) |
| A>T     | 20:45222766| SMEG2  | Low              | Synonymous    | ISC       | NR2 (D21) |
| G>A     | 20:45222776| SMEG2  | Moderate         | Missense      | ISC       | NR2 (D21) |
| C>T     | 5:168666663| SLIT3  | High             | Stop gained   | ISC       | NR2 (D21) |
| C>T     | 3:37520998 | LANCL3 | Moderate         | Missense      | ISC       | NR2 (D21) |
| C>T     | 16:2530887 | CEMP1  | Low              | Synonymous    | ISC       | OS1 (T21) |
| C>T     | 2:179945716| CWC22  | High             | Splice acceptor & intron | ISC       | OS1 (T21) |
| G>A     | 2:179945716| PER3   | Moderate         | Missense      | ISC       | NR1 (D21) |
| T>G     | 8:680060689| PREX2  | Moderate         | Missense      | ISC       | NR1 (D21) |

The table shows all mutations located in protein coding regions and the estimated effect based on snpEff. The mutations in MYC and PREX2 are considered as potential cancer driver mutations because these were annotated with a moderate effect and were present in the cancer gene census.
generated iPSCs and human embryonic stem cells (hESCs), suggesting that the reprogramming was successful (Figure S3) (Kuijk et al., 2020). We generated clonal iPSC cultures (see STAR Methods), which were analyzed by WGS to catalog the mutations present in the parental cell. These clonal iPSC lines were kept in culture for approximately 3 months to allow somatic mutations to accumulate. Cells were cultured under hypoxic conditions (3% O2), and all procedures, such as medium preparation, medium change, and passaging, were performed under hypoxic conditions to reduce the acquisition of mutations in culture (Kuijk et al., 2020; Li and Marbán, 2010; Thompson et al., 2020). Subsequently, a second clonal expansion step was performed to generate sub-clonal iPSC lines, which were also subjected to WGS analysis. By subtracting the somatic mutations present in the parental clone from those observed in the matching sub-clone, we were able to accurately assess the mutation burden associated with \( \textit{in vitro} \) expansion of iPSCs (Jager et al., 2017). We subjected seven clone-subclone pairs to WGS analysis, which were derived from one of the T21 and the two D21 fetuses. In these clonal iPSC cell lines, we identified 723 culture-associated base substitutions and 99 culture-associated insertions-deletions (indels) (Figure 1A). In one of the clonal iPSC lines, which was derived from a D21 fetus, a part of chromosome 18 was deleted (Figure S4A). This deletion was sub-clonally present in the culture and was absent in the sub-clone. Besides that, we did not observe any karyotype abnormalities in the other iPSC lines (Figure S4A). Of all mutations that accumulated between the two subsequent clonal expansion steps, 10 mutations were located in protein-coding sequences (Figure 1B and Table 1). Of these, one was considered as a potential cancer driver gene and was located in the oncogene \( \text{MYC} \).

To accurately estimate the culture-associated mutation rate, we determined the cell cycle length of the iPSCs by pulse chase of 5-ethyl-2′-deoxyuridine (EdU) followed by measuring co-staining with DAPI and the mitosis marker phosphorylated histone 3 (pH3) (see STAR Methods). Using this approach, we found that iPSCs divide every 18 hours (Figure S4B), indicating that iPSCs acquire 1.03 \( \pm \) 0.14 (SD) mutations per cell division under reduced oxygen conditions (Figures 3B and S5A). The observed cell division time is in line with a previous study, where the cell cycle length of iPSCs under normoxic culture conditions was determined to be 16–18 hours (Ghule et al., 2011). This finding suggests that the hypoxic culture conditions we applied had no effect on the cell division time.
To compare the culture-associated mutation rate of iPSCs with the culture-associated mutation rate of their source cells, we performed sequential single cell expansion steps of the ISC clones (Figure S1). For this, we used the clonal ISCs of the same fetuses to determine the culture-associated mutation rate under normoxic culture conditions. Before the generation of sub-clonal fetal ISC lines, clonal ISC cultures were kept in culture for approximately 1.5 months in order to accumulate sufficient culture-associated mutations to accurately determine an in vitro mutation burden. The sub-clonal fetal ISC lines were subjected to WGS and a similar filtering method was applied as described above. In total, we subjected 10 sub-clones to WGS derived from 8 clonal ISCs of 2 D21 fetuses and 2 T21 fetuses. We identified 1,017 culture-associated base substitutions and 101 cultured-associated indels (Figure 1A). Of all culture-associated mutations, eight mutations were located in protein-coding sequences (Figure 1B and Table 1). One mutation was located in the oncogene PREX2 and therefore might represent a potential cancer driver mutation. The cell division length of ISCs in culture has been previously determined to be 26 hours, indicating that fetal ISCs acquire 2.95 ± 1.23 (SD) mutations per cell division under normoxic culture conditions (Drost et al., 2017) (Figures 3C and S5B).

Overall, we conclude that cultured iPSCs display a similar mutation rate in vitro as their pluripotent and totipotent cells do in vivo (Wilcoxon rank-sum test p = 0.12), suggesting that the culture system has no or only limited additive mutagenic effects. Moreover, iPSCs derived from fetal ISCs display a lower mutation rate in culture compared with the culture-associated mutation rate of their fetal source cells (Wilcoxon rank-sum test p = 0.0001).

**Active mutational processes**

Although the mutation rate in cultured iPSCs was similar to the mutation rate pre-gastrulation, the mutation spectra were slightly different (Figures 4A, 4B, and S6A), suggesting that there is a difference in mutational processes acting on these cells. To obtain more insight in the mutational processes causing these mutations, we performed mutational signature analysis (see STAR Methods). The predominant mutational signature in both cultured iPSCs and embryonic cells in vivo was COSMIC signature SBSS (Figures 4C, 4D, S6B, and S6C). More than 60% of the observed mutations in embryonic cells pre-gastrulation and more than 80% of culture-associated mutations in iPSCs could be attributed to this signature (Figures 4C, 4D, S6B, and S6C). The remainder of the mutations in both cell types could be attributed to COSMIC signature SBS1 (Figures 4C, 4D, S6B, and S6C). This signature is caused by the spontaneous deamination of methylated cytosines and may represent a mutational clock in terms of cell division or time (Alexandrov et al., 2015; Blokzijl et al., 2016; Koh et al., 2021). Embryonic cells pre-gastrulation show more activity of the mutational process underlying this mutational signature, which is line with the observed increased proportion of C>T changes at CpG (Figure 4A). During embryonic development cells display high levels of proliferation to give rise to cells of all germ layers in a short time, which may explain the increased contribution of SBSS1. In conclusion, these findings indicate that the main mutational processes in cultured iPSCs and embryonic cells pre-gastrulation are similar, which support our initial observation that the mutation burden in hypoxic-cultured iPSCs and totipotent and pluripotent cells in vivo is highly similar. However, the contribution of SBS1 and SBSS5 to the mutation spectra of embryonic cells pre-gastrulation and iPSCs is slightly different.

**DISCUSSION**

Previous studies have shown that both multipotent stem cells and iPSCs acquire mutations during culture, which may compromise the clinical use of these cells (Kuijk et al., 2020; Laurent et al., 2011; Martins-Taylor et al., 2011; Merkle et al., 2017). Indeed, it has been demonstrated that multipotent adult stem cells can have an increased mutation rate in vitro as compared with their in vivo counterpart (Kuijk et al., 2020). For example, adult ISCs in vitro have a nearly 40-fold increased mutation rate as compared to the mutation rate in vivo. In contrast, our study demonstrates that iPSCs, when cultured under hypoxic conditions, display a similar mutation burden as their in vivo totipotent and pluripotent counterparts. In addition, the culture-associated mutation rate of iPSCs is lower compared to the culture-associated mutation rate of their fetal source cells. These results indicate that iPSCs maintain their natural genomic stability during hypoxic cell culture. The observed culture-associated mutation rate of iPSCs cultured in hypoxic conditions are in line with two other studies measuring the mutation rate of iPSCs and hESC during hypoxic cell culture (Kuijk et al., 2020; Thompson et al., 2020). These and our results demonstrate that hypoxic cell culture reduces the number of culture-associated mutations (Kuijk et al., 2020; Rouhani et al., 2016; Thompson et al., 2020). The levels of oxygen have also been reported to reduce the methylation status of hESC, decrease spontaneous differentiation, and promote the differentiation capacity toward other cell types.
Together, this suggests that hypoxic cell culture is beneficial for the genomic integrity of pluripotent cells. Of note, we show that karyotypically abnormal (T21) iPSCs remain stable during culture, whereas their parental fetal ISCs are less stable in vivo compared with karyotypically normal ISCs (Hasaart et al., 2020). This finding implies that iPSCs generated from a disease model remain stable during culture as well. Of note, the reprogramming process of somatic cells toward pluripotency is a stressful process and a G1/S cell cycle checkpoint deficiency during the initial phases of reprogramming causes an increase in base substitutions during this limited time frame (Araki et al., 2020; Bhutani et al., 2016; Kida et al., 2015; Sugiura et al., 2014). Therefore, the initial phase of

**Figure 4. Active mutational processes**

(A) Relative contribution of the indicated base substitutions to the mutational spectra of embryonic cells pre-gastrulation. Dots represent the four fetuses (T21; n = 2, D21; n = 2, 23 ISCs and 38 HSPCs) (left) Relative contribution of the indicated base substitutions to the mutational spectra of iPSCs. Dots represent the seven iPSC clones derived from the fetal ISCs of three fetuses (T21; n = 1, D21; n = 2) (right). Data are represented as mean ± SD. The mutational spectra differ significantly (chi-squared test p = 0.017).

(B) The 96-trinucleotide profiles of embryonic cells pre-gastrulation and iPSCs. Cosine similarity between the two profiles is 0.74.

(C) The relative contribution of each mutational signature to the spectra of base substitutions.

(D) The cosine similarity between the mutational spectra and the mean reconstructed profiles (Figure S6).
iPSC differentiation toward other lineages might also result in an increased mutational burden during a limited time frame.

Oncogenic mutations, such as TP53, occur once in every $\sim 2.0 \times 10^9$ cultured iPSCs according to previous research (Kuijk et al., 2020; Merkle et al., 2017). Indeed, we observed a mutation in oncogene MYC and a subclonal deletion of a part of chromosome 18 in the matching clonal iPSC cell line. Nevertheless, during fetal development and even during early embryogenesis, cells can also acquire oncogenic mutations (Custers et al., 2021; Greaves, 2018; Williams et al., 2020). For example, sacrococcygeal teratomas, the most common tumor type in newborns, arise from germ cells and can recapitulate cells of all three germ layers (Manes, 1976; Moore et al., 2003). Germ cells arise around the time of gastrulation and are induced from the epiblast, suggesting that the cell of origin of these tumors arose at the time of pluripotency (Nicholls et al., 2019). This suggests that the acquisition of oncogenic mutations in pluripotent cells in vivo and in vitro can be a “natural” event.

Overall, we conclude that cultured iPSCs maintain a genomic integrity during hypoxic cell culture at a similar degree as embryonic cells pre-gastrulation in vivo. Therefore, iPSCs may safely be used for regenerative medicine in terms of genome stability. We emphasize to culture iPSCs under hypoxic culture conditions to reduce the number of culture-associated mutations. However, since it is still possible for a cancer driver mutation to occur, we recommend that parental somatic cells and cultured iPSCs are screened for oncogenic mutations before they are used for stem cell therapies in patients.

Limitations of the study
In this study, we estimated the somatic mutation rate pre-gastrulation of four human fetuses using a phylogenetic inference approach. Some branch points are converted to polytomies, suggesting that cell divisions have occurred without the acquisition of any somatic mutation. Importantly, mutations that were acquired in lineages, which were lost during development, cannot be annotated. To estimate the somatic mutation rate pre-gastrulation as accurately as possible, we only included cell divisions in which the VAF of the mutations in the daughter cells was equal to the VAF of the mutations in the mother cell. However, we stress that the obtained somatic mutation rate pre-gastrulation is an estimation. In addition, the iPSCs in this study are cultured under hypoxic culture conditions, because previous studies have shown that this reduces the number of culture-associated mutations. Nevertheless, the vast majority of laboratories and cell therapy companies make use of normoxic culture conditions. Therefore, the results obtained in this study might not be applicable to iPSCs cultured in normoxic culture conditions.

STAR METHODS
Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHOD DETAILS
  - Isolation and clonal cultures of HSPCs
  - Intestinal organoid cultures
  - iPSC culture
  - Cell cycle length of iPSCs
  - RNA sequencing
  - Genomic DNA isolation
  - Whole genome sequencing and read alignment
  - Somatic mutation calling and filtering
  - Targeted deep re-sequencing
  - Construction of developmental lineage tree
  - Mutation rate during early embryogenesis
  - Mutation rate of iPSCs and ISCs
  - Mutational profiles and signature analyses
- QUANTIFICATION AND STATISTICAL ANALYSIS
SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2022.103736.

ACKNOWLEDGMENTS

We would like to thank Arianne Brandsma for providing feedback on our manuscript and the FACS and Imaging facility of the Princess Máxima Center for Pediatric Oncology for their support. This study was financially supported by a VIDI grant from the Netherlands Organisation for Scientific Research (NWO no. 016.Vidi.171.023) to R.v.B. and a grant from the Dutch Cancer Society (KWF no. 11307) to R.v.B. and S.M.C.d.S.L.

AUTHOR CONTRIBUTIONS

K.A.L.H. performed sample preparation, FACS, RNA isolation, and ISC and HSPC cultures. K.A.L.H. and M.V. performed iPSC culture and DNA isolations. K.A.L.H. and N.M.G. performed EdU-PH3 stainings. M.V. performed library preparation and capture for targeted deep sequencing. E.K. generated iPSC cultures. K.A.L.H. and M.V. performed library preparation and capture for targeted deep sequencing. E.K. generated iPSC cultures. K.A.L.H. and M.V. performed iPSC culture and DNA isolations. K.A.L.H. performed sample preparation, FACS, RNA isolation, and ISC and HSPC cultures. K.A.L.H. and N.M.G. performed EdU-PH3 stainings. K.A.L.H. performed sample preparation, FACS, RNA isolation, and ISC and HSPC cultures. K.A.L.H. and N.M.G. performed EdU-PH3 stainings. K.A.L.H. and N.M.G. performed EdU-PH3 stainings.

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES

Alexanderrov, L.B., Jones, P.H., Wedge, D.C., Sale, J.E., Campbell, P.J., Nik-Zainal, S., and Stratton, M.R. (2015). Clock-like mutational processes in human somatic cells. Nat. Genet. 47, 1402–1407.

Alexanderrov, L.B., Kim, J., Haradhvala, N.J., Huang, M.N., Ng, A.W.T., Wu, Y., Boot, A., Covington, K.R., Gordenin, D.A., Bergstrom, E.N., et al. (2020). The repertoire of mutational signatures in human cancer. Nature 578, 94–101.

Araki, R., Hoki, Y., Suga, T., Obara, C., Sunayama, M., Imadome, K., Fujita, M., Kamimura, S., Nakamura, M., Wakayama, S., et al. (2020). Genetic aberrations in iPSCs are introduced by a transient G1/S cell cycle checkpoint deficiency. Nat. Commun. 11, 197.

Behjati, S., Huch, M., van Boxtel, R., Karthaus, W., Wedge, D.C., Tamuri, A.U., Martincorena, I., Petitljak, M., Alexandrov, L.B., Gundem, G., et al. (2014). Genome sequencing of normal cells reveals developmental lineages and mutational processes. Nature 513, 422–425.

Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. R. Stat. Soc. Series B Methodol. 57, 289–300.

Bhutani, K., Nazor, K.L., Williams, R., Tran, H., Dai, H., Dizakula, Z., Cho, E.H., Pang, A.W.C., Rao, M., Cao, H., et al. (2016). Whole-genome mutational burden analysis of three pluripotency induction methods. Nat. Commun. 7, 1–8.

Blomkvist, F., de Ligt, J., Jager, M., Sasselli, V., Roerink, S., Sasaki, N., Huch, M., Boymans, S., Kuijk, E., Prins, P., et al. (2016). Tissue-specific mutation accumulation in human adult stem cells during life. Nature 538, 260–264.

Blomkvist, F., Janssen, R., van Boxtel, R., and Cuppen, E. (2018). MutationalPatterns: comprehensive genome-wide analysis of mutational processes. Genome Med. 10, 33.

Boeva, V., Popova, T., Bleakley, K., Chiche, P., Cappo, J., Schleiermacher, G., Janoueix-Lerosey, I., Delattre, O., and Barillot, E. (2012). Control-FREEC: a tool for assessing copy number and allelic content using next-generation sequencing data. Bioinformatics 28, 423.

Cingolani, P., Platts, A., Wang, L.L., Coon, M., Nguyen, T., Wang, L., Land, S.J., Lu, X., and Ruden, D.M. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. Fly 6, 80–92.

Coorens, T.H.H., Moore, L., Robinson, P.S., Sanghvi, R., Christopher, J., Hewinson, J., Przybilka, M.J., Lawson, A.R.J., Spach, M., Zugg, M., and Bennaceur-Griscelli, A. (2014). Morphological analysis of human induced pluripotent stem cells during induced differentiation and reverse programming. Biores. Open Access 3, 206–216.

Custers, L., Khabirova, E., Coorens, T.H.H., Oliver, T.R.W., Calandrini, C., Young, M.D., Viera Braga, F.A., Ellis, P., Mamanova, L., Segers, H., et al. (2021). Somatic mutations and single-cell transcriptomes reveal the root of malignant rhabdoid tumours. Nat. Commun. 12, 1407.

DePristo, M.A., Banks, E., Poplin, R., Garimella, K.V., Maguire, J.R., Hartl, C., Philippakis, A.A., del Angel, G., Rivas, M.A., Hanna, M., et al. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat. Genet. 43, 491–498.

Drost, J., Van Boxtel, R., Bloxjift, F., Mizutani, T., Sasaki, N., Sasselli, V., de Ligt, J., Behjati, S., Grolleman, J.E., van Wezel, T., et al. (2017). Use of CRISPR-modified human stem cell organoids to study the origin of mutational signatures in cancer. Science 358, 234–238.

Ewels, P., Mans, M., Lundin, S., and Kaller, M. (2016). MultiQC: summarize analysis results for multiple tools and samples in a single report. Bioinformatics 32, 3047–3048.

Ezashi, T., Das, P., and Roberts, R.M. (2005). Low O2 tensions and the prevention of differentiation of hES cells. Proc. Natl. Acad. Sci. U S A 102, 4783–4788.

Fynes, K., Tostoes, R., Ruban, L., Weil, B., Mason, C., and Veratch, F.S. (2014). The differential effects of 2% oxygen preconditioning on the subsequent differentiation of mouse and human
pluripotent stem cells. Stem Cells Dev. 23, 1910–1922.

Gerstung, M., Papamann, E., and Campbell, P.J. (2014). Subclonal variant calling with multiple samples and prior knowledge. Bioinformatics 30, 1198.

Ghule, P.N., Medina, R., Lengner, C.J., Mander, M., Qiao, M., Dominski, Z., Lian, J.B., Stein, J.L., van Wijnen, A.J., and Stein, G.S. (2011). Reprogramming the pluripotent cell cycle: restoration of an abbreviated G1 phase in human induced pluripotent stem (iPS) cells. J. Cell Physiol. 225, 1149–1156.

Greaves, M. (2018). A causal mechanism for childhood acute lymphoblastic leukaemia. Nat. Rev. Cancer 18, 471–484.

Hassar, K.A.L., Manders, F., van der Hoorn, M.L., Verheul, M., Poplonski, T., Kuijk, E., de Sousa Lopes, S.M.C., and van Boxtel, R. (2020). Mutation accumulation and developmental lineages in normal and down syndrome human fetal haematopoiesis. Sci. Rep. 10, 12991.

Huber, A.R., Maders, F., Oka, R., and van Boxtel, R. (2019). Characterizing mutational load and clonal composition of human blood. J. Vis. Exp. https://doi.org/10.3791/59846.

Jager, M., Blokzijl, F., Sasselli, V., Boymans, S., Janssen, R., Besselink, N., Clevens, H., van Boxtel, R., and Cuppen, E. (2017). Measuring mutation accumulation in single human adult stem cells by whole-genome sequencing of organoid cultures. Nat. Protoc. 13, 59–78.

Ju, Y.S., Martinezcorena, I., Gerstung, M., Petjak, M., Alexandrov, L.B., Rahbari, R., Wedge, D.C., Ju, Y.S., Martincorena, I., Gerstung, M., Petljak, Janssen, R., Besselink, N., Clevers, H., van Boxtel, R., et al. (2020). The clonal dynamics of haematopoiesis across the human lifespan. bioRxiv, 2021.08.16.456475.

Kida, Y.S., Kawamura, T., Downes, M., Evans Correspondence, R.M., Article, S., Wei, Z., Sogo, T., Jacinto, S., Shigeto, A., Kushige, H., et al. (2015). ERRs mediate a metabolic switch required for somatic cell reprogramming to pluripotency. Cell Stem Cell 28, 1178–1185.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). The sequence alignment/map format and SAMtools. Bioinformatics 25, 2078–2079.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 15, 1–21.

Ludecke, D. (2018). Ggfects: tidy data frames of marginal effects from regression models. J. Open Source Softw. 3, 772.

Mans, C. (1976). Summation: model systems for the study of oncodevelopmental gene expression—murine teratocarcinoma. Cancer Res. 36, 4238.

Martino-Taylor, K., Nisler, B.S., Taapkema, S.M., Compton, T., Crandall, L., Montgomery, K.D., Lalande, M., and Xu, R.H. (2011). Recurrent copy number variations in human induced pluripotent stem cells. Nat. Biotechnol. 29, 488–491.

Merkle, F.T., Ghosh, S., Kamitaki, N., Mitchell, J., Aivior, Y., Mello, C., Koshin, S., Meckhoubad, S., Jilic, D., Charlton, M., et al. (2017). Human pluripotent stem cells recurrently acquire and expand dominant negative P53 mutations. Nature 545, 229–233.

Mitchell, E., Chapman, M.S., Williams, N., Dawson, K., Mende, N., Calderbank, E.F., Jung, H., Mitchell, T., Cooren, T., Spencer, D., et al. (2021). Clonal dynamics of haematopoiesis across the human lifespan. bioRxiv 2021.08.16.456475.

Moore, L., Cagan, A., Cooren, T.H.H., Neville, M.D.C., Sanghvi, R., Sanders, M.A., Oliver, T.R.W., Leongamornlert, D., Ellis, P., Noorani, A., et al. (2021). The mutational landscape of human somatic and germline cells. Nature 597, 381–386.

Moore, S.W., Satge, D., Sasco, A.J., Zimmermann, A., and Plaschkes, J. (2003). The epidemiology of neonatal tumours. Pediatr. Surg. Int. 19, 509–519.

Nicholls, P.K., Schorle, H., Naqui, S., Hu, Y.C., Fan, Y., Carmell, M.A., Dobriniski, I., Watson, A.L., Carlson, D.F., Faehrenkrug, S.C., et al. (2019). Mammalian germ cells are determined after PGC colonization of the nascent gonad. Proc. Natl. Acad. Sci. U S A 116, 25567–25568.

Nik-Zainal, S., Van Loo, P., Wedge, D.C., Alexandrov, L.B., Greenman, C.D., Lau, K.W., Rane, K., Jones, D., Marshall, J., Ramakrishna, M., et al. (2012). The life history of 21 breast cancers. Cell 149, 994–1007.

Osorio, F.G., Rosendahl Huber, A., Oka, R., Verheul, M., Patel, S.H., Haskaat, K., de la Fonteinje, L., Varela, I., Camargo, F.D., and van Boxtel, R. (2018). Somatic mutations reveal lineage relationships and age-related mutagenesis in human hematopoiesis. Cell Rep. 25, 2308–2316.e4.

Parais, E., Claude, J., and Strimmer, K. (2004). APE: Analyses of Phylogenetics and Evolution in R language. Bioinformatics 20, 289–290.

Park, S., Mali, N.M., Kim, R., Choi, J.-W., Lee, J., Lim, J., Park, J.M., Park, J.W., Kim, J., Kund, T., et al. (2021). Clonal dynamics in early human embryogenesis inferred from somatic mutation. Nature 597, 393–397.

Rouhani, F.J., Nik-Zainal, S., Wuster, A., Li, Y., Conte, N., Koke-Yusa, H., Kumasaka, N., Valler, L., Yusa, K., and Bradley, A. (2016). Mutational history of a human cell lineage from somatic to induced pluripotent stem cells. PLoS Genet. 12, e1005932.

Rowe, R.G., and Daley, G.Q. (2019). Induced pluripotent stem cells in disease modelling and drug discovery. Nat. Rev. Genet. 20, 377–388.

Smigielski, E.M., Sirokin, K., Ward, M., and Sherry, S.T. (2000). dbSNP: a database of single nucleotide polymorphisms. Nucleic Acids Res. 28, 352.

Sondka, Z., Barnford, S., Cole, C.G., Ward, S.A., Dunham, I., and Forbes, S.A. (2018). The COSMIC Cancer Gene Census: describing genetic dysfunction across all human cancers. Nat. Rev. Cancer 18, 696–705.

Sugiura, M., Kasama, Y., Araki, R., Hoki, Y., Sunayama, M., Udo, M., Nakamura, M., Ando, S., and Abe, M. (2014). Induced pluripotent stem cell generation-associated point mutations arise during the initial stages of the conversion of these cells. Stem Cell Rep. 2, 52–63.

Takashashi, K., and Yamanaka, S. (2006). Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell 126, 663–676.

Tarasov, A., Vilella, A.J., Cuppen, E., Nijman, I.J., and Prins, P. (2015). Sambamba: fast processing of NGS alignment formats. Bioinformatics 31, 2032–2034.

Thompson, O., von Meyenn, F., Hewitt, Z., Alexander, J., Wood, A., Weightman, R., Gregory, S., Krueger, F., Andrews, S., Barbaric, I., et al. (2020). Low rates of mutation in clinical grade human pluripotent stem cells under different culture conditions. Nat. Commun. 11, 1–14.

Thorvaldsdottir, H., Robinson, J.T., and Mesirov, J.P. (2013). Integrative genomics viewer (IGV):
high-performance genomics data visualization and exploration. Brief Bioinform. 14, 178–192.

Warlich, E., Kuehle, J., Cantz, T., Brugman, M.H., Maetzig, T., Galla, M., Filipczyk, A.A., Halle, S., Klump, H., Schöler, H.R., et al. (2011). Lentiviral vector design and imaging approaches to visualize the early stages of cellular reprogramming. Mol. Ther. 19, 782–789.

Wickham, H. (2016). ggplot2 (Springer International Publishing).

Williams, N., Lee, J., Moore, L., Baxter, E.J., Hewinson, J., Dawson, K.J., Menzies, A., Godfrey, A.L., Green, A.R., Campbell, P.J., et al. (2020). Phylogenetic reconstruction of myeloproliferative neoplasm reveals very early origins and lifelong evolution. bioRxiv, 2020.11.09.374710.

Yamanaka, S. (2020). Pluripotent stem cell-based cell therapy-promise and challenges. Cell Stem Cell 27, 523–531.

Yates, A., Beal, K., Keenan, S., McLaren, W., Pignatelli, M., Ritchie, G.R.S., Ruffier, M., Taylor, K., Vullo, A., and Flicek, P. (2015). The Ensembl REST API: Ensembl Data for Any Language. Bioinformatics 31, 143–145.
## Key Resources Table

### Antibodies

| REAGENT or RESOURCE | SOURCE | IDENTIFIER |
|---------------------|--------|------------|
| CD34-BV421, clone 561 | Biolegend | Cat# 343609; RRID: AB_2561358 |
| CD38-PE, clone H1T2 | Biolegend | Cat# 303505; RRID: AB_314357 |
| CD45RA-PerCp/Cy5.5, clone HI100 | Biolegend | Cat# 304121; RRID: AB_893358 |
| CD49f-PE/Cy7, clone GoH3 | Biolegend | Cat# 313622; RRID: AB_2561705 |
| CD90-APC, clone 5E10 | Biolegend | Cat# 328113; RRID: AB_893440 |
| Lineage(CD3/CD14/CD19/CD20/CD56)-FITC, clones UCHT1, HCD14,HiB9, HCD56) | Biolegend | Cat# 348701; RRID: AB_10644012 |
| CD11c-FITC, clone 3.9 | Biolegend | Cat# 301603; RRID: AB_314173 |
| CD16-FITC, clone 3G8 | Biolegend | Cat# 302005; RRID: AB_314205 |
| Anti-Rabbit IgG AF594 | Invitrogen | Cat# 855; RRID: 2165334 |
| Rabbit anti-pH3 | Merck Milipore | Cat# 06-570; RRID: AB_310177 |

### Biological samples

| REAGENT or RESOURCE | SOURCE | IDENTIFIER |
|---------------------|--------|------------|
| Fetal tissue | Leiden University Medical center | Not applicable |

### Chemicals, peptides, and recombinant proteins

| REAGENT or RESOURCE | SOURCE | IDENTIFIER |
|---------------------|--------|------------|
| Recombinant human thrombopoietin (TPO) | Preprotech | Cat# 300-18 |
| Recombinant human stem cell factor (SCF) | Preprotech | Cat# 300-07 |
| Recombinant human FLT3-L | Preprotech | Cat# 300-19 |
| Recombinant human IL-6 | Preprotech | Cat# 200-06 |
| Recombinant human IL-3 | Preprotech | Cat# 160-01 |
| BSA | Sigma- Aldrich | Cat# A7030-10G |
| HEPES | Thermo-Fisher Scientific | Cat# 15630106 |
| Penicillin-streptomycin (10,000U/mL) | Thermo-Fisher Scientific | Cat# 15140122 |
| Glutamax supplement | Thermo-Fisher Scientific | Cat# 35050061 |
| PBS | Thermo-Fisher Scientific | Cat# 14190 |
| EDTA | Sigma- Aldrich | Cat# T4049 |
| Primocin | Invivogen | Cat# ant-pm-1 |
| SB 202190 monohydrochloride hydrate | Sigma- Aldrich | Cat# 57076-5MG |
| B27 supplement | Thermo-Fisher Scientific | Cat# 175044 |
| N-Acetyl-cysteine | Sigma- Aldrich | Cat# A9165 |
| A83-01 | Tocoris | Cat# 239/10 |
| Y-27632 dihydrochloride | Abmole | Cat# M1817 |
| Human EGF | Preprotech | Cat# AF-100-15 |
| Nicotinamide | Sigma- Aldrich | Cat# N0636 |
| WNT surrogate-fc fusion protein | U-protein express | Cat# N001 |
| FBS | Sigma- Aldrich | Cat# A4766801 |
| Clostridium histolyticum type IA | Sigma- Aldrich | Cat# C9891 |
| DNase I | Sigma- Aldrich | Cat# DN25 |
| EBSS | Thermo-Fisher Scientific | Cat# 24010043 |
| Revitacell supplement | Thermo-Fisher Scientific | Cat# A2644501 |
| Triton X-100 | Thermo-Fisher Scientific | Cat# 85111 |

(Continued on next page)
Continued

| REAGENT or RESOURCE | SOURCE | IDENTIFIER |
|---------------------|--------|------------|
| Fluormount-G Mounting Medium | Thermo-Fisher Scientific | Cat#00-4958-02 |
| Advanced DMEM/F-12 | Thermo-Fisher Scientific | Cat#12634028 |

Critical commercial assays

| REAGENT or RESOURCE | SOURCE | IDENTIFIER |
|---------------------|--------|------------|
| QiAamp DNA Micro kit | QiAgen | Cat#65304 |
| Genomic tip 20/G | QiAgen | Cat#10223 |
| Dneasy blood & tissue kit | QiAgen | Cat#65904 |
| Rneasy mini kit | QiAgen | Cat#74004 |
| Essential 8 flex medium kit | Thermo-Fisher Scientific | Cat#A2858501 |
| Click-iT EdU Alexa Fluor 488 cell proliferation kit for imaging | Thermo-Fisher Scientific | Cat#10337 |

Deposited data

| REAGENT or RESOURCE | SOURCE | IDENTIFIER |
|---------------------|--------|------------|
| Whole-genome sequence data from this article | This paper | European Genome-Phenome Archive (EGA); https://ega-archive.org/ega/home. Accession Number EGA: EGAS00001005939 |
| Deep targeted re-sequencing | This paper | European Genome-Phenome Archive (EGA); https://ega-archive.org/ega/home. Accession Number EGA: EGAS00001005939 |
| Whole-genome sequencing data | (Hasaart et al., 2020) | European Genome-Phenome Archive (EGA); https://ega-archive.org/ega/home. Accession Number EGA: EGAS00001003982 and EGAS00001002886 |
| RNA sequencing data from this article | This paper | European Genome-Phenome Archive (EGA); https://ega-archive.org/ega/home. Accession Number EGA: EGAS00001005939 |
| RNA sequencing data | (Kuijk et al., 2020) | European Genome-Phenome Archive (EGA); https://ega-archive.org/ega/home. Accession Number EGA: EGAS00001002955, EGAS00001000881 and EGAS00001001682 |

Software and algorithms

| REAGENT or RESOURCE | SOURCE | IDENTIFIER |
|---------------------|--------|------------|
| Whole genome sequencing read alignment and mutation calling pipeline | Not applicable | https://github.com/UMCUGenetics/IAP https://github.com/UMCUGenetics/NF-IAP |
| RNA sequencing read alignment and mutation calling pipeline | Not applicable | https://github.com/UMCUGenetics/RNASeq-NF |
| SNV filtering pipeline | Not applicable | https://github.com/ToolsVanBox/SMuRF/ |
| Indel filtering pipeline | Not applicable | https://github.com/ToolsVanBox/SMuRF/ |
| R (v3.6) | Not applicable | https://www.r-project.org/ |
| ggeffects R package (v0.14.2) | (Lüdecke, 2018) | https://cran.r-project.org/web/packages/ggeffects/index.html |
| ggplot2 R package (v3.2.1) | (Wickham, 2016) | https://cran.r-project.org/web/packages/ggplot2/index.html |
| Burrows-Wheeler Aligner mapping tool (v0.7.5a and v0.7.17) | (Li and Durbin, 2010) | https://github.com/lh3/bwa |
| DESeq2 | (Love et al., 2014) | https://bioconductor.org/packages/release/bioc/html/DESeq2.html |
| GATK HaplotypyCaller (v3.8.1-0 and v4.1.3.0) | (DePristo et al., 2011) | https://gatk.broadinstitute.org/hc/en-us |
| Single Nucleotide Polymorphism Database (v146) | (Smigielski et al., 2000) | https://www.ncbi.nlm.nih.gov/snp/ |
| Bedtools | (Smigielski et al., 2000) | https://bedtools.readthedocs.io/en/latest/ |
| NimbleDesign Software | Not applicable | https://sequencing.roche.com/en/support-resources/discontinued-products/nimble-design-software.html |

(Continued on next page)
RESOURCE AVAILABILITY

**Lead contact**
Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Ruben van Boxtel (R.vanboxtel@prinsesmaximacentrum.nl).

**Materials availability**
This study did not generate new unique reagents.

**Data and code availability**
Data: WGS, RNAseq and targeted re-sequencing data have been deposited at EGA (https://www.ebi.ac.uk/ega/), and accession numbers are listed in the key resources table. They are available upon request if access is granted. To request access, contact EGA.

In addition, tools used to process this data have been deposited at https://github.com/ToolsVanBox/ and are publicly available as the date of publication. The accession numbers are also listed in the key resources table. Microscopy data will be shared upon request. To request access, Ruben van Boxtel, R.vanboxtel@prinsesmaximacentrum.nl.

Code: Most code has been deposited at https://github.com/ToolsVanBox/ and in the MutationalPatterns R package and is publicly available as of the date of publication. DOIs are listed in the key resources table. Additional scripts will be shared upon request.

Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

**EXPERIMENTAL MODEL AND SUBJECT DETAILS**
For this study, we collected two T21 fetuses (gestational age (GA) 12 and 14.5) and two karyotypically normal fetuses (GA 14 and 16). The same fetuses were used for our previous study (Hasaart et al., 2020). The GA in weeks was determined by the measurement of first-trimester crown-rump length by ultrasonography. The age in weeks after conception was determined by subtracting 2 weeks of the GA. Fetal material was collected and disassociated as described before (Hasaart et al., 2020). Collection of material for this study (PO8.087) was approved by the Medical Ethical Committee of the Leiden University Medical Center. The study was performed in accordance with the guidelines and regulations of the Helsinki declaration and
its later amendments or comparable ethical standards. Details of all samples are depicted in Table S1. Fetal ISC and iPSC cultures were generated from the fetuses described above.

METHOD DETAILS
Isolation and clonal cultures of HSPCs
Mononuclear cells were flushed out from fetal bone marrow. Single liver cells and mononuclear fetal bone marrow cells were stained with an antibody cocktail as described before (Hasaart et al., 2020; Huber et al., 2019). CD34+ and lineage- cells were single cell sorted in a 384 wells plate and cultured for 3-4 weeks in StemSpan SFEM medium supplemented with growth factors as described before (Hasaart et al., 2020). Cells were spun down and DNA was isolated as described in the section ‘Genomic DNA isolation’ in order to perform WGS.

Intestinal organoid cultures
Fetal intestine was dissociated into a single cell solution using collagenase type 1A (Sigma-Aldrich) and 0.1mg/ml DNaseI (Sigma-Aldrich) as described before (Hasaart et al., 2020). Single intestinal cells were plated in Matrigel droplets (Corning) and plated in limited dilution. Clonal ISC cultures were generated as described before and cultured in human ISC organoid (CHIO) medium (Hasaart et al., 2020). After a culture period of 1.5 month, sub-clonal ISC cultures were generated as described before (Jager et al., 2017). In short, ISC organoid were made single cell with trypLE, single ISC were sorted with the Sony SH800S in a tube based on FSC and SSC and plated in limited dilution in Matrigel (Corning) droplets.

iPSC culture
Bulk ISC cultures were made single cell using trypLE (thermo fisher). 20.000 Single ISCs were resuspended in 1 ml CHIO medium supplemented with 10 μM Rho kinase inhibitor (Abmole) and 4 mg/ml polybrene. 10.000 cells were plated in Matrigel coated plates diluted in Advanced DMEM/F-12 supplemented with 1% penicillin/streptomycin, 1% GlutaMAX, and 1% HEPES 10 mM. Cells were transduced with a third generation lentiviral vector containing Oct4, Klf4, Sox2, c-Myc, dTomato (Warlich et al., 2011). Cells were cultured in CHIO medium supplemented with 10 μM Rho kinase inhibitor and polybrene. After 3 days, medium was replaced with Essential E8 Flex medium (Thermo Scientific). RNA seq analysis confirmed that reprogramming was successful.

To make clonal iPSC lines, cells were made single cell with trypLE. After, single iPSCs were plated in limited dilution in a 96-well cell culture plate coated with 100ul Matrigel diluted in Advanced DMEM/F-12 supplemented with 1% penicillin/streptomycin, 1% GlutaMAX, and 1% HEPES 10 mM (1:40). To enhance cell survival after the clonal steps, Essential E8 Flex medium was supplemented with RevitaCell (Thermo Scientific). Clonal iPSC lines were cultured and passaged under reduced oxygen conditions (3% O2) to allow mutations to accumulate. To prevent changes in oxygen levels, cells were placed in CondoCells. Cells were passaged 2 times a week with 0.5mM EDTA. After a culture period of approximately 3 months, a second clonal step was performed.

Cell cycle length of iPSCs
The length iPSC cultures was determined by pulse-chase of 5-ethynyl-2'-deoxyuridine (EdU) and co-staining for the mitosis phosphorylated histone 3 (pH3). iPSCs were labeled with 10 mM EdU (Alexa Fluor 488 Imaging kit, Life Technologies EdU kit) for 30 minutes. After, cells were washed two times with PBSO and medium was replaced with fresh Essential E8 Flex medium. Chase was performed 2, 6, 9.5, 15, 19, 23, 28, 33.5, 39, 43 and 47 hours after washing. Cells were made single cell with trypLE. For each time point two times 50.000 single cells were cytospun (1000 rpm, 5minutes). Cells were fixed with 4%PFA, permeabilized with 0.1% Triton X100/TBS and blocked with 1% BSA TBST. Cells were stained with rabbit anti-pH3 (Merck Millipore, 1:500) overnight at 4 degree and with Alexa 594-conjugated donkey anti rabbit (Invitrogen, 1:500). Afterward, EdU detection was performed with Click-iT assay following the manufacturer’s manual. iPSCs were mounted with DAPI and Fluoromount, and imaged with the Leica Thunder microscope. Per condition the full slide was imaged and analyzed with Fiji software and Rstudio.

RNA sequencing
iPSC cell pellets and ISC cell pellets were snap frozen and RNA was isolated using the RNeasy mini kit (Qiagen). RNA extractions were performed according to manufacturers’ instructions. RNA quality was
measured with the bioanalyzer. polyA tail libraries were generated from 200ng input RNA. Libraries were sequenced using the NextSeq2000 (2x 50 bp). RNA sequencing reads were aligned to human reference genome GRCh38 using the STAR algorithm. Full pipeline description and settings are available at https://github.com/UMCUGenetics/RNASeq-NF. Read counts were normalized using DESeq2 package version 1.32.0 (Love et al., 2014). Downstream analysis, including differential expression analysis and hierarchical clustering, were performed with DESeq2 package version 1.32.0 as well. The dds object was transformed using the function vst and heatmaps were produced by normalizing the expression mean to 0. Count tables of H9 ES cells were obtained from a previously published study and used to compare the transcriptional profile of the generated iPSCs with the transcriptional profile of H9 ES cells (Kuijk et al., 2020).

Genomic DNA isolation
Genomic tip 20/G (Qiagen) was used to extract DNA from bulk skin tissue, ISC organoid cultures and spleen. DNA from clonal HSPCs cultures was extracted using Qiap DNA Micro Kit (Qiagen). The blood and tissue kit (Qiagen) was used to extract DNA from clonal iPSC and cultures. All DNA isolations were performed according to manufacturers’ instructions.

Whole genome sequencing and read alignment
In addition to the data described below, WGS data from a previously published study was used (Hasaart et al., 2020). DNA libraries for Illumina sequencing were generated using standard protocols (Illumina). Libraries were generated from 300-500 ng of genomic DNA isolated from clonal iPSC cultures and ISC organoid cultures. For the clonal HSPC culture, 50-100 ng genomic DNA was used as input. All samples were sequenced (2 x 150 bp) using NovaSeq 6000 sequencers to 15X base coverage except for the sub-clonal ISC cultures and DNA of bulk skin these were sequenced 30X base coverage. DNA of bulk skin was sequenced twice 30X base coverage. WGS data was mapped against human reference genome GRCh38 using Burrows-Wheeler Aligner v0.7.5a mapping tool (Li and Durbin, 2010) with settings ‘bwa mem -c 100 -M’. Sequence reads were marked for duplicates using Sambamba v0.6.8 markdup. Full pipeline description and settings are available at: https://github.com/UMCUGenetics/IAP.

For the clonal iPSC and sub-clonal ISC cultures, Burrows-Wheeler Aligner v0.7.17 was used (https://github.com/ToolsVanBox/NF-IAP/).

Somatic mutation calling and filtering
Raw variants were multisample-called using the GATK HaplotypeCaller v3.8-1-0 (DePristo et al., 2011) and GATK Queue v3.8-1-0 with default settings and the additional option ‘EMIT_ALL_CONFIDENT_SITES’. The quality of variants and reference positions was evaluated by GATK VariantFiltration v3.8-1-0 with options -snpFilterName SNP_LowQualityDepth -snpFilterExpression “QD < 2.0” -snpFilterName SNP_MappingQuality -snpFilterExpression “MQ < 40.0” -snpFilterName SNP_StrandBias -snpFilterExpression “FS > 60.0” -snpFilterName SNP_HaplotypeScoreHigh -snpFilterExpression “HaplotypeScore > 13.0” -snpFilterName SNP_MQRankSumLow -snpFilterExpression “MQRankSum < -12.5” -snpFilterName SNP_ReadPosRankSumLow -snpFilterExpression “ReadPosRankSum < -8.0” -snpFilterName SNP_HardToValidate -snpFilterExpression “MQ0 >= 4 && ((MQ0 / (1.0 * DP)) > 0.1)” -snpFilterName SNP_LowCoverage -snpFilterExpression “DP < 5” -snpFilterName SNP_VeryLowQual -snpFilterExpression “QUAL >= 30.0 && QUAL < 50.0” -snpFilterName SNP_SOR -snpFilterExpression “SOR > 4.0” -cluster 3 -window 10 -indelType INDEL -indelType MIXED -indelFilterName INDEL_LowQualityDepth -indelFilterExpression “QD < 2.0” -indelFilterName INDEL_ReadPosRankSumLow -indelFilterExpression “ReadPosRankSum < -20.0” -indelFilterName INDEL_HardToValidate -indelFilterExpression “MQ0 >= 4 && ((MQ0 / (1.0 * DP)) > 0.1)” -indelFilterName INDEL_LowCoverage -indelFilterExpression “DP < 5” -indelFilterName INDEL_VeryLowQual -indelFilterExpression “QUAL < 30.0” -indelFilterName INDEL_LowQual -indelFilterExpression “QUAL >= 30.0 && QUAL < 50.0” -indelFilterName INDEL_SOR -indelFilterExpression “SOR > 10.0”. For the clonal iPSC and sub-clonal fetal ISC cultures, GATK v4.1.3.0 was used with the same settings. To obtain high-quality somatic mutation catalogs, we applied postprocessing filters as described (Blokzijl et al., 2016). Briefly, we considered variants at autosomal chromosomes without any evidence from a paired control sample (skin); passed by VariantFiltration with a GATK phred-scaled quality score R 100; a base coverage of at least 5X in the clonal and paired control sample; mapping quality (MQ) of 60; no overlap with single nucleotide polymorphisms (SNPs) in the Single Nucleotide Polymorphism Database v146; and absence of the variant...
in a panel of unmatched normal human genomes (BED-file available upon request). We additionally filtered base substitutions with a GATK genotype score (GQ) lower than 99 or 10 in clonal or paired control sample, respectively. For indels, we filtered variants with a GQ score lower than 99 in both clonal and paired control sample (Blokzijl et al., 2016; Jager et al., 2017). We used Bayesian Dirichlet modeling to check the clonality of the clones, as described before (Hasaart et al., 2020; Nik-Zainal et al., 2012). Mutations with a VAF > 0.3 were assumed to be clonal. Mutations located in protein coding regions were annotated with snpeff (v4.1).

A mutation was considered as potential cancer driver if it met two requirements. First it needed to be annotated with “MODERATE” or “HIGH” effect based on the effect impact criteria of snpEff v4.1 (Cingolani et al., 2012) and second it needed to be annotated in the Cosmic cancer gene census (v88).

To assess somatic mutations from our previous published study, we converted the VCFs to human reference genome GRCh38 (Hasaart et al., 2020), using Ensembl REST API (https://rest.ensembl.org/documentation/info/assembly_map).

To determine the mutations acquired during iPSC or ISC culture we subtracted all clonal mutations in the parental clone from the clonal mutations in the sub-clone.

Mutations were manually inspected using IGV to exclude false positives calls and include false negative calls.

**Targeted deep re-sequencing**

Genomic DNA of bulk skin, bulk ISC organoid cultures passage 0 and spleen were subjected to targeted deep re-sequencing. A bait set, containing all somatic mutations identified with WGS was designed using NimbleDesign Software. DNA libraries were generated from 200ng genomic DNA using TruSeq Nano DNA library prep kit according to manufacturers’ instructions. DNA libraries were pooled and DNA regions of interest were captured from a 1000 ng pooled library using SeqCap EZ Hypercap (Roche). DNA libraries were sequenced (2 x 150 bp) using Nextseq500.

WGS data was aligned against the human reference genome GRCh37 with the same methods as described previously. Further bioinformatic analyses were performed using the R language. Reads were loaded using the loadAllData function from the deepSNV (v 1.32.0) R package, using q=30 and mq=30 (Gerstung et al., 2014). After, the betabinLRT function from the same package was used with rho=1e-4, maxtruncate=1 and truncate=1 to run a maximum likelihood version of the Shearwater algorithm. Next, Benjami and Hochberg multiple testing correction was performed on the resulting p-values (Benjamini and Hochberg, 1995). The qvals2Vcf function with standard settings was used to generate a vcf file. Next, variants were filtered in several steps. First, variants were the reference and alternative alleles were the same were removed. After this, variants were the reference allele was set to “-” were filtered out. Variants within 10bp of an indel were removed. Next, only variants that were present in both the bulk skin and bulk ISC of a single sample were retained. Finally, for each fetus we only took the variants that matched the variants found in the WGS data of the clones.

After we converted the obtained VCFs to human reference genome GRCh38, using Ensembl REST API (https://rest.ensembl.org/documentation/info/assembly_map).

**Construction of developmental lineage tree**

Developmental lineage trees were constructed as described before (Hasaart et al., 2020). Somatic mutations sub clonally present in bulk skin were included and represent early embryonic mutations. The relationship between clones deducted from the shared and unique somatic mutations were visualized with the ape package.

**Mutation rate during early embryogenesis**

The number of SNVs per branch of the developmental lineage tree were counted. To ensure we assessed mutations that occurred pre-gastrulation, only the early branches whose contribution summed to 100% were used. Because not all cells were assessed, either because they were not sampled or because they were lost during development, a single branch can contain multiple cell divisions. The VAF was used to determine if the mutations in a branch occurred during a single cell division. In ambiguous cases, a Chi-square test with a simulated p-value (k=2000), was performed on the reference- and alternative allele counts of all mutations in a branch. If the VAFs between mutations were significantly different, then the
mutations occurred during multiple cell divisions. Based on the number of mutations per branch, the average number of mutations per cell division was calculated. This analysis was done separately for both the targeted mutations and the mutations that were identified in the fetal clones.

**Mutation rate of iPSCs and ISCs**
Culture-associated SNVs and indels were identified by subtracting the clonal mutations present in de sub clone from the clonal mutations present in the matching-clone. Culture-associated SNVs and indels were pooled per clone. The culture-associated mutation rate of iPSCs and ISCs in each clone was calculated by the total number of culture-associated mutations, the number of days in culture and the average cell cycle length determined by pulse-chase of 5-ethynyl-2’-deoxyuridine (EdU) and co-staining for the mitosis phosphorylated histone 3 (pH3). The cell cycle length of ISCs in culture was obtained from a previously published study, which measured the cell cycle length with a similar approach (Drost et al., 2017).

**Mutational profiles and signature analyses**
We determined all somatic mutations, which are shared between tissues or cells from at least 2 different germ layers as somatic mutations acquired pre-gastrulation. For this we included: 1) shared somatic mutations between ISC and HSPC clones 2) somatic mutations present in bulk skin identified with WGS 3) somatic mutations shared between skin and bulk ISC passage 0 identified with targeted re-sequencing. Somatic mutations acquired pre-gastrulation of the 4 fetuses are pooled together, because of the low number of mutations. All somatic mutations were counted once. In addition, all culture-associated mutations in iPSCs were pooled together. Mutational profile and mutational signature analyses were performed on pooled samples with the new functions of the R package MutationalPatterns (release 3.13) https://github.com/ToolsVanBox/MutationalPatterns (Blokzijl et al., 2018). A chi squared test was performed on the 7-type mutation spectra to determine differences between the mutation spectra of cultured iPSCs and pluripotent embryonic cells in vivo. Mutational signature analyses was performed with the newest version of the COSMIC signatures (v 3.21) (Alexandrov et al., 2020). De novo mutational signature extraction was performed using the data reported here, in combination with genome-wide mutation data of healthy stem cells of human small intestine, colon and liver as described previously (Blokzijl et al., 2018; Osorio et al., 2018). We applied nonnegative matrix factorization and extracted 2 mutational signatures based on the residual sum of squares (RSS) plot and plot_original_vs_reconstructed function. The 2 de novo extracted signatures were compared to the COSMIC signatures and based on a cosine similarity being >0.85 were identified as SBS1 and SBS5. With these 2 signatures we performed a strict refit on the pooled samples with the fit_to_signatures_strict function with the setting max_delta=0.05 and best fit option.

**QUANTIFICATION AND STATISTICAL ANALYSIS**
Sample numbers used for statistical analysis are depicted in the figures.

To determine the somatic mutation rate pre-gastrulation, we used the VAF to determine if the mutations in a branch occurred during a single cell division. In ambiguous cases, a Chi-square test with a simulated p-value (k=2000), was performed on the reference- and alternative allele counts of all mutations in a branch. If the VAFs between mutations were significantly different, then the mutations occurred during multiple cell divisions. A Wilcoxon rank-sum test was used to compare the mutation rate of iPSCs and embryonic cells pre-gastrulation, and to compare the culture-associated mutation rate of iPSCs and fetal ISCs.

A Chi-squared test was used to determine differences in the 7-mutation spectra of cultured iPSC and embryonic cell pre-gastrulation. We used the cosine similarity to compare the 96-mutational spectra of iPSCs and embryonic cells pre-gastrulation.