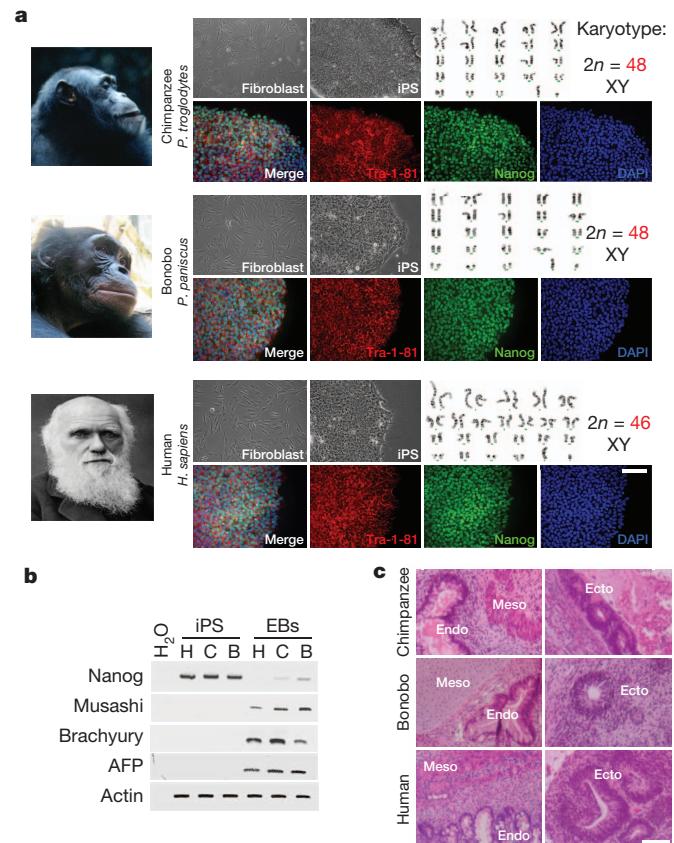


# Differential L1 regulation in pluripotent stem cells of humans and apes

Maria C. N. Marchetto<sup>1\*</sup>, Iñigo Narvaiza<sup>1\*</sup>, Ahmet M. Denli<sup>1</sup>, Christopher Benner<sup>1</sup>, Thomas A. Lazzarini<sup>1</sup>, Jason L. Nathanson<sup>2</sup>, Apuā C. M. Paquola<sup>1</sup>, Keval N. Desai<sup>3</sup>, Roberto H. Herai<sup>4</sup>, Matthew D. Weitzman<sup>5</sup>, Gene W. Yeo<sup>2</sup>, Alysson R. Muotri<sup>4,6</sup> & Fred H. Gage<sup>1,6</sup>

Identifying cellular and molecular differences between human and non-human primates (NHPs) is essential to the basic understanding of the evolution and diversity of our own species. Until now, preserved tissues have been the main source for most comparative studies between humans, chimpanzees (*Pan troglodytes*) and bonobos (*Pan paniscus*)<sup>1,2</sup>. However, these tissue samples do not fairly represent the distinctive traits of live cell behaviour and are not amenable to genetic manipulation. We propose that induced pluripotent stem (iPS) cells could be a unique biological resource to determine relevant phenotypical differences between human and NHPs, and that those differences could have potential adaptation and speciation value. Here we describe the generation and initial characterization of iPS cells from chimpanzees and bonobos as new tools to explore factors that may have contributed to great ape evolution. Comparative gene expression analysis of human and NHP iPS cells revealed differences in the regulation of long interspersed element-1 (L1, also known as LINE-1) transposons. A force of change in mammalian evolution, L1 elements are retrotransposons that have remained active during primate evolution<sup>3–5</sup>. Decreased levels of L1-restricting factors APOBEC3B (also known as A3B)<sup>6</sup> and PIWIL2 (ref. 7) in NHP iPS cells correlated with increased L1 mobility and endogenous L1 messenger RNA levels. Moreover, results from the manipulation of A3B and PIWIL2 levels in iPS cells supported a causal inverse relationship between levels of these proteins and L1 retrotransposition. Finally, we found increased copy numbers of species-specific L1 elements in the genome of chimpanzees compared to humans, supporting the idea that increased L1 mobility in NHPs is not limited to iPS cells in culture and may have also occurred in the germ line or embryonic cells developmentally upstream to germline specification during primate evolution. We propose that differences in L1 mobility may have differentially shaped the genomes of humans and NHPs and could have continuing adaptive significance.

Humans, chimpanzees and bonobos are genetically very similar, sharing nearly 98% of their alignable genomic sequence<sup>1–3</sup>. However, cellular and molecular phenotypes, especially at identical stages of development, are difficult to establish, mainly owing to limited access to embryonic material from humans and NHPs<sup>8</sup>. We reprogrammed fibroblasts from two bonobos and two chimpanzees into iPS cells as previously described<sup>9,10</sup> (Extended Data Fig. 1a). After culture in human embryonic stem (ES) cell-supporting conditions, NHP iPS cell colonies could be distinguished by the high nucleus-to-cytoplasm ratio morphology. iPS cell clones from both species continuously expressed pluripotency markers, retained an undifferentiated morphology in culture, and maintained a normal karyotype (Fig. 1a). After



**Figure 1 | Characterization of iPS cells derived from the three primate species.** **a**, Morphology of fibroblasts and iPS cells. No karyotypic abnormalities were observed in iPS cell clones. Immunofluorescence for the pluripotency markers Tra-1-81 and Nanog in iPS cells is shown. DAPI, 4',6-diamidino-2-phenylindole. **b**, Reverse transcription PCR (RT-PCR) for undifferentiation (Nanog) and for the three germ cell layers (musashi, brachyury and  $\alpha$ -fetoprotein (AFP)) markers in human (H), chimpanzee (C) and bonobo (B) iPS cells, and in differentiated embryoid bodies (EBs). **c**, Haematoxylin and eosin staining of teratoma sections showing differentiation into three germ layers: goblet cells in gastrointestinal tract (endo), neuroretinal epithelium (ecto), and muscle and cartilage/bone (meso). Scale bars, 100  $\mu$ m (a) and 150  $\mu$ m (c).

<sup>1</sup>Laboratory of Genetics, The Salk Institute for Biological Studies, 10010 North Torrey Pines Road, La Jolla, California 92037, USA. <sup>2</sup>University of California San Diego, Department of Cellular and Molecular Medicine, Stem Cell Program, Institute for Genomic Medicine, Sanford Consortium for Regenerative Medicine, 2880 Torrey Pines Scenic Drive, La Jolla, California 92037, USA. <sup>3</sup>University of California San Diego, Division of Biological Sciences, 9500 Gilman Drive, La Jolla, California 92093, USA. <sup>4</sup>University of California San Diego, School of Medicine, Department of Pediatrics/Rady Children's Hospital San Diego, Department of Cellular & Molecular Medicine, Stem Cell Program, Sanford Consortium, MC 0695, 2880 Torrey Pines Scenic Drive, La Jolla, California 92093, USA. <sup>5</sup>Department of Pathology and Laboratory Medicine, University of Pennsylvania Perelman School of Medicine and Center for Cellular and Molecular Therapeutics, The Children's Hospital of Philadelphia, Philadelphia, Pennsylvania 19104-4318, USA. <sup>6</sup>Center for Academic Research and Training in Anthropogeny (CARTA), 9500 Gilman Drive, La Jolla, California 92093, USA.

\*These authors contributed equally to this work.

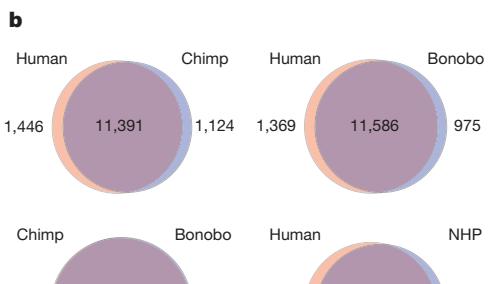
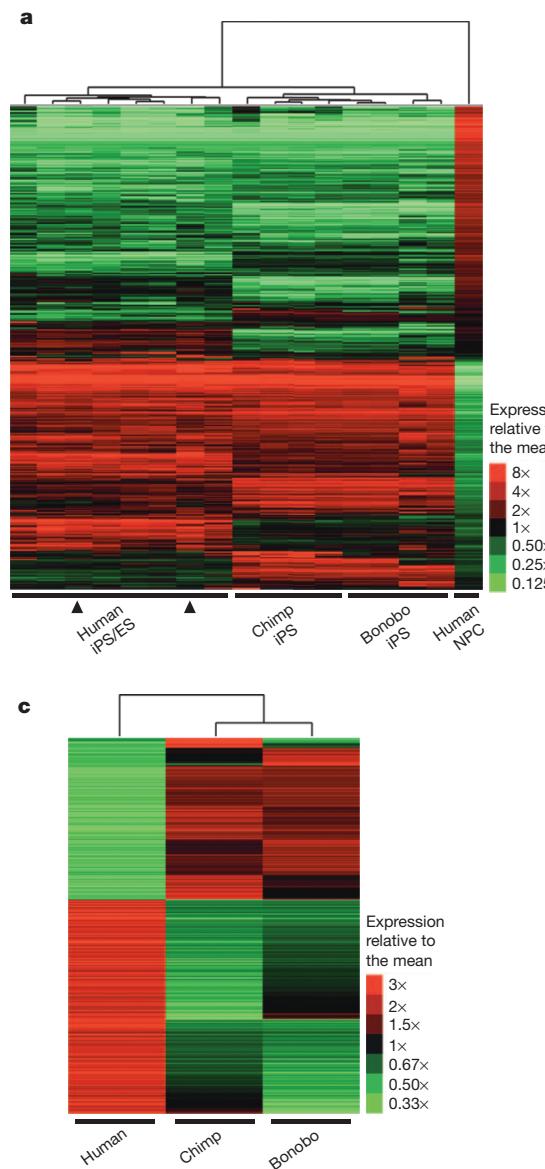
embryoid-body-mediated differentiation *in vitro*, clones contained tissue derivatives from the three embryonic germ layers and down-regulated expression of pluripotency markers (Fig. 1b). iPS-cell-selected clones were also able to differentiate into the three embryonic germ layers *in vivo*, as shown by analysis of teratomas in nude mice (Fig. 1c). Together, these data demonstrate that NHP iPS cell clones re-established pluripotency at the molecular and cellular levels.

To gain insight into differences in gene expression between human and NHP iPS cells, we performed high-throughput RNA sequencing (RNA-seq) analyses on four human, two chimpanzee and two bonobo iPS cell lines (Extended Data Fig. 1b). The expression profiles of iPS cells from the three species clustered together with human ES cells (HUES6 and H1), and were distinguishable from ES-cell-derived neural progenitor cells (Fig. 2a); chimpanzee and bonobo iPS cells clustered closer to each other than to human iPS cells (Fig. 2a). We then performed pairwise comparisons of protein-coding gene expression levels (Fig. 2b). Venn diagrams in Fig. 2b represent expressed genes with non-significant differences between species (purple), and upregulated genes with estimated false discovery rates (FDR) of less than 5% and a fold change greater than twofold (pale orange and blue). Comparison between humans and NHPs (Fig. 2b, bottom right) revealed 1,376 genes with increased expression in human iPS cells, and 1,042 common

genes with increased expression in NHP iPS cells, whereas no significant differences were observed in 11,585 protein-coding genes. Next, we focused on genes differentially expressed between human and NHP iPS cells (Fig. 2c–e and Extended Data Fig. 1c, d), and found, among the top 50 genes with increased expression in human compared to NHP iPS cells, two genes involved in the restriction of L1 retrotransposition, namely *A3B* and *PIWIL2* (Fig. 2d).

Active, full-length L1 elements have the ability to move from one location in the genome to another by a copy-paste mechanism known as retrotransposition<sup>11</sup>. Active L1 elements have been detected in both germline and somatic tissues, and can affect genome integrity<sup>12,13</sup>. As uncontrolled retrotransposition activity can be deleterious to the host<sup>14</sup>, organisms have evolved mechanisms to control L1 mobility<sup>11</sup>. *A3B* is a member of the APOBEC3 family of cytidine deaminases that can inhibit L1 mobility in different cell types, including human ES and iPS cells, via a still unclear mechanism<sup>6,15,16</sup>. *PIWIL2* is an effector of the Piwi-interacting RNA (piRNA) pathway involved in L1 silencing mainly in the germ line<sup>7</sup>.

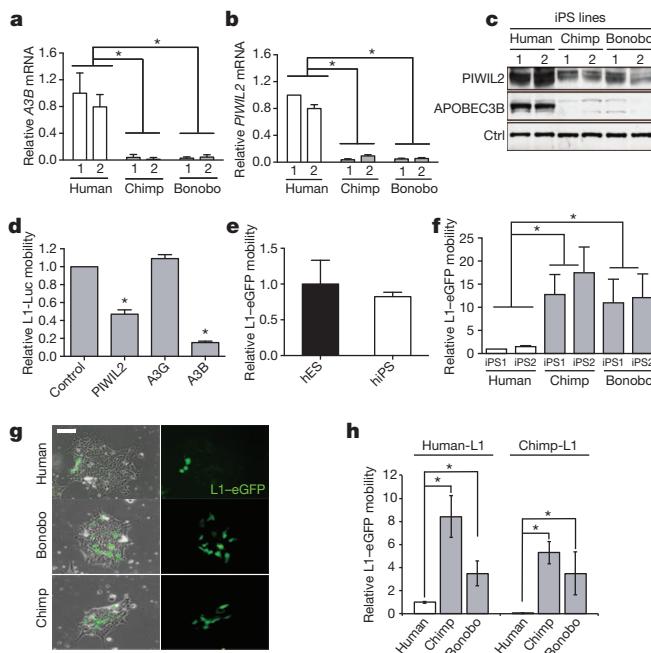
To confirm differences in *A3B* and *PIWIL2* in human versus NHP iPS cells, we first cloned their complementary DNAs from the three species, and found a high degree of conservation between humans and NHPs (Extended Data Fig. 2). Quantification of *A3B* mRNA levels by



#	Name	Log FC	FDR
1	<i>RPL12</i>	5.45	$8.54 \times 10^{-168}$
2	<i>NPF2</i>	6.82	$2.20 \times 10^{-53}$
3	<i>MM24</i>	5.83	$1.52 \times 10^{-118}$
4	<i>PCAM1</i>	5.41	$6.52 \times 10^{-93}$
5	<i>IDC1</i>	6.50	$2.26 \times 10^{-90}$
6	<i>EIF3F</i>	2.88	$8.94 \times 10^{-83}$
7	<i>PBDHD</i>	5.41	$2.50 \times 10^{-82}$
8	<i>PIWIL2</i>	6.91	$4.40 \times 10^{-82}$
9	<i>COL6A2</i>	3.88	$1.79 \times 10^{-61}$
10	<i>COL14A1</i>	3.35	$6.05 \times 10^{-58}$
11	<i>CBS</i>	2.30	$3.21 \times 10^{-57}$
12	<i>VSNL1</i>	4.91	$3.53 \times 10^{-55}$
13	<i>SMPD3B</i>	2.88	$1.54 \times 10^{-51}$
14	<i>DPEP3</i>	6.88	$4.63 \times 10^{-43}$
15	<i>MILPH</i>	4.92	$1.02 \times 10^{-42}$
16	<i>DPIVSL3</i>	2.77	$2.00 \times 10^{-40}$
17	<i>KLK1B1</i>	6.19	$1.18 \times 10^{-40}$
18	<i>SLC38A4</i>	4.33	$8.48 \times 10^{-40}$
19	<i>COL6A1</i>	4.09	$4.97 \times 10^{-38}$
20	<i>GPR114</i>	5.03	$5.77 \times 10^{-38}$
21	<i>TUBB3</i>	2.42	$1.11 \times 10^{-36}$
22	<i>DCLK1</i>	2.83	$2.04 \times 10^{-36}$
23	<i>SCN4A</i>	4.70	$2.23 \times 10^{-36}$
24	<i>SLC17A9</i>	3.67	$2.52 \times 10^{-35}$
25	<i>TMEM64</i>	3.33	$6.49 \times 10^{-35}$
26	<i>PFA1</i>	5.51	$1.48 \times 10^{-35}$
27	<i>CDP4</i>	2.00	$4.17 \times 10^{-35}$
28	<i>IHLA2</i>	5.07	$3.91 \times 10^{-35}$
29	<i>PRSA</i>	1.73	$1.29 \times 10^{-34}$
30	<i>PDLIM1</i>	2.24	$1.60 \times 10^{-34}$
31	<i>PBLT</i>	2.31	$3.10 \times 10^{-34}$
32	<i>VLDLR</i>	2.55	$4.60 \times 10^{-34}$
33	<i>PRAS2</i>	1.85	$4.80 \times 10^{-34}$
34	<i>TRIM61</i>	7.13	$3.11 \times 10^{-33}$
35	<i>PLA2G16</i>	2.68	$4.78 \times 10^{-33}$
36	<i>MEDAG</i>	4.55	$6.76 \times 10^{-33}$
37	<i>CCNB1IP1</i>	1.83	$9.31 \times 10^{-33}$
38	<i>APOBEC3B</i>	3.83	$2.21 \times 10^{-31}$
39	<i>MXRA7</i>	2.64	$2.53 \times 10^{-31}$
40	<i>BOLA2</i>	7.48	$1.65 \times 10^{-29}$
41	<i>HK1</i>	1.53	$2.25 \times 10^{-28}$
42	<i>FAM71F1</i>	5.31	$6.22 \times 10^{-28}$
43	<i>TMEM158</i>	3.33	$1.93 \times 10^{-28}$
44	<i>DBNDD1</i>	2.02	$2.27 \times 10^{-28}$
45	<i>SNX25</i>	2.96	$6.07 \times 10^{-28}$
46	<i>PAQR5</i>	3.77	$1.30 \times 10^{-27}$
47	<i>NIP1PB3</i>	6.93	$1.81 \times 10^{-27}$
48	<i>FAM20A</i>	5.86	$1.82 \times 10^{-27}$
49	<i>NID1</i>	1.89	$2.08 \times 10^{-27}$
50	<i>NNMT2</i>	2.70	$4.95 \times 10^{-27}$

**Figure 2 | RNA-seq profiling of human and NHP iPS cells. a**, High-throughput sequencing of 14 RNA samples corresponding to four human, two chimp and two bonobo iPS cell lines. Expression profiles of human ES cells (H1 and HUES6, arrowheads) and ES-cell-derived neural progenitor cells (NPCs) are shown. Heat-map representation of mapped reads corresponds to protein-coding genes. **b**, Venn diagrams showing pairwise comparison of protein-coding genes. Pale orange and blue denote significantly upregulated genes (FDR < 0.05 and fold change > 2); purple denotes expressed genes with no significant differences in mRNA levels between compared species.

**c**, Heat-map representation of differentially expressed protein-coding genes with FDR < 0.05 and fold change > 2 between human and NHP iPS cells. **d**, e, List of the top 50 differentially expressed genes in human compared to NHP iPS cells, with increased expression in human (d) and NHP (e) iPS cells. *PIWIL2* and *APOBEC3B* were expressed at significantly higher levels in human than in NHP iPS cells (positions 8 and 38 in d, respectively). Rank, gene name, logarithmic base 2 fold change (log FC), and FDR are shown.



**Figure 3 | Reduced levels of A3B and PIWIL2 and increased L1 mobility in NHP iPS cells.** **a, b**, qPCR analysis of A3B (**a**) and PIWIL2 (**b**) expression in human and NHP iPS cells (Extended Data Figs 2 and 9). **c**, Immunoblot for A3B and PIWIL2. **d**, Effect of A3B and PIWIL2 on L1-expressing firefly luciferase (L1-Luc) retrotransposition in 293T cells. 293T cells were cotransfected with L1-luciferase plasmid (pYX017)<sup>19</sup> plus control (ctrl), PIWIL2-, A3B- or A3G-expressing plasmid. L1-luciferase mobility was calculated as firefly luciferase units relative to *Renilla* luciferase units. L1 activity is shown relative to control. **e**, Comparable levels of L1-eGFP<sup>18</sup> mobility in human ES (hES) and iPS (hiPS) cells. L1-eGFP mobility is shown as a percentage of eGFP-positive cells by fluorescence-activated cell sorting (FACS) relative to human ES cells. **f**, L1-eGFP retrotransposition in human, chimp and bonobo iPS cells. L1 mobility was calculated as a percentage of eGFP-positive cells and shown as relative L1 mobility to human iPS cell line 1 (iPS1). **g**, Representative images of human, chimpanzee and bonobo iPS cells transfected with L1-eGFP. Scale bar, 50 µm. **h**, Retrotransposition quantification of species-specific L1 elements. The mobility of human and chimp reporter L1-eGFP elements (human-L1 and chimp-L1, respectively) was quantified in transfected human, chimp and bonobo iPS cells. Retrotransposition activity is shown relative to human-L1 activity in human iPS cells. Error bars denote s.e.m. \*P < 0.01 between indicated groups using t-test (*n* = 3 (**a**, **b**, **e** and **f**) and 4 (**d** and **h**) biological replicates).

quantitative PCR (qPCR) confirmed significantly higher levels (~30-fold) of A3B in both human iPS cell lines compared to NHP iPS cells (Fig. 3a). Levels of PIWIL2 mRNA were 16-fold higher in human than in NHP iPS cell lines (Fig. 3b). PIWIL2-mediated control of transposons is most active in the germ line, and we observed that levels of PIWIL2 mRNA are 20–40-fold lower in human iPS cells than in the testis (Extended Data Fig. 3a). The increased expression observed in human iPS cells seems to be specifically restricted to A3B and PIWIL2 compared to other members of these protein families (Extended Data Fig. 3b, c). Differences in A3B and PIWIL2 mRNA levels reflected higher A3B and PIWIL2 protein levels in human versus NHP iPS cells (Fig. 3c).

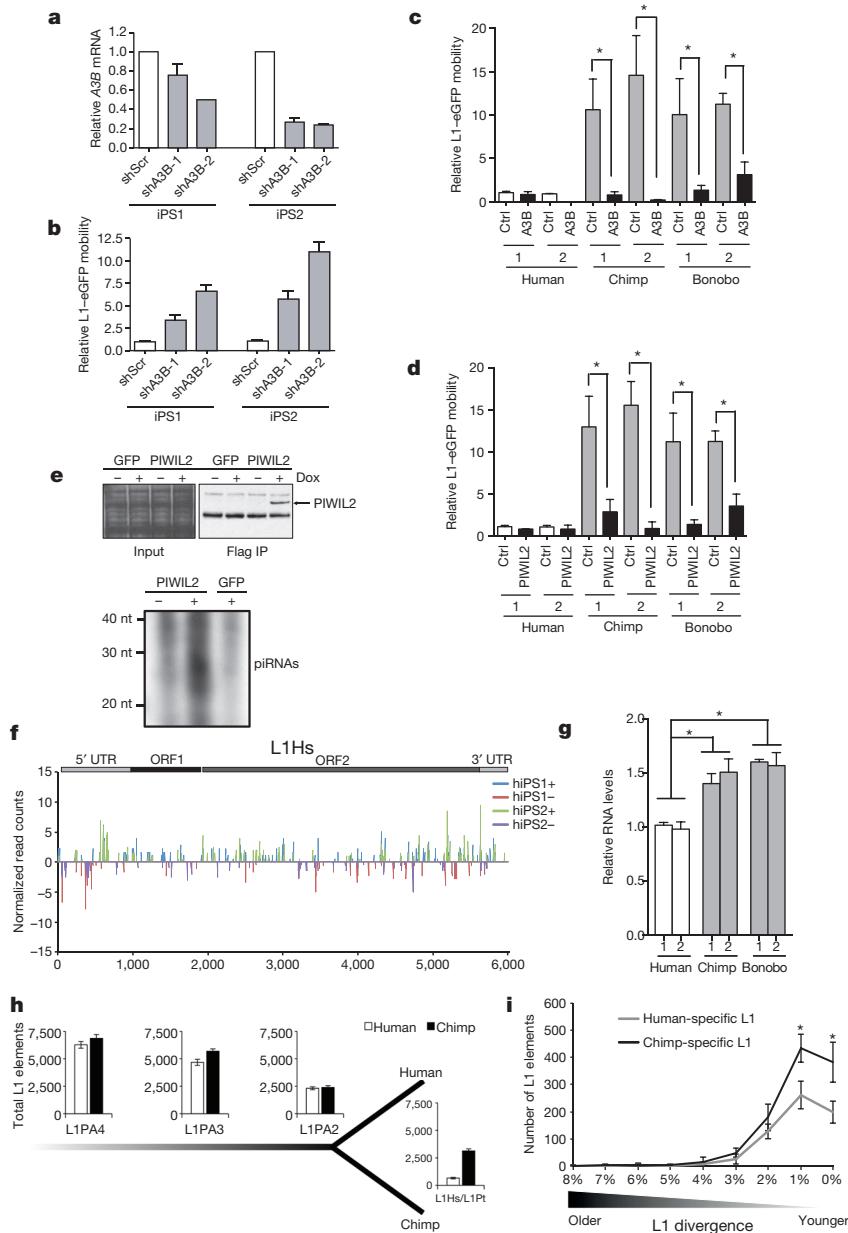
Ectopic expression of A3B has been shown to inhibit the mobility of human L1 reporter elements<sup>6,17–19</sup> (Extended Data Fig. 4a). In 293T cells, ectopic expression of human A3B significantly reduced L1-expressing firefly luciferase<sup>18,19</sup> mobility by fivefold compared to control plasmid or a plasmid expressing A3G, another APOBEC3 protein that lacks anti-L1 activity<sup>17</sup> (Fig. 3d). We also found a significant decrease in L1-luciferase retrotransposition in cells overexpressing PIWIL2 compared to control transfected cells (Fig. 3d). We then confirmed that human L1 can retrotranspose in human ES and iPS cells under our

culture conditions, as previously shown<sup>15,20,21</sup> (Fig. 3e). Because we found reduced levels of L1 restriction factors A3B and PIWIL2 in NHPs, we compared L1 activity in human versus NHP iPS cells using human L1 tagged with the enhanced green fluorescent protein (eGFP) reporter element<sup>18,19</sup>. L1 retrotransposition was significantly higher in NHP compared to human iPS cell lines, with 10- and 8-fold increases in eGFP-positive cells in chimpanzee and bonobo iPS cells, respectively (Fig. 3f, g). To test whether the differential L1 regulation in iPS cells is specific to human L1, we measured the activity of an NHP L1 element in iPS cells. We generated a retrotransposition-competent chimpanzee L1-eGFP reporter element (chimp-L1) (Extended Data Fig. 5), and observed that chimp-L1 was significantly more active in NHP than in human iPS cells (Fig. 3h), suggesting that the decreased L1 activity in human iPS cells is not specific to the human L1 element, and that human iPS cells are more efficient in repressing L1 retrotransposition than NHP iPS cells.

To analyse the contribution of endogenous A3B to the differential L1 activity observed among primate cells, we generated human ES and iPS cells with decreased levels of A3B (Fig. 4a and Extended Data Fig. 4b, c). Stable short hairpin RNA (shRNA)-mediated knockdown of A3B (shA3B) resulted in a significant increase in L1-eGFP activity compared to scramble (shScr) control cells in human iPS cells (Fig. 4a, b). Knockdown was specific to A3B, not affecting other APOBEC3 proteins (Extended Data Fig. 4d–f). As expected, L1 mobility was significantly decreased in both chimpanzee and bonobo iPS cell lines after A3B overexpression (Fig. 4c). Similarly, overexpression of chimpanzee or bonobo PIWIL2 in NHP iPS cells reduced L1 mobility to levels detected in human iPS cells (Fig. 4d). We did not observe differences in the levels of L1-eGFP mRNA expressed from transfected L1 plasmid or in L1 promoter activity between human and NHP iPS cells (Extended Data Fig. 4g–i). Together, these results suggest that differences in A3B and PIWIL2 expression levels contribute to higher L1 retrotransposition in NHP than in human iPS cells.

PIWIL2 repression of transposons is mediated through piRNAs<sup>22</sup>. Thus, we analysed the presence of PIWIL2-bound piRNAs in doxycycline-inducible human iPS cells expressing Flag-tagged PIWIL2 by immunoprecipitation and subsequent 5' end labelling. Analysis of PIWIL2-associated small RNAs revealed the presence of ~26–30-nucleotide RNAs only in cells expressing Flag-PIWIL2 but not in control cells or in pulldowns with control antibody (Fig. 4e and Extended Data Fig. 6a, b). Next, to probe for the presence of L1-targeting-piRNAs, we characterized the small RNA populations in human iPS cells by small RNA-seq analysis (Extended Data Fig. 6c, d and Supplementary Tables 1 and 2). We detected 272 and 229 annotated piRNAs in human iPS cell lines 1 and 2, respectively (Extended Data Fig. 6d–f and Supplementary Table 2). In addition, we observed a number of 26–33-nucleotide small RNAs mapping to the consensus human-specific L1 element (L1 *Homo sapiens*; L1Hs) sequence (Fig. 4 and Extended Data Fig. 7a, b), including 12 and 10 of the 37 annotated piRNAs mapping to L1Hs in piRNABank (<http://pirnabank.ibab.ac.in/>) in hiPS cell lines 1 and 2, respectively (Extended Data Figs 6e, g and 7a–c). Together, these results demonstrate the presence of piRNAs complementary to L1Hs in human iPS cells.

We then asked whether different levels of L1 reporter mobility between human and NHP iPS cells reflect differences in endogenous L1 activity. First, we analysed endogenous L1 RNA levels by qPCR, and found higher levels of endogenous L1 mRNA in chimpanzee and bonobo than in human iPS cells (Fig. 4g and Extended Data Fig. 8a–c). Next, we examined the number of L1 elements in human and chimpanzee genomes to assess differences in recent L1 mobility. We did not observe major differences in the number of L1 elements for older families (L1PA4, L1AP3 and L1PA2; approximately 18, 12.5 and 7.6 million years old, respectively)<sup>23,24</sup> (Fig. 4h). However, we did observe significantly higher numbers of chimpanzee-specific L1 elements (L1 *Pan troglodytes*; L1Pt) compared to L1Hs elements<sup>25,26</sup> (Fig. 4h). Differences in the expression of A3B and PIWIL2 suggest that L1 mobility may



**Figure 4 | Species-specific L1 elements are more abundant in chimpanzee genomes than in human genomes, correlating with decreased levels of A3B and PIWIL2.** **a**, Stable shRNA-mediated knockdown of A3B (shA3B-1, shA3B-2) or control (shScr) in human iPS cells. A3B expression was normalized to GAPDH and shown relative to shScr. **b**, L1-eGFP mobility in shA3B iPS cells. eGFP-positive cells were quantified by FACS analysis and shown relative to shScr control. **c, d**, Overexpression of A3B (**c**) and PIWIL2 (**d**) decreases L1-eGFP retrotransposition in NHP iPS cells. Cells were electroporated with L1-eGFP plus control, A3B- or PIWIL2-expressing plasmids. L1-eGFP mobility is shown relative to human iPS cell-1 control. **e**, Immunoprecipitation (IP) of piRNAs associated with PIWIL2 ribonucleoproteins (RNPs) from Tet-inducible GFP and Flag-tagged PIWIL2 human iPS cells after addition of doxycycline (dox). Bottom, [ $\gamma$ -<sup>32</sup>P]ATP 5'-end labelling of RNA associated with Flag-PIWIL2 RNPs. Size markers are indicated (nt, nucleotides). **f**, Mapping of 26–33-nucleotide RNA reads (containing uracil at the 5' end and/or adenine at position +10) detected by small RNA-seq from human iPS cell lines 1 and 2 to consensus L1Hs (Repbase). Positive and negative values indicate sense and antisense piRNAs, respectively. Schematic representation of L1 is shown on top. Read counts were normalized to  $10^7$  reads per experiment. **g**, qRT-PCR analysis of endogenous L1 RNA in human and NHP iPS cells. Values represent average of relative levels for L1 RNA (5' untranslated region (UTR), open reading frame (ORF) 1 and 2), normalized to ACTB mRNA. L1 levels are shown relative to iPS cell line 1. **h**, Comparative quantitative analysis of L1 elements in human and chimpanzee genomes for L1 families L1PA4, L1PA3, L1PA2, L1Pt and L1Hs. **i**, Number of species-specific L1 insertions (L1PA2, L1Hs and L1Pt) relative to their divergence. L1 elements plotted as a histogram relative to their divergence (number of mutations relative to the canonical element). Error bars denote s.d. (**a–d** and **g**). \* $P < 0.001$  (**h**; between human and chimpanzee; Mann-Whitney *U* test) and \* $P < 0.01$  (**c, d** and **g**; between indicated groups; *t*-test).  $n = 3$  (**a–d**) and 4 (**g**) biological replicates.

have been altered at a relatively recent evolutionary divergence. Therefore, using divergence as a measurement of L1 age, we estimated the number of species-specific L1 loci, and found that the number of chimpanzee-specific loci was significantly higher than the number of human-specific loci (Fig. 4*i* and Extended Data Fig. 8d–g). This increased number of species-specific L1 loci in chimpanzee suggests that endogenous L1 has been more active in NHP genomes, correlating with the decreased levels of A3B and PIWIL2.

Here we show that iPS cells from both chimpanzees and bonobos have increased L1 mobility. Different rates of L1 activity could lead to considerable changes in genomic structure and function, and could potentially affect adaptation. The human population has gone through one or more bottlenecks throughout evolution that might have contributed to decreased genetic diversity<sup>27</sup>. Chimpanzees and bonobos, in contrast, have increased levels of genetic diversity when compared to humans<sup>28</sup>. This idea is also supported by data showing that there is substantially more genetic difference among individuals within chimpanzee troops in West Africa than among all living humans<sup>29</sup>. Although it remains unclear what the main generators of the phenotypic differences between us and our closest living relatives are (despite the

extreme genetic similarity), we propose that L1 mobility could be involved in differentially shaping the genomes of humans and NHPs, providing an extra layer of variability to the latter. In fact, recent studies have suggested that ongoing L1 retrotransposition may contribute to inter-individual genetic variation<sup>30</sup>. In this work we also present a new perspective on the use of iPS cell technology as a powerful tool for the study of early stages of development and possible validation of evolutionary genomic and transcriptomic modifications that identify humans as outliers among primates. The iPS cells from great apes that we describe here can also be used for comparative studies of any derivative pluripotent or terminally differentiated cell types, limited solely by the availability of differentiation protocols.

## METHODS SUMMARY

Reprogramming of fibroblast was performed by transduction with retroviral vectors expressing OCT4 (also known as POU5F1), MYC, KLF4 and SOX2 human cDNAs. For RNA-seq, libraries from polyA<sup>+</sup> RNA and small RNA were generated using the Illumina TruSeq RNA and Small RNA TruSeq Sample Prep kits, respectively, and analysed on an Illumina HiSeq 2000 sequencer. L1 reporter assays were performed as previously described<sup>18,19</sup>. Quantification of L1-derived genomic

sequences was based on Repbase defined elements annotated by RepeatMasker (<http://www.repeatmasker.org>). L1 genomic positions for human (hg19, GRCh37) and chimpanzee (panTro3, CGSC 2.1.3) genomes were downloaded from the UCSC Genome Browser annotation database. To identify reference L1 elements that were inserted into the genome after the last common ancestor for human and chimpanzee, L1 elements were mapped between homologous regions of each genome using the UCSC LiftOver tool.

**Online Content** Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 3 October 2012; accepted 17 September 2013.

Published online 23 October 2013.

1. The Chimpanzee Sequencing and Analysis Consortium. Initial sequence of the chimpanzee genome and comparison with the human genome. *Nature* **437**, 69–87 (2005).
2. Prüfer, K. et al. The bonobo genome compared with the chimpanzee and human genomes. *Nature* **486**, 527–531 (2012).
3. Lander, E. S. et al. Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921 (2001).
4. Waterston, R. H. et al. Initial sequencing and comparative analysis of the mouse genome. *Nature* **420**, 520–562 (2002).
5. Kazazian, H. H. Jr. Mobile elements: drivers of genome evolution. *Science* **303**, 1626–1632 (2004).
6. Bogerd, H. P. et al. Cellular inhibitors of long interspersed element 1 and Alu retrotransposition. *Proc. Natl Acad. Sci. USA* **103**, 8780–8785 (2006).
7. Aravin, A. A., Sachidanandam, R., Girard, A., Fejes-Toth, K. & Hannon, G. J. Developmentally regulated piRNA clusters implicate MILI in transposon control. *Science* **316**, 744–747 (2007).
8. Varki, A. & Altheide, T. K. Comparing the human and chimpanzee genomes: searching for needles in a haystack. *Genome Res.* **15**, 1746–1758 (2005).
9. Takahashi, K. et al. Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* **131**, 861–872 (2007).
10. Yu, J. et al. Induced pluripotent stem cell lines derived from human somatic cells. *Science* **318**, 1917–1920 (2007).
11. Burns, K. H. & Boeke, J. D. Human transposon tectonics. *Cell* **149**, 740–752 (2012).
12. Kazazian, H. H. Jr. Mobile elements and disease. *Curr. Opin. Genet. Dev.* **8**, 343–350 (1998).
13. Muotri, A. R., Marchetto, M. C., Coufal, N. G. & Gage, F. H. The necessary junk: new functions for transposable elements. *Hum. Mol. Genet.* **16**, R159–R167 (2007).
14. Beck, C. R., Garcia-Perez, J. L., Badge, R. M. & Moran, J. V. LINE-1 elements in structural variation and disease. *Annu. Rev. Genomics Hum. Genet.* **12**, 187–215 (2011).
15. Wissing, S., Montano, M., Garcia-Perez, J. L., Moran, J. V. & Greene, W. C. Endogenous APOBEC3B restricts LINE-1 retrotransposition in transformed cells and human embryonic stem cells. *J. Biol. Chem.* **286**, 36427–36437 (2011).
16. Chiu, Y. L. & Greene, W. C. The APOBEC3 cytidine deaminases: an innate defensive network opposing exogenous retroviruses and endogenous retroelements. *Annu. Rev. Immunol.* **26**, 317–353 (2008).
17. Chen, H. et al. APOBEC3A is a potent inhibitor of adeno-associated virus and retrotransposons. *Curr. Biol.* **16**, 480–485 (2006).
18. Ostertag, E. M., Prak, E. T., DeBerardinis, R. J., Moran, J. V. & Kazazian, H. H. Jr. Determination of L1 retrotransposition kinetics in cultured cells. *Nucleic Acids Res.* **28**, 1418–1423 (2000).
19. Xie, Y., Rosser, J. M., Thompson, T. L., Boeke, J. D. & An, W. Characterization of L1 retrotransposition with high-throughput dual-luciferase assays. *Nucleic Acids Res.* **39**, e16 (2011).
20. Garcia-Perez, J. L. et al. LINE-1 retrotransposition in human embryonic stem cells. *Hum. Mol. Genet.* **16**, 1569–1577 (2007).
21. Wissing, S. et al. Reprogramming somatic cells into iPS cells activates LINE-1 retroelement mobility. *Hum. Mol. Genet.* **21**, 208–218 (2012).
22. De Fazio, S. et al. The endonuclease activity of Mili fuels piRNA amplification that silences LINE1 elements. *Nature* **480**, 259–263 (2011).
23. Mathews, L. M., Chi, S. Y., Greenberg, N., Ovchinnikov, I. & Swergold, G. D. Large differences between LINE-1 amplification rates in the human and chimpanzee lineages. *Am. J. Hum. Genet.* **72**, 739–748 (2003).
24. Khan, H., Smit, A. & Boissinot, S. Molecular evolution and tempo of amplification of human LINE-1 retrotransposons since the origin of primates. *Genome Res.* **16**, 78–87 (2006).
25. Lee, J. et al. Different evolutionary fates of recently integrated human and chimpanzee LINE-1 retrotransposons. *Gene* **390**, 18–27 (2007).
26. Mills, R. E. et al. Recently mobilized transposons in the human and chimpanzee genomes. *Am. J. Hum. Genet.* **78**, 671–679 (2006).
27. Campbell, M. C. & Tishkoff, S. A. African genetic diversity: implications for human demographic history, modern human origins, and complex disease mapping. *Annu. Rev. Genomics Hum. Genet.* **9**, 403–433 (2008).
28. Gagneux, P. et al. Mitochondrial sequences show diverse evolutionary histories of African hominoids. *Proc. Natl Acad. Sci. USA* **96**, 5077–5082 (1999).
29. Bowden, R. et al. Genomic tools for evolution and conservation in the chimpanzee: *Pan troglodytes elliotti* is a genetically distinct population. *PLoS Genet.* **8**, e1002504 (2012).
30. Beck, C. R. et al. LINE-1 retrotransposition activity in human genomes. *Cell* **141**, 1159–1170 (2010).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** The work was supported by funds from the National Institutes of Health (NIH) (TR01: MH095741 and Eureka: MH08848 to F.H.G.), the Mathers Foundation and the Helmsley Foundation. This work was also partially supported by funds from the NIH to A.R.M. (MH094753), M.D.W. (AI074967) and G.W.Y. (NS075449, HG004659 and GM084317). G.W.Y. is a recipient of the Alfred P. Sloan Research Fellowship. We thank J. V. Moran and W. An for reagents. We would like to thank A. Varki, P. Gagneux, L. Fourgeaud and I. Guimont for discussions, N. Varki for help with teratoma analysis, R. Keithley, I. Gallina and Y. Nunez for technical assistance, and M. L. Gage for editorial comments.

**Author Contributions** M.C.N.M. and I.N. are the leading authors. M.C.N.M., I.N. and A.M.D. contributed to the concept, designed and performed the experiments, and analysed the data. M.C.N.M. reprogrammed NHP fibroblasts and performed iPS cell cultures and transduction assays. I.N. and M.C.N.M. performed L1 assays. I.N. designed and performed biochemical experiments. A.M.D., C.B. and I.N. designed and performed comparative analysis of L1 insertions in the human and NHP genomes. T.A.L. produced lentiviruses and provided tissue culture assistance. I.N. and K.N.D. generated the chimp-L1 reporter plasmid. C.B., A.C.M.P. and R.H.H. performed bioinformatics analysis. I.N., C.B. and J.L.N. contributed to the generation of libraries and analysis of RNA-seq data. M.D.W., G.W.Y. and A.R.M. contributed to concept and financial support. F.H.G. is the senior author. He contributed to the concept, analysed the data, revised the manuscript and provided financial support. I.N., M.C.N.M., A.M.D. and F.H.G. wrote the manuscript. All the authors read and approved the final manuscript.

**Author Information** RNA-seq and small RNA-seq data have been deposited in the Gene Expression Omnibus under accession number GSE47626. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to F.H.G. (gage@salk.edu).

## METHODS

**Cell culture and retrovirus infection.** Human ES cells HUES6 and H1, human iPS cell lines WT-33, ADRC-40 (human iPS cell lines 1 and 2 in this work, respectively) and WT-126 were previously described<sup>31</sup>. Fibroblasts from human GM22159 (WT-9), *P. troglodytes* (chimpanzees: PR00818 and PR01209) and *P. paniscus* (bonobos: AG05253 and PR01086) were from Coriell Cell Repositories (NJ) (Extended Data Table 1). All fibroblasts were cultured in MEM (Invitrogen) supplemented with 10% FBS (HyClone Laboratories). Retroviral vectors expressing *OCT4* (also known as *POU5F1*), *MYC*, *KLF4* and *SOX2* human cDNAs from Yamanaka's group<sup>9</sup> were obtained from Addgene. Recombinant viruses were produced by transient transfection in 293T cells, as previously described<sup>32</sup>. Two days after infection, cells were plated on mitotically inactivated mouse embryonic fibroblasts (Chemicon) with human ES cell medium. After 2–4 weeks, iPS cell colonies were picked manually and directly transferred to feeder-free conditions on matrigel-coated dishes (BD) using mTeSR1 (StemCell Technologies). Established iPS cell colonies were kept in feeder-free conditions indefinitely, and passed using mechanical dissociation. Embryoid-body-mediated differentiation in suspension was carried out for 10 days in the absence of growth factors. The use of chimpanzee and bonobo fibroblast samples was approved by the US Fish and Wildlife Service, under the permit MA206206. Protocols describing the use of iPS and human ES cells were previously approved by the University of California, San Diego (UCSD), the Salk Institute Institutional Review Board and the Embryonic Stem Cell Research Oversight Committee<sup>31</sup>. To generate stable shA3B cell lines, HUES6, WT-33 and ADRC-40 cells were transduced with lentiviruses expressing shRNAs and selected for puromycin resistance. pLKO.1-based lentiviral plasmids encoding shRNAs against *A3B* (RHS3979-99216651 and RHS3979-99216658) were obtained from Open Biosystems. Recombinant lentiviruses were produced by transient transfection on 293T cells as previously described<sup>33</sup>.

**Teratoma formation in nude mice.** Around  $1 \times 10^6$ – $3 \times 10^6$  cells were injected subcutaneously into the dorsal flanks of nude mice (CByJ.Cg-Foxn1nu/J) anaesthetized with isoflurane. Five to six weeks after injection, teratomas were dissected, fixed overnight in 10% buffered formalin phosphate and embedded in paraffin. Tissues were then prepared for histopathologic analysis by the UCSD Mouse Phenotyping Services (<http://mousepheno.ucsd.edu>). In brief, the tissue was sectioned and stained with haematoxylin and eosin. Control mice injected with fibroblasts failed to form teratomas.

**Karyotyping.** Standard G-banding chromosome analysis was performed by Cell Line Genetics. Diploid human cells with  $2n = 46$  chromosomes; bonobo and chimpanzee with  $2n = 48$  chromosomes.

**RNA extraction and RT-PCR.** Total cellular RNA was extracted from  $\sim 5 \times 10^6$  cells using the RNeasy Protect Mini kit or RNeasy Plus kit (Qiagen), according to the manufacturer's instructions, and was reverse transcribed using the SuperScript III First-Strand Synthesis System RT-PCR from Invitrogen. For iPS cell markers, cDNA was amplified by PCR using Accuprime Taq DNA polymerase system (Invitrogen). Primer sequences are shown in Extended Data Table 1. PCR products were separated by electrophoresis on a 2% agarose gel, stained with ethidium bromide and visualized by ultraviolet illumination. Total RNA samples from human testis were obtained from Clontech. Small RNA was extracted using the mirVana kit (Ambion).

**Quantitative RT-PCR.** RNA was extracted using a QIAGEN RNeasy Plus kit or TRIzol (Life Technologies) and then retrotranscribed to cDNA with the Super-script III First-Strand synthesis system (Invitrogen). qRT-PCR reactions were carried out using SYBR Green mix (Roche) or TaqMan Assays (Life Technologies) using ABI Prism 7900HT sequence detection system (Applied Biosystems). The primers and Taqman sets used in this work are described in Extended Data Table 1. Data analysis was performed with SDS 2.3 software (Applied Biosystems). Primer efficiency was verified by linear regression to the standard curve. Values were normalized to *GAPDH*, *HPRT* or *ACTB*. Reactions were carried out in triplicate and data were analysed using the comparative ( $\Delta\Delta C_t$ ) method. For *A3B* and *PIWIL2*, RNA levels were normalized to *GAPDH* or *ACTB* and represented as relative to iPS cell line 1. Relative *A3B* and *PIWIL2* mRNA levels normalized to *GAPDH* for each individual iPS cell line and fibroblasts are shown in Extended Data Fig. 9. The reduced levels of *A3B* in NHP iPS cell were not due to an *A3B* deletion polymorphism previously described in human individuals<sup>34</sup> (data not shown). For L1 RNA qRT-PCR, values representing the average of relative levels for L1 RNA (5' UTR, ORF1 and ORF2) were calculated and normalized to actin mRNA levels. L1 levels are shown relative to iPS cell line 1. qRT-PCR analysis of L1-reporter expression in iPS cell lines transfected with L1-eGFP plasmid was carried out 60–72 h after transfection. At this time after transfection, eGFP RNA expressed from retrotransposed L1-eGFP will be insignificant compared to L1-eGFP plasmid-driven expression. eGFP levels were normalized to *GAPDH* or puromycin. L1-eGFP contains a puromycin expression cassette under *PGK* promoter control. Thus, puromycin expression can be used as normalizer for transfection.

iPS cells from two different individuals per species were transfected, and eGFP levels are shown relative to human iPS cells.

**Plasmids.** Human *A3B* cDNA from WT-33 and ADRC-40 iPS cells was amplified using Phusion high-fidelity polymerase (New England BioLabs), and primers are described in Extended Data Table 1. *A3B* cDNA fused to a haemagglutinin (HA) tag was then inserted into KpnI/XbaI-digested pcDNA3.1+ (pcDNA3-A3B) as previously described<sup>35</sup>. Similarly, *PIWIL2* cDNAs were amplified from human and NHP iPS cells as described above and inserted into pEF-BOS-EX using EcoRI/Sall (ref. 36). The plasmid expressing APOBEC3G was previously described<sup>35</sup>.

**RNA library generation and deep sequencing.** PolyA<sup>+</sup> RNA was fragmented and prepared into sequencing libraries using the Illumina TruSeq RNA sample preparation kit and analysed on an Illumina HiSeq 2000 sequencer at the UCSD Biomedical Genomics Laboratory (BIOGEM). cDNA libraries were prepared from four human, two chimpanzee and two bonobo iPS cell lines derived from fibroblasts (two clones each, except for human WT-9 and WT-126), and two human ES cell lines (HUES6 and H1). Libraries were sequenced using paired-end 2×100-bp (base pair) reads at a depth of 15–30 million reads per library (250 ± 25 bp (mean ± s.d.) fragments) (Extended Data Fig. 1b). Paired end reads from all libraries were mapped to both the human (hg19, GRCh37) and chimpanzee (panTro3, CGSC 2.1.3) genomes using STAR (v2.2.0c)<sup>37</sup>. To compare gene expression between human and NHP iPS cells, we first mapped paired end reads from all libraries to both human and chimpanzee genomes and then calculated gene expression read counts relative to human RefSeq transcripts. Owing to the lack of annotation in the chimpanzee genome, human gene models (RefSeq) were used to quantify gene expression. To avoid bias introduced by genome insertions and deletions, only reads mapping to both the human and chimpanzee genomes uniquely were used from each sample when comparing gene expression values (~4% of reads mapped to only one genome per sample). To calculate gene expression, read counts in the exons of RefSeq transcripts where calculated using HOMER<sup>38</sup>. Gene expression clustering was carried out using Gene Cluster 3.0 and visualized with Java Tree View<sup>39,40</sup>. EdgeR was used to identify differentially expressed genes comparing human samples with NHPs, and pairwise between bonobo, chimpanzee and human<sup>41</sup>. Functional enrichment analysis was restricted to differentially expressed coding genes with false discovery rates less than 5% and a fold change greater than twofold. We further restricted genes to contain an average of ten normalized reads across sample groups to remove genes with very low expression. Gene Ontology functional enrichment for biological processes (level 2) was carried out using DAVID<sup>42</sup>. *Homo sapiens* whole genome was set as background.

**Small RNA library generation and deep sequencing.** Small RNA (15–40-nucleotide) libraries were prepared using the Illumina TruSeq Small RNA sample preparation kit and analysed on an Illumina HiSeq 2000 sequencer at the Beijing Genomics Institute. Libraries were sequenced using single-end reads at a depth of 15–25 million reads per library. Adaptor sequences were clipped from the 3' end of each read and then aligned to the human (hg19, GRCh37) genome or to the L1s consensus sequence (Rephase<sup>43</sup>) using Bowtie2 (v4.1.2)<sup>44</sup>. Reads aligning to miRBase-defined microRNA transcripts were quantified using HOMER. Matches to previously identified human piRNAs were restricted to small RNAs with lengths between 26 and 33 nucleotides with 5' ends within 2 nucleotides of previously identified piRNA 5' ends based on piRNABank (<http://pirnabank.ibab.ac.in/>)<sup>45</sup>.

**L1 retrotransposition.** Reporter L1 elements are tagged with a reporter gene (eGFP or firefly luciferase) such that only cells that complete a round of retrotransposition will express the reporter gene<sup>46</sup>. Three L1 reporter plasmids were used in this work. L1-eGFP<sup>18,46</sup> was previously described and was a gift from J. V. Moran. L1-luciferase-tagged plasmids (pYX014 and pYX017)<sup>19</sup> were obtained from W. An. In pYX014, L1 is regulated by its native promoter (5' UTR) and, in pYX017, by the heterologous promoter CAG. pYX014 and pYX017 plasmids contain a *Renilla* luciferase expression cassette that allows for control of transfection efficiency. L1 assays in 293T cells were carried out as previously described<sup>47</sup>. 293T cells were transfected with L1 reporter plasmid together with control plasmid or plasmids expressing *A3B*, *A3G* or *PIWIL2* using polyethylenimine (PEI). L1-luciferase retrotransposition was measured by quantification of luciferase activity using the Dual-Glo luciferase reporter assay (Promega) and normalized to *Renilla* luciferase. L1-luciferase inhibition in the presence of *A3B* or *PIWIL2* was independent of the promoter driving L1 expression (data not shown). Inhibition levels of L1 retrotransposition by *A3B* and *PIWIL2* were comparable between the three L1 reporter plasmids used in this study.

Plasmid transfections of iPS cells were performed by electroporation of L1-eGFP plasmid following the manufacturer's instructions (Lonza/Amaxa Nucleofactor, Kit V). The cells were then cultured under normal conditions for 10 days and the percentage of retrotransposition was measured by FACS of eGFP-positive cells. Electroporation efficiency of the L1-eGFP plasmid in human and NHP iPS cell was controlled by transfecting a cassette expressing eGFP and analysed by FACS after

48 h. Human and NHP iPS cell lines had similar transfection efficiency rates. To test the effect of A3B and PIWIL2 overexpression on L1 activity in NHP iPS cells, human *A3B* and human, chimpanzee or bonobo *PIWIL2* cDNAs were electroporated. All experiments were performed at least three times independently. L1 mobility assays are shown as relative value compared to control plasmid transfections or human iPS cell line 1 and represented as mean  $\pm$  s.e.m. of at least three independent experiments.

**Identification and cloning of a retrotransposition-competent chimpanzee L1.** To clone an intact L1 and generate a chimpanzee L1-eGFP reporter plasmid, we followed a modification of the strategy previously described<sup>48</sup>. Intact L1Pt elements were identified in the chimpanzee genome (CSAC 2.1.4/panTro4, UCSC) through Blat and L1Xplorer analyses<sup>49</sup>. Among the identified intact full-length L1Pt elements, we amplified the L1 element located in chromosome 7:11771100–11777132 of the chimpanzee genome from 0.2 ng of genomic DNA extracted from chimpanzee iPS cell 1. Primers were designed to match unique sequences flanking 5' and 3' of the L1Pt and PCR reactions were performed using Phusion High-Fidelity polymerase (NEB). PCR product was sequenced to confirm intactness (Extended Data Fig. 5). A second PCR was performed using the first PCR product as template to introduce a NotI site upstream of the 5' end of L1Pt. The second PCR product was digested with NotI/BstZ17I (New England Biolabs) and inserted into NotI/BstZ17I-digested pL1-eGFP replacing the human-L1 element to generate L1IN71 using the Rapid Ligation kit (Roche). L1IN71 contains a full-length L1Pt element tagged with the eGFP retrotransposition reporter cassette. Primers used for cloning L1IN71 are show in Extended Data Table 1.

**L1 promoter activity.** Human and chimp L1 promoters (L1 5' UTR) were amplified by PCR from L1-eGFP and L1IN71 plasmids, and inserted into XbaI/HindIII digested pGL4.10 (Promega) upstream firefly luciferase cDNA (L1 5' UTR plasmids). To quantify L1 promoter activity, L1 5' UTR plasmids were co-transfected into human and NHP iPS cell lines with a plasmid expressing *Renilla* luciferase. Seventy-two hours after transfection, luciferase activity was quantified and firefly luciferase signal was normalized to the *Renilla* luciferase signal. Results are shown as relative to human L1 5' UTR activity in human iPS cells. Two iPS cell lines from different individuals (iPS cell 1 and 2) per species were transfected. Primers used for cloning L1 5' UTRs are show in Extended Data Table 1.

**Quantification of reference genome-encoded L1 insertions.** Quantification of L1-derived genomic sequences was based on Repbase defined elements annotated by RepeatMasker (<http://www.repeatmasker.org>). L1 genomic positions for human (hg19, GRCh37) and chimpanzee (panTro3, CGSC 2.1.3) genomes were downloaded from the UCSC Genome Browser annotation database<sup>50</sup>. Owing to the large number of unfinished gaps in the chimpanzee genome assembly greater than 2 kilobases (kb) in size, only truncated L1 elements between 100 bp and 1 kb in length were considered in this analysis. Most of these represented the 3' end of L1 elements. L1 elements were separated based on their annotation as L1HS, L1Pt, L1PA2, L1PA3 or L1PA4 and were plotted as a histogram relative to their divergence values, which indicates the fractions of nucleotides that are mutated relative to the consensus element for each family. To estimate the variability of L1 coverage across the genome, each genome was fragmented into 1 megabase (Mb) sections and then was randomly sampled in ten separate groups to calculate the standard deviation in number of L1 elements across different regions of the genome. A strong concentration of L1 elements in a few specific regions of the genome would result in a very high variance between groups, whereas uniform insertion across the genome would result in a low variance. This standard deviation between each sampling was reported as a function of divergence for each class of L1 elements.

To identify reference L1 elements that were inserted into the genome after the last common ancestor for human and chimpanzee, L1 elements were mapped between homologous regions of each genome using the UCSC LiftOver tool. If an element failed to map between genomes, the 100 bp regions immediately upstream and downstream of the L1 element were also mapped between genomes using the LiftOver tool. If the upstream and downstream regions both mapped to the other genome, then the L1 element was mostly likely a result of a recent insertion. If only one or neither of the upstream and downstream regions mapped between genomes, the region was more likely to be the result of a genomic duplication or deletion and was discarded from the analysis. Error bars (s.d.) represent the differences in L1 density based on the sampling of different genomic regions and represents the variability of L1 coverage across the genomes.

**Immunocytochemistry.** Cells were fixed in 4% paraformaldehyde and then permeabilized with 0.5% Triton X-100 in PBS. Cells were then blocked in 5% donkey serum for 1 h before incubation with primary antibody overnight at 4 °C. After three washes with PBS, cells were incubated with secondary antibodies conjugated to fluorophors (Jackson Immuno Research) for 1 h at room temperature. Fluorescence was detected using a Zeiss inverted microscope.

**Immunoblotting.** Immunoblotting was performed as previously described<sup>35</sup>. Cell pellets were lysed in lysis buffer supplemented with Complete protease inhibitor

cocktail (Roche) for 30 min on ice<sup>35</sup>. Protein concentrations from whole cell lysates were quantified by BCA assay (Bio-Rad). Proteins were separated in 4–12% Acrylamide Bis-Tris NuPage gels in MOPS buffer (Invitrogen) and transferred onto Hybond nitrocellulose membranes (Amersham Biosciences).

**Antibodies.** Primary antibodies used in this study were: Tra-1-81 (1:100, Millipore, MAB4381), Nanog (1:500, R&D Systems, AF1997), APOBEC3B (D-15) (1:500, Santa Cruz, sc-86289), PIWIL2 (1:1,000, R&D Systems, AF6558), GFP (1:200, Molecular Probes-Invitrogen, A-6455), Flag (1:1,000 Sigma, F7425) and HA (1:1,000 Covance, MMS-101R). All secondary antibodies were purchased from Jackson ImmunoResearch.

**PIWIL2 RNPs immunoprecipitation and end labelling.** Tetracycline-inducible human iPS cells expressing flag-tagged PIWIL2 were generated by transduction with lentiviruses (Lv)<sup>51</sup>. Cells were first transduced with an Lv-expressing tetracycline transactivator rTAA (LvXETO). After 10 days of culture in growth media with neomycin (neo), neo-resistant colonies were then transduced with a lentivirus expressing Flag-PIWIL2 under the control of a tetracycline-inducible promoter (LvXTP-FlagPIWIL2) and selected for resistance to puromycin. For PIWIL2 RNP immunoprecipitation,  $\sim 3 \times 10^7$  human iPS cells were treated with doxycycline for 72 h, and pelleted cells were resuspended in 1 ml lysis buffer 1 (20 mM Tris-HCl, pH 7.4, 150 mM NaCl, 1 mM MgCl<sub>2</sub>, 0.5% NP40, 1% glycerol, 1 mM dithiothreitol (DTT), 0.1 U  $\mu$ l<sup>-1</sup> RNase inhibitor (Ambion), Complete EDTA-free protease inhibitor (Roche)). Cell lysates were cleared by centrifugation at 20,000g for 20 min at 4 °C. Cleared lysates were incubated with EZview Red FLAG M2 Affinity Gel (Sigma) for 3 h at 4 °C and washed five times with wash buffer (lysis buffer 1 without glycerol). Co-immunoprecipitated RNAs were extracted with Trizol, followed by precipitation with isopropanol and glycogen (Ambion). Isolated RNA was 5' labelled with [ $\gamma$ -<sup>32</sup>P]ATP using T4 polynucleotide kinase (NEB), resolved on 15% PAGE TBE urea gels along with radiolabelled Decade size marker (Ambion) and visualized in a Typhoon phosphorimager (Amersham Biosciences). Control immunoprecipitations were carried out with lysates from cells without doxycycline induction, from doxycycline-induced eGFP-expressing human iPS cells or with control antibody (anti-HA, Roche, 3F10).

**Data deposition.** RNA-seq and small RNA-seq data have been deposited in the GEO under accession number GSE47626. GenBank accession numbers: KF651164 (*P. paniscus* PIWIL2), KF651165: (*H. sapiens* PIWIL2), KF651166 (*P. troglodytes* PIWIL2), KF651167 (*H. sapiens* APOBEC3B), KF651168 (*P. troglodytes* APOBEC3B), KF651169 (*P. paniscus* APOBEC3B) and KF661301 (L1Pt in chimp-L1 plasmid).

31. Marchetto, M. C. et al. A model for neural development and treatment of Rett syndrome using human induced pluripotent stem cells. *Cell* **143**, 527–539 (2010).
32. Muotri, A. R., Nakashima, K., Toni, N., Sandler, V. M. & Gage, F. H. Development of functional human embryonic stem cell-derived neurons in mouse brain. *Proc. Natl. Acad. Sci. USA* **102**, 18644–18648 (2005).
33. Landry, S., Narvaiza, I., Linfesty, D. C. & Weitzman, M. D. APOBEC3A can activate the DNA damage response and cause cell-cycle arrest. *EMBO Rep.* **12**, 444–450 (2011).
34. Kidd, J. M., Newman, T. L., Tuzun, E., Kaul, R. & Eichler, E. E. Population stratification of a common APOBEC gene deletion polymorphism. *PLoS Genet.* **3**, e6 (2007).
35. Narvaiza, I. et al. Deaminase-independent inhibition of parvoviruses by the APOBEC3A cytidine deaminase. *PLoS Pathog.* **5**, e1000439 (2009).
36. Mizushima, S. & Nagata, S. pEF-BOS, a powerful mammalian expression vector. *Nucleic Acids Res.* **18**, 5322 (1990).
37. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
38. Heinz, S. et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576–589 (2010).
39. de Hoon, M. J., Imoto, S., Nolan, J. & Miyano, S. Open source clustering software. *Bioinformatics* **20**, 1453–1454 (2004).
40. Saldanha, A. J. Java Treeview-extensible visualization of microarray data. *Bioinformatics* **20**, 3246–3248 (2004).
41. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).
42. Dennis, G. Jr et al. DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biol.* **4**, P3 (2003).
43. Jurka, J. et al. Repbase Update, a database of eukaryotic repetitive elements. *Cytogenet. Genome Res.* **110**, 462–467 (2005).
44. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nature Methods* **9**, 357–359 (2012).
45. Sai Lakshmi, S. & Agrawal, S. piRNABank: a web resource on classified and clustered Piwi-interacting RNAs. *Nucleic Acids Res.* **36**, D173–D177 (2008).
46. Moran, J. V. et al. High frequency retrotransposition in cultured mammalian cells. *Cell* **87**, 917–927 (1996).
47. Bulliard, Y. et al. Structure-function analyses point to a polynucleotide-accommodating groove essential for APOBEC3A restriction activities. *J. Virol.* **85**, 1765–1776 (2011).

48. Brouha, B. *et al.* Hot L1s account for the bulk of retrotransposition in the human population. *Proc. Natl Acad. Sci. USA* **100**, 5280–5285 (2003).
49. Penzkofer, T., Dandekar, T. & Zemojtel, T. L1Base: from functional annotation to prediction of active LINE-1 elements. *Nucleic Acids Res.* **33**, D498–D500 (2005).
50. Fujita, P. A. *et al.* The UCSC Genome Browser database: update 2011. *Nucleic Acids Res.* **39**, D876–D882 (2011).
51. Ladewig, J. *et al.* Small molecules enable highly efficient neuronal conversion of human fibroblasts. *Nature Methods* **9**, 575–578 (2012).

**a**

Species	Sex	Name/Source
<i>Homo sapiens</i>	Female	hES HUES6 / (HSCI) Embryonic Stem Cell
<i>Homo sapiens</i>	Male	hES H1 / (WiCell) Embryonic Stem Cell
<i>Homo sapiens</i>	Female	WT-33 (iPS1*) / Fibroblast (Marchetto et. al/Cell 2010)
<i>Homo sapiens</i>	Male	ADRC-40 (iPS2*) / Fibroblast (Marchetto et. al/Cell 2010)
<i>Homo sapiens</i>	Male	GM22159 (WT-9* iPS) / Fibroblast (Coriell Cell Repositories)
<i>Homo sapiens</i>	Male	WT-126 (iPS) / Fibroblast (Marchetto et. al/Cell 2010)
<i>Pan paniscus</i> (Bonobo, pigmy chimp)	Male	PR01086 (iPS1*) / Fibroblast (Coriell Cell Repositories)
<i>Pan paniscus</i> (Bonobo, pigmy chimp)	Male	AG05253 (iPS2*) / Fibroblast (Coriell Cell Repositories)
<i>Pan troglodytes</i> (Chimpanzee)	Male	PR01209 (iPS1*) / Fibroblast (Coriell Cell Repositories)
<i>Pan troglodytes</i> (Chimpanzee)	Female	PR00818 (iPS2*) / Fibroblast (Coriell Cell Repositories)

\* nomenclature used in this study

**b**

RNA-Seq Samples	Mapped Read Totals
Human WT-33 iPS1 A	25430637
Human WT-33 iPS1 B	13839658
Human ADRC-40 iPS2 A	30337397
Human ADRC-40 iPS2 B	22682712
Human WT-126 iPS	28193749
Human GM22159 (WT-9) iPS	25613490
Bonobo PR01086 iPS1 A	13137905
Bonobo PR01086 iPS1 B	15666314
Bonobo AG05253 iPS2 A	11520455
Bonobo AG05253 iPS2 B	62143394
Chimp PR01209 iPS1 A	22705535
Chimp PR01209 iPS1 B	10917561
Chimp PR00818 iPS2 A	24052999
Chimp PR00818 iPS2 B	28197740
Human ES HUES6	26114155
Human ES H1	29868106
Human NPC (derived from HUES6)	26814443

**c**

Gene Ontology Term	Count	PValue	Benjamini
0007154:cell communication	91	4.72 x10 <sup>-8</sup>	9.39 x10 <sup>-6</sup>
0007610:behavior	60	3.61 x10 <sup>-7</sup>	3.59 x10 <sup>-5</sup>
0007155:cell adhesion	80	3.81 x10 <sup>-7</sup>	2.52 x10 <sup>-5</sup>
0042221:response to chemical stimulus	124	1.73 x10 <sup>-6</sup>	8.61 x10 <sup>-4</sup>
0051239:regulation of multicellular organismal process	94	9.70 x10 <sup>-6</sup>	3.86 x10 <sup>-4</sup>
0009605:response to external stimulus	89	5.34 x10 <sup>-5</sup>	1.77 x10 <sup>-3</sup>
0007275:multicellular organismal development	230	5.70 x10 <sup>-5</sup>	1.62 x10 <sup>-3</sup>
0032879:regulation of localization	64	8.96 x10 <sup>-5</sup>	2.23 x10 <sup>-3</sup>
0048856:anatomical structure development	202	2.55 x10 <sup>-4</sup>	4.60 x10 <sup>-3</sup>
0065008:regulation of biological quality	126	3.67 x10 <sup>-4</sup>	6.08 x10 <sup>-3</sup>

**d**

Gene Ontology Term	Count	PValue	Benjamini
0006323:DNA packaging	13	5.89 x10 <sup>-3</sup>	6.76 x10 <sup>-1</sup>
0007586:digestion	11	7.14 x10 <sup>-3</sup>	4.96 x10 <sup>-1</sup>
0008037:cell recognition	8	1.06 x10 <sup>-2</sup>	4.94 x10 <sup>-1</sup>
0022414:reproductive process	48	1.44 x10 <sup>-2</sup>	5.00 x10 <sup>-1</sup>
0006950:response to stress	94	1.58 x10 <sup>-2</sup>	4.56 x10 <sup>-1</sup>
0009605:response to external stimulus	54	2.91 x10 <sup>-2</sup>	6.10 x10 <sup>-1</sup>
0051093:negative regulation of developmental process	19	4.14 x10 <sup>-2</sup>	6.84 x10 <sup>-1</sup>
0009791:post-embryonic development	8	4.34 x10 <sup>-2</sup>	6.53 x10 <sup>-1</sup>
0042445:hormone metabolic process	10	4.67 x10 <sup>-2</sup>	6.37 x10 <sup>-1</sup>
0050878:regulation of body fluid levels	12	5.02 x10 <sup>-2</sup>	6.26 x10 <sup>-1</sup>

**Extended Data Figure 1 | Cell lines used, number of mapped reads per sample in RNA-seq and gene ontology enrichment analysis for differentially expressed genes.** **a**, Origin of iPS cells used or generated in this study. **b**, Total number of mapped reads per sample in RNA-seq. **c, d**, Gene ontology (GO) enrichment analysis of differentially expressed genes. **c**, Top 10 enriched GO terms for genes with higher expression in human versus NHP iPS

cells. **d**, Top 10 enriched GO terms for genes highly expressed in NHP versus human iPS cells. GO analysis was restricted to differentially expressed protein-coding genes (FDR < 0.05 and fold change > 2). GO enrichment for biological processes (level 2) was performed using DAVID. Figure shows GO term, number of genes (count), and P values for EASE score and Benjamini adjustment.

## APOBEC3B

**a**

	20	40	60	80	100
Human	MNPQIRNPMERMYRDTFYDNFENEPILYGRSYTWLCYEVKIKRGRSNLLWDTGVFRGQVYFEFOYHAEMCFLSWFCGNQLPAYKCFQITWFVSWTPCPDC				
Chimp	MNPQIRNPMEMWYQRTFYNNFENEPILYGRSYTWLCYEVKIRRGSNLLWDTGVFRGQMYSPPEHHAEMCFLSWFCGNQLSAYKCFQITWFVSWTPCPDC				
Bonobo	MNPQIRNPMEMWYQRTFYNNFENEPILYGRSYTWLCYEVKIRRGSNLLWDTGVFRGQMYSPPEHHAEMYFLSWFCGNQLSAYKCFQITWFVSWTPCPDC				

	120	140	160	180	200
Human	VAKLAEFLSEHPNVTLTISAARLYYYWERDYRRALCRLSQAGARVKIMDYEFFAYCWENFVYNEGQQFMPWYKFEDENYAFLHRTLKEILRYLMDPDTFTF				
Chimp	VAKLAEFLAEPNVTLTISAARLYYYWERDYRRALCRLSQAGARVKIMDDEEFAYCWENFVYNEGQPFMPWYKFDDNYAFLHRTLKEIIRHLMDPDTFTF				
Bonobo	VAKLAEFLAEPNVTLTISAARLYYYWERDYRRALCRLSQAGARVKIMDYEFFAYCWENFVYNEGQPFMPWYKFDDNYAFLHRTLKEIIRHLMDPDTFTF				

	220	240	260	280	300
Human	NFNNDPLVLRRRQTYLCYEVERLDNGTWLMDQHMGFLNEAKNLLCGFYGRHAEELRFLLDVLPSLQDPAQIYRVTFWISWSPCFSWGCAEVRAFLQEN				
Chimp	NFNNDPLVLRRRQTYLCYEVERLDNGTWLMDQHMGFLNEAKNLLCGFYGRHAEELRFLLDVLPSLQDPAQIYRVTFWISWSPCFSWGCAEVRAFLQEN				
Bonobo	NFNNDPLVLRRRQTYLCYEVERLDNGTWLMDQHMGFLNEAKNLLCGFYGRHAEELRFLLDVLPSLQDPAQIYRVTFWISWSPCFSWGCAEVRAFLQEN				

	320	340	360	380
Human	THVRLRIFAARIYDYPFLYKALQMLRDAGAQVSIMTYDEFYCWDTFVYRQGCPFPWDGLEEHHSQALSGRLRAILQNQGN			
Chimp	THVRLRIFAARIYDYPFLYKALQMLRDAGAQVSIMTYDEFYCWDTFVYRQGCPFPWDGLEEHHSQALSGRLRAILQNQGN			
Bonobo	THVRLRIFAARIYDYPFLYKALQMLRDAGAQVSIMTYDEFYCWDTFVYRQGCPFPWDGLEEHHSQALSGRLRAILQNQGN			

**b**

	20	40	60	80	100
Human	MDPFRPSFRGQSPIHPSQCQAVRMPGCPQASKPLDPAAGRGAQGRGHVFGRPEEPSTQRGPQRESVGVLVSMFRGLGIETVSKTPLKREMLPSGRGIL				
Chimp	MDPFRPSFRGQSPIHPSQCQAVRMPGCPQASKPLDPAAGRGAQGRGHVFGRPEEPSTQRGPQRESVGVLVSMFQGLGIETVSKTPLKREMLPSGRGIL				
Bonobo	MDPFRPSFRGQSPIHPSQCQAVRMPGCPQASKPLDPAAGRGAQGRGHVFGRPEEPSTQRGPQRESVGVLVSMFQGLGIETVSKTPLKREMLPSGRGIL				

	120	140	160	180	200
Human	GRGLSANLVRKDREELSPTFWDPKVLAAGDSKMAETSGVWSRTLGRGSSDAEPLGRGAPAGRGAQGRGHVFGRPEEPSTQRGPQRESVGVLVSMFRGLGIETVSKTPLKREMLPSGRGIL				
Chimp	GRGLSANLVRKDREELSPTFWDPKVLAAGDSKMAASVGWSRTLGRGSSDAEPLGRGAPAGRGAQGRGHVFGRPEEPSTQRGPQRESVGVLVSMFQGLGIETVSKTPLKREMLPSGRGIL				
Bonobo	GRGLSANLVRKDREELSPTFWDPKVLAAGDSKMAASVGWSRTLGRGSSDAEPLGRGAPAGRGAQGRGHVFGRPEEPSTQRGPQRESVGVLVSMFQGLGIETVSKTPLKREMLPSGRGIL				

	220	240	260	280	300
Human	PLVLTVEHKEKELIVKQGSKGTTPQSLGLNLVKIQCHNEAVYQYHVTFSPNVECKSMRFGMLKDHQAVTGNVTAFDGSIYLPLVKLQQVLELKQSRKTDSA				
Chimp	PLVLTVEHKEKELIVKQGSKGTTPQSLGLNLVKIQCHNEAVYQYHVTFSPNVECKSMRFGMLKDHQAVTGNVTAFDGSIYLPLVKLQQVLELKQSRKTDSA				
Bonobo	PLVLTVEHKEKELIVKQGSKGTTPQSLGLNLVKIQCHNEAVYQYHVTFSPNVECKSMRFGMLKDHQAVTGNVTAFDGSIYLPLVKLQQVLELKQSRKTDSA				

	320	340	360	380	400
Human	EISIKIQMTKILEPCSDLCIPFVNFRVMKLDMKLVGRNFYDPTSAMVLLQOHLRQIWPYGAASIRRTDGGFLLADVSHKVIRNDCVLDDVMHAIYQQ				
Chimp	EISIKIQMTKILEPCSDLCIPFVNFRVMKLDMKLVGRNFYDPTSAMVLLQOHLRQIWPYGAASIRRTDGGFLLADVSHKVIRNDCVLDDVMHAIYQQ				
Bonobo	EISIKIQMTKILEPCSDLCIPFVNFRVMKLDMKLVGRNFYDPTSAMVLLQOHLRQIWPYGAASIRRTDGGFLLADVSHKVIRNDCVLDDVMHAIYQQ				

	420	440	460	480	500
Human	NKEHFQDECCKLLVGNIVITRYNNTYRIDDVWNKTPKDSFTMSDGKEITFLEYYSKNYGITVKEEDQPLLIHRPQSERQDNHGMLLKGEILLPELSFM				
Chimp	NKEHFQEECCTKLGVNIVITRYNNTYRIDDVWNKTPKDSFTMSDGKEITFLEYYSKNYGITVKEEDQPLLIHRPQSERQNNHGMLLKGEILLPELSFM				
Bonobo	NKEHFQEECCTKLGVNIVITRYNNTYRIDDVWNKTPKDSFTMSDGKEITFLEYYSKNYGITVKEEDQPLLIHRPQSERQNNHGMLLKGEILLPELSFM				

	520	540	560	580	600
Human	TGIPERMKKKDFRAMKDLAQIQLSPKQHHSALECLLQRIAKNEEAATNELMRWGLRLQKDVKHIEGRVLPMERINLNKTSFITSQELNWVKEVTRDPSILT				
Chimp	TGIPERMKKKDFRAMKDLAQIQLSPKQHHSALECLLQRIAKNEEAATNELMRWGLRLQKDVKHIEGRVLPMERINLNKTSFITSQELNWVKEVTRDPSILT				
Bonobo	TGIPERMKKKDFRAMKDLAQIQLSPKQHHSALECLLQRIAKNEEAATNELMRWGLRLQKDVKHIEGRVLPMERINLNKTSFITSQELNWVKEVTRDPSILT				

	620	640	660	680	700
Human	IPMHFWALFYPKRAMDQARELVNMLEKIAGPIGRMSPPAWVELKDDRIETYVRTIQLSTLGAEKGIKMVCVCIIMGPRDDLYGAIKKLCCVQSPVPSQVNN				
Chimp	IPMHFWALFYPKRAMDQARELVNMLEKIAGPIGRMSPPAWVELKDDRIETYVRTIQLSTLGAEKGIKMVCVCIIMAPRDDLYGAIKKLCCVQSPVPSQVNN				
Bonobo	IPMHFWALFYPKRAMDQARELVNMLEKIAGPIGRMSPPAWVELKDDRIETYVRTIQLSTLGAEKGIKMVCVCIIMAPRDDLYGAIKKLCCVQSPVPSQVNN				

	720	740	760	780	800
Human	VRTIGQPTRLRSVAQKILLQINCCKLGGELWGVDFPLKQLMVIAMDVYHDPSRGMRSSVGFVASINLTLTKWYSRUVFQMPHQEIVDSLKLCLVGSLLKKFY				
Chimp	VRTIGQPTRLRSVAQKILLQINCCKLGGELWGVDFPLKQLMVIAMDVYHDPSRGMRSSVGFVASINLTLTKWYSRUVFQMPHQEIVDSLKLCLVGSLLKKFY				
Bonobo	VRTIGQPTRLRSVAQKILLQINCCKLGGELWGVDFPLKQLMVIAMDVYHDPSRGMRSSVGFVASINLTLTKWYSRUVFQMPHQEIVDSLKLCLVGSLLKKFY				

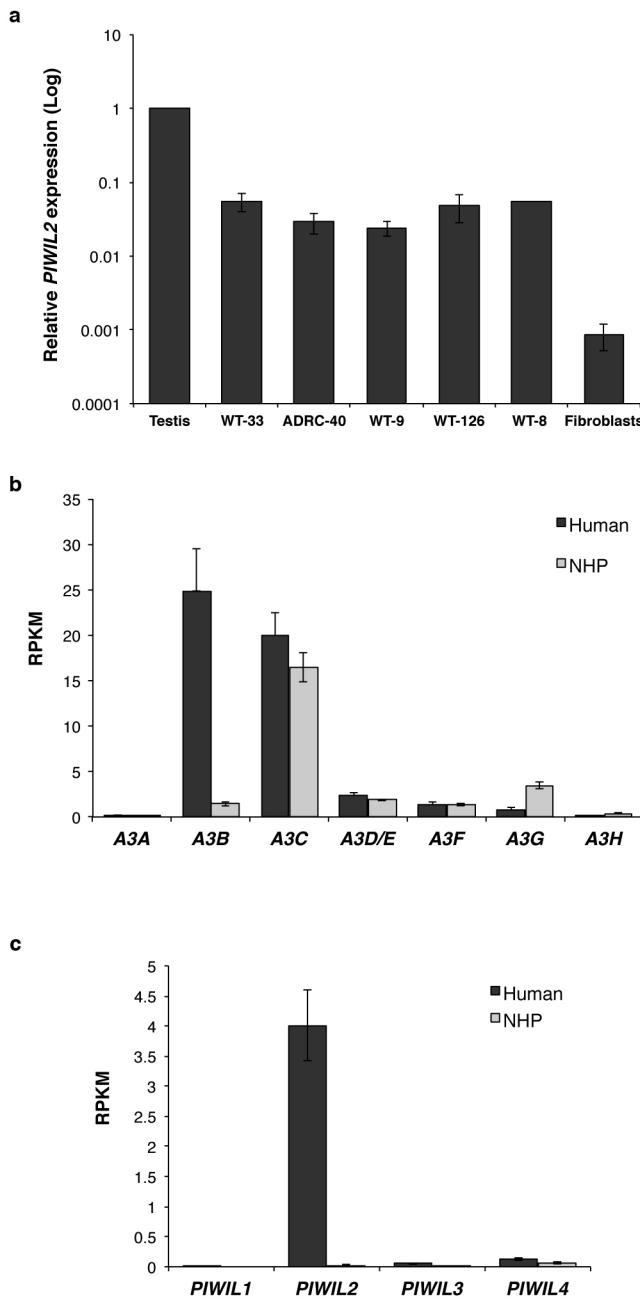
	820	840	860	880	900
Human	EVNHCLPEKIVVYRDGVSDGQLKTVANEYIPLQLQKCFEAFAENYQPKMVVFVVKKKISTNLYLAAPQNFVTPTPGTUVVDHTITSCEWVDFYLLAHHVRQGC				
Chimp	EVNHCLPEKIVVYRDGVSDGQLKTVANEYIPLQLQKCFEAFAENYQPKMVVFVVKKKISTNLYLAAPQNFVTPTPGTUVVDHTITSCEWVDFYLLAHHVRQGC				
Bonobo	EVNHCLPEKIVVYRDGVSDGQLKTVANEYIPLQLQKCFEAFAENYQPKMVVFVVKKKISTNLYLAAPQNFVTPTPGTUVVDHTITSCEWVDFYLLAHHVRQGC				

	920	940	960	
Human	GIPTHYVCVLNTANLSPDHMQRLTFKLCMHWYNNWPGTIRVPAPCKYAHKLAFLSGHILHHEPAIQLCENLFFL			
Chimp	GIPTHYVCVLNTANLSPDHMQRLTFKLCMHWYNNWPGTIRVPAPCKYAHKLAFLSGHILHHEPAIQLCENLFFL			
Bonobo	GIPTHYVCVLNTANLSPDHMQRLTFKLCMHWYNNWPGTIRVPAPCKYAHKLAFLSGHILHHEPAIQLCENLFFL			

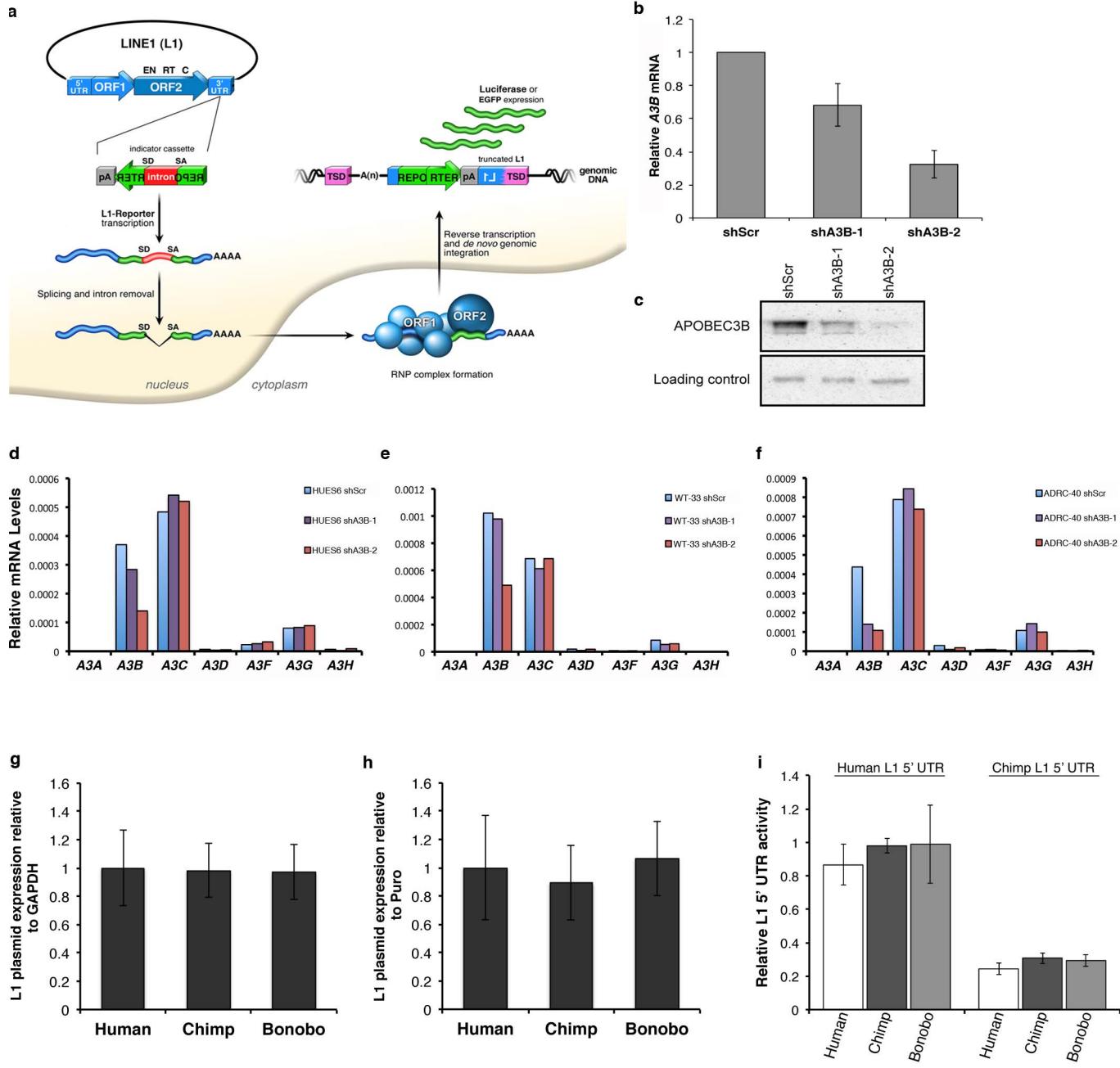
## Extended Data Figure 2 | Amino acid alignment of A3B and PIWI2.

**a, b**, Protein sequences of human, chimp and bonobo A3B (a) or PIWI2 (b) were aligned using ClustalW. **a**, Alignment of A3B showing >93% identity

between human and NHP proteins. **b**, Alignment of PIWI2 showing >98% identity between human and NHP proteins.



**Extended Data Figure 3 | mRNA levels of APOBEC3 and PIWI-like protein family members in iPS cells.** **a**, Comparative analysis of *PIWIL2* mRNA levels. qRT-PCR analysis of *PIWIL2* mRNA levels in human testis, human iPS cell lines, and available fibroblasts from which the iPS cell lines were derived. mRNA levels were normalized to *GAPDH* and shown relative to human testis (mean  $\pm$  s.e.m.;  $n = 3$  biological replicates). Compared to testis, *PIWIL2* levels are 20–40 fold lower in iPS cells and  $\sim$ 1,100-fold lower in fibroblasts. **b, c**, Quantification of mRNA levels of APOBEC3 and PIWI-like family members in human and NHP iPS cells by RNA-seq. Increased mRNA levels in human iPS cells are restricted for *APOBEC3B* and *PIWIL2*. *y* axes in **b** and **c** denote the reads per kilobase per million mapped reads (RPKM).



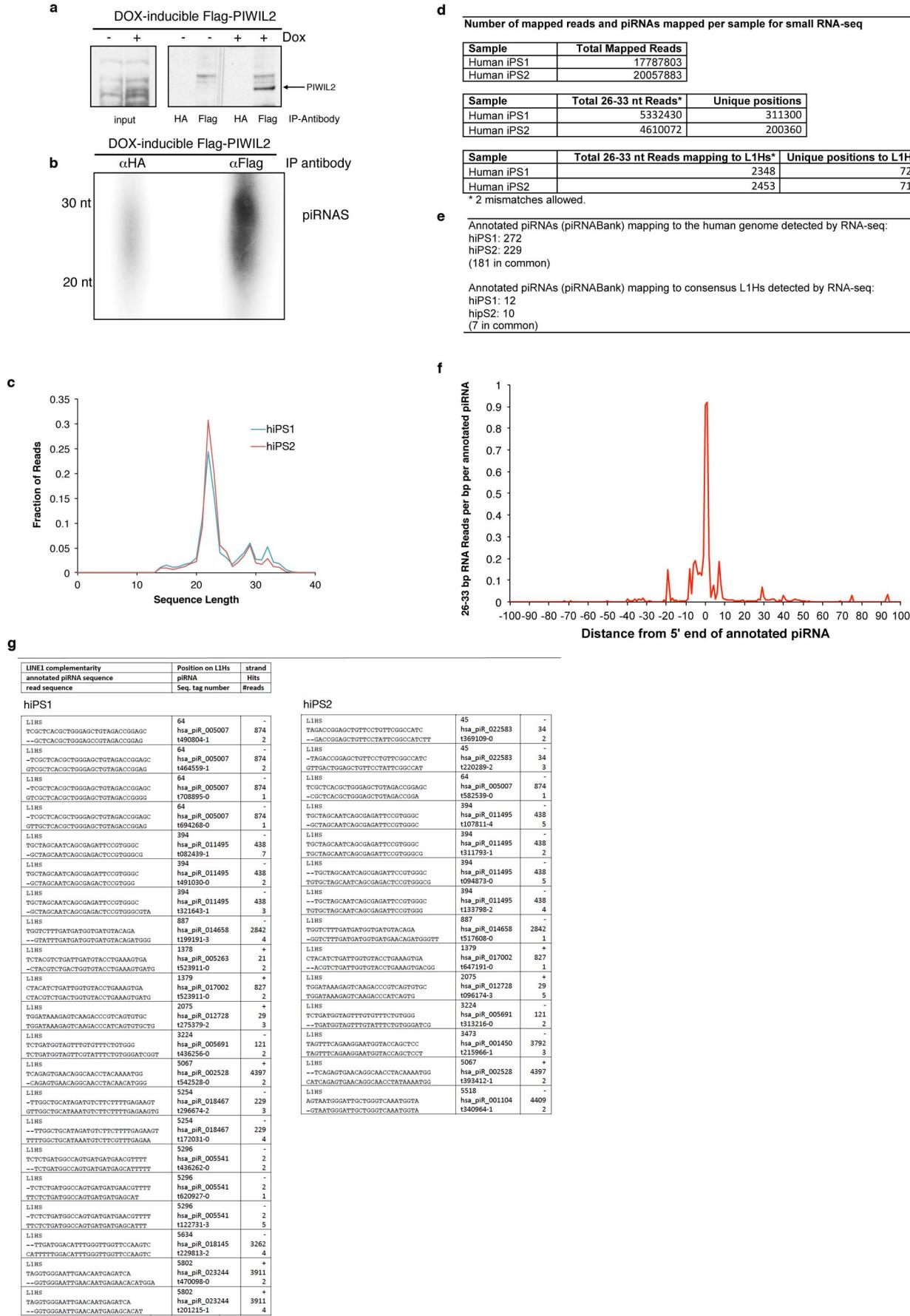
**Extended Data Figure 4 | L1 reporter activity in iPS cells.** **a**, L1 retrotransposition reporter system. The L1-reporter plasmid contains a retrotransposition-competent human L1 element and carries either an eGFP or a luciferase reporter construct in its 3' UTR region. The reporter gene is interrupted by an intron in the same transcriptional orientation as the L1 transcript. This arrangement ensures that eGFP/luciferase-positive cells will arise only when a transcript initiated from the promoter driving L1 expression is spliced, reverse transcribed, and integrated into chromosomal DNA, thereby allowing expression of the reporter gene from a heterologous promoter. **b–f**, Efficient A3B knockdown in human ES and iPS cells. **b**, Stable shRNA-mediated knockdown of A3B in human ES cells (HUES6) using lentivirus expressing different shRNAs against A3B (shA3B-1 and shA3B-2) or scrambled control (shScr). Levels of A3B expression were normalized to GAPDH and shown relative to shScr (mean  $\pm$  s.e.m.;  $n = 3$  biological replicates). **c**, Western blot confirming stable A3B knockdown in human ES cells. **d–f**, shRNA-mediated knockdown in human ES cells (HUES6) and iPS cell lines 1 and 2 (WT-33 and ADRC-40, respectively) was specific for A3B.

**g–h**, qRT–PCR analysis of plasmid expression in iPS cell lines transfected with L1-eGFP plasmid. Total RNA samples were obtained 60–72 h after transfection. L1 plasmid expression was normalized to GAPDH (g) or puromycin (h). L1-eGFP contains a puromycin expression cassette under PGK promoter control. Thus, puromycin expression was used as normalizer for transfection. iPS cells from two different individuals per species were transfected, and eGFP levels are shown as relative to human iPS cells. No significant differences were observed for L1 plasmid expression between human and NHP iPS cell lines (mean  $\pm$  s.e.m.;  $n = 3$  biological replicates).

**i**, Relative L1 5' UTR promoter activity. Human and chimp L1 promoters (L1 5' UTR) controlling firefly luciferase were transfected into human and NHP iPS cell lines. *Renilla* luciferase was co-transfected as control. Luciferase activity was quantified as firefly luciferase units relative to *Renilla* luciferase units. Results are shown as normalized to human L1 5' UTR activity in human iPS cell. iPS cells from two different individuals per species were transfected. No significant differences were observed for L1 promoter activities between human and NHP iPS cell lines (mean  $\pm$  s.e.m.;  $n = 4$  biological replicates).

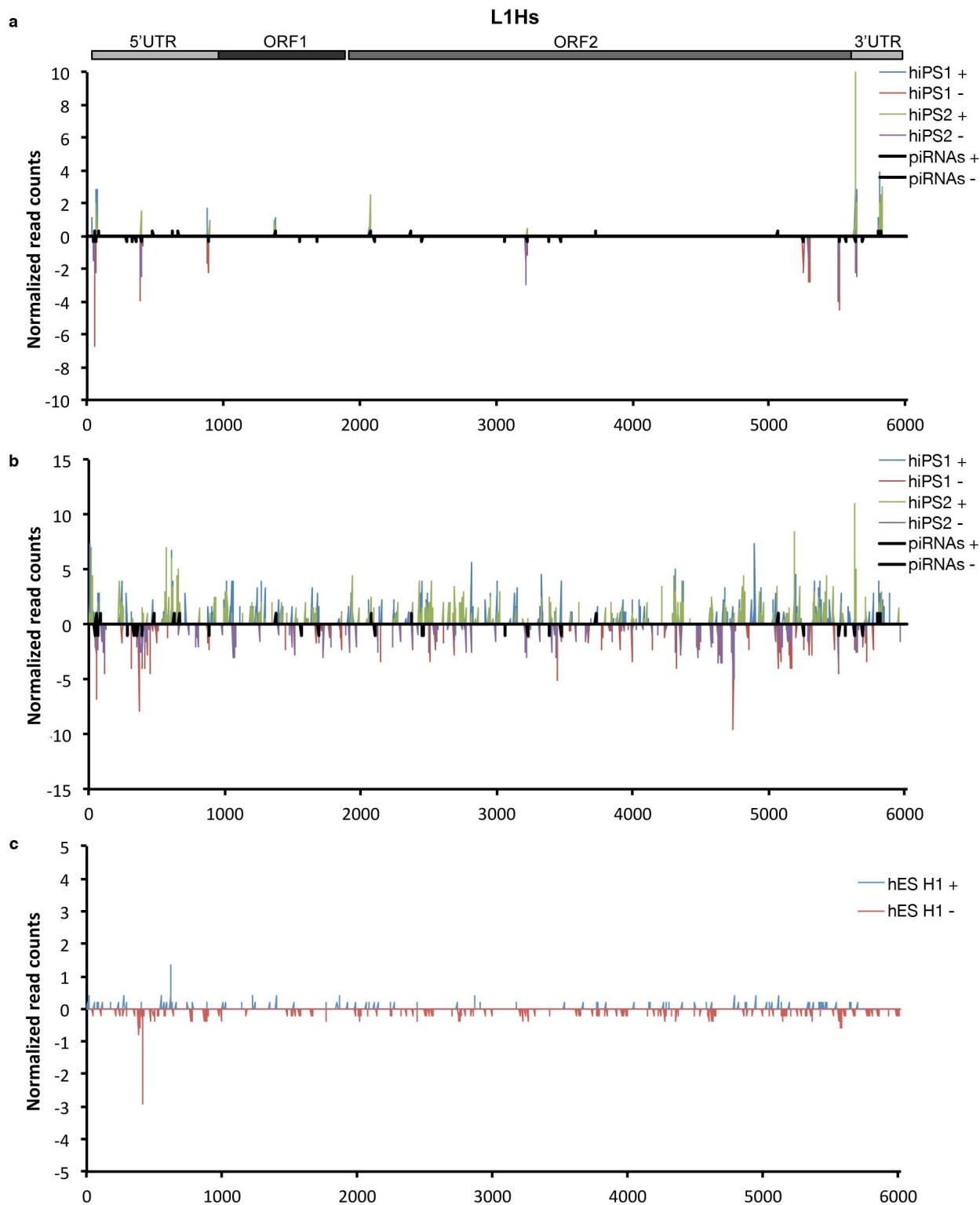
**Extended Data Figure 5 | Nucleic acid alignment of human and chimpanzee L1 elements.** Sequence of the chimpanzee L1P element cloned and used to

generate the chimpanzee L1-eGFP tagged reporter plasmid (L1IN71) (*top sequence*). *LRE2*-human L1 (*bottom sequence*)



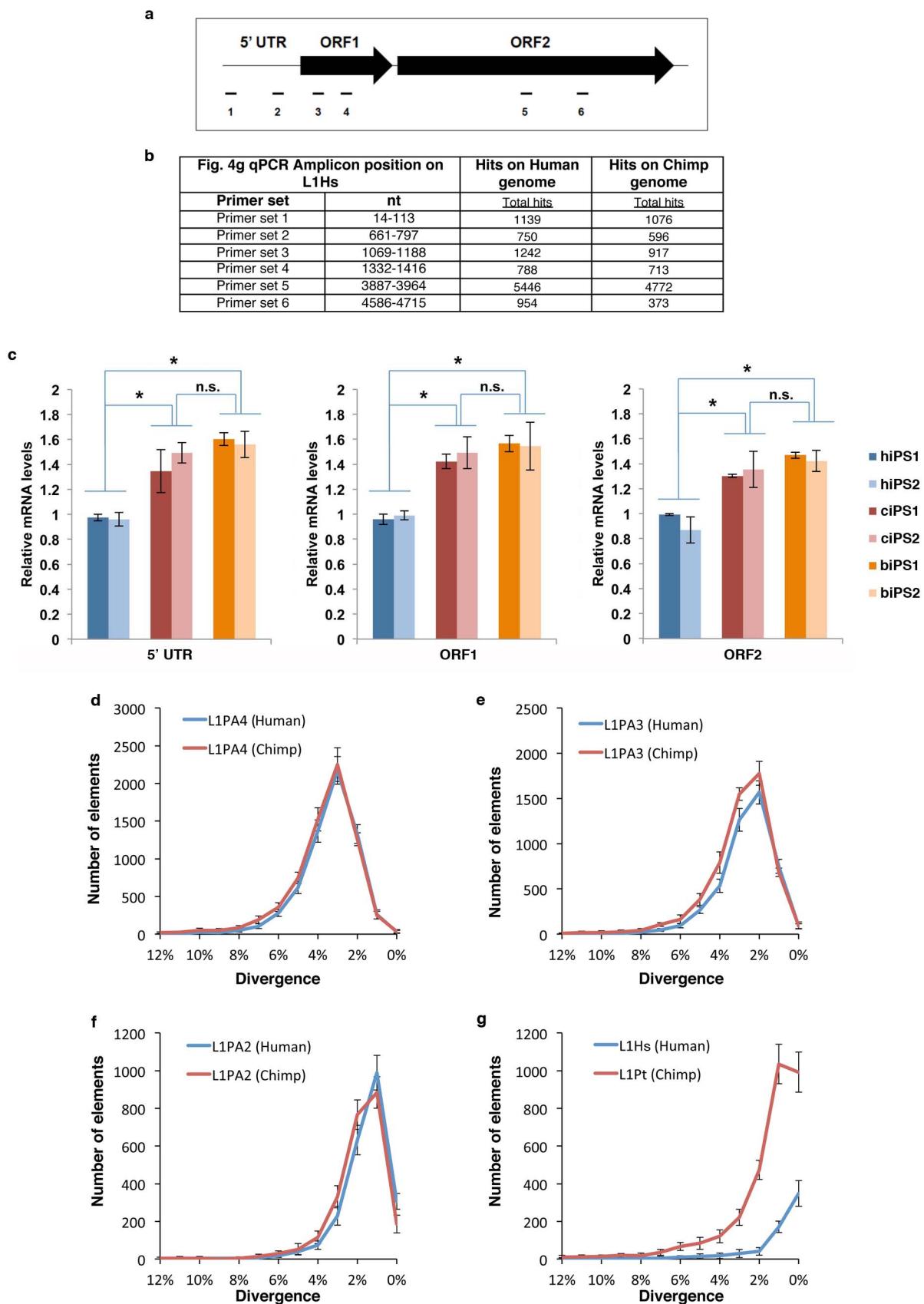
**Extended Data Figure 6 | Immunoprecipitation of piRNAs associated with PIWIL2 in human iPS cells and annotated piRNAs mapping to consensus L1Hs in iPS cells.** **a**, Immunoprecipitation of PIWIL2 RNPs using Flag-tag antibodies from Tet-inducible Flag-tagged PIWIL2 human iPS cells after addition of doxycycline to the culture media. HA-tag antibody was used as control. **b**, [ $\gamma$ - $^{32}$ P]ATP end-labelling of RNAs associated with Flag-PIWIL2 RNPs. Signal in the piRNAs size range is detected only in anti-Flag but not in control antibody anti-HA immunoprecipitates. **c**, Size distribution of RNA reads detected by small RNA-seq from small RNAs samples extracted from

human iPS cell lines. **d**, Number of mapped reads per sample in small RNA-seq. **e**, Number of annotated piRNAs (piRNAbank) detected by RNA-seq in human iPS cells 1 and 2. **f**, Characterization of 5' end of piRNAs detected in human iPS cells relative to annotated piRNAs. Read count distribution relative to piRNA 5' ends (piRNAbank). **g**, Sequences of annotated piRNAs (piRNAbank) mapping to consensus L1Hs detected in human iPS cells 1 and 2. The 26–33-nucleotide RNA reads from human iPS cell lines 1 and 2 characterized by RNA-seq are aligned to annotated piRNAs mapping to the consensus L1Hs sequence. Analysis of mapping sequences was performed allowing two mismatches.



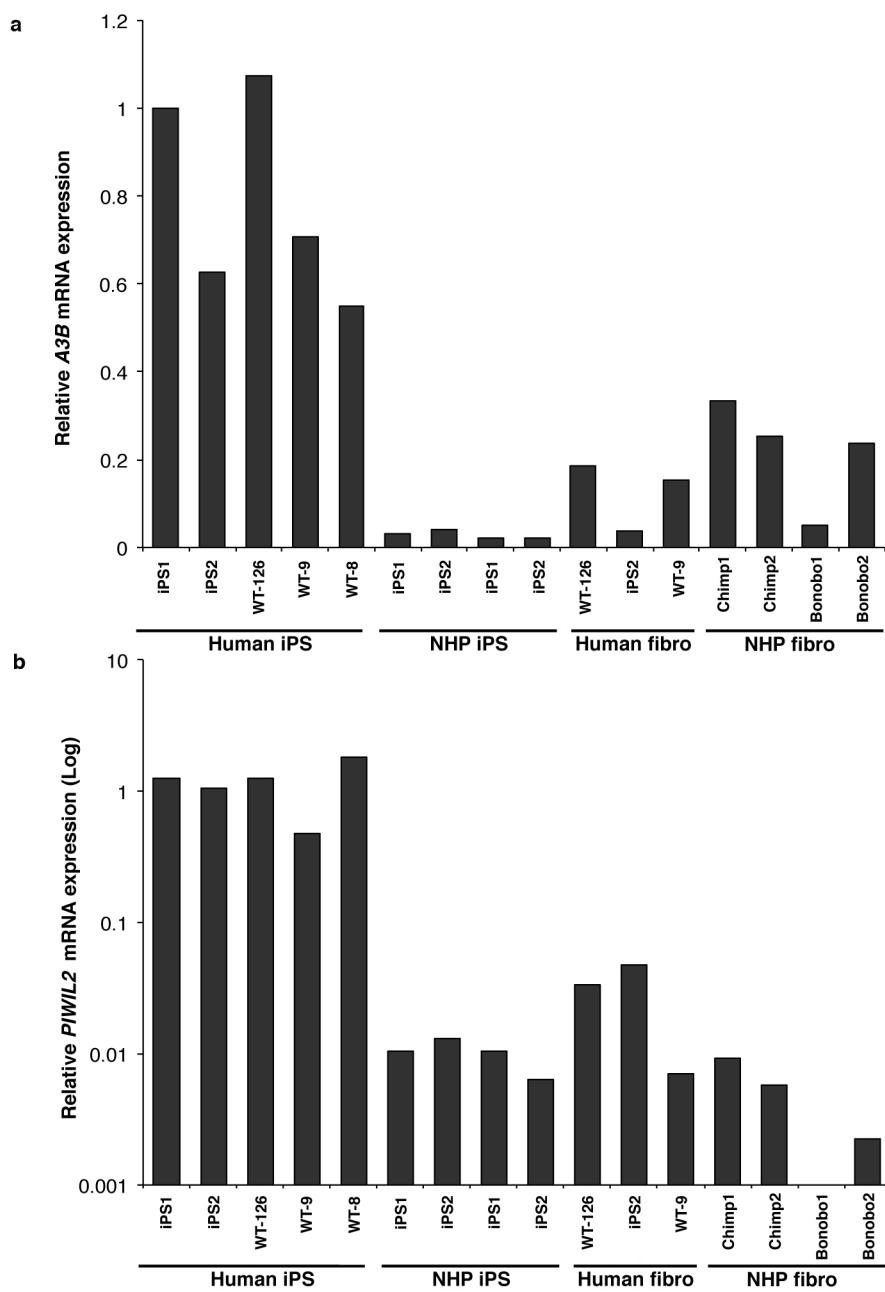
**Extended Data Figure 7 | Mapping of 26–33-nucleotide RNAs in human iPS cells to consensus L1Hs.** **a**, Mapping of annotated piRNAs (piRNAbank) detected by RNA-seq from human iPS cell lines to the consensus sequence for L1Hs (from Repbase). All annotated piRNAs (piRNAbank) complementary to L1Hs are indicated (black bars). **b**, Total 26–33-nucleotide RNA reads

characterized by small RNA-seq mapped to L1Hs. **c**, Similar analysis as in **b** of ENCODE data for small RNAs from H1 cells. Positive and negative values indicate sense (+) and antisense (−) piRNAs, respectively. Schematic representation of the L1Hs element is shown (top). *y* axes represent read counts normalized to  $10^7$  reads per experiment.



**Extended Data Figure 8 | Higher levels of endogenous L1 RNA and recent species-specific L1 elements in chimpanzee.** **a**, Scheme of amplicons mapped to the L1Hs consensus sequence. Six primer pairs (two per region) were used for quantification of 5' UTR, ORF1 and ORF2. The primers were designed to recognize both species-specific and common families. **b**, Positions of the amplicons in L1Hs consensus sequence and the number of *in silico* PCR hits on the human and chimp genomes. **c**, qRT–PCR analysis using primers for different regions of L1 element show higher levels of L1 RNA in NHP iPS cells regardless of the L1 region tested: 5' UTR, ORF1 and ORF2 (mean  $\pm$  s.e.m.;

$n = 3$  biological replicates;  $*P < 0.01$  between indicated groups, *t*-test). **d–g**, Quantification of L1 elements in human and chimpanzee genomes using a population divergence model. Number of L1 elements found in the human and chimpanzee genomes for families: L1PA4 (**d**), L1PA3 (**e**), L1PA2 (**f**) and L1Pt and L1Hs (**g**) plotted as a histogram relative to their divergence (number of mutations relative to the canonical element). The standard deviation describes the differences in L1 density based on the sampling of different genomic regions and represents the variability of L1 coverage across the genomes (see Methods).



**Extended Data Figure 9 | Relative *A3B* and *PIWIL2* mRNA levels in iPS cells and fibroblasts.** Relative expression of *A3B* (a) and *PIWIL2* (b) in human and NHP iPS cell lines, and the available source fibroblasts from which iPS cells

were derived. mRNA levels were normalized to *GAPDH* and shown relative to human iPS cell line 1.

Extended Data Table 1 | List of primers used in this study

Primers	Sequence	Use
Nanog-F	5'- CCTATGCCTGTGATTGTGG -3'	PCR
Nanog-R	5'- CTGGGACCTTGTCTTCCTTT -3'	PCR
AFP-F	5'- AAAAGCCC ACTCCAGC ATC -3'	PCR
AFP-R	5'- CAGACAATCCAGCACATCTC -3'	PCR
Musashi-F	5'- AAAGGAGGTGATGTCGCCAA -3'	PCR
Musashi-R	5'- TGGTCCGTAGGCAGTGAGA -3'	PCR
Brachury-F	5'- GCCCTCTCCCTCCCCCTCCACGCACAG -3'	PCR
Brachury-R	5'- CGGCGGCCGTTGCTCACAGACCACAGG -3'	PCR
B-Actin-F	5'- TGTTTTCTGCGCAAGTTAGGTTT -3'	PCR
B-Actin-R	5'- GCCGACAGGATGCAAGGAGAT -3'	PCR
APOBEC3B (20-40)	5'- GCGGGACAGGGACAAGCGTAT -3'	Cloning
APOBEC3B (1250-1228)	5'- CTGCTCAACCCAGGCTCTGCCT -3'	Cloning
APOBEC3B (19-41)	5'- AGCGGGACAGGGACAAGC GTATC -3'	Cloning
APOBEC3B(1309-1288)	5'- AGCTGGAGATGGTGGTGAACGG -3'	Cloning
L1Pt ch7 11 F	5'- TTGCAGGTACTCTGAGCTTCAC -3'	Cloning
L1Pt ch7 11 R	5'- AAGGAGAACGACCTGCATGA -3'	Cloning
Not-L1 F	5'- ATAAGAATGCGGCCGCGGGGGAGGAGCCAAGATG -3'	Cloning
XhoNotI L1 5'UTR	5'- CCGCTCGAGCGGCCGCGGGGGAGGG -3'	Cloning
L1 5'UTRHindIIIATG	5'- TTTTAAGCTTCCATCTTGTGGTTTATCTAC -3'	Cloning
APOBEC3B-F	5'- CGCCAGACCTACTTGCTAT -3'	qPCR
APOBEC3B-R	5'- CATTGCAAGCGCTCCTTAT -3'	qPCR
GAPDH-F	5'- CATGTTCAATATGATTCCACC -3'	qPCR
GAPDH-R	5'- CTCCACGACGTACTCAGCG -3'	qPCR
PIWIL2-F	5'- TTGTGGACAGCCTGAAGCTA -3'	qPCR
PIWIL2-R	5'- CCATCAGACACTCCATCACG -3'	qPCR
L1 5'UTR set1-R	5'- AAGATGGCGAATAGGAACA -3'	qPCR
L1 5'UTR set1-R	5'- GATGAACCCGGTACCTCAGA -3'	qPCR
L1 5'UTR set2-R	5'- GAGATCTGAGAACGGCAGA -3'	qPCR
L1 5'UTR set2-R	5'- AGCTGCAGGTCTGTTGAAT -3'	qPCR
L1 ORF1 set1-F	5'- GCTACGGGAGGACATTCAA -3'	qPCR
L1 ORF1 set1-R	5'- TTCAGCTCCATCAGCTCCTT -3'	qPCR
L1 ORF1 set2-F	5'- ATGAGCAAAGCCTCCAAGAA -3'	qPCR
L1ORF1 set2-R	5'- TTCTCCCCCATCACTTCTGAGG -3'	qPCR
L1 ORF2 set1-F	5'- TGACAAACCCACAGCCAATA -3'	qPCR
L1 ORF2 set1-R	5'- CCCCTGCTTGTGCCAGTTT -3'	qPCR
L1 ORF2 set2-F	5'- TGGAGGCATCACACTACCT -3'	qPCR
L1 ORF2 set2-R	5'- ATGCGGCATTATTCAGG -3'	qPCR
Actin-F	5'- TACAATGAGCTGCGTGTGG -3'	qPCR
Actin-R	5'- TAGCACAGCCTGGATAGCAA -3'	qPCR
GFP F2	5'- GGGTGTCTGCTGGTAGTGG -3'	qPCR
GFP R2	5'- TATATCATGGCCGACAAGCA -3'	qPCR
PURO F	5'- CTCGACATCGGCAAGGTGTG -3'	qPCR
PURO R	5'- GCCTTCCATCTGTTGCTGCC -3'	qPCR
APOBEC3A	TaqMan Assay (Life Technologies) Hs00377444	qPCR
APOBEC3B	TaqMan Assay (Life Technologies) Hs00358981	qPCR
APOBEC3C	TaqMan Assay (Life Technologies) Hs00828074	qPCR
APOBEC3D	TaqMan Assay (Life Technologies) Hs00537163	qPCR
APOBEC3F	TaqMan Assay (Life Technologies) Hs01665324	qPCR
APOBEC3G	TaqMan Assay (Life Technologies) Hs00222415	qPCR
APOBEC3H	TaqMan Assay (Life Technologies) Hs00962174	qPCR
PIWIL2	TaqMan Assay (Life Technologies) Hs01032720	qPCR
GAPDH	TaqMan Assay (Life Technologies) Hs03929097	qPCR
HPRT	TaqMan Assay (Life Technologies) Hs01003267	qPCR