F e e f e e g CRISPR-Ca 9–ed ed ea a e b

G AH -L , J Z , A M C , N M. E. F , N K , E H H , M G , D W HH , J M. A. T , K K. N[1]

[a]Human Embryo and Stem Cell Laboratory, The Francis Crick Institute, NW1 1AT London, United Kingdom; [b]Sex Chromosome Biology Laboratory, The Francis Crick Institute, NW1 1AT London, United Kingdom; [c]Centre for Stem Cells and Regenerative Medicine, Guy's Campus, King's College London, SE1 9RT London, United Kingdom; [d]Nuffield Department of Women's and Reproductive Health, John Radcliffe Hospital, University of Oxford, OX3 9DU Oxford, United Kingdom; [e]Jesus College, University of Oxford, Odso3tm[(heteriW290.15Oxford)-11.5(,)]T90.18United

when they targeted exons (20%). Consistent with this, Kosicki et al. observed large deletions (up to 6 kb) and other complex genomic lesions at frequencies of 5–20% of their clones after targeting the *PigA* and *Cd9* loci in two mouse embryonic stem cell (mESC) lines and primary mouse cells from the bone marrow, as well as the *PIGA*

overall size to chromosome 6, but the frequency of whole chromosome abnormalities is similar to that observed for chromosome 6, suggesting that genome editing does not exacerbate the rates of whole chromosome errors (*SI Appendix*, Fig. S2C). -41T5.4(e)-h.1(9)-0.7.8(4)-2(hroompariTJ-0.0157480.6-0317139.1255TD[().s)6(o)-15.7(c)4n11.6(,)

the samples that we analyzed, we cannot exclude the possibility that they inherited a homozygous genotype. Therefore, we required the presence of heterozygous SNPs in at least one additional cell from the same embryo to call putative LOH events.

The variant-calling pipeline that we implemented was specifically adjusted for MiSeq data from single cell amplified DNA and includes stringent preprocessing and filtering of the MiSeq reads (*Methods*). To have sufficient depth of coverage and to construct reliable SNP profiles, we only considered samples with ≥5× coverage in at least two-thirds of the amplicons across the *POU5F1* locus (*edsd oningeep6-31dusou696NPsesidg eedsied*

e dchr6No LOHLOH at locusLOH book

the amplicons covering exons 1 and 5 of the *POU5F1* gene (fragments E1-2, G1, and E4 in Fig. 2*A*) and homozygous SNPs in between (50% of control and 2.4% of targeted samples). These putative LOH samples would have had to have a cell isolated from the same embryo that had a detectable SNP(s) anywhere in between these flanking exons (e.g., see samples G_8.03 versus G_8.04 in *SI Appendix*, Fig. S7). Interestingly, this was the most prevalent pattern in Cas9 control samples (Fig. 2*B* and *SI Appendix*, Fig. S7), which may indicate the possibility of technical issues due to sequencing or overamplification of one parental allele (see below). Bookended samples have two heterozygous SNPs flanking the cut site but in fragments outside the *POU5F1* locus (20% of control and 23.8% of targeted samples). These LOH events could represent deletions of lengths between ∼7 kb (G_C12.03, *SI Appendix*, Fig. S10) and ∼12 kb (W_C11.04, *SI Appendix*, Fig. S9). Finally, in open-ended samples (10% of control and 61.9% of targeted samples), it was not possible to find heterozygous SNPs in any of the amplified fragments (G_C12.07, Fig. 2*A*) or there was one or a few heterozygous SNPs on only one side of the region of interest (G_C16.02, *SI Appendix*, Fig. S12). This was the most common pattern in targeted samples (Fig. 2*B* and *SI Appendix*, Figs. S8–S12) and could represent large deletions of ∼20 kb in length (the size of the region explored) or larger.

As mentioned aboved3(s)-331u2Seq(s)-31.3at5(f5)19.2mus4(l6)23.6b3(s5oved)interpret.6(ab-325.wi(e8346-19998Td[As)-256cau-265(a78.6(.iv

Disc ssion

In all, we reveal unexpected on-target complexity following CRISPR-Cas9 genome editing of human embryos. Our data suggest ~16% of samples exhibit segmental losses/gains adjacent to the *POU5F1* locus and LOH events that span 4 kb to at least 20 kb. Chromosome instability, including whole or segmental chromosome gain or loss, is common in human preimplantation embryos (27, 28). However, in contrast to Cas9 control embryos, we noted a significantly higher frequency of CRISPR-Cas9–targeted embryos with a segmental gain or loss that was directly adjacent to the *POU5F1* on-target site. The segmental errors were observed in embryos from distinct genetic backgrounds and donors. Therefore, together with their on-target location, this suggests that the errors may have been an unintended consequence of CRISPR-Cas9 genome editing. This is supported by the higher frequency of

respect to OCT4 function, are interpretable. Moreover, our transcriptome-based digital karyotypes and differential gene expression analysis indicate biallelic transcripts and gene expression upstream and downstream of the *POU5F1*

derived from targeted embryos, suggests that repair from the maternal chromosome by IH-HR results in a stretch of LOH. Of note, due to the selection bias that occurs during ESC derivation and the mosaicism observed following genome editing, it is not possible to draw definitive conclusions about the extent of LOH or its cause in an embryo context, whereby cells with complex mutations may be preferentially excluded from ESC derivation. By contrast, another study by Zuccaro et al., using the same microinjection method, suggests that the LOH observed following CRISPR-Cas9–mediated genome editing is a consequence of whole chromosome or segmental loss adjacent to the on-target site and that microhomology-mediated end-joining (MMEJ) is the dominant repair pathway in this context (37). This corroborates our previous findings in human embryos targeted postfertilization, where we noted a stereotypic pattern to the type of indel mutations and speculated that this was likely due MMEJ (17). Although microhomologies can promote gene conversion by, for example, interchromosomal template switching in a RAD51-dependent manner (38), based on our previous transcriptome analysis, we found that components of the MMEJ pathway (i.e., *POLQ*) are transcribed in early human embryos, while factors essential for HDR (i.e., *RAD51*) are not appreciably expressed. This suggests that MMEJ-derived large deletions (14, 37) are more likely than microhomology-mediated gene conversion in this context, although protein expression has yet to be fully characterized. Consistent with this, a significant fraction of somatic structural variants arises from MMEJ in human cancer (39). Moreover, microhomology-mediated break-induced replication underlies copy number variation in mammalian cells (40) and microhomology/microsatellite-induced replication leads to segmental anomalies in budding yeast (41). The discrepancy between the Liang et al. and Zuccaro et al. studies could be due to locus-dependent differences of CRISPR-Cas9 genome editing fidelity. For example, Przewrocka et al. demonstrate that the proximity of the CRISPR-Cas9–targeted locus to the telomere significantly increases the possibility of inadvertent chromosome arm truncation (16). To fully elucidate the LOH that has occurred at the on-target site in our study, and to resolve the controversy over the IH-HR reported by others (8, 9, 36, 37), will require the development of a pipeline to enrich for the region of interest and then perform deep (long-read) sequencing to evaluate the presence and extent of on-target damage. By bookending SNPs on either side of an LOH event, primers could be designed to incorporate the SNPs and bo87.8(ough)ee-1.1065TD[the)-475..7(haucidabhh1.82730.p)-5l8risesehntthef

from the same DNA sample were pooled to generate 137 libraries that were sequenced by Illumina MiSeq v3. See *SI Appendix* for more details.

SNP Typing. We trimmed the MiSeq paired-end reads with DADA2 (46), corrected substitution errors in the trimmed reads with RACER (47), and mapped the corrected reads to the human genome hg38 with BWA v0.7.17 (44). Subsequently, SAM files were converted to the BAM format and postprocessed using Samtools v1.3.1 (48). SNP calling was performed with BCFtools v1.8 (49) using mpileup and call. SNPs supported by less than 10 reads and with mapping quality below 50 were filtered out. To control for allele overamplification, homozygous SNPs were changed to heterozygous if the fraction of reads supporting the reference allele was at least 6% of the total (21). This threshold corresponds to the median of the distribution of the fraction of reads supporting the reference allele across samples. See *SI Appendix* for more details.

scRNA-Seq Data Analysis. scRNA-seq reads from G&T-seq samples were processed as previously described (17). Samples with a breadth of sequencing below 0.05 were not considered for any downstream analysis (*SI Appendix*, Fig. S13 *A*–*C*). Differential gene expression analysis was carried out with DESeq2 v1.10.1 (50). For digital karyotyping based on gene expression, we adapted the method described in ref. 24 to identify gains or losses of chromosomal arms (z-score-karyotyping). For digital karyotyping based on SNP expression, we applied the eSNP-Karyotyping pipeline with default parameters (26). See *SI Appendix* for more details.

Data and Software Availability. All data supporting the findings of this study are available within the article and its *SI Appendix*. MiSeq and low-pass WGS data have been deposited to the Sequence Read Archive under accession no. PRJNA637030 (51). scRNA-seq data were extracted from the Gene Expression Omnibus using accession no. GSE100118 (52). A detailed analysis pipeline is available at the following site: https://github.com/galanisl/loh_scripts (53).

ACKNOWLEDGMENTS. We thank the generous donors whose contributions have enabled this research. We thank Robin Lovell-Badge, James Haber, Alexander Frankell, Aska Przewrocka, Charles Swanton, Maxime Tarabichi, and the K.K.N. and J.M.A.T. laboratories for discussion, advice, and feedback; the Francis Crick Institute's core facilities including Jerome Nicod and Robert Goldstone at the Advanced Sequencing Facility; D.W. was supported by the National Institute for Health Research Oxford Biomedical Research Centre Programme. N.K. was supported by the University of Oxford Clarendon Fund and Brasenose College Joint Scholarship. Work in the K.K.N. and J.M.A.T. laboratories was supported by the Francis Crick Institute, which receives its core funding from Cancer Research UK Grants FC001120 and FC001193, UK Medical Research Council Grants FC001120 and FC001193, and Wellcome Trust Grants FC001120 and FC001193. Work in the K.K.N. laboratory was also supported by the Rosa Beddington Fund. For the purpose of Open Access, the author has applied a CC BY public copyright licence to any Author Accepted Manuscript version arising from this submission.

1. M. Adli, The CRISPR tool kit for genome editing and beyond. *Nat. Commun.* **9**, 1911 (2018).