

# Haplotype-phased *Callithrix jacchus* embryonic stem cell line for genome editing using CRISPR/Cas9

Bo Zhou<sup>1,2</sup>, Steve S. Ho<sup>1,2</sup>, Louis C. Leung<sup>1,3</sup>, Thomas R. Ward<sup>1,2</sup>, Marcus Ho<sup>1,2</sup>, Melanie J. Plastini<sup>1,2</sup>, Scott C. Vermilyea<sup>4,5</sup>, Marina E. Emborg<sup>4,5,6</sup>, Thaddeus G. Golos<sup>4,7</sup>, Philippe Mourrain<sup>1,3</sup>, Dimitri Perrin<sup>8</sup>, Karen J. Parker<sup>1,\*</sup>, Alexander E. Urban<sup>1,2,9,\*</sup>

<sup>1</sup>Department of Psychiatry and Behavioral Sciences, Stanford University School of Medicine, Stanford, CA 94305, USA

<sup>2</sup>Department of Genetics, Stanford University School of Medicine, Stanford, California 94305, USA

<sup>3</sup>Stanford Center for Sleep Sciences and Medicine, Stanford University School of Medicine, Stanford, CA 94305, USA

<sup>4</sup>Wisconsin National Primate Research Center, University of Wisconsin-Madison, Madison, WI 53715, USA

<sup>5</sup>Neuroscience Training Program, University of Wisconsin-Madison, Madison, WI 53705, USA

<sup>6</sup>Department of Medical Physics, University of Wisconsin-Madison, Madison, WI 53705, USA

<sup>7</sup>Departments of Comparative Biosciences and Obstetrics and Gynecology, University of Wisconsin-Madison, Madison, WI 53715, USA

<sup>8</sup>Science and Engineering Faculty, Queensland University of Technology, Brisbane, QLD 4001, Australia

<sup>9</sup>Program on Genetics of Brain Function, Stanford Center for Genomics and Personalized Medicine, Tasha and John Morgridge Faculty Scholar, Stanford Child Health Research Institute, Stanford University, Stanford, CA, USA

\*co-corresponding authors

## Corresponding authors:

Alexander E. Urban, Ph.D  
3165 Porter Drive, Room 2180  
Palo Alto, CA 94304-1213  
USA  
[aeurban@stanford.edu](mailto:aeurban@stanford.edu)

Karen J. Parker, Ph.D  
1201 Welch Road  
MSLS, Room P-113  
Stanford, CA 94305-5485  
USA  
[kjparker@stanford.edu](mailto:kjparker@stanford.edu)

**Running title:** Haplotype-phased *Callithrix jacchus* ESC line

**Keywords:** *Callithrix jacchus*, transgenics, embryonic stem cells, haplotype phasing, CRISPR/Cas9, model organisms, marmoset, induced neurons (iNs)

## 1 ABSTRACT

2 Due to anatomical and physiological similarities to humans, the common  
3 marmoset (*Callithrix jacchus*), a non-human primate, is an ideal organism for the study  
4 of many human diseases. Researchers are currently leveraging genome-editing  
5 technologies such as CRISPR/Cas9 to genetically engineer marmosets for the *in vivo*  
6 biomedical modeling of human neuropsychiatric and neurodegenerative diseases.  
7 Several marmoset embryonic stem cell (ESC) lines and induced pluripotent stem cell  
8 (iPSC) lines established for these purposes are now available and being distributed to  
9 various research institutions. The genome characterization of these cell lines can  
10 greatly reinforce these transgenic efforts and serve as a valuable resource in many  
11 ways. For example, knowing the genomic sequence variants and haplotype phasing  
12 information for a particular marmoset cell line allows for the better control of off-target  
13 effects in CRISPR/Cas9 experiments and also the ability to make allele-specific edits. It  
14 also provides the genomic contexts required for the accurate interpretation of functional  
15 genomics data. Here, we performed whole-genome characterization in a haplotype-  
16 resolved manner for marmoset ESC line cj367 from the Wisconsin National Primate  
17 Research Center. This is the first haplotype-resolved analysis of a marmoset genome  
18 and the first whole-genome characterization of any marmoset ESC line. We identified  
19 and phased single-nucleotide variants (SNVs) and Indels across the genome. By  
20 leveraging this haplotype information, we then compiled a list of cj367 ESC allele-  
21 specific CRISPR targeting sites. Furthermore, we demonstrated successful Cas9  
22 Endonuclease Dead (dCas9) expression and targeted localization in cj367 as well as  
23 sustained pluripotency after dCas9 transfection by teratoma assay. Lastly, we show  
24 that these ESCs can be directly induced into functional neurons in a rapid, single-step  
25 process. Our study provides a valuable set of genomic resources for primate  
26 transgenics in this post-genome era.

## 1 **Introduction**

2           Biomedical research has been propelled by the ability to genetically modify  
3 conventional laboratory animals such as mice, fish, flies, and worms that have  
4 traditionally served as model organisms. These species, however, are evolutionarily  
5 distant to, and physiologically distinct from, humans, and are ill-suited for modeling  
6 complex brain functions and neuropsychiatric illness (Nelson and Winslow 2009; Phillips  
7 et al. 2014; Camus et al. 2015). Although there is a critical need for developing  
8 sophisticated animal models with greater homology to human disease (Capitano and  
9 Emborg 2008; Emborg 2017; Parker et al. 2018), until recently, transgenic non-human  
10 primate (NHP) research endeavors were a prohibitively inefficient and expensive  
11 undertaking (Grow et al. 2016). Recent breakthroughs in NHP genome-editing  
12 technologies, however, have advanced the possibility of gaining mechanistic insights  
13 into the genetic basis of human brain development and function.

14           The common marmoset (*Callithrix jacchus*) is an ideal model organism by which  
15 to advance the NHP engineering objective. Common marmosets are small New World  
16 monkeys and thrive in captivity under standard laboratory settings. Like humans,  
17 common marmosets are highly social, raise young biparentally, and demonstrate  
18 complex cognitive abilities (Schultz-Darken et al. 2016; Ausderau et al. 2017). Their  
19 metabolic, immunological, endocrinological, and anatomical characteristics are likewise  
20 quite similar to those of humans (Sasaki 2015). Importantly, marmosets and humans  
21 share conserved brain structures, such as enlarged cerebral neocortices (Wise 2008;  
22 Izpisua-Belmonte et al. 2015), as well as complex neural circuitry (Okano et al. 2016;  
23 Okano et al. 2012), rendering marmosets a well-suited model organism for the

1 development of *in vivo* biomedical modeling for human neurological, psychiatric, and  
2 neurodegenerative diseases. Finally, marmosets also have inherent advantages over  
3 other NHPs, such as macaques, due to their small size, high reproductive efficiency,  
4 year-round ovarian cycles, ease of handling, and absence of serious zoonotic viral  
5 infections which can be fatal to humans (Mansfield 2003).

6 Several marmoset embryonic stem cell (ESC) lines and induced pluripotent stem cell  
7 (iPSC) lines with the capabilities of differentiating into specified cell lineages are now  
8 available (Thomson et al. 1996; Sasaki et al. 2005; Tomioka et al. 2010; Vermilyea et al.  
9 2017). The first transgenic primate showing successful germline transmission of  
10 transgenes was accomplished using the marmoset in which lentiviruses expressing  
11 enhanced green fluorescent protein (EGFP) were injected into preimplantation embryos  
12 (Sasaki et al. 2009). A similar technology was also used to generate transgenic  
13 marmosets expressing genetically encoded calcium indicators ((Sasaki et al. 2009; Park  
14 et al. 2016; Tomioka et al. 2017). Shortly after the reporting of the first genetically  
15 modified marmosets, the marmoset genome was sequenced and assembled using  
16 various strategies including Sanger sequencing at 6x coverage (Marmoset Genome  
17 Sequencing and Analysis Consortium 2014; Sato et al. 2015). Sequencing of the  
18 marmoset transcriptome across various tissues has also been undertaken in an effort to  
19 provide high-quality gene annotations for its recent genome assembly (Maudhoo et al.  
20 2014). Furthermore, gene-editing technologies based on homologous recombination  
21 and programmable nucleases, such as zinc-finger nucleases (ZFNs) and transcription  
22 activator-like effector nucleases (TALENs), have been successfully applied in marmoset  
23 ESCs and embryos (Shiozawa et al. 2011; Sato et al. 2016).

1 Advances in CRISPR/Cas9 technology have enabled greater possibilities for NHP  
2 transgenic research including the introduction of copy number variations (CNVs) in  
3 animal models. The ability to create transgenic NHP animals using CRISPR/Cas9 was  
4 successfully demonstrated in cynomolgus monkeys in which multiple genes were  
5 altered in a single step (Niu et al. 2014). Efforts to engineer marmoset cells using  
6 CRISPR/Cas9 are currently underway in multiple labs around the world using the  
7 already-established marmoset ESC and iPSC lines (Vermilyea et al. 2018), and  
8 advances are expected in the coming years (Izpisua-Belmonte et al. 2015; Sato and  
9 Sasaki 2018).

10 Genomic editing strategies stand to benefit from characterizing the genome of these  
11 cell lines. Whole genome characterization can assist the design of CRISPR gRNAs to  
12 gain sequence specificity and to decrease off target effects. Whole genome  
13 characterization also provides genomic contexts for accurate interpretation of functional  
14 genomics studies, such as measuring genome-wide expression levels and methylation  
15 states. Haplotype phasing across the genome allows for the identification of CRISPR  
16 sites suitable for allele-specific targeting. Whole-genome characterization in the context  
17 of copy-number variation (CNV) analysis also allows for the identification of large CNVs  
18 not evident from standard karyotyping analysis. This is especially important for studies  
19 that utilize *in vitro* differentiation of neurons, as they are known to harbor mosaic ploidy  
20 changes during development via a yet to be understood mechanism (McConnell et al.  
21 2013; Cai et al. 2014; Knouse et al. 2016). Lastly, whole-genome characterization  
22 provides a reference genetic background for the transgenic animals or other cell lines  
23 derived from the ESCs.

1        Here, we performed whole-genome characterization in a haplotype-resolved manner  
2        for marmoset ESC line cj367 from the Wisconsin National Primate Research Center  
3        (Vermilyea et al. 2017). To the best of our knowledge, this is the first haplotype-  
4        resolved analysis of a marmoset genome and the first whole-genome characterization  
5        of any marmoset ESC line. Specifically, we identified single-nucleotide variants (SNVs)  
6        and Indels by performing deep-coverage whole-genome sequencing. We then phased  
7        these SNV and Indel haplotypes by performing linked-read sequencing at 128.5×  
8        physical coverage. Leveraging these phased haplotypes, we then compiled a list of  
9        cj367 ESC allele-specific CRISPR targeting sites. Furthermore, we demonstrated  
10       successful Cas9 Endonuclease Dead (dCas9) expression and targeted localization in  
11       cj367 as well as sustained pluripotency after dCas9 transfection by teratoma assay. We  
12       also performed independent confirmation of cj367 ESC pluripotency and also show that  
13       these ESCs can be directly induced into functional neurons in a rapid, single-step  
14       process. Our datasets provide a valuable set of resources for primate transgenic  
15       research.

## 16       **Results**

### 17       ***Characteristics of marmoset ESC line cj367 and direct induction to neurons***

18       Characteristics of the cj367 ESCs were similar to that of human iPSCs, growing  
19       in dense, circular colonies adhered to the plate (Figure 1). Following the thawing and  
20       first passaging of the ESCs after arrival from the Wisconsin National Primate Research  
21       Center, removal of differentiated cells was only needed once. The ESCs required a  
22       daily change of media and to be passaged every 4 to 5 days (~50-80% confluency).  
23       ESC colonies have clearly defined outlines, identical optical properties, and clear nuclei

1 (Figure 1a). Extra care is needed when dissociating the cells from the plate because  
2 they readily detach with little to no agitation after adding a cell dissociation agent such  
3 as Accutase. Instead of aspirating the Accutase before adding PBS to fully dissociate  
4 the cells, DMEM/F12 media was added directly to the Accutase to wash the cells and  
5 then spun down to remove the Accutase/media mixture. The cj367 ESCs were also  
6 confirmed to be pluripotent by immunostaining for pluripotency markers NANOG,  
7 SSEA4, and OCT3/4 (Figure 1d-f) as was initially demonstrated elsewhere (Vermilyea  
8 et al. 2017).

9 We differentiated the ESCs into immature cortical glutaminergic neurons using a  
10 two-week protocol originally designed and optimized for use in human iPSC lines  
11 (Zhang et al. 2013). To test the effectiveness of this protocol, the ESCs were plated at  
12 different densities based on total cell count (100k, 150k, 200k, 250k cells). Each well of  
13 cells received the same amount of the three lentiviral vectors described in (Zhang et al.  
14 2013), and the protocol was followed for the next nine days. Cell density did not alter  
15 the results of the differentiation; however, the ESCs behaved differently from human  
16 iPSCs in two major ways. First, compared to human iPSCs, the cells that survived the  
17 differentiation process grouped entirely around the edge of the well, and only a few cells  
18 were observed in the middle. Second, the marmoset ESCs required 3 to 4 more days to  
19 reach complete differentiation compared to human iPSCs. The protocol was stopped  
20 short at day 9/10 for the marmoset ESCs, i.e. the human day-seven-equivalent (Zhang  
21 et al. 2013), and GFP fluorescence was used as a marker to visualize the differentiation  
22 process (Figure 1b, c).

23 ***Identification of SNVs and Indels***

1           From deep-coverage (~60×) short-insert Illumina WGS of the cj367 genome, we  
2 identified SNVs and indels using GATK Haplotypecaller (McKenna et al. 2010). In  
3 standard GATK Best Practices workflow (DePristo et al. 2011) for human genomes,  
4 high-confidence SNVs and indels are filtered through a statistical learning approach  
5 using training datasets of variants derived from population-scale data from the 1000  
6 Genomes Project (The 1000 Genomes Project Consortium et al. 2010) or the  
7 International HapMap Project (International HapMap Consortium 2005). Since similar  
8 datasets are not available for marmosets, to identify high-confidence SNVs and indels,  
9 we adopted our own statistical approach. Briefly, we analyzed the distributions of the  
10 following parameters (QD, FS, MQ, MQRankSum, ReadPosRankSum) in the raw  
11 outputs from GATK Haplotypecaller and determined a series of appropriate filtering  
12 thresholds for SNVs and indels separately (see Materials and Methods).

13           Through this approach, we identified a total of ~3.75M SNVs (2.23M  
14 heterozygous, 1.52M homozygous) and 0.82M indels (0.43M heterozygous, 0.38M  
15 homozygous) (Table 1, Figure 2a, EVA accession PRJEB27676); 33% of SNVs and 35%  
16 of indels are intragenic, and 0.73% of SNVs and 0.41% of indels are within protein  
17 coding exons (Table 1). Furthermore, we also identified long genomic stretches  
18 exhibiting continuous homozygosity by using a Hidden Markov Model (HMM) adopted  
19 from Adey and colleagues (Adey et al. 2013). In cj367, continuous stretches of  
20 chromosome 1, 2, 4, 6, 7, 9, 12, 13, 14, 15, and 22 show long continuous stretches of  
21 homozygosity that likely resulted from inbreeding (Figure 2a, Table S1).

## 22 ***Haplotype Phasing***



1           We then haplotype-phased the identified SNVs and indels in cj367 by performing  
2 10X Genomics Chromium linked-read library preparation and sequencing (Zheng et al.  
3 2016; Marks et al. 2018). Post sequencing analysis showed that approximately 1.26 ng  
4 or 381 genomic equivalents of high molecular weight (HMW) genomic DNA fragments  
5 (average size 36 kb, 66.2% > 20kb, 14.3% >100kb) were partitioned into approximately  
6 1.27 million droplets and uniquely barcoded (16 bp) (Table 1). The linked-read library  
7 was sequenced (2×151 bp) to 30.4× genome coverage with half of the reads coming  
8 from HMW DNA molecules with at least 30 linked-reads (N50 Linked-Reads per  
9 Molecule) (Table 1). We estimate the actual physical coverage ( $C_F$ ) to be 128.5×.  
10 Coverage of the mean insert by sequencing ( $C_R$ ) is 8,520 bp (284bp × 30 linked-reads)  
11 or 23.7% of the average input DNA fragment size, thus the overall sequencing coverage  
12  $C = C_R \times C_F = 30.4\times$ . In 3115 haplotype blocks (Table 1, EVA accession  
13 PRJEB27676), 2.16M (96.9%) of heterozygous SNVs and 0.79M (83.7%) of indels were  
14 phased. The longest phased haplotype block is 23.4 Mbp (N50 = 2.49Mbp) (Figure 2b,  
15 Table 1, EVA accession PRJEB27676). Haplotype block sizes vary widely across  
16 chromosomes (Figure S1, Figure 2a), and poorly phased regions correspond to regions  
17 exhibiting continuous homozygosity (Table S1, Figure 2a, EVA accession  
18 PRJEB27676).

### 19 ***Cj367 genome-specific and allele-specific CRISPR targets***

20           We identified 44,590 targets in cj367 genome suitable for allele-specific CRISPR  
21 targeting (Table S2). Phased variant sequences (including reverse complement) that  
22 differ by >1 bp between the alleles were extracted to identify all possible CRISPR  
23 targets by pattern searching for [G, C, or A]<sub>N</sub>GG (see Materials and Methods). Only

1 conserved high-quality targets were retained by using a selection method previously  
2 described and validated (Sunagawa et al. 2016). We took the high-quality target  
3 filtering process further by taking the gRNA function and structure into account. Targets  
4 with multiple exact matches, extreme GC fractions, and those with TTTT sequence  
5 (which might disrupt the secondary structure of gRNA) were removed. Furthermore, we  
6 used the Vienna RNA-fold package (Lorenz et al. 2011) to identify gRNA secondary  
7 structure and eliminated targets for which the stem loop structure for Cas9 recognition is  
8 not able to form (Nishimasu et al. 2014). Finally, we calculated the off-target risk score  
9 using the tool as described for this purpose (Ran et al. 2013). A very strict off-target  
10 threshold score was chosen in which candidates with a score below 75 were rejected to  
11 ensure that all targets are as reliable and as specific as possible.

## 12 ***CRISPR/Cas9 in marmoset ESC***

13 In order to test whether CRISPR/Cas9 technology would be suitable to create  
14 transgenic marmoset models, we tested whether sgRNA-guided CRISPR entry was  
15 possible in cj367. After transfecting cj367 marmoset ESCs with plasmids carrying  
16 *dCas9-GFP* (Chen et al. 2013) and using sgRNA sequences that target telomeres, we  
17 successfully located dCas9-GFP fluorescence from telomeres within the nuclei of live  
18 marmoset ESCs (three nuclei shown; Fig. 3a-c), demonstrating that the Cas9 nuclease  
19 can indeed enter the nucleus and possibly edit the genome in a targeted fashion.  
20 Teratoma assay demonstrate that pluripotency of ESC remains unchanged after dCas9-  
21 GFP transfection (data not shown).

## 22 **Discussion**

1           One of the many advantages of using the common marmoset as a model to  
2 study human disease is that its physiology, brain structure, and aspects of behavior  
3 more closely resemble that of humans (Okano et al. 2012; Kishi et al. 2014). The  
4 creation of pluripotent marmoset SC lines provides a platform on which to study  
5 molecular and cellular patterns typical of the species. Traditionally, marmoset ESC  
6 lines have been derived from the inner cell mass of blastocysts, and all of the surviving  
7 lines had been female (Thomson et al. 1996; Sasaki et al. 2005; Debowski et al. 2016).  
8 Recently, three novel marmoset ESC lines have been derived from the natural morula  
9 stage of preimplantation embryos, and one novel male ESC line has been derived from  
10 the expanded blastocyst stage (Debowski et al. 2016). Marmoset iPSCs from fetal liver  
11 cells and adult fibroblasts have also been reported (Tomioka et al. 2010; Farnsworth et  
12 al. 2013; Qiu et al. 2015; Vermilyea et al. 2017). The capacity of marmoset ESC and  
13 iPSC lines to differentiate into cardiomyocytes and hematopoietic cells by induction has  
14 been demonstrated (Sasaki et al. 2005; Tomioka et al. 2010). They can also be  
15 directed to neural cell differentiation using the stromal cell-derived inducing activity  
16 method and to dopaminergic neurons using a dual-SMAD-inhibition induction method  
17 (Sasaki et al. 2005; Tomioka et al. 2010; Vermilyea et al. 2017). Here, we  
18 independently validate the pluripotency of the cj367 cell line (Figure d-f) and show that  
19 ESC line cj367 can be directly induced to neurons in a rapid, single step using a  
20 protocol developed for human pluripotent stem cells (Zhang et al. 2013) (Figure 1b, c).  
21 In addition to revealing these biological characteristics of cj367 that are especially  
22 applicable for neuroscience research, our results allow researchers to perform genome-  
23 editing in cj367 using CRISPR/Cas9 in a “personalized” and allele-specific manner. Our

1 haplotype-resolved whole-genome characterization of cj367 also provides a reference  
2 for all future functional genomic and epigenomics studies, such as RNA-seq, DNA  
3 methylation analysis, or ChIP-seq, carried out using cj367 or derivative cell  
4 lines/animals. Taken together, our data provide a valuable resource for the primate  
5 transgenic research community in this post-genome era.

6         The recent successes in using CRISPR-Cas9 to disrupt specific genes such as  
7 *Ppar-γ*, *Rag1*, and *Dax1* in cynomolgus (Niu et al. 2014; Kang et al. 2015) and *DMD* in  
8 rhesus monkeys (Chen et al. 2015) forecast that the use of CRISPR-Cas9 to genetically  
9 edit marmoset cells or embryos is within reach (Kishi et al. 2014; Izpisua-Belmonte et al.  
10 2015; Kropp et al. 2017; Sato and Sasaki 2018). Gene-editing technologies that  
11 leverage programmable nucleases, such as ZFNs and TALENs have already been  
12 demonstrated in marmoset ESCs (Shiozawa et al. 2011; Sato et al. 2016). Here, we  
13 provide further evidence on this exciting possibility by demonstrating the successful  
14 sgRNA-guided CRISPR entry and expected dCas9-GFP localization in marmoset  
15 ESCs, which suggests that the Cas9 nuclease can indeed enter the nucleus and edit  
16 the genome in a sequence-directed manner. Furthermore, this not only demonstrates  
17 the relative ease of porting existing gene-editing methodologies but also the exciting  
18 future feasibility of creating genetic models of human disorders that provide new  
19 insights only possible in NHP models (Capitanio and Emborg 2008; Zahs and Ashe  
20 2010).

21         With the imminent application of CRISPR-Cas9 for marmosets on the horizon, it  
22 is important to have genome characterizations of the NHP ESC and iPSC lines that are  
23 being distributed to various research labs across the world from primate cell repositories.

1 It is now widely recognized that sequence variants can confound the intended on-target  
2 and off-target sites of the gRNAs especially when appropriate controls are lacking  
3 (Pinello et al. 2016). The CRISPR system relies on Watson and Crick base pairing to  
4 mediate genomic editing, thus it is not surprising that genetic sequence variation can  
5 widely affect its efficiency. By taking advantage of the extensive documentation of  
6 human genetic variation, efforts have been made to elucidate the impact of genetic  
7 variation on CRISPR targeting (Canver et al. 2018). Importantly, it has also been  
8 recognized that this effect on CRISPR-targeting efficiency due to genetic sequence  
9 variation can be leveraged to make cell line-specific or allele-specific CRISPR edits for  
10 various studies (Zhou et al. 2018). The need for such resources to facilitate the use of  
11 CRISPR-Cas9 in primate research has already been voiced in the literature (Luo et al.  
12 2016). Here, we performed the first whole-genome characterization for an established  
13 NHP ESC line and the first haplotype-phasing for a marmoset genome. We also  
14 determined the CRISPR-targeting sites suitable for allele-specific editing for cj367  
15 (Table S2). Our haplotype-resolved whole-genome analysis as well as a list of  
16 CRISPR-targeting sites will not only be useful for cj367 in particular but also for future  
17 cell lines and animal models derived from the same genome as cj367.

18 Possessing the unique property among mammals of routinely producing dizygotic  
19 twins that exchange hematopoietic stem cells *in utero*, all marmoset animals are  
20 naturally chimeric (Benirschke et al. 1962). While such chimerism will affect analyses of  
21 DNA samples obtained from developed animals, it does not affect our whole-genome of  
22 ESCs since the cj367 ESC line was derived before the developmental stage where the  
23 differentiation and exchange of hematopoietic stem cells can occur.

1           Another important aspect of whole-genome characterization is the analyses of  
2 structural and copy-number variation. The current quality of the marmoset reference  
3 genome assembly does not allow for confident mapping of structural variants  
4 (Marmoset Genome Sequencing and Analysis Consortium 2014; Sato et al. 2015).  
5 Improvements in this regard are greatly needed. This also applies to other NHP  
6 genome assemblies as well (Luo et al. 2016).

7           Our whole-genome analysis of marmoset ESC line cj367 also allows for the  
8 accurate and confident interpretation of functional genomics datasets. Recently, RNA-  
9 seq experiments have been carried out in marmoset ESCs that show that overall  
10 expression patterns differ between cell lines derived from different animals under the  
11 same culturing conditions (Debowski et al. 2016), which could be a result of different  
12 genetic backgrounds. Furthermore, the expression patterns of various marmoset ESC  
13 lines more closely resemble primed pluripotency states rather than naïve states even  
14 though these cells passed all criteria for expression of established pluripotent markers  
15 (Debowski et al. 2016). This suggests that the culturing methods established for human  
16 pluripotent cells are not directly applicable to marmoset cells. Importantly, these results  
17 also suggest that functional genomic and epigenomics studies such as RNA-seq,  
18 ATAC-seq, and whole-genome bisulfite sequencing could be adopted as a tool to  
19 screen for and identify marmoset ESC lines that may exhibit true naïve pluripotency  
20 states.

## 21 **Materials and Methods**

### 22 ***Cell Culture***

1 The frozen marmoset ESC starter cultures were removed from the liquid nitrogen tank (-  
2 80 °C) and placed in a 37 °C water bath until thawed. Cells were transferred to a 15 mL  
3 centrifuge tube and 10 mL DMEM/F12 (Thermo Scientific, SH30023.01) was added  
4 drop-wise in order to minimize bubble formation. The mixture was centrifuged at 1000  
5 RPM for 3 min. The supernatant was aspirated and the process was repeated allowing  
6 the cells to be washed once more with another 10 mL DMEM/F12. Cells were re-  
7 suspended in 1 mL of medium, which was added to a single well of Matrigel  
8 hESC-qualified Matrix (BD Biosciences, 356234)-coated BioLite 6-well clear multi-dish  
9 (Thermo Scientific, 130184) and incubated at 37 °C and 5% CO<sub>2</sub>. The medium in each  
10 well was changed daily and the cells were passaged every 4 to 5 days. For passaging,  
11 cells were washed with DPBS (Life Technologies, 14190-144), dissociated with  
12 Accutase (Stem Cell Technologies, 7920), and added to fresh medium on a new plate  
13 and/or added to freezing medium for storage.

14 Media ingredients:

- 15 • E8 Medium (Thermo Scientific, A15169-01)
- 16 • E8 Supplement (Thermo Scientific, A15171-01)
- 17 • Glutamax (Life Technologies, 35050-061)
- 18 • Lipid Concentrate (Life Technologies, 11905-031)
- 19 • Nodal (R&D Systems, 3218-ND)
- 20 • Glutathione (Sigma, G4251)

21 Freezing medium: 20% DMSO (Thermo Fischer, 20688) solution (DMSO +  
22 DMEM/F12).

23 ***Differentiating marmoset ESCs into iNs***

1           When the marmoset ESCs reached confluence, the cells were dissociated using  
2   Accutase and plated at 250,000, 200,000, 150,000, and 100,000 cells per well. One-  
3   day post-passage, a single well was transfected with the 3 viral vectors as described in  
4   the two-week iN protocol (Zhang et al. 2013). The vectors were acquired from the  
5   Genome Virus and Vector Core (GVVC) at Stanford Neuroscience Institute  
6   ([neuroscience.stanford.edu](http://neuroscience.stanford.edu)). The protocol was stopped prior to plating on mouse glia in  
7   order to achieve immature cortical glutamatergic neurons. Following transfection, cells  
8   were subjected to puromycin and doxycyclin selection and then monitored for a  
9   minimum of 7 days. For human iPSCs, full differentiation is usually completed at day 7.  
10   The marmoset ESCs required 10 to 11 days (the day of passage is defined as “day 0”)  
11   to achieve full differentiation.

### 12   ***CRISPR-Cas9 Transfection***

13   Marmoset cj367 embryonic stem cells were seeded onto a 35mm glass bottom  
14   microwell dish (MatTek) that was coated with 0.083  $\mu$ g of BD Matrigel (BD Biosciences)  
15   per well. While the cells were maintained for approximately 48 hours to recover and  
16   return to their typical growth rate, they were visually scanned under a light microscope  
17   and manually dissected to remove any patches of spontaneously differentiated cells. At  
18   70% confluence on the 35mm dish, the cells were transfected simultaneously with  
19   Addgene plasmids pSLQ1658-dCas9-EGFP and pSLQ1651-sgTelomere(F+E). We  
20   followed Invitrogen’s protocol for the Lipofectamine LTX transfection reagent (Thermo  
21   Fisher Scientific) using 5 $\mu$ L with 1 $\mu$ g of DNA from each plasmid for one 35mm dish.  
22   The lipid-DNA complexes were adjusted to a final volume of 100 $\mu$ L using OptiMEM  
23   (Thermo Fisher Scientific) and added to the marmoset cells for six hours. Following six



1 hours of incubation, medium was changed to its original E8 formula. Five hours before  
2 imaging, cells were incubated with a photostable DMEM (MEMO EMD Millipore)  
3 supplemented with ROCK inhibitor to reduce photobleaching effects of the E8 media.  
4 Finally, an additional medium change using MEMO was done immediately before  
5 imaging.

### 6 ***Nuclear imaging***

7 Marmoset cell nuclei were imaged with an OMX BLAZE 3D-Structured Illumination  
8 Microscope (Applied Precision, GE). Cells were transferred to a modified stage adaptor  
9 held at 37°C and 5% CO<sub>2</sub>. 1µm z-stacks (0.125µm intervals) were taken at 100x  
10 magnification (U-PLANAPO 100X SIM objective N.A. 1.42) with a 488nm laser using  
11 DeltaVision OMX acquisition software (version 3.70.9622.0, GE) and deconvolved with  
12 SoftWoRx 7.0.0 software (Deltavision) before maximum projections were made for the  
13 figures (Fiji/ImageJ).

### 14 ***Illumina WGS***

15 Marmoset ESC (cj367) genomic DNA was extracted using the Qiagen DNeasy Blood &  
16 Tissue Kit (Cat.D No. 69504) and quantified using the Qubit dsDNA HS Assay Kit  
17 (Invitrogen, Waltham, MA, USA). DNA purity was verified (OD<sub>260/280</sub> > 1.8;  
18 OD<sub>260/230</sub> > 1.5) using NanoDrop (Thermo Scientific, Waltham, MA, USA). Afterwards,  
19 DNA was sheared to 200 bp to 850 bp (average size = 400 bp) using the Covaris S2  
20 instrument (Covaris, Woburn, MA, USA). DNA was then concentrated to a volume of 50  
21 µL using the DNA Clean & Concentrator kit (Cat. No. D4013) from Zymo Research  
22 (Irvine, CA, USA). One short-insert WGS library was constructed using the Kapa Hyper  
23 Prep Kit (KR0961) from Kapa Biosystems (Woburn, MA, USA) with 150 ng of sheared

1 DNA as input following standard manufacturer's protocol with 10 cycles of PCR.  
2 Library-size selection (420 bp to 1 kb) was carried out on the BluePippin instrument  
3 using the R2 marker on a 1.5% dye-free cassette from Sage Science (Beverly, MA,  
4 USA). WGS library-size selection was then verified using the 2100 Bioanalyzer DNA  
5 1000 Kit (Agilent Technologies, Santa Clara, CA, USA). The WGS library was then  
6 sequenced (2×150 bp) on two lanes of the Illumina HiSeq 4000 to achieve >600 million  
7 pass-filter paired-end reads.

### 8 ***Determining SNVs and Indels***

9 Paired-end reads from the two lanes of Illumina HiSeq 4000 were combined and aligned  
10 to the *Callithrix jacchus* genome (GenBank accession GCA\_000004665.1) using BWA-  
11 MEM version 0.7.5 (Li and Durbin 2009) followed by marking of duplicates using Picard  
12 tools (version 1.129) (<http://broadinstitute.github.io/picard/>). SNVs and Indels were  
13 called using by GATK Haplotypecaller (version 3.7) (McKenna et al. 2010) and hard  
14 filtered for quality (QD > 2.0, FS < 60.0, MQ > 40.0, MQRankSum > -12.5,  
15 ReadPosRankSum > -8.0).

### 16 ***Identifying genomic regions exhibiting continuous homozygosity***

17 A Hidden Markov Model (HMM) was used to identify genomic regions exhibiting  
18 continuous homozygosity. The HMM is designed with two states: homozygous and  
19 heterozygous. We used the hard-filtered SNVs we derived from GATK HaplotypeCaller,  
20 and filtered again for GQ > 30 and QUAL > 50. The genome was split into 25 kb bins;  
21 heterozygous and homozygous SNVs were tallied for each bin, and bins with <25 SNVs  
22 were removed. A bin was classified as heterozygous if >50% of the SNVs within the bin  
23 are heterozygous, otherwise it was classified as homozygous. This classification was

1 used as the HMM emission sequence. The HMM was initialized with the same initiation  
2 and transition probabilities (Prob=10E-8) (Adey et al., 2013), the Viterbi algorithm was  
3 used to estimate a best path, and adjacent homozygous intervals were merged.

#### 4 ***10X Genomics linked-read library and sequencing***

5 Marmoset ESC (cj367) genomic DNA was extracted using the MagAttract HMW DNA  
6 Kit from Qiagen (Cat. No. 67563) (Hilden, Germany). The extracted genomic DNA was  
7 verified to be of HMW (average size >30 kb) using using field-inversion gel  
8 electrophoresis on the Pippin Pulse System (Sage Science, Beverly, MA, USA). DNA  
9 was then diluted to 1 ng/μl to be used as input for the 10x Genomics (Pleasanton, CA,  
10 USA) Chromium system (Zheng et al. 2016; Marks et al. 2018) in which HMW DNA  
11 fragments are partitioned into >1 million droplets in emulsion, uniquely barcoded (16 bp)  
12 within each droplet, and subjected to random priming and isothermal amplification  
13 following standard manufacturer's protocol. The barcoded DNA molecules were then  
14 released from the droplets and converted to a linked-read library in which each library  
15 molecule has a 16 bp "HMW fragment barcode". The final linked-read library (8 cycles  
16 of PCR amplification) was quantified using the Kapa Library Quantification Kit (Cat. No.  
17 KK4824) from Kapa Biosystems (Woburn, MA, USA) and diluted to 4 nM. Sequencing  
18 (2×151 bp) was performed on the Illumina NextSeq 500 using the NextSeq 500/550  
19 High Output v2 kit (Cat. No. FC-404-2004) to achieve >30× genomic coverage.

#### 20 ***Haplotype phasing using linked-reads***

21 Read-pairs i.e. linked-reads, that come from the same HMW DNA fragment carry the  
22 same "HMW fragment barcode" on Read 1. Virtual long-reads that are representative of  
23 the sequences of the original HMW genomic DNA fragments can be generated by

1 clustering of linked-reads with identical barcodes. This process allows for the haplotype  
2 phasing of heterozygous SNVs and Indels. Paired-end linked-reads (median insert size  
3 368 bp, duplication rate 5.39%, Q30 Read1 75.6%, Q30 Read2 68.4%) were aligned to  
4 *Callithrix jacchus* genome (GenBank accession GCA\_000004665.1) (alignment rate  
5 87.3%, mean coverage 30.4x, zero coverage 0.938%) and analyzed using the Long  
6 Ranger Software (version 2.1.3) from 10x Genomics (Zheng et al. 2016; Marks et al.  
7 2018) (Pleasanton, CA, USA). Genome indexing was performed prior to alignment  
8 using the Long Ranger *mkref* module. Since Long Ranger *mkref* requires less than 500  
9 contigs, we concatenated contigs in GCA\_000004665.1 to meet this requirement.  
10 Reference gaps and repeat regions, downloaded from UCSC Genome Browser  
11 (Karolchik et al. 2004; Kuhn et al. 2013), were excluded from the analysis. Phasing was  
12 performed by specifying the set of pre-determined and filtered heterozygous SNVs and  
13 Indels from using GATK (see above) and formatted using *mkvcf* from Long Ranger  
14 (version 2.1.5). Phasing analysis was performed using the Long Ranger *wgs* module.

### 15 ***Allele-specific CRISPR targets***

16 To identify allele-specific CRISPR targets, we started by extracting variants that  
17 satisfy the following properties from Dataset S2 (phased variants):

- 18 1. They passed quality control (VCF field 'Filter' is equal to "PASS")
- 19 2. They are phased (VCF field 'GT' uses "|" as separator rather than "/")
- 20 3. The alleles are heterozygous (e.g. "0|1" or "1|0", but not "1|1")
- 21 4. The difference between the alleles is > 1 bp (to ensure target specificity)

22 For the 704,571 variants that satisfy these four properties, we extracted the two  
23 haplotype sequences (maximum length: 595). We only worked with the sequences that

1 were present in the phased genotype. Extracted sequences were tagged them  
2 according to their haplotype, for instance:

- 3 • If the sequence in the 'Ref' field was "GTA", the 'Alt' field sequence was "TA",  
4 and phasing was "0|1" i.e. Haplotype\_1|Haplotype\_2", the sequence containing  
5 "GTA" was tagged "1" for Haplotype 1 and the sequence containing "TA" was  
6 tagged "2" for Haplotype 2.
- 7 • If the 'Ref' field was sequence "GTA", the 'Alt' field was "TA,GCTA", phasing was  
8 "1|2", the sequence containing "TA" is tagged "1", the sequence containing  
9 "GCTA" was tagged "2" (and the sequence containing "GTA" was not used)

10 A regular expression, RegEx, was used to extract all potential CRISPR targets from  
11 these sequences (i.e. all sequences that matched a [G, C, or A]N<sub>20</sub>GG pattern and  
12 those for which the reverse-complement matched this pattern). This yielded 720,239  
13 candidates, which were then filtered to retain only high-quality targets. The process is  
14 adapted from a selection method previously described and validated (Sunagawa et al.  
15 2016) and has already been used for more than 20 genes (Tatsuki et al. 2016). A high-  
16 quality candidate gRNA needs to target a unique site. All the candidates that have  
17 multiple exact matches in hg19 (irrespective of location) were identified using Bowtie2  
18 (Langmead and Salzberg 2012) and removed. We also removed targets with extreme  
19 GC content (>80% or <20%), and targets that contain TTTT, which tends to break the  
20 gRNA's secondary structure. We also used the Vienna RNA-fold package (Lorenz et al.  
21 2011) to compute the gRNA's secondary structure. We eliminated all candidates for  
22 which the stem loop structure cannot fold correctly for Cas9 recognition (Nishimasu et al.  
23 2014), except if the folding energy was above -18 (indicating that the "incorrectly-folded"

1 structure is unstable). Finally, we evaluated the off-target risk score using our own  
2 implementation of the Zhang tool (Ran et al. 2013). To ensure that all targets are as  
3 reliable and specific as possible, we used a very strict threshold and rejected  
4 candidates with an off-target risk score less than 75. Candidates that satisfy all these  
5 requirements are considered high quality. For each candidate, we report the location of  
6 the variant (chromosome and position), the haplotype ( “1” or “2” from the “Phased  
7 Genotype“ column i.e. “Haplotype\_1 | Haplotype\_2”), the gRNA target sequence, its  
8 position relative to the start of the variant, its orientation, its off-target score, and the  
9 genomic element targeted (gene or enhancer). Note that the position relative to the start  
10 of the variant is for the 5'-most end of the target respective to the genome: if the target  
11 is 5'-3', it is its 5' end; if a target was extracted on the reverse complement, it is its 3'  
12 end.

### 13 ***Data access***

14 All sequencing data are available via NCBI SRA accession SRP141443. Variant data is  
15 available via European Variation Archive accession PRJEB27676.

### 16 **Acknowledgements**

17 This research was supported by a Tashia and John Morgridge Faculty Scholar Award  
18 from the Stanford Child Health Research Institute (AEU), a the Stanford Neurosciences  
19 Institute Seed Grant (AEU & KJP), the Stanford Child Health Research Institute Grant  
20 (AEU & KJP), and the Stanford Medicine Faculty Innovation Program (AEU). BZ was  
21 funded by NIH grant T32 HL110952. LCL was funded by the Stanford School of  
22 Medicine Dean’s Fellowship and NIH grants NS062798, DK090065, and  
23 MH099647. We are thankful for the support of the Stanford CSIF core facility – the

1 imaging with the DeltaVision OMX system described was supported, in part, by award  
2 1S10OD01227601 from the National Center for Research Resources (NCRR). This  
3 research was partially supported by NIH grants R24OD019803 to TGG and MEE, and  
4 P51OD011106 to the Wisconsin National Primate Research Center. The sequencing  
5 data was generated on an Illumina HiSeq 4000 that was purchased with funds from NIH  
6 award S10OD018220. Lastly, we thank Arineh Khechaduri for performing marmoset  
7 ESC genomic DNA extraction.

#### 8 **Disclosure Declaration**

9 The authors of this manuscript declare no conflicts of interest.

#### 10 **References**

- 11 Adey A, Burton JN, Kitzman JO, Hiatt JB, Lewis AP, Martin BK, Qiu R, Lee C, Shendure  
12 J. 2013. The haplotype-resolved genome and epigenome of the aneuploid HeLa  
13 cancer cell line. *Nature* **500**: 207–211.  
14 <http://www.nature.com/doi/10.1038/nature12064>.
- 15 Ausderau KK, Dammann C, McManus K, Schneider M, Emborg ME, Schultz-Darken N.  
16 2017. Cross-species comparison of behavioral neurodevelopmental milestones in  
17 the common marmoset monkey and human child. *Dev Psychobiol* **59**: 807–821.  
18 <http://www.ncbi.nlm.nih.gov/pubmed/28763098>.
- 19 Benirschke K, Anderson JM, Brownhill LE. 1962. Marrow Chimerism in Marmosets.  
20 *Science* **138**: 513–5. <http://www.ncbi.nlm.nih.gov/pubmed/17753948>.
- 21 Cai X, Evrony GD, Lehmann HS, Elhosary PC, Mehta BK, Poduri A, Walsh CA. 2014.  
22 Single-cell, genome-wide sequencing identifies clonal somatic copy-number  
23 variation in the human brain. *Cell Rep* **8**: 1280–9.

- 1 <http://www.ncbi.nlm.nih.gov/pubmed/25159146>.
- 2 Camus S, Ko WKD, Pioli E, Bezard E. 2015. Why bother using non-human primate  
3 models of cognitive disorders in translational research? *Neurobiol Learn Mem* **124**:  
4 123–9. <http://www.ncbi.nlm.nih.gov/pubmed/26135120>.
- 5 Canver MC, Joung JK, Pinello L. 2018. Impact of Genetic Variation on CRISPR-Cas  
6 Targeting. *Cris J* **1**: 159–170.  
7 <http://www.liebertpub.com/doi/10.1089/crispr.2017.0016>.
- 8 Capitanio JP, Emborg ME. 2008. Contributions of non-human primates to neuroscience  
9 research. *Lancet* **371**: 1126–35. <http://www.ncbi.nlm.nih.gov/pubmed/18374844>.
- 10 Chen B, Gilbert LA, Cimini BA, Schnitzbauer J, Zhang W, Li G-W, Park J, Blackburn EH,  
11 Weissman JS, Qi LS, et al. 2013. Dynamic imaging of genomic loci in living human  
12 cells by an optimized CRISPR/Cas system. *Cell* **155**: 1479–91.  
13 <http://www.ncbi.nlm.nih.gov/pubmed/24360272>.
- 14 Chen Y, Zheng Y, Kang Y, Yang W, Niu Y, Guo X, Tu Z, Si C, Wang H, Xing R, et al.  
15 2015. Functional disruption of the dystrophin gene in rhesus monkey using  
16 CRISPR/Cas9. *Hum Mol Genet* **24**: 3764–74.  
17 <http://www.ncbi.nlm.nih.gov/pubmed/25859012>.
- 18 Debowski K, Drummer C, Lentjes J, Cors M, Dressel R, Lingner T, Salinas-Riester G,  
19 Fuchs S, Sasaki E, Behr R. 2016. The transcriptomes of novel marmoset monkey  
20 embryonic stem cell lines reflect distinct genomic features. *Sci Rep* **6**: 29122.  
21 <http://www.ncbi.nlm.nih.gov/pubmed/27385131>.
- 22 DePristo M a, Banks E, Poplin R, Garimella K V, Maguire JR, Hartl C, Philippakis A a,  
23 del Angel G, Rivas M a, Hanna M, et al. 2011. A framework for variation discovery



- 1 and genotyping using next-generation DNA sequencing data. *Nat Genet* **43**: 491–8.
- 2 Emborg ME. 2017. Nonhuman Primate Models of Neurodegenerative Disorders. *ILAR J*
- 3 **58**: 190–201. <http://www.ncbi.nlm.nih.gov/pubmed/28985333>.
- 4 Farnsworth SL, Qiu Z, Mishra A, Hornsby PJ. 2013. Directed neural differentiation of
- 5 induced pluripotent stem cells from non-human primates. *Exp Biol Med (Maywood)*
- 6 **238**: 276–84. <http://www.ncbi.nlm.nih.gov/pubmed/23598973>.
- 7 Grow DA, McCarrey JR, Navara CS. 2016. Advantages of nonhuman primates as
- 8 preclinical models for evaluating stem cell-based therapies for Parkinson’s disease.
- 9 *Stem Cell Res* **17**: 352–366.
- 10 <http://linkinghub.elsevier.com/retrieve/pii/S187350611630112X>.
- 11 International HapMap Consortium. 2005. A haplotype map of the human genome.
- 12 *Nature* **437**: 1299–320. <http://www.ncbi.nlm.nih.gov/pubmed/16255080>.
- 13 Izpisua-Belmonte JC, Callaway EM, Caddick SJ, Churchland P, Feng G, Homanics GE,
- 14 Lee K-F, Leopold DA, Miller CT, Mitchell JF, et al. 2015. Brains, genes, and
- 15 primates. *Neuron* **86**: 617–31. <http://www.ncbi.nlm.nih.gov/pubmed/25950631>.
- 16 Kang Y, Zheng B, Shen B, Chen Y, Wang L, Wang J, Niu Y, Cui Y, Zhou J, Wang H, et
- 17 al. 2015. CRISPR/Cas9-mediated Dax1 knockout in the monkey recapitulates
- 18 human AHC-HH. *Hum Mol Genet* **24**: 7255–7264.
- 19 <https://academic.oup.com/hmg/article-lookup/doi/10.1093/hmg/ddv425>.
- 20 Karolchik D, Hinrichs AS, Furey TS, Roskin KM, Sugnet CW, Haussler D, Kent WJ.
- 21 2004. The UCSC Table Browser data retrieval tool. *Nucleic Acids Res* **32**: D493-6.
- 22 <http://www.ncbi.nlm.nih.gov/pubmed/14681465>.
- 23 Kishi N, Sato K, Sasaki E, Okano H. 2014. Common marmoset as a new model animal

- 1 for neuroscience research and genome editing technology. *Dev Growth Differ* **56**:  
2 53–62. <http://www.ncbi.nlm.nih.gov/pubmed/24387631>.
- 3 Knouse KA, Wu J, Amon A. 2016. Assessment of megabase-scale somatic copy  
4 number variation using single-cell sequencing. *Genome Res* **26**: 376–84.  
5 <http://www.ncbi.nlm.nih.gov/pubmed/26772196>.
- 6 Kropp J, Di Marzo A, Golos T. 2017. Assisted reproductive technologies in the common  
7 marmoset: an integral species for developing nonhuman primate models of human  
8 diseases†. *Biol Reprod* **96**: 277–287.  
9 <https://academic.oup.com/biolreprod/article/96/2/277/2948764>.
- 10 Kuhn RM, Haussler D, Kent WJ. 2013. The UCSC genome browser and associated  
11 tools. *Brief Bioinform* **14**: 144–161. [https://academic.oup.com/bib/article-](https://academic.oup.com/bib/article-lookup/doi/10.1093/bib/bbs038)  
12 [lookup/doi/10.1093/bib/bbs038](https://academic.oup.com/bib/article-lookup/doi/10.1093/bib/bbs038).
- 13 Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat*  
14 *Methods* **9**: 357–9. <http://www.ncbi.nlm.nih.gov/pubmed/22388286>.
- 15 Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler  
16 transform. *Bioinformatics* **25**: 1754–60.  
17 <http://www.ncbi.nlm.nih.gov/pubmed/19451168>.
- 18 Lorenz R, Bernhart SH, Höner Zu Siederdisen C, Tafer H, Flamm C, Stadler PF,  
19 Hofacker IL. 2011. ViennaRNA Package 2.0. *Algorithms Mol Biol* **6**: 26.  
20 <http://www.ncbi.nlm.nih.gov/pubmed/22115189>.
- 21 Luo X, Li M, Su B. 2016. Application of the genome editing tool CRISPR/Cas9 in non-  
22 human primates. *Zool Res* **37**: 214–9.  
23 <http://www.ncbi.nlm.nih.gov/pubmed/27469252>.

- 1 Mansfield K. 2003. Marmoset models commonly used in biomedical research. *Comp*  
2 *Med* **53**: 383–92. <http://www.ncbi.nlm.nih.gov/pubmed/14524414>.
- 3 Marks P, Garcia S, Barrio AM, Belhocine K, Bernate J, Bharadwaj R, Bjornson K,  
4 Catalanotti C, Delaney J, Fehr A, et al. 2018. Resolving the Full Spectrum of  
5 Human Genome Variation using Linked-Reads. *bioRxiv* 230946.  
6 <https://www.biorxiv.org/content/early/2018/01/09/230946>.
- 7 Marmoset Genome Sequencing and Analysis Consortium. 2014. The common  
8 marmoset genome provides insight into primate biology and evolution. *Nat Genet*  
9 **46**: 850–7. <https://www.nature.com/articles/ng.3042>.
- 10 Maudhoo MD, Ren D, Gradnigo JS, Gibbs RM, Lubker AC, Moriyama EN, French JA,  
11 Norgren RB. 2014. De novo assembly of the common marmoset transcriptome  
12 from NextGen mRNA sequences. *Gigascience* **3**: 14.  
13 <https://academic.oup.com/gigascience/article-lookup/doi/10.1186/2047-217X-3-14>.
- 14 McConnell MJ, Lindberg MR, Brennand KJ, Piper JC, Voet T, Cowing-Zitron C,  
15 Shumilina S, Lasken RS, Vermeesch JR, Hall IM, et al. 2013. Mosaic copy number  
16 variation in human neurons. *Science* **342**: 632–7.  
17 <http://www.ncbi.nlm.nih.gov/pubmed/24179226>.
- 18 McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K,  
19 Altshuler D, Gabriel S, Daly M, et al. 2010. The Genome Analysis Toolkit: a  
20 MapReduce framework for analyzing next-generation DNA sequencing data.  
21 *Genome Res* **20**: 1297–303. <http://www.ncbi.nlm.nih.gov/pubmed/20644199>.
- 22 Nelson EE, Winslow JT. 2009. Non-human primates: model animals for developmental  
23 psychopathology. *Neuropsychopharmacology* **34**: 90–105.

- 1 <http://www.ncbi.nlm.nih.gov/pubmed/18800061>.
- 2 Nishimasu H, Ran FA, Hsu PD, Konermann S, Shehata SI, Dohmae N, Ishitani R,  
3 Zhang F, Nureki O. 2014. Crystal structure of Cas9 in complex with guide RNA and  
4 target DNA. *Cell* **156**: 935–49. <http://www.ncbi.nlm.nih.gov/pubmed/24529477>.
- 5 Niu Y, Shen B, Cui Y, Chen Y, Wang J, Wang L, Kang Y, Zhao X, Si W, Li W, et al.  
6 2014. Generation of gene-modified cynomolgus monkey via Cas9/RNA-mediated  
7 gene targeting in one-cell embryos. *Cell* **156**: 836–43.  
8 <http://www.ncbi.nlm.nih.gov/pubmed/24486104>.
- 9 Okano H, Hikishima K, Iriki A, Sasaki E. 2012. The common marmoset as a novel  
10 animal model system for biomedical and neuroscience research applications.  
11 *Semin Fetal Neonatal Med* **17**: 336–40.  
12 <http://www.ncbi.nlm.nih.gov/pubmed/22871417>.
- 13 Park JE, Zhang XF, Choi S-H, Okahara J, Sasaki E, Silva AC. 2016. Generation of  
14 transgenic marmosets expressing genetically encoded calcium indicators. *Sci Rep*  
15 **6**: 34931. <http://www.ncbi.nlm.nih.gov/pubmed/27725685>.
- 16 Parker KJ, Garner JP, Oztan O, Tarara ER, Li J, Sclafani V, Del Rosso LA, Chun K,  
17 Berquist SW, Chez MG, et al. 2018. Arginine vasopressin in cerebrospinal fluid is a  
18 marker of sociality in nonhuman primates. *Sci Transl Med* **10**.  
19 <http://www.ncbi.nlm.nih.gov/pubmed/29720452>.
- 20 Phillips KA, Bales KL, Capitanio JP, Conley A, Czoty PW, 't Hart BA, Hopkins WD, Hu  
21 S-L, Miller LA, Nader MA, et al. 2014. Why primate models matter. *Am J Primatol*  
22 **76**: 801–27. <http://www.ncbi.nlm.nih.gov/pubmed/24723482>.
- 23 Pinello L, Canver MC, Hoban MD, Orkin SH, Kohn DB, Bauer DE, Yuan G-C. 2016.

- 1 Analyzing CRISPR genome-editing experiments with CRISPResso. *Nat Biotechnol*  
2 **34**: 695–697. <http://www.nature.com/doi/10.1038/nbt.3583>.
- 3 Qiu Z, Mishra A, Li M, Farnsworth SL, Guerra B, Lanford RE, Hornsby PJ. 2015.  
4 Marmoset induced pluripotent stem cells: Robust neural differentiation following  
5 pretreatment with dimethyl sulfoxide. *Stem Cell Res* **15**: 141–50.  
6 <http://www.ncbi.nlm.nih.gov/pubmed/26070112>.
- 7 Ran FA, Hsu PD, Wright J, Agarwala V, Scott DA, Zhang F. 2013. Genome engineering  
8 using the CRISPR-Cas9 system. *Nat Protoc* **8**: 2281–2308.  
9 <http://www.nature.com/doi/10.1038/nprot.2013.143>.
- 10 Sasaki E. 2015. Prospects for genetically modified non-human primate models,  
11 including the common marmoset. *Neurosci Res* **93**: 110–5.  
12 <http://www.ncbi.nlm.nih.gov/pubmed/25683291>.
- 13 Sasaki E, Hanazawa K, Kurita R, Akatsuka A, Yoshizaki T, Ishii H, Tanioka Y, Ohnishi Y,  
14 Suemizu H, Sugawara A, et al. 2005. Establishment of novel embryonic stem cell  
15 lines derived from the common marmoset (*Callithrix jacchus*). *Stem Cells* **23**: 1304–  
16 13. <http://www.ncbi.nlm.nih.gov/pubmed/16109758>.
- 17 Sasaki E, Suemizu H, Shimada A, Hanazawa K, Oiwa R, Kamioka M, Tomioka I,  
18 Sotomaru Y, Hirakawa R, Eto T, et al. 2009. Generation of transgenic non-human  
19 primates with germline transmission. *Nature* **459**: 523–7.  
20 <http://www.ncbi.nlm.nih.gov/pubmed/19478777>.
- 21 Sato K, Kuroki Y, Kumita W, Fujiyama A, Toyoda A, Kawai J, Iriki A, Sasaki E, Okano H,  
22 Sakakibara Y. 2015. Resequencing of the common marmoset genome improves  
23 genome assemblies and gene-coding sequence analysis. *Sci Rep* **5**: 16894.

- 1 <http://www.readcube.com/articles/10.1038/srep16894> (Accessed April 22, 2016).
- 2 Sato K, Oiwa R, Kumita W, Henry R, Sakuma T, Ito R, Nozu R, Inoue T, Katano I, Sato  
3 K, et al. 2016. Generation of a Nonhuman Primate Model of Severe Combined  
4 Immunodeficiency Using Highly Efficient Genome Editing. *Cell Stem Cell* **19**: 127–  
5 38. <http://linkinghub.elsevier.com/retrieve/pii/S1934590916301539>.
- 6 Sato K, Sasaki E. 2018. Genetic engineering in nonhuman primates for human disease  
7 modeling. *J Hum Genet* **63**: 125–131.  
8 <http://www.ncbi.nlm.nih.gov/pubmed/29203824>.
- 9 Schultz-Darken N, Braun KM, Emborg ME. 2016. Neurobehavioral development of  
10 common marmoset monkeys. *Dev Psychobiol* **58**: 141–58.  
11 <http://www.ncbi.nlm.nih.gov/pubmed/26502294>.
- 12 Shiozawa S, Kawai K, Okada Y, Tomioka I, Maeda T, Kanda A, Shinohara H, Suemizu  
13 H, James Okano H, Sotomaru Y, et al. 2011. Gene targeting and subsequent site-  
14 specific transgenesis at the  $\beta$ -actin (ACTB) locus in common marmoset embryonic  
15 stem cells. *Stem Cells Dev* **20**: 1587–99.  
16 <http://www.ncbi.nlm.nih.gov/pubmed/21126169>.
- 17 Sunagawa GA, Sumiyama K, Ukai-Tadenuma M, Perrin D, Fujishima H, Ukai H,  
18 Nishimura O, Shi S, Ohno R-I, Narumi R, et al. 2016. Mammalian Reverse  
19 Genetics without Crossing Reveals Nr3a as a Short-Sleeper Gene. *Cell Rep* **14**:  
20 662–677. <http://www.ncbi.nlm.nih.gov/pubmed/26774482>.
- 21 Tatsuki F, Sunagawa GA, Shi S, Susaki EA, Yukinaga H, Perrin D, Sumiyama K, Ukai-  
22 Tadenuma M, Fujishima H, Ohno R, et al. 2016. Involvement of Ca(2+)-Dependent  
23 Hyperpolarization in Sleep Duration in Mammals. *Neuron* **90**: 70–85.

- 1 <http://www.ncbi.nlm.nih.gov/pubmed/26996081>.
- 2 The 1000 Genomes Project Consortium, Abecasis GR, Altshuler D, Auton A, Brooks LD,  
3 Durbin RM, Gibbs RA, Hurles ME, McVean GA. 2010. A map of human genome  
4 variation from population-scale sequencing. *Nature* **467**: 1061–73.  
5 <http://www.nature.com/doi/10.1038/nature09534>.
- 6 Thomson JA, Kalishman J, Golos TG, Durning M, Harris CP, Hearn JP. 1996.  
7 Pluripotent cell lines derived from common marmoset (*Callithrix jacchus*)  
8 blastocysts. *Biol Reprod* **55**: 254–9. <http://www.ncbi.nlm.nih.gov/pubmed/8828827>.
- 9 Tomioka I, Ishibashi H, Minakawa EN, Motohashi HH, Takayama O, Saito Y, Popiel HA,  
10 Puentes S, Owari K, Nakatani T, et al. 2017. Transgenic Monkey Model of the  
11 Polyglutamine Diseases Recapitulating Progressive Neurological Symptoms.  
12 *eneuro* **4**: ENEURO.0250-16.2017.  
13 <http://eneuro.sfn.org/lookup/doi/10.1523/ENEURO.0250-16.2017>.
- 14 Tomioka I, Maeda T, Shimada H, Kawai K, Okada Y, Igarashi H, Oiwa R, Iwasaki T,  
15 Aoki M, Kimura T, et al. 2010. Generating induced pluripotent stem cells from  
16 common marmoset (*Callithrix jacchus*) fetal liver cells using defined factors,  
17 including Lin28. *Genes to Cells* **15**: 959–969. <http://doi.wiley.com/10.1111/j.1365-2443.2010.01437.x>.
- 18
- 19 Vermilyea SC, Babinski A, Guthrie S, Golos TG, Emborg ME. 2018. LRRK2 Genomic  
20 Editing in Common Marmoset Stem Cells. *Cell Transplant* **27**: 682–722.  
21 <https://doi.org/10.1177/0963689718765742>.
- 22 Vermilyea SC, Guthrie S, Meyer M, Smuga-Otto K, Braun K, Howden S, Thomson JA,  
23 Zhang S-C, Emborg ME, Golos TG. 2017. Induced Pluripotent Stem Cell-Derived

- 1 Dopaminergic Neurons from Adult Common Marmoset Fibroblasts. *Stem Cells Dev*  
2 **26**: 1225–1235. <http://www.ncbi.nlm.nih.gov/pubmed/28635509>.
- 3 Wise SP. 2008. Forward frontal fields: phylogeny and fundamental function. *Trends*  
4 *Neurosci* **31**: 599–608.  
5 <http://linkinghub.elsevier.com/retrieve/pii/S0166223608002075>.
- 6 Zahs KR, Ashe KH. 2010. “Too much good news” - are Alzheimer mouse models trying  
7 to tell us how to prevent, not cure, Alzheimer’s disease? *Trends Neurosci* **33**: 381–  
8 9. <http://www.ncbi.nlm.nih.gov/pubmed/20542579>.
- 9 Zhang Y, Pak C, Han Y, Ahlenius H, Zhang Z, Chanda S, Marro S, Patzke C, Acuna C,  
10 Covy J, et al. 2013. Rapid single-step induction of functional neurons from human  
11 pluripotent stem cells. *Neuron* **78**: 785–98.  
12 <http://www.ncbi.nlm.nih.gov/pubmed/23764284>.
- 13 Zheng GXY, Lau BT, Schnall-Levin M, Jarosz M, Bell JM, Hindson CM, Kyriazopoulou-  
14 Panagiotopoulou S, Masquelier DA, Merrill L, Terry JM, et al. 2016. Haplotyping  
15 germline and cancer genomes with high-throughput linked-read sequencing. *Nat*  
16 *Biotechnol* **34**: 303–11. <http://www.ncbi.nlm.nih.gov/pubmed/26829319> (Accessed  
17 July 14, 2016).
- 18 Zhou B, Ho SS, Zhu X, Zhang X, Spies N, Byeon S, Arthur JG, Pattni R, Ben-Efraim N,  
19 Haney MS, et al. 2018. Comprehensive, integrated and phased whole-genome  
20 analysis of the primary ENCODE cell line K562. *bioRxiv*.  
21 <http://biorxiv.org/content/early/2018/01/22/192344.abstract>.

22

23



1 **FIGURE AND TABLE LEGENDS**

2 **Figure 1.** (a) cj367 ESC colony at 20× magnification. Days 5 (b) and 9 (c) respectively  
3 of cj367 directly induced into immature cortical glutamatergic neurons; white bar = 400  
4 microns. Cells are marked by EGFP. Immunostaining of pluripotency markers OCT3/4  
5 (d), NANOG (e), and SSEA4 (f) in cj367; white bar = 200 microns.

6 **Figure 2.** (a) Circos visualization of cj367 genome variants with the following tracks: 1.  
7 phased haplotype blocks (demarcated with 4 colors for clearer visualization); 2. SNV  
8 density in 1 Mb windows; 3. Indel density in 1 Mb windows; 4. zygosity (heterozygous or  
9 homozygous > 50%) in 1 Mb windows; regions with continuous homozygosity. (b) Violin  
10 plot of size distribution of haplotype blocks in cj367.

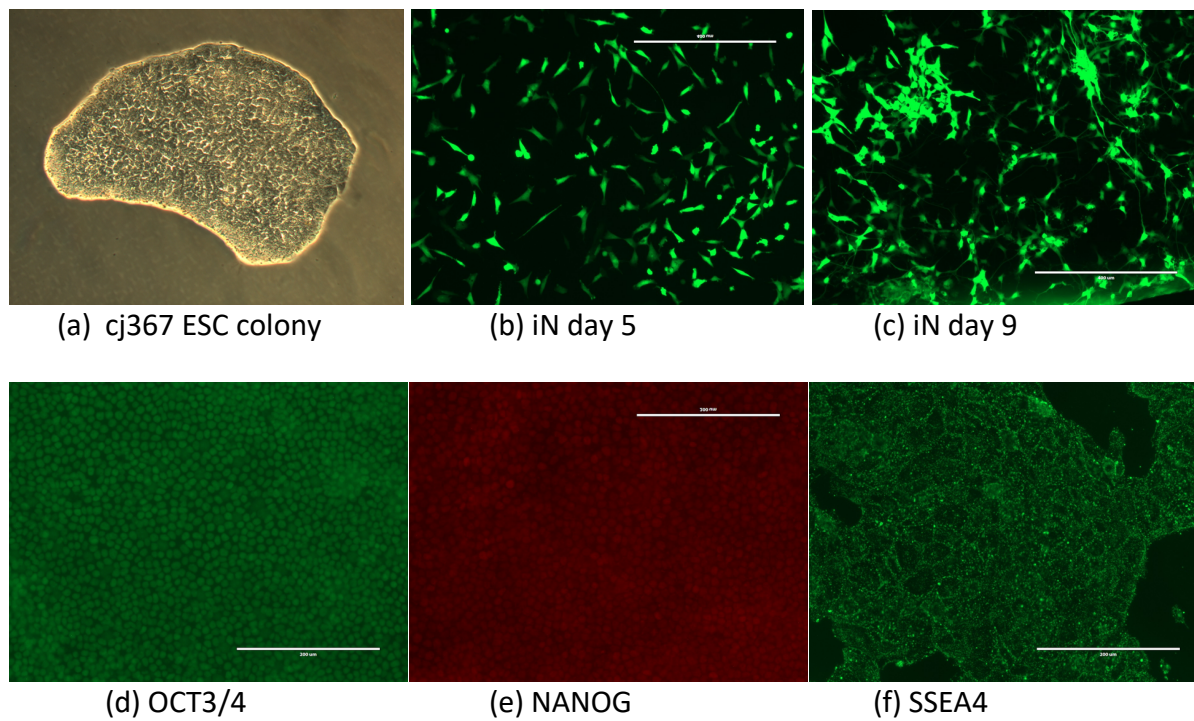
11 **Figure 3.** (a-c) Three examples of marmoset ESCs transfected with dCas9-GFP that  
12 are targeted telomeric sequences revealing telomeres as fluorescent puncta (white  
13 arrowheads). Images are maximum projections of the nuclear volume and the dashed  
14 outlines indicate the nuclear boundaries. Scale bar, 5 μm.

15 **Figure S1.** Violin plot of size distribution of haplotype blocks in cj367 by chromosomes

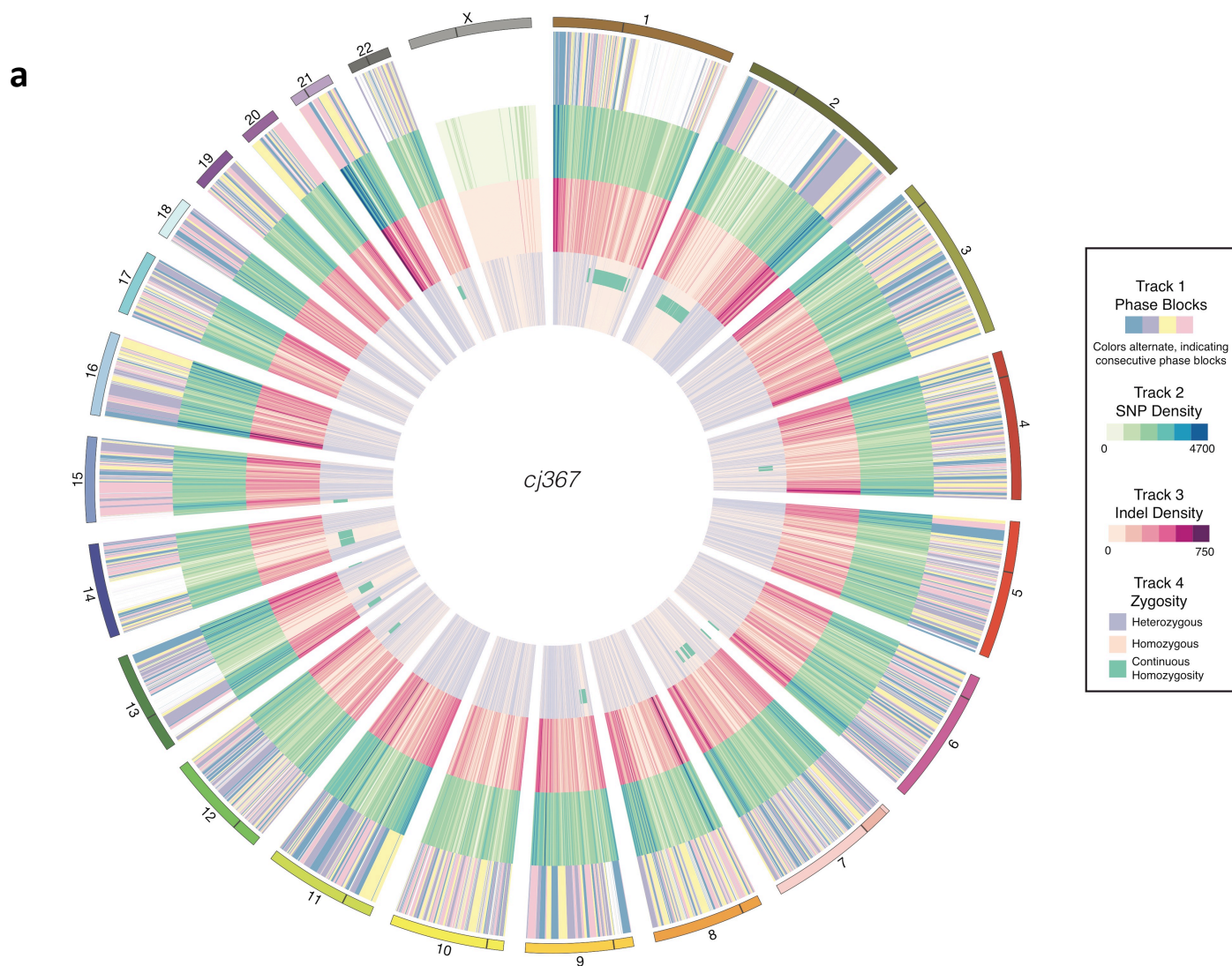
16 **Table 1.** Summary of cj367 SNVs and Indels

17 **Table S1.** cj367 genomic regions exhibiting continuous homozygosity.

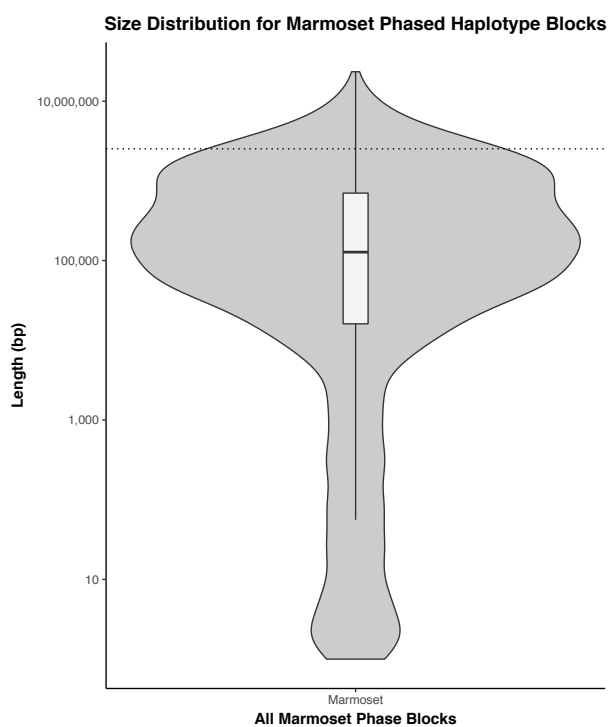
18 **Table S2.** List of CRISPR sites suitable for allele-specific targeting in cj367.

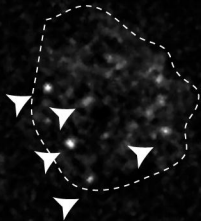
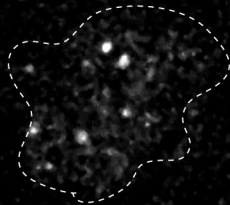
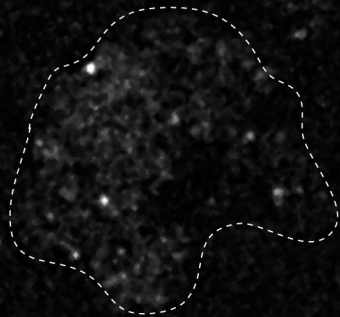


bar = 400 um (b, c); bar = 200 um (d-f).



**b**



**a** dCas9-GFP**b** dCas9-GFP**c** dCas9-GFP

# Size Distribution for Marmoset Phased Haplotype Blocks

