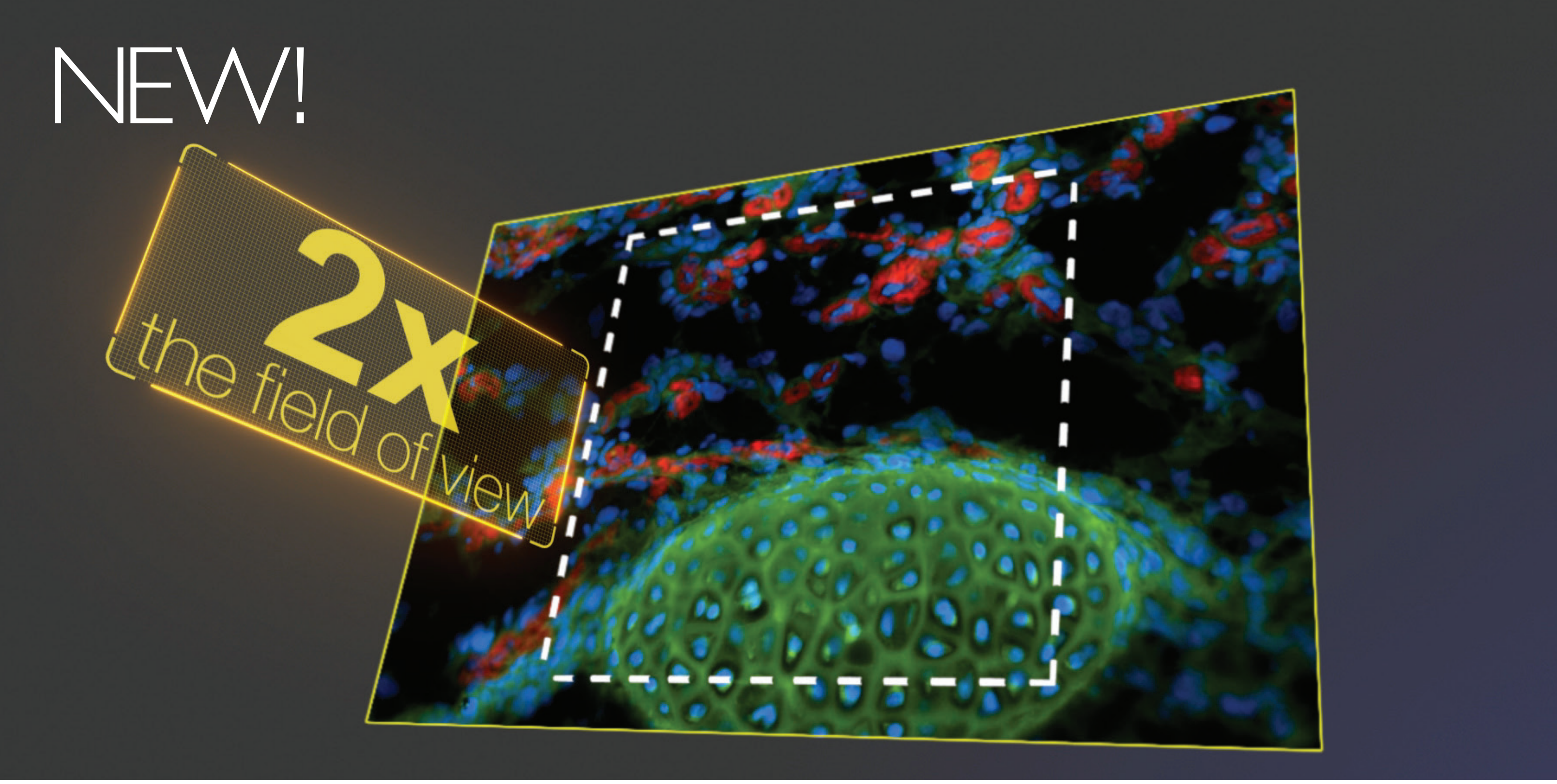
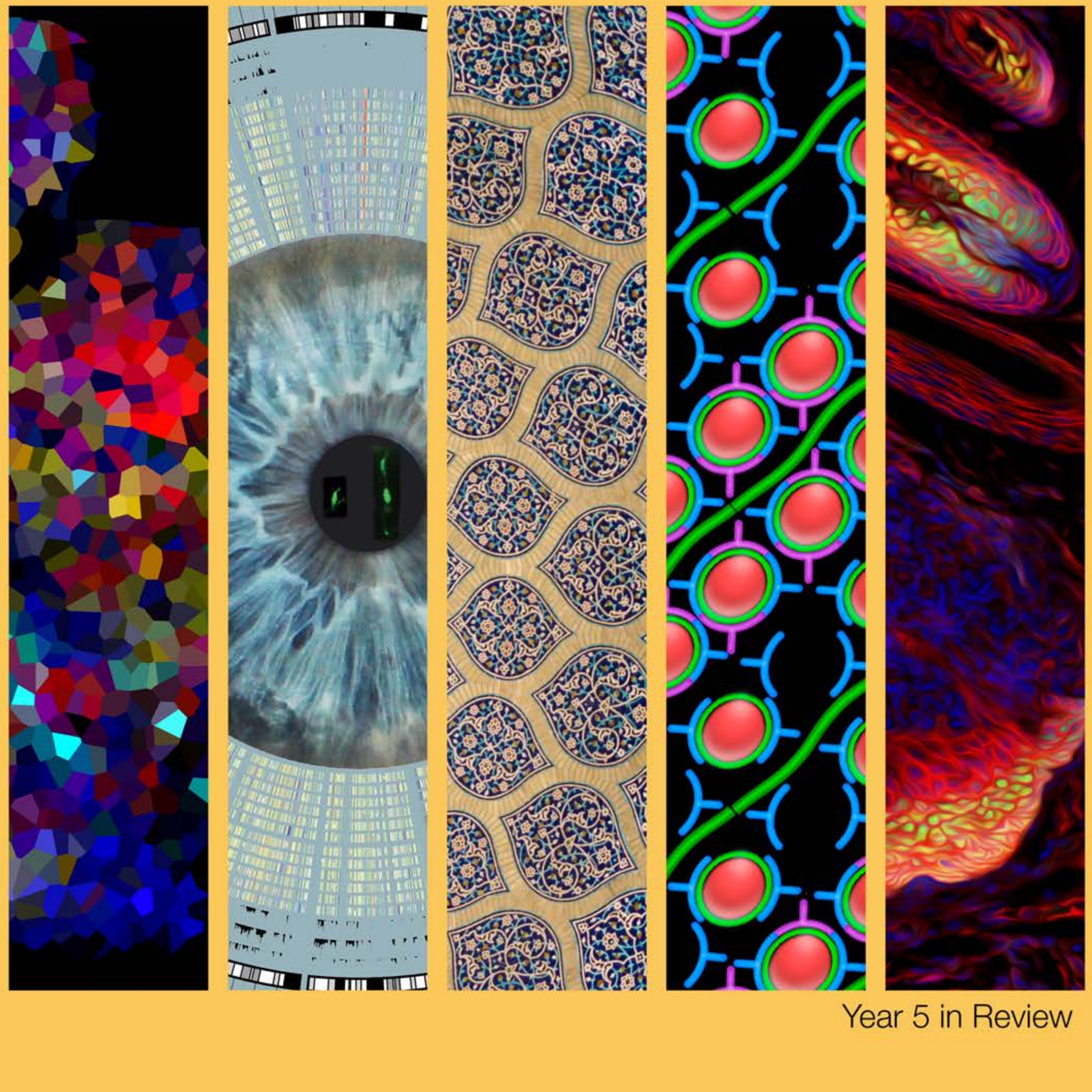

NEW!
2x
the field of view

Cell Reports

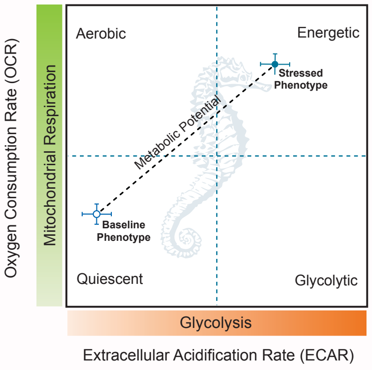
Best of 2016



Year 5 in Review

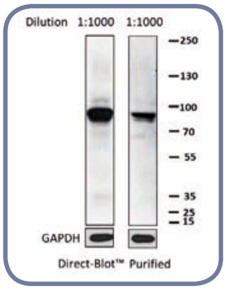
Cell Press
www.cell.com
CELLBOREP2016


Seahorse Bioscience
XFp


Oxygen Consumption Rate (OCR)
Mitochondrial Respiration
Aerobic
Energetic
Stressed Phenotype
Metabolic Potential
Baseline Phenotype
Quiescent
Glycolytic
Glycolysis
Extracellular Acidification Rate (ECAR)


QUALIFIED ANTIBODIES | MADE IN THE USA | VALIDATED IN THE USA
OH MY
#Y$&!


Vacuole
Mitochondria
Signaling proteins
Receptor
Nucleus
Rough ER
Ribosomes
Cytoskeletal Filaments
Golgi
Smooth ER
Vacuole
Lysosomes
Centrosome
Mitochondria
Cell Biology


Dilution 1:1000 1:1000
250
130
100
70
55
35
25
15
GAPDH
Direct-Blot™ Purified

Foreword

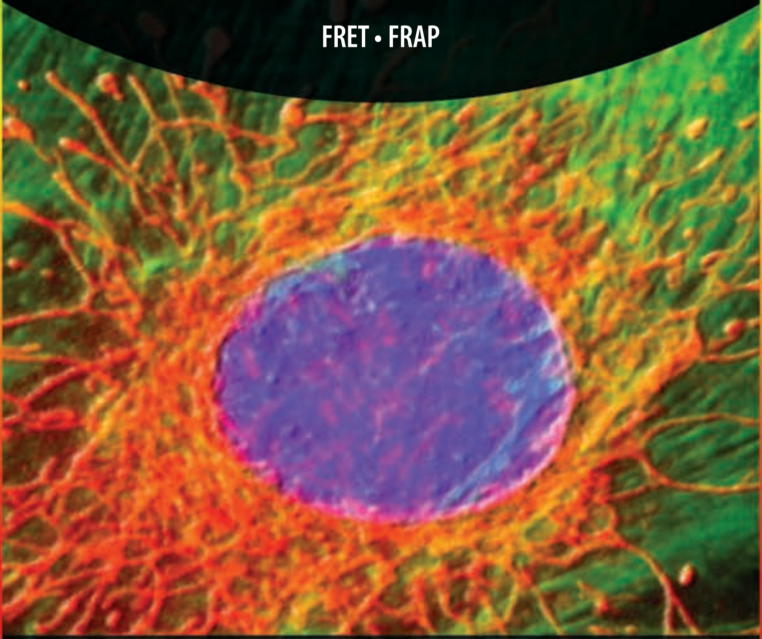
It is with great pleasure that we present the *Best of Cell Reports 2016*. From intrepid analysis of the hibernating bear microbiome to the targeting of HIV-1 with CRIPSR/Cas-9 editing, we published cutting-edge papers from across the life sciences. This made it extraordinarily difficult to pull together 11 papers from the annals of the year, and there were tough choices to make. The final content in this supplement summarizes the breadth of biology covered by *Cell Reports* and was chosen on the basis of reader downloads, Altmetric scores, and citation information. We invite you to browse www.cell.com/cell-reports/home to explore the full range of science we publish, in your field and beyond, with all papers immediately available to readers upon publication.

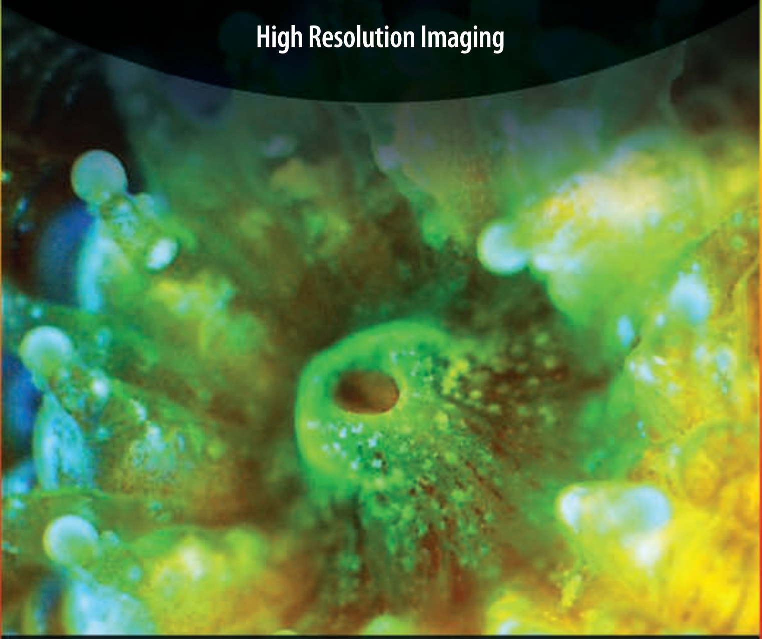
We are really looking forward to 2017, as we reach 5 years of publishing great science. And we would like to thank the scientists who are the authors, reviewers, readers, and editorial and advisory board members that make *Cell Reports* possible. With an expanding editorial and production team internally, and editorial board externally, 2017 holds much promise. But as the year closes, first sit back and enjoy this collection of some of our best papers.

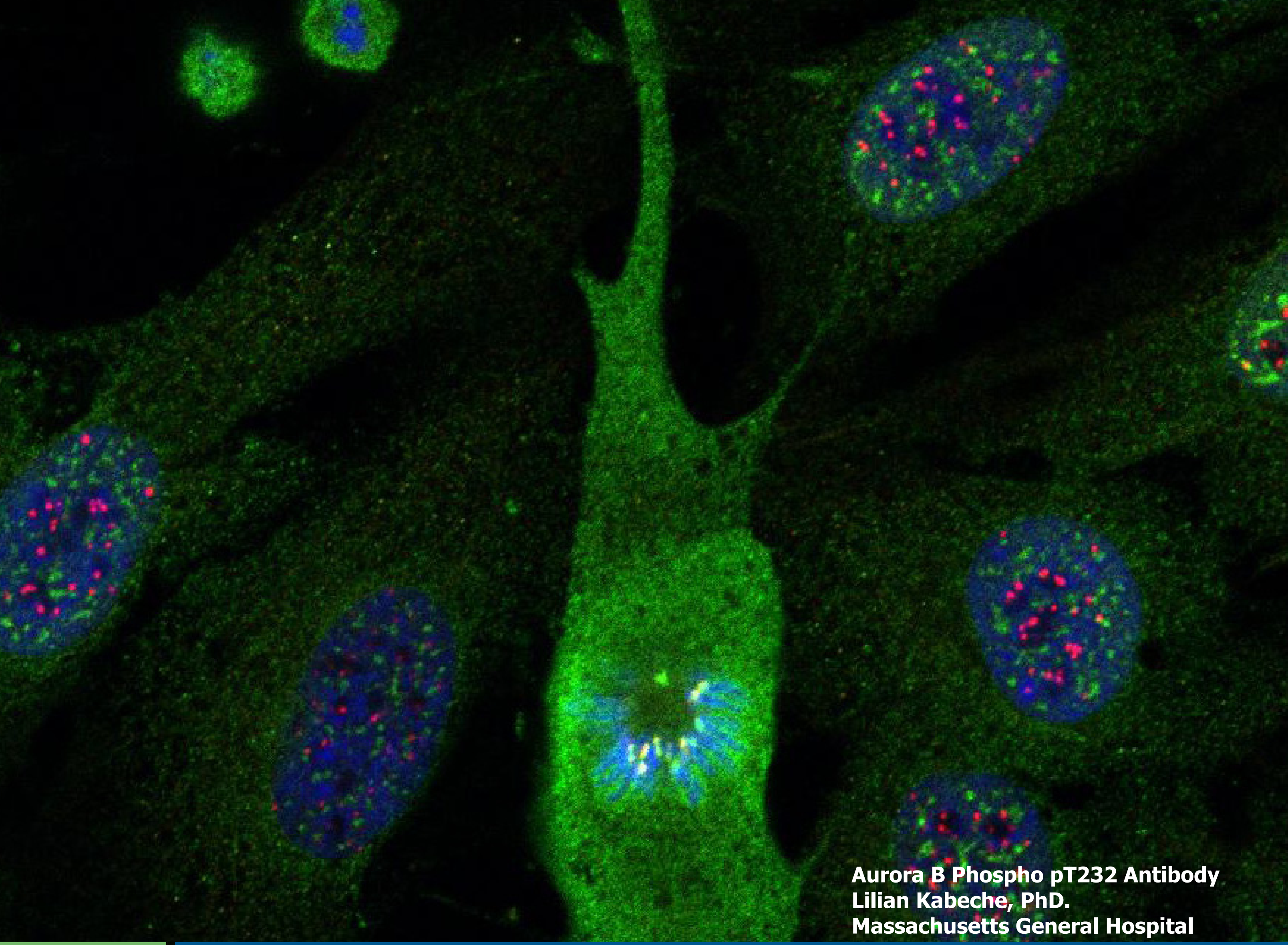
Finally, we are grateful for the generosity of our sponsors, who helped to make this reprint collection possible.

Cell Reports

For information for the Best of Series, please contact:
Jonathan Christison
Program Director, Best of Cell Press
e: jchristison@cell.com
p: 617-397-2893
t: @cellPressBiz







Cell Reports

Best of 2016

Reports

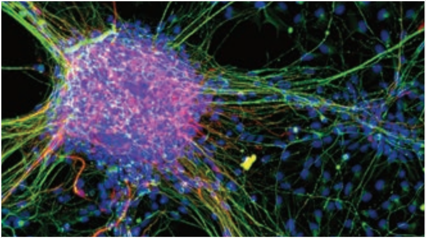
CRISPR/Cas9-Derived Mutations Both Inhibit HIV-1 Replication and Accelerate Viral Escape	*Zhen Wang, Qinghua Pan, Patrick Gendron, Weijun Zhu, Fei Guo, Shan Cen, Mark A. Wainberg, and Chen Liang*
Genetic Isolation of Hypothalamic Neurons that Regulate Context-Specific Male Social Behavior	*Marta E. Soden, Samara M. Miller, Lauren M. Burgeno, Paul E.M. Phillips, Thomas S. Hnasko, and Larry S. Zweifel*
The Adipose Transcriptional Response to Insulin Is Determined by Obesity, Not Insulin Sensitivity	*Mikael Rydén, Olga Hrydziuszko, Enrichetta Mileti, Amitha Raman, Jette Bornholdt, Mette Boyd, Eva Toft, Veronica Qvist, Erik Näslund, Anders Thorell, Daniel P. Andersson, Ingrid Dahlman, Hui Gao, Albin Sandelin, Carsten O. Daub, and Peter Arner*

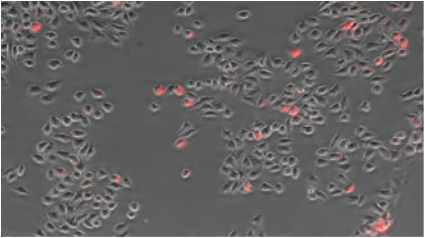
Articles and Resources

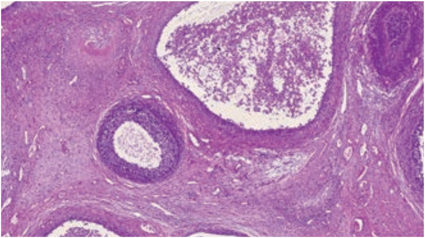
Parent-of-Origin DNA Methylation Dynamics during Mouse Development	*Yonatan Stelzer, Hao Wu, Yuelin Song, Chikdu S. Shivalila, Styliani Markoulaki, and Rudolf Jaenisch*
The Gut Microbiota Modulates Energy Metabolism in the Hibernating Brown Bear *Ursus arctos*	*Felix Sommer, Marcus Ståhlman, Olga Ilkayeva, Jon M. Arnemo, Jonas Kindberg, Johan Josefsson, Christopher B. Newgard, Ole Fröbert, and Fredrik Bäckhed*
Zika Virus Infection in Mice Causes Panuveitis with Shedding of Virus in Tears	*Jonathan J. Miner, Abdoulaye Sene, Justin M. Richner, Amber M. Smith, Andrea Santeford, Norimitsu Ban, James Weger-Lucarelli, Francesca Manzella, Claudia Rückert, Jennifer Govero, Kevin K. Noguchi, Gregory D. Ebel, Michael S. Diamond, and Rajendra S. Apte*
Comprehensive RNA Polymerase II Interactomes Reveal Distinct and Varied Roles for Each Phospho-CTD Residue	*Kevin M. Harlen, Kristine L. Trotta, Erin E. Smith, Mohammad M. Mosaheb, Stephen M. Fuchs, and L. Stirling Churchman*
Macrophage Ontogeny Underlies Differences in Tumor-Specific Education in Brain Malignancies	*Robert L. Bowman, Florian Klemm, Leila Akkari, Stephanie M. Pyonteck, Lisa Sevenich, Daniela F. Quail, Surajit Dhara, Kenishana Simpson, Eric E. Gardner, Christine A. Iacobuzio-Donahue, Cameron W. Brennan, Viviane Tabar, Philip H. Gutin, and Johanna A. Joyce*

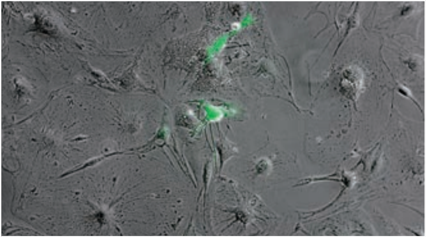
(continued)

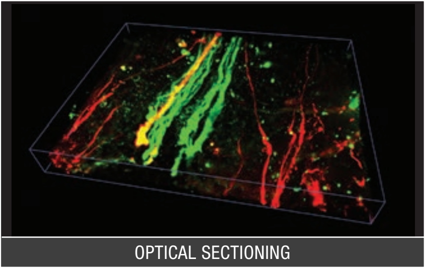

CONVENTIONAL SETUP
2015 R&D 100 FINALIST
BZ-X700
ALL-IN-ONE
FLUORESCENCE
MICROSCOPE

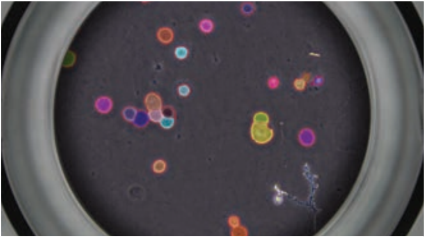










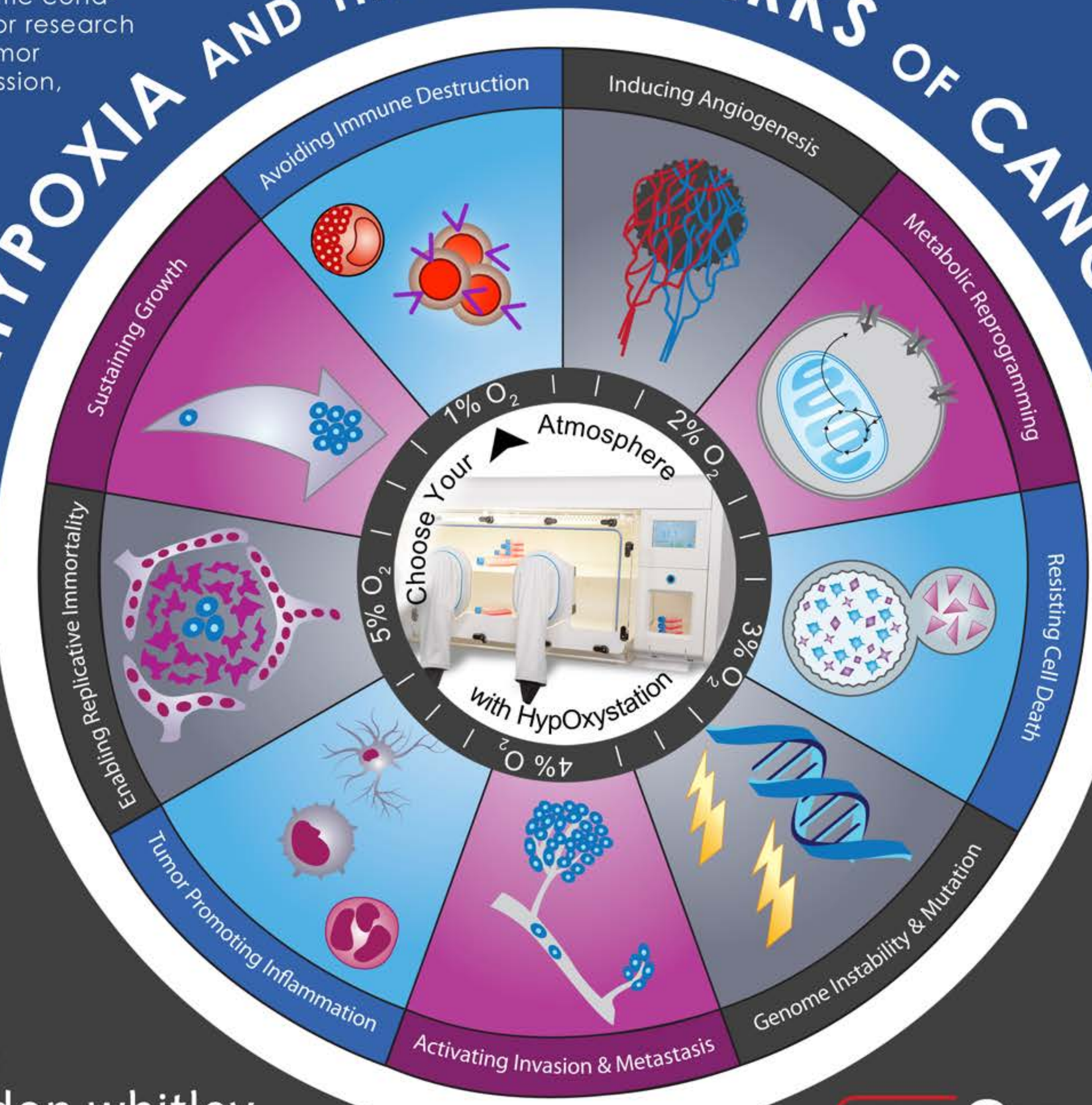


The iconic "Hallmarks of Cancer" first described by Douglas Hanahan and Robert Weinberg in 2000 continue to develop explosively as efforts to illuminate the "acquired capabilities of cancer" evolve. Hypoxia in the fast-growing, poorly perfused tumor setting is one of the main factors determining survival of cancer cells and driving mutagenesis and metastasis. Oxygen in the tumor microenvironment and within the cells is central to metabolism, immunology, epigenetics and therapy resistance of all the cancers.

Physiological low oxygen and hypoxia in cell culture exert effects on metabolism, maintenance, cell yield, and cell survival. Cell culture in the **HypOxystation** establishes authentic conditions for research into tumor progression, differentiation and self-renewal of cancer stem cells, response to chemotherapeutic drugs and other events which determine malignancy and outcome of cancer diseases. The **HypOxystation** enables glove-less access to cultivate and manipulate cells under physiological and hypoxic conditions, in a HEPA-clean environment.

Wigerup and Pahlman, **HypOxystation** users in Sweden, recently published a brilliant review on "Therapeutic Targeting of Hypoxia and Hypoxia-Inducible Factors in Cancer" (Pharm & Ther. 2016, 164: 152-169). They describe how "tumor hypoxia affects most of the Cancer Hallmarks" and give an overview of strategies to target the hypoxic phenotype to eradicate cancer.

HYPOXIA AND THE HALLMARKS OF CANCER



don whitley scientific

www.dwscientific.co.uk

HYPOXYGEN

Define your environment

www.HypOxygen.com

Cell Reports
Report

CRISPR/Cas9-Derived Mutations Both Inhibit HIV-1 Replication and Accelerate Viral Escape

Zhen Wang,[1,2] Qinghua Pan,[1] Patrick Gendron,[3] Weijun Zhu,[4] Fei Guo,[4] Shan Cen,[5] Mark A. Wainberg,[1,2] and Chen Liang[1,2,*]
[1]McGill University AIDS Centre, Lady Davis Institute, Jewish General Hospital, Montreal H3T 1E2, Canada
[2]Departments of Medicine, Microbiology and Immunology, McGill University, Montreal H3A 2B4, Canada
[3]Institute for Research in Immunology and Cancer, University of Montreal, Montreal H3C 3J7, Canada
[4]MOH Key Laboratory of Systems Biology of Pathogens and Center for AIDS Research, Institute of Pathogen Biology, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing 100730, China
[5]Institute of Medicinal Biotechnology, Chinese Academy of Medical Sciences, Beijing 100050, China
*Correspondence: chen.liang@mcgill.ca
http://dx.doi.org/10.1016/j.celrep.2016.03.042

SUMMARY

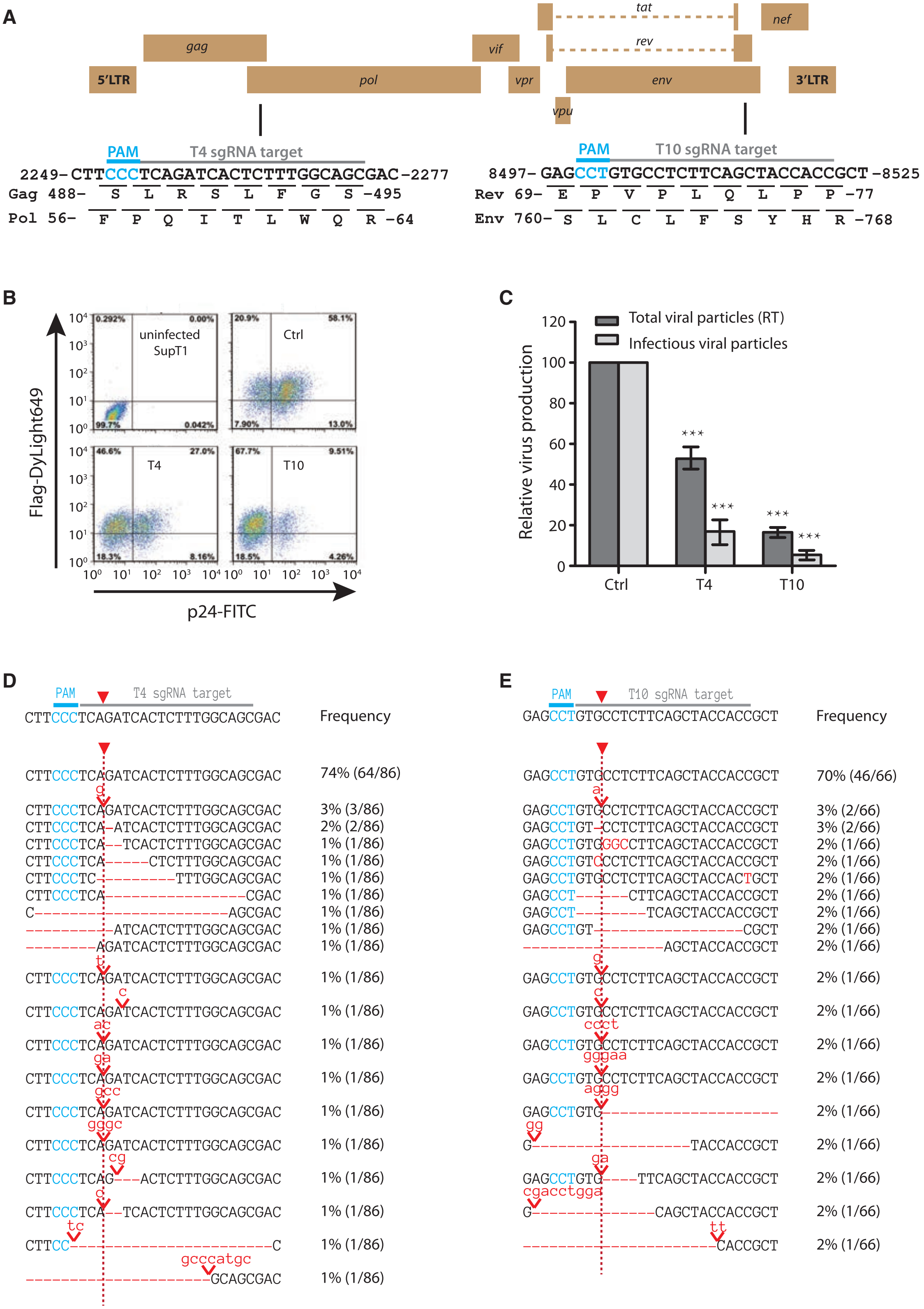
Cas9 cleaves specific DNA sequences with the assistance of a programmable single guide RNA (sgRNA). Repairing this broken DNA by the cell's error-prone non-homologous end joining (NHEJ) machinery leads to insertions and deletions (indels) that often impair DNA function. Using HIV-1, we have now demonstrated that many of these indels are indeed lethal for the virus, but that others lead to the emergence of replication competent viruses that are resistant to Cas9/sgRNA. This unexpected contribution of Cas9 to the development of viral resistance is facilitated by some indels that are not deleterious for viral replication, but that are refractory to recognition by the same sgRNA as a result of changing the target DNA sequences. This observation illustrates two opposite outcomes of Cas9/sgRNA action, i.e., inactivation of HIV-1 and acceleration of viral escape, thereby potentially limiting the use of Cas9/sgRNA in HIV-1 therapy.

INTRODUCTION

CRISPR (clustered regularly interspaced short palindromic repeats)/Cas9 was originally discovered in bacteria as a mechanism of adaptive immunity against infection by bacteriophages (Barrangou et al., 2007; Bolotin et al., 2005; Mojica et al., 2005; Pourcel et al., 2005). This machinery utilizes RNA as a guide to target the Cas9 endonuclease to the bacteriophage DNA that is subsequently cleaved and degraded (Barrangou et al., 2007). Successful adaptation of the CRISPR/Cas9 system to mammalian cells marks a new era of genome engineering (Cho et al., 2013; Cong et al., 2013; Jinek et al., 2013; Mali et al., 2013). By modifying the sequence of the single guide RNA (sgRNA) that associates with Cas9, it may be possible to target and edit virtually any DNA sequence.

The 20-nucleotide sequence at the 5′ end of sgRNA binds to target DNA through Watson-Crick base-pairing. In addition to this RNA-DNA hybrid-mediated recognition, Cas9 also needs to detect a multi-nucleotide sequence that is adjacent to the target DNA, which is called the protospacer adjacent motif (PAM). After Cas9 cleaves the target DNA at a position three nucleotides from the PAM, the double-stranded DNA break is often repaired by the non-homologous end-joining (NHEJ) machinery (Hsu et al., 2014). NHEJ repair is error-prone and introduces short insertions and deletions (indels) that survive Cas9/sgRNA cleavage and often disturb the function of target DNA, which empowers Cas9/sgRNA to play a role in genome engineering.

Among many applications, Cas9/sgRNA has been used to inactivate the viral DNA genome and thereby inhibit viral infection (Kennedy and Cullen, 2015). A number of viruses have been tested in this regard, including adenovirus (Bi et al., 2014), herpes simplex virus type 1 (HSV-1) (Bi et al., 2014; Suenaga et al., 2014), human papillomavirus (Kennedy et al., 2014, 2015; Zhen et al., 2014), Epstein-Barr virus (EBV) (Wang and Quake, 2014; Yuen et al., 2015), vaccinia virus (Yuan et al., 2015), hepatitis B virus (HBV) (Dong et al., 2015; Karimova et al., 2015; Lin et al., 2014; Ramanan et al., 2015; Seeger and Sohn, 2014; Wang et al., 2015; Zhen et al., 2015), and HIV type 1 (HIV-1) (Dampier et al., 2014; Ebina et al., 2013; Hu et al., 2014; Li et al., 2015; Liao et al., 2015; Zhu et al., 2015). sgRNAs have been designed to target viral promoter DNA or viral genes that are essential for viral replication. Indels that are created by the NHEJ machinery following Cas9 cleavage ablate viral gene function and thus lead to suppression of viral replication. Dual sgRNAs have also been designed to eliminate a fragment of viral DNA (Wang et al., 2015). Taking advantage of the two long terminal repeats (LTRs) at the 5′ and 3′ termini of the integrated HIV-1 DNA, a single sgRNA targeting the LTR has been shown to cause the deletion of nearly full-length HIV-1 DNA (Ebina et al., 2013; Hu et al., 2014). Not only can Cas9/sgRNA inhibit new viral infection, it can also participate in the clearance of latent viral DNA including that of HBV, herpes viruses, and HIV-1 (Hu et al., 2014; Lin et al., 2014; Wang and Quake, 2014). In spite of successful suppression of viral replication in multiple studies, however, it is not known whether viruses might be able to escape from Cas9/sgRNA inhibition. In order to answer this important question, we have utilized HIV-1 as a model and performed viral evolution experiments in CD4+



(legend on next page)

T cells that express Cas9/sgRNA that targets the HIV-1 genome. We now show viral escape from Cas9/sgRNA on the basis of mutations in the viral DNA sequence that is targeted by sgRNA.

RESULTS

HIV-1 Escapes from Suppression Mediated by Cas9/sgRNA

To investigate whether HIV-1 is able to escape from Cas9/sgRNA-mediated inhibition, we first generated CD4+ SupT1 cell lines that stably expressed both Cas9 and sgRNA that we previously showed could inhibit HIV-1 production in transient transfection experiments (Zhu et al., 2015). These Cas9 and sgRNA genes were stably transduced into SupT1 cells using a lentiviral vector (Sanjana et al., 2014). These Cas9/sgRNA-expressing cells showed growth capacity similar to that of the control cells (Figure S1A). The T4 sgRNA targets the overlapping open reading frames (ORFs) of HIV-1 *gag/pol* genes, while T10 targets the overlapping ORFs of HIV-1 *env/rev* genes (Figure 1A). Both viral targets are very conserved in HIV-1 sequences that are registered in the HIV database (Figure S1B). Since each of these two sgRNAs targets two specific viral genes, we conjectured that the genetic barrier should be high for HIV-1 to mutate and escape from inhibition. A control SupT1 cell line expressed Cas9 only.

We first tested these SupT1 cell lines by exposing them to the NL4-3 HIV-1 strain for a short term of infection. The results showed that T4 or T10 sgRNA together with Cas9 reduced the number of HIV-1 infected cells (Figure 1B) and diminished the production of infectious viruses (Figure 1C). To demonstrate that these reductions had resulted from the action of Cas9/sgRNA that causes indels, we extracted total cellular DNA from the infected cells, amplified the viral DNA region that was targeted by the T4 or T10 sgRNA, cloned the PCR products, and sequenced the DNA clones. Although no mutations were detected in the targeted viral DNA that was extracted from the infected control SupT1 cells, rich arrays of indels were identified in viral DNA from the infected SupT1 cells that expressed T4 or T10 sgRNA (Figures 1D and 1E). The percentages of indels for the T4 and T10 sgRNAs were approximately 25% and 30%, respectively. We also tested a number of these indels by inserting them into the HIV-1 DNA and observed that the majority of them abolished the production of infectious viruses in addition to the two substitution mutations that produced as much infectious viruses as the wild-type viral DNA did (Figure S1C). In addition to the NL4-3 HIV-1 strain, we further tested the T4 sgRNA against two primary HIV-1 isolates 89.6 and YU-2, as well as three transmitted founder viruses CH040, CH077, and CH106. The results showed that Cas9/T4 sgRNA caused indels in these viral DNA and strongly inhibited the production of each of these latter viruses (Figures S1D and S1E). Together, these results confirm that Cas9/sgRNA inhibits HIV-1 infection by introducing various mutations into viral DNA.

We next performed HIV-1 evolution experiments and monitored viral growth over prolonged times by measuring viral reverse transcriptase (RT) activity in culture supernatants. The results showed that HIV-1 replication was delayed in SupT1 cells expressing T4 or T10 sgRNA compared to viral replication in control SupT1 cells (Figure 2A). Nonetheless, viral production eventually peaked in the T4 and T10 SupT1 cells, showing that HIV-1 had escaped from suppression by Cas9/sgRNA. To further demonstrate viral escape, we collected viruses at the peaks of viral RT levels in the control, T4, and T10 cells, and then utilized the same RT levels of each virus to infect the corresponding SupT1 cell line. The T4 and T10 viruses displayed even moderately faster replication kinetics than the control virus in this second round of replication (Figure 2B), which suggests that the escape viruses might have gained mutations that improve viral infectivity.

The Cas9/sgRNA-Resistant HIV-1 Bears Mutations in the Viral DNA Region that Is Targeted by sgRNA

To identify the mutations that led to viral escape, we extracted RNA from relevant viruses, amplified the sgRNA target sequence in the viral genome by RT-PCR, cloned the PCR products, and sequenced the DNA clones. No mutations were found in the genome of viruses that were produced by the control SupT1 cells. In contrast, five different mutations were identified in the 47 DNA clones that were sequenced for the T4 escape virus (Figure 2C), with a dominant substitution being G2258A (81%) that causes a R490K change in Gag. Less frequent mutations include A2259G (13%), C2256A/A2257C/A2259G (2%), C2252T/G2258A (2%), and a 2258-ACCCCCC-2263 (2%) insertion. It is noted that all mutations are located adjacent to where Cas9 cleaves the target DNA, i.e., 3 nt from the PAM (Figure 2C), which suggests that these mutations are likely the result of Cas9 activity. In the case of T10 sgRNA, 19 different mutations were identified in 42 DNA clones that were sequenced (Figure 2D),

Figure 1. Cas9/sgRNA Inhibits HIV-1 Production

(A) Illustration of the T4 and T10 sgRNAs that target HIV-1 genes. The positions of the nucleotides refer to the HIV-1(NL4-3) proviral DNA. PAM, protospacer adjacent motif.

(B) T4 and T10 sgRNAs diminished the number of SupT1 cells that were infected by HIV-1 over 36 hr. Cas9-FLAG and viral p24 were stained with anti-FLAG and anti-p24 antibodies and positive cells were scored by flow cytometry. The controls (Ctrl) are cells that expressed Cas9-FLAG alone. The uninfected SupT1 cells served as a negative control. The representative results of three independent experiments are shown.

(C) Levels of viruses that were produced by SupT1 cells after a 36 hr infection by HIV-1. The total amounts of viruses were determined by measuring levels of viral RT activity in the supernatants. The levels of infectious HIV-1 in the culture supernatants were determined by infecting the TZM-bl indicator cells. The results shown are the average of three independent experiments. The levels of viruses that were produced by the control cells were arbitrarily set as 100 (***$p < 0.0001$).

(D) Indels in viral DNA from T4 sgRNA-expressing SupT1 cells that were infected with HIV-1(NL4-3) for 36 hr. 86 DNA clones were sequenced. The frequencies of wild-type and mutated sequences are shown.

(E) Indels in the T10 sgRNA target site from the infected T10 sgRNA-expressing SupT1 cells. 66 DNA clones were sequenced. The red arrow indicates the cleavage site of Cas9 in the sgRNA target.

See also Figure S1.

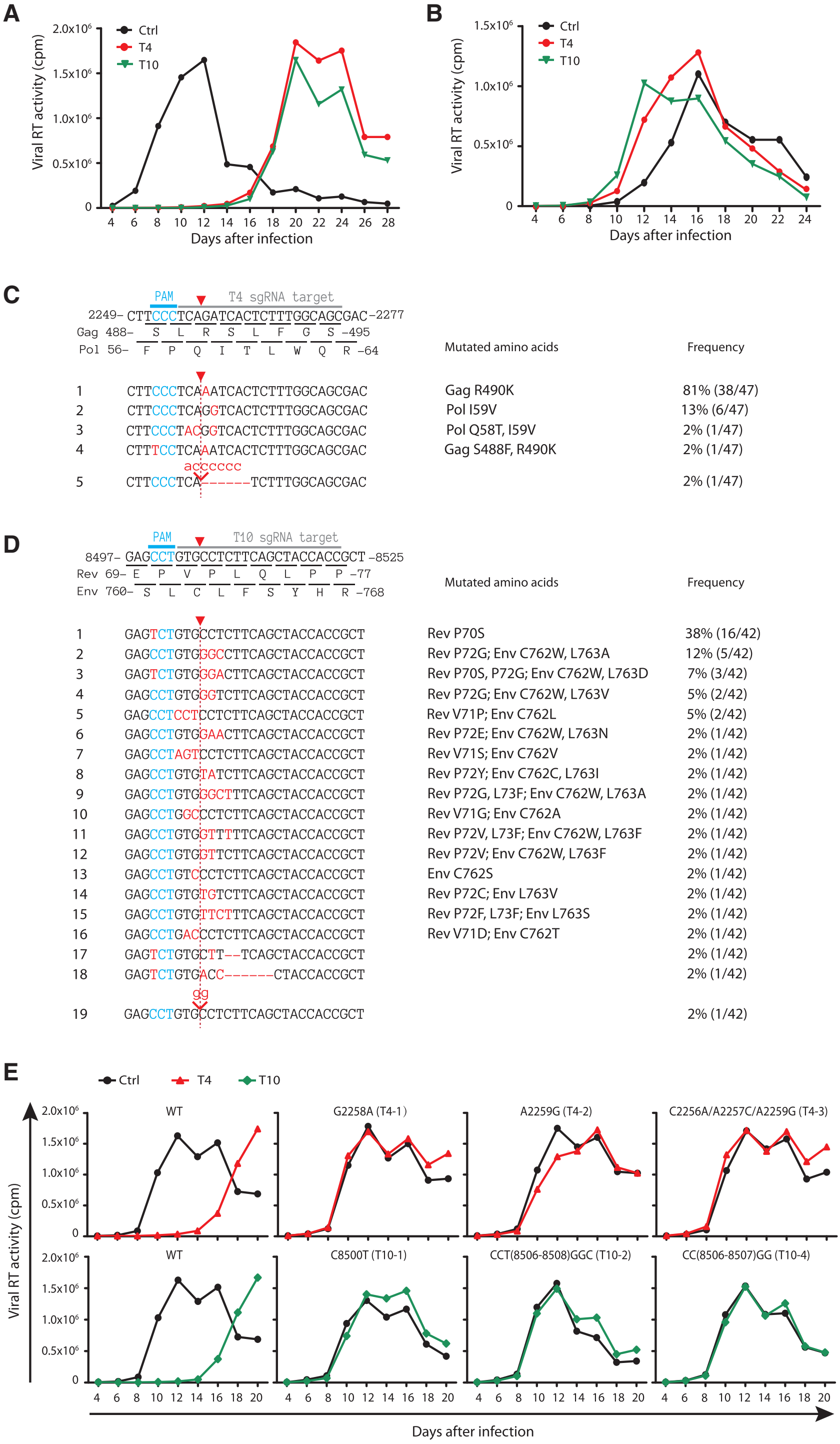


Figure 2. HIV-1 Escapes from Cas9/sgRNA Inhibition

(A) Replication of HIV-1(NL4-3) in SupT1 cell lines that express T4 or T10 sgRNA together with Cas9. The control is SupT1 cells that express Cas9 only. The levels of viruses in the culture supernatants were monitored by measuring viral RT activity over time.

(B) Viruses that were collected at the peaks of RT values shown in (A) were used to infect the corresponding cell lines at the same RT values, i.e., viruses that were produced from T4 SupT1 cells were used to re-infect the same fresh T4 SupT1 cells. The viral replication was monitored by measuring viral RT activity in the culture supernatants. The representative results of two independent infection experiments are shown.

(C) Mutations in the escape virus DNA that is targeted by the T4 sgRNA. The T4 sgRNA-targeted viral DNA was amplified by RT-PCR and cloned into a plasmid vector. 47 DNA clones were sequenced. The frequency of each mutation is shown. The cleavage site in the target region by Cas9 is indicated by a red arrow. The mutated nucleotides are highlighted in red. The changed amino acids in Gag and/or Pol are also listed.

(D) Mutations in the T10 sgRNA target site in the T10 escape viruses.

(E) Replication of the wild-type or mutated HIV-1 in the control, T4, or T10 SupT1 cell lines. Three mutations that were identified in the T4 escape viruses were tested for resistance to T4 sgRNA. They were G2258A (T4-1 in C), A2259G (T4-2), and C2256A/A2257C/A2259G (T4-3). Three mutations in the T10 escape viruses were also tested for resistance to T10 sgRNA. They were C8500T (T10-1 in D), CCT(8506-8508)GGC (T10-2), and CC(8506-8507)GG (T10-4). The virus production was monitored by measuring levels of viral RT activity in culture supernatants over different time intervals.

See also Figure S2.

with a dominant mutation C8500T present in 38% of DNA clones. Other mutations were clustered at the Cas9 cleavage site in the target DNA. It is noted that three mutations (i.e., 17, 18, and 19) contain small indels that cause frameshifts in the *env* and *rev* genes and are expected to impair the generation of functional Env and Rev proteins (Figure 2D).

We next tested which of these mutations confer resistance to Cas9/sgRNA by inserting them into the parental HIV-1 DNA. These viral mutants were as infectious as the wild-type virus (Figure S2A). We used the same levels of mutated viruses to infect control or corresponding sgRNA-expressing SupT1 cells. In one-round infection experiments, similar levels of the HIV-1 mutants were produced from the sgRNA-expressing SupT1 cells and control cells, as opposed to the marked inhibition of wild-type HIV-1 by the sgRNA (Figure S2B). This resistant phenotype was confirmed by documenting similar replication kinetics of these viral mutants in the sgRNA-expressing cells and in control cells (Figure 2E). Together, these results demonstrate that HIV-1 is able to escape from Cas9/sgRNA suppression by accumulating mutations in the PAM/sgRNA target sequence.

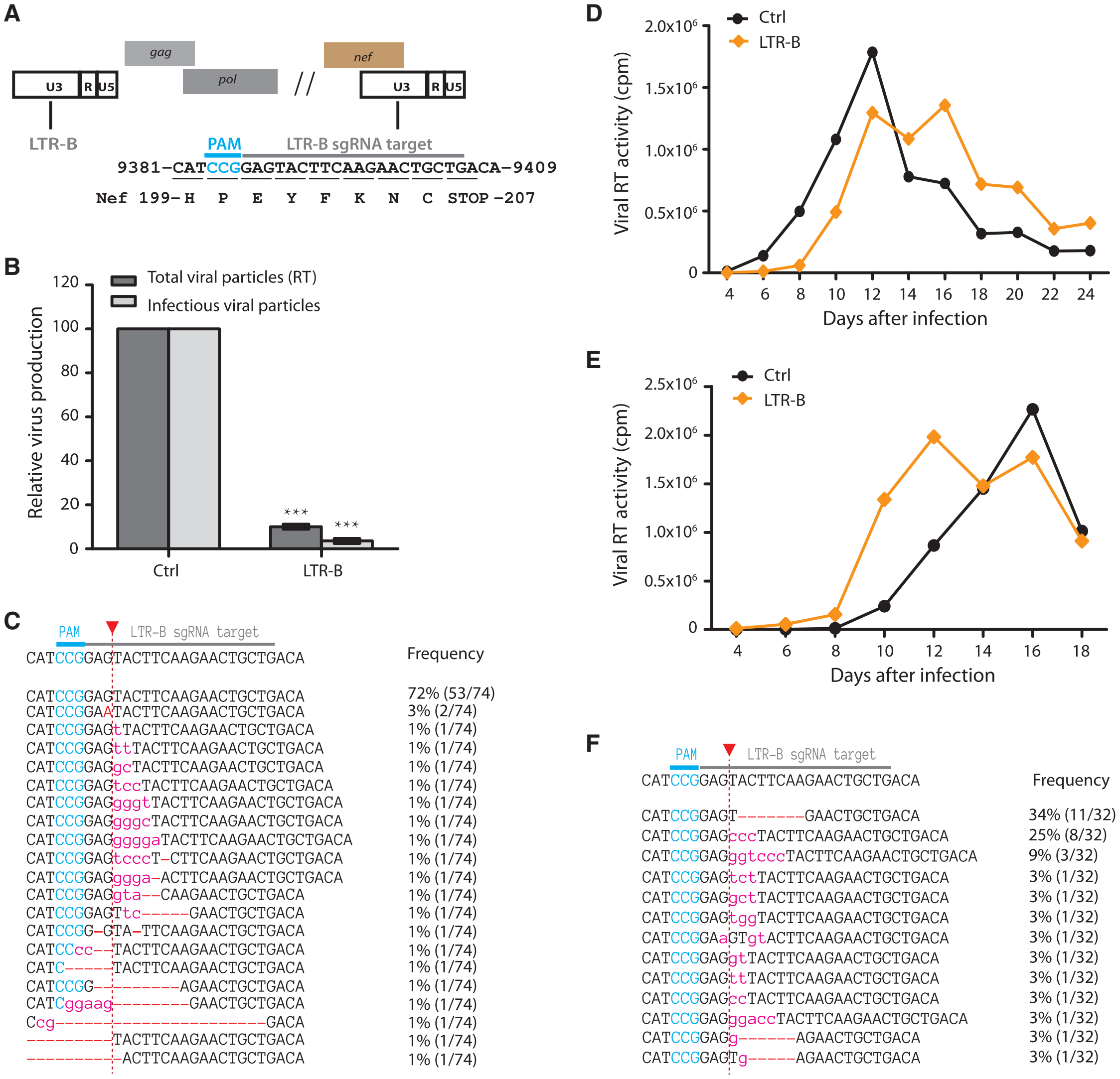


Figure 3. HIV-1 Escapes from the LTR-B sgRNA that Targets the U3 Region of HIV-1 LTR

(A) Illustration of the target site of LTR-B sgRNA. Both the 5′ and 3′ U3 regions in provirus are targeted.

(B) Inhibition of HIV-1 production by Cas9/LTR-B sgRNA. The levels of viruses were determined either by measuring viral RT activity or by infecting the TZM-bl indicator cells (***$p < 0.0001$).

(C) Indels in HIV-1 DNA that were caused by Cas9/LTR-B sgRNA in the infected SupT1 cells. The frequency of each indel is shown.

(D) Replication of HIV-1 in the control and LTR-B SupT1 cell lines. The virus production was monitored by measuring viral RT activity in the culture supernatants.

(E) Replication of the escape viruses. The viruses that were collected at the RT peaks in (D) were used to re-infect the corresponding cell lines. The viruses of the same RT amounts were used to initiate the infection.

(F) Mutations in the LTR-B escape viruses. The LTR-B sgRNA target region in the escape viruses was amplified by RT-PCR and sequenced. The frequency of each mutation is shown.

See also Figure S3.

HIV-1 Escapes from sgRNA that Targets the U3 Region

In addition to ablating HIV-1 genes, sgRNAs have been designed to target the HIV-1 U3 region (Hu et al., 2014). This latter type of sgRNA together with Cas9 leads to cleavage at the two U3 regions that are located at the 5′ and 3′ ends of HIV-1 proviral DNA and thereby results in removal of the viral genome that is integrated into cellular DNA. Deletion of the entire viral genome should preclude HIV-1 escape from sgRNA-mediated inhibition. To test this, we generated SupT1 cell lines that stably expressed Cas9 and the LTR-B sgRNA that had been designed to target the HIV-1 U3 region (Figure 3A) (Hu et al., 2014). The results of one-round infection experiments showed strong inhibition of HIV-1 production from LTR-B SupT1 cells (Figure 3B) and that Cas9/LTR-B led to indels in the U3 region of HIV-1 DNA (Figure 3C). Although Cas9/LTR-B initially suppressed HIV-1 replication (Figure 3D), viral escape was subsequently observed (Figure 3E). Sequencing the genome of the escape viruses revealed a variety of indels in the U3 region that is targeted by LTR-B (Figure 3F). In contrast to the T4 and T10 escape mutations that are nucleotide substitutions and thus maintain the ORFs of the targeted viral genes, the LTR-B escape viruses contained indels that are reminiscent of indels caused by the cleavage of Cas9. This difference is mainly because LTR-B targets a sequence within U3 that is not essential for efficient HIV-1 replication in SupT1 cells.

In order to further demonstrate that the HIV-1 escapes from Cas9/sgRNA inhibition is not limited to specific sgRNAs, we tested five more sgRNAs that target different regions in the viral *env* gene. These include the EnvT1, EnvV2, EnvT3, EnvV4, and EnvV5 (Figure S3A), among which the EnvT1 and EnvT3 sgRNAs have previously been shown to inhibit HIV-1 infection (Liao et al., 2015). The results of viral evolution experiments showed that HIV-1 escaped from each of these five sgRNAs (Figure S3B). It was noted that HIV-1 escaped from some sgRNAs (such as EnvT1 and EnvV4) faster than from others (such as EnvT3, EnvV2, and EnvV5). These different rates of viral escape from different sgRNAs can be a result of different degrees of conservation of the target sequences, as suggested by a recent study (Wang et al., 2016), or a result of different efficiencies of Cas9 cleavage at different target sites. Mutations were identified in the sgRNA target region in the genome of the escape viruses (Figure S3C). Together, these data support the conclusion that HIV-1 is able to escape from Cas9/sgRNA inhibition as a result of mutations in viral DNA sequence that is targeted by sgRNA.

Cas9/sgRNA-Derived Indels Represent One Source of HIV-1 Escape Mutations

The location of some of the escape mutations to the cleavage site of Cas9 within the target DNA suggests that these mutations might have been derived from the action of Cas9/sgRNA. To further investigate this possibility, we utilized the Illumina MiSeq technique to achieve a deeper coverage of the indels that are caused by Cas9/sgRNA in HIV-1 DNA. To this end, Cas9/sgRNA-expressing SupT1 cells were infected by HIV-1 for 36 hr, which permits one round of HIV-1 infection and minimizes re-infection (Holmes et al., 2015). The recorded indels in viral DNA within this relatively short time window (36 hr) thus represent a snapshot of the numerous indels that are generated by Cas9/sgRNA in long-term HIV-1 evolution experiments. Given the relatively high error rate of the Illumina MiSeq technique, single nucleotide substitutions were excluded from further analysis if frequency was <5%. We also registered mutations in the escape viruses using the MiSeq technique and compared the data to indels from the transiently infected SupT1 cells. Our expectation was that some of the mutations in the escape viruses would be attributable to indels that were introduced by Cas9/sgRNA into viral DNA, and that enrichment of some indels in the escape viral population would be observed if these indels had been selected by the virus for escape.

In the infected control SupT1 cells that did not express sgRNA, very few indels were identified in HIV-1 DNA (Table S1). The frequencies of these indels were often <0.1%. Consistently, viruses that had been passaged in the control cells in the evolution experiments contained very few indels, and these indels occurred at very low frequencies (Table S1). These results suggest that neither HIV-1 RT nor the Illumina MiSeq technique introduces frequent indels.

In the transiently infected T4 sgRNA-expressing SupT1 cells, 164 different indels were identified across the PAM/T4 sgRNA target region, with 38 located at the cleavage site of Cas9 (between positions 3 and 4) (Figures 4A and 4B; Table S2). The majority of these indels had frequencies below 0.1%. Twenty-two of the 164 indels (13%) cause amino acid substitutions without disrupting the ORFs of Gag and/or Pol and thus may produce viable viruses. In the T4 escape viruses, two single nucleotide substitutions were identified at frequencies of 80% and 7% (Figure 4B; Table S2). Although these two substitutions were also observed in the viral DNA from the transiently infected T4 SupT1 cells, their very low frequencies in the latter setting suggest that they might be MiSeq errors. Thirty-seven indels were identified in the escape viruses, of which one had a frequency of 7% (Figure 4B; Table S2). Twenty-eight of the 37 (76%) indels cause amino acid substitutions and do not disrupt the ORFs of Gag and/or Pol (Table S2), which shows an enrichment for these potentially viable indels in the escape virus population compared to the 13% in the infected cells. Seven common indels were identified between the T4 escape viruses and the transiently infected T4 SupT1 cells (Table S2). These data suggest that Cas9/T4 sgRNA-derived indels were archived in HIV-1 genomes, even though they did not dominate the population of the escape viruses. This is likely because the single nucleotide mutation G2258A (80%) is much superior in terms of viral fitness and resistance to Cas9/T4 sgRNA.

When we examined the T10 sgRNA indels in the transiently infected T10 SupT1 cells, 207 different indels were identified that disrupt the PAM/T10 target sequence, the majority of which were present at <0.1% frequency (Figure 4C; Table S3). Fifty-seven of these indels were located directly at the Cas9 cleavage site (between positions 3 and 4) and 39 of the 207 (19%) indels cause amino acid substitutions. In the T10 escape viruses, 69 different indels were identified, with three having frequencies of 11%, 14%, and 28%, which shows enrichment of these indels during HIV-1 evolution toward escape (Figure 4C; Table S3). Fifty-five of these 69 indels (80%) cause amino acid substitutions without disrupting the ORFs of the coded genes, which again shows viral selection of potentially viable indels during evolution toward escape from Cas9/sgRNA inhibition. Nineteen common indels were identified between the T10 escape viruses and the transiently infected T10 cells, of which one was present at 11% in the escape virus population (Figure 4C; Table S3).

Last, 285 different indels of low frequency were identified in LTR-B SupT1 cells that were infected by HIV-1 for 36 hr (Figure 4D; Table S4). In the LTR-B escape viruses, the results of MiSeq revealed 136 different indels, of which two were present at frequencies of 17% and 37% (Figure 4D; Table S4). Sixty-five common indels were identified between these two indel groups, including those with frequencies of 17% and 37% (Figure 4D; Table S4). Together, these results indicate that indels that were derived from the action of Cas9/sgRNA in the infected cells were represented in the viral population, and that some of these became enriched during viral evolution and contributed to viral escape from the inhibition by Cas9/sgRNA.

DISCUSSION

Here, we report viral escape from Cas9/sgRNA-mediated inhibition of replication. In some ways, it may not seem surprising that HIV-1 is able to escape from Cas9/sgRNA since HIV-1 can

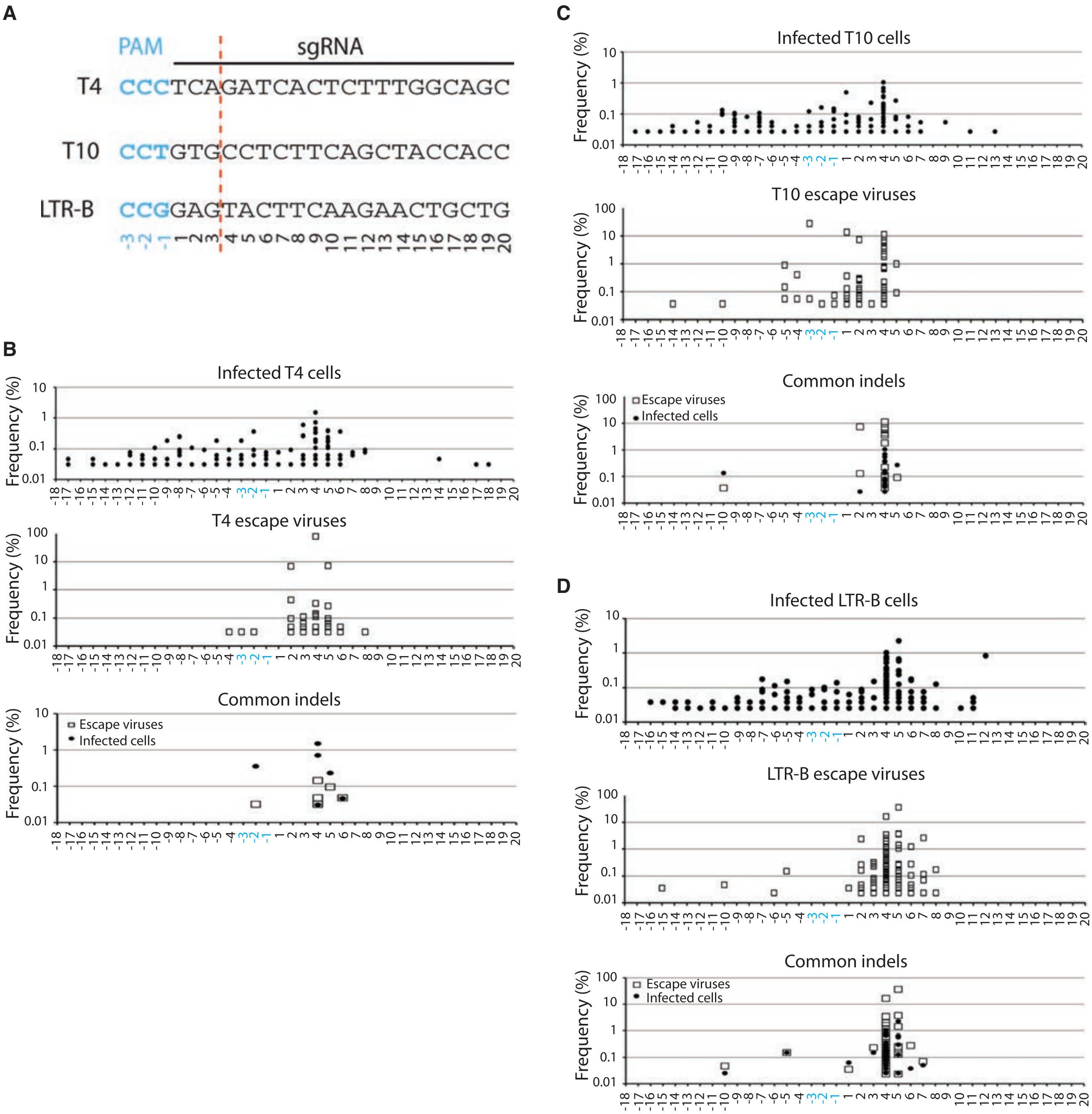


Figure 4. Enrichment of Cas9/sgRNA-Derived Indels in the Escape Viruses

(A) Illustration of the viral target sequences for sgRNA T4, T10, and LTR-B. The Cas9 cleavage site is indicated with a red dash line. The nucleotide positions refer to the junction of PAM and the sgRNA target.

(B) Frequency and distribution of indels in viral DNA from the T4 SupT1 cells that were infected with HIV-1 for 36 hr as well as indels in the T4 sgRNA escape viruses. The indels that alter the PAM/sgRNA target sequences are summarized. The common indels in the escape viruses and in viral DNA from the infected T4 cells are also shown. The indels were identified by Illumina MiSeq. The positions of the indels are denoted as in (A). The details of the indels are presented in Table S2.

(C) Indels in viral DNA from the T10 SupT1 cells that were infected with HIV-1 for 36 hr as well as in the T10 sgRNA escape viruses. The details of the indels are presented in Table S3.

(D) Indels in viral DNA from the LTR-B SupT1 cells that were infected with HIV-1 for 36 hr as well as indels in the LTR-B sgRNA escape viruses. The details of the indels are presented in Table S4.

develop resistance to antiviral drugs, immune responsiveness, and other types of pressure. However, resistance in these cases is mostly due to an accumulation of resistance mutations that derive from the error-prone RT and certain cellular factors such as APOBEC3G (Jern et al., 2009). In contrast, Cas9/sgRNA is itself a DNA sequence-specific mutagen and provides an independent source of resistance mutations. An analogy is inhibition of HIV-1 by RNAi (Manjunath et al., 2013), since both Cas9/sgRNA and RNAi utilize RNA as a guide to impede HIV-1 infection, even though RNAi leads to cleavage of viral RNA and leaves viral DNA intact. Thus, RNAi does not itself introduce mutations into the viral genome, whereas Cas9/sgRNA can directly alter viral DNA sequences. One common mechanism behind viral escape from the Cas9/sgRNA and RNAi involves mutating the viral sequence that is targeted by sgRNA or small interfering (si)RNA.

We have shown that the indels generated by Cas9/sgRNA confer resistance against Cas9/sgRNA. Following recognition of PAM by Cas9, the adjacent target DNA unwinds and initially binds to the first 10-nt seed sequence of sgRNA (Jiang et al., 2015). Cas9 then cleaves the target DNA at a position three nucleotides away from PAM. The NHEJ machinery is then recruited to the double-stranded DNA break. While repairing this DNA lesion, NHEJ often introduces insertion or deletion mutations (Hsu et al., 2014). These indels result in a change in the target DNA sequence, thus preventing sgRNA from binding and leading to resistance to Cas9/sgRNA. If the sgRNA targets a viral DNA sequence that is not essential for viral replication, then the indels that are generated should quickly lead to the emergence of Cas9/sgRNA-resistant, replication-competent viruses, as we observed with the LTR-B sgRNA (Figure 4D). When essential viral genes are targeted by sgRNA, the resistance-conferring indels should contribute to viral escape if they minimally affect the functions of the targeted viral genes. These latter indels should maintain the ORFs of viral genes and lead to only minimal changes in numbers of amino acids (one or two). The results of our MiSeq experiments reveal that these types of indels do exist in transiently infected cells as well as in the escape viruses (Figures 4B and 4C). Results of our study do not exclude the possibility that, when cells contain two or more copies of proviral DNA, homologous repair may contribute to the generation of escape mutations. Our findings are corroborated by a recent report showing HIV-1 escapes from Cas9/sgRNA inhibition by mutating the sgRNA target sequence (Wang et al., 2016).

The indels that are compatible with viral viability should be taken into consideration if Cas9/sgRNA is used to treat virus infection and genetic diseases. We expect that such indels would contribute to virus escape not only when Cas9/sgRNA is utilized to control new infections, but also in the context of eliminating latent viral DNA of herpes viruses, HBV, and HIV, among others. This is because introduction of a viable indel into latent viral DNA should lead to the mutated viral DNA being resistant to Cas9/sgRNA, but still able to produce infectious viruses upon activation. One potential solution might be to simultaneously target two or multiple sites in the viral genome with an array of sgRNAs in the way that multiple siRNAs have been used to durably suppress HIV-1 replication (Schopman et al., 2010).

EXPERIMENTAL PROCEDURES

Generation of Stably Transduced SupT1 Cell Lines

The lentiCRISPR(v2) DNA, with or without the T4, T10, LTR-B, EnvT1, EnvT3, EnvV2, EnvV4, and EnvV5 sgRNA sequences, was transfected into HEK293T cells together with DNA constructs that express the glycoprotein G of vesicular stomatitis virus (VSV), HIV-1 Gag/Gag-Pol, and HIV-1 Rev as described (Sanjana et al., 2014). Viruses in the culture supernatants were used to infect the SupT1 CD4+ T cell line. The stably transduced cell lines were selected with puromycin (2 μg/ml). Expression of Cas9 (containing the FLAG tag) in these cell lines was detected by immunostaining with anti-FLAG antibody and flow cytometry. Reagents used and cell culture protocols are described in the Supplemental Information.

HIV-1 Evolution Assay

The SupT1 cell lines that were stably transduced with lentiCRISPR(v2) were infected with HIV-1 equivalent to 3 ng of viral p24. Cells were passaged every 4 days until cytopathic effect (CPE) and cell death were observed. Levels of viruses in the culture supernatants were monitored by measuring viral RT activity.

Identifying Escape Mutations

Viral RNA was extracted from escape viruses using a QIAamp Viral RNA Mini Kit (QIAGEN). The viral RNA regions that are targeted by sgRNAs were amplified with the Superscript III One-Step RT-PCR Platinum Taq HiFi Kit (Invitrogen). Primers are described in the Supplemental Information. RT-PCR products were cloned into the PCR blunt vector and sequenced.

Illumina MiSeq

The sgRNA-targeted HIV-1 genomic regions were amplified using the primers described in the Supplemental Information. The PCR products were sequenced using the Illumina MiSeq platform at the McGill University and Genome Quebec Innovation Centre. Raw fragments from the MiSeq sequencer were first preprocessed to remove any trace of PhiX sequence and then trimmed to remove sequencing adapters as well as low quality bases from the 3′-end (using Trimmomatic version 0.32) (Bolger et al., 2014). Numbers of reads that underwent further analysis for each sample are presented in Table S1. Resulting fragments were aligned to reference sequences using a dynamic programming algorithm. This algorithm allows for more sensibility in aligning reads that potentially include large indels with regards to the reference, as compared to traditionally used next-generation sequencing mappers. Finally, the aligned fragments were post-processed to classify each variation, identify recurrent mutations in each sample, and quantify their occurrences. Indels of each experiment are presented in Tables S1, S2, S3, and S4.

Statistical Analysis

p values were calculated with Student's t test.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, three figures, and four tables and can be found with this article online at http://dx.doi.org/10.1016/j.celrep.2016.03.042.

AUTHOR CONTRIBUTIONS

C.L. conceived the study. Z.W., Q.P., and W.Z. performed the experiments. All authors contributed to experimental design and data analysis. Z.W., C.L., and M.A.W. prepared the manuscript. All authors read and approved the manuscript.

ACKNOWLEDGMENTS

This work was supported by funding from the Canadian Institutes of Health Research to C.L., the Ministry of Science and Technology of China (2012CB911103 and 2013ZX10001005-002 to F.G.), and the Nature Science Foundation of China (81371808 and 81528012 to F.G.).

Received: November 9, 2015
Revised: February 3, 2016
Accepted: March 10, 2016
Published: April 7, 2016

REFERENCES

Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D.A., and Horvath, P. (2007). CRISPR provides acquired resistance against viruses in prokaryotes. Science *315*, 1709–1712.

Bi, Y., Sun, L., Gao, D., Ding, C., Li, Z., Li, Y., Cun, W., and Li, Q. (2014). High-efficiency targeted editing of large viral genomes by RNA-guided nucleases. PLoS Pathog. *10*, e1004090.

Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics *30*, 2114–2120.

Bolotin, A., Quinquis, B., Sorokin, A., and Ehrlich, S.D. (2005). Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. Microbiology *151*, 2551–2561.

Cho, S.W., Kim, S., Kim, J.M., and Kim, J.S. (2013). Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease. Nat. Biotechnol. *31*, 230–232.

Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A., and Zhang, F. (2013). Multiplex genome engineering using CRISPR/Cas systems. Science *339*, 819–823.

Dampier, W., Nonnemacher, M.R., Sullivan, N.T., Jacobson, J.M., and Wigdahl, B. (2014). HIV excision utilizing CRISPR/Cas9 technology: attacking the proviral quasispecies in reservoirs to achieve a cure. MOJ Immunol. *1*, 00022.

Dong, C., Qu, L., Wang, H., Wei, L., Dong, Y., and Xiong, S. (2015). Targeting hepatitis B virus cccDNA by CRISPR/Cas9 nuclease efficiently inhibits viral replication. Antiviral Res. *118*, 110–117.

Ebina, H., Misawa, N., Kanemura, Y., and Koyanagi, Y. (2013). Harnessing the CRISPR/Cas9 system to disrupt latent HIV-1 provirus. Sci. Rep. *3*, 2510.

Holmes, M., Zhang, F., and Bieniasz, P.D. (2015). Single-cell and single-cycle analysis of HIV-1 replication. PLoS Pathog. *11*, e1004961.

Hsu, P.D., Lander, E.S., and Zhang, F. (2014). Development and applications of CRISPR-Cas9 for genome engineering. Cell *157*, 1262–1278.

Hu, W., Kaminski, R., Yang, F., Zhang, Y., Cosentino, L., Li, F., Luo, B., Alvarez-Carbonell, D., Garcia-Mesa, Y., Karn, J., et al. (2014). RNA-directed gene editing specifically eradicates latent and prevents new HIV-1 infection. Proc. Natl. Acad. Sci. USA *111*, 11461–11466.

Jern, P., Russell, R.A., Pathak, V.K., and Coffin, J.M. (2009). Likely role of APOBEC3G-mediated G-to-A mutations in HIV-1 evolution and drug resistance. PLoS Pathog. *5*, e1000367.

Jiang, F., Zhou, K., Ma, L., Gressel, S., and Doudna, J.A. (2015). Structural Biology. A Cas9-guide RNA complex preorganized for target DNA recognition. Science *348*, 1477–1481.

Jinek, M., East, A., Cheng, A., Lin, S., Ma, E., and Doudna, J. (2013). RNA-programmed genome editing in human cells. eLife *2*, e00471.

Karimova, M., Beschorner, N., Dammermann, W., Chemnitz, J., Indenbirken, D., Bockmann, J.H., Grundhoff, A., Lüth, S., Buchholz, F., Schulze zur Wiesch, J., and Hauber, J. (2015). CRISPR/Cas9 nickase-mediated disruption of hepatitis B virus open reading frame S and X. Sci. Rep. *5*, 13734.

Kennedy, E.M., and Cullen, B.R. (2015). Bacterial CRISPR/Cas DNA endonucleases: A revolutionary technology that could dramatically impact viral research and treatment. Virology *479-480*, 213–220.

Kennedy, E.M., Kornepati, A.V., Goldstein, M., Bogerd, H.P., Poling, B.C., Whisnant, A.W., Kastan, M.B., and Cullen, B.R. (2014). Inactivation of the human papillomavirus E6 or E7 gene in cervical carcinoma cells by using a bacterial CRISPR/Cas RNA-guided endonuclease. J. Virol. *88*, 11965–11972.

Kennedy, E.M., Bassit, L.C., Mueller, H., Kornepati, A.V., Bogerd, H.P., Nie, T., Chatterjee, P., Javanbakht, H., Schinazi, R.F., and Cullen, B.R. (2015). Suppression of hepatitis B virus DNA accumulation in chronically infected cells using a bacterial CRISPR/Cas RNA-guided DNA endonuclease. Virology *476*, 196–205.

Li, C., Guan, X., Du, T., Jin, W., Wu, B., Liu, Y., Wang, P., Hu, B., Griffin, G.E., Shattock, R.J., and Hu, Q. (2015). Inhibition of HIV-1 infection of primary CD4+ T-cells by gene editing of CCR5 using adenovirus-delivered CRISPR/Cas9. J. Gen. Virol. *96*, 2381–2393.

Liao, H.K., Gu, Y., Diaz, A., Marlett, J., Takahashi, Y., Li, M., Suzuki, K., Xu, R., Hishida, T., Chang, C.J., et al. (2015). Use of the CRISPR/Cas9 system as an intracellular defense against HIV-1 infection in human cells. Nat. Commun. *6*, 6413.

Lin, S.R., Yang, H.C., Kuo, Y.T., Liu, C.J., Yang, T.Y., Sung, K.C., Lin, Y.Y., Wang, H.Y., Wang, C.C., Shen, Y.C., et al. (2014). The CRISPR/Cas9 system facilitates clearance of the intrahepatic HBV templates in vivo. Mol. Ther. Nucleic Acids *3*, e186.

Mali, P., Yang, L., Esvelt, K.M., Aach, J., Guell, M., DiCarlo, J.E., Norville, J.E., and Church, G.M. (2013). RNA-guided human genome engineering via Cas9. Science *339*, 823–826.

Manjunath, N., Yi, G., Dang, Y., and Shankar, P. (2013). Newer gene editing technologies toward HIV gene therapy. Viruses *5*, 2748–2766.

Mojica, F.J., Díez-Villaseñor, C., García-Martínez, J., and Soria, E. (2005). Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. J. Mol. Evol. *60*, 174–182.

Pourcel, C., Salvignol, G., and Vergnaud, G. (2005). CRISPR elements in Yersinia pestis acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. Microbiology *151*, 653–663.

Ramanan, V., Shlomai, A., Cox, D.B., Schwartz, R.E., Michailidis, E., Bhatta, A., Scott, D.A., Zhang, F., Rice, C.M., and Bhatia, S.N. (2015). CRISPR/Cas9 cleavage of viral DNA efficiently suppresses hepatitis B virus. Sci. Rep. *5*, 10833.

Sanjana, N.E., Shalem, O., and Zhang, F. (2014). Improved vectors and genome-wide libraries for CRISPR screening. Nat. Methods *11*, 783–784.

Schopman, N.C., ter Brake, O., and Berkhout, B. (2010). Anticipating and blocking HIV-1 escape by second generation antiviral shRNAs. Retrovirology *7*, 52.

Seeger, C., and Sohn, J.A. (2014). Targeting hepatitis B virus with CRISPR/Cas9. Mol. Ther. Nucleic Acids *3*, e216.

Suenaga, T., Kohyama, M., Hirayasu, K., and Arase, H. (2014). Engineering large viral DNA genomes using the CRISPR-Cas9 system. Microbiol. Immunol. *58*, 513–522.

Wang, J., and Quake, S.R. (2014). RNA-guided endonuclease provides a therapeutic strategy to cure latent herpesviridae infection. Proc. Natl. Acad. Sci. USA *111*, 13157–13162.

Wang, J., Xu, Z.W., Liu, S., Zhang, R.Y., Ding, S.L., Xie, X.M., Long, L., Chen, X.M., Zhuang, H., and Lu, F.M. (2015). Dual gRNAs guided CRISPR/Cas9 system inhibits hepatitis B virus replication. World J. Gastroenterol. *21*, 9554–9565.

Wang, G., Zhao, N., Berkhout, B., and Das, A.T. (2016). CRISPR-Cas9 can inhibit HIV-1 replication but NHEJ repair facilitates virus escape. Mol. Ther. *24*, 522–526.

Yuan, M., Zhang, W., Wang, J., Al Yaghchi, C., Ahmed, J., Chard, L., Lemoine, N.R., and Wang, Y. (2015). Efficiently editing the vaccinia virus genome by using the CRISPR-Cas9 system. J. Virol. *89*, 5176–5179.

Yuen, K.S., Chan, C.P., Wong, N.H., Ho, C.H., Ho, T.H., Lei, T., Deng, W., Tsao, S.W., Chen, H., Kok, K.H., and Jin, D.Y. (2015). CRISPR/Cas9-mediated genome editing of Epstein-Barr virus in human cells. J. Gen. Virol. *96*, 626–636.

Zhen, S., Hua, L., Takahashi, Y., Narita, S., Liu, Y.H., and Li, Y. (2014). In vitro and in vivo growth suppression of human papillomavirus 16-positive cervical cancer cells by CRISPR/Cas9. Biochem. Biophys. Res. Commun. *450*, 1422–1426.

Zhen, S., Hua, L., Liu, Y.H., Gao, L.C., Fu, J., Wan, D.Y., Dong, L.H., Song, H.F., and Gao, X. (2015). Harnessing the clustered regularly interspaced short palindromic repeat (CRISPR)/CRISPR-associated Cas9 system to disrupt the hepatitis B virus. Gene Ther. *22*, 404–412.

Zhu, W., Lei, R., Le Duff, Y., Li, J., Guo, F., Wainberg, M.A., and Liang, C. (2015). The CRISPR/Cas9 system inactivates latent HIV-1 proviral DNA. Retrovirology *12*, 22.

OPEN
ACCESS
CellPress

Cell Reports
Report

Genetic Isolation of Hypothalamic Neurons that Regulate Context-Specific Male Social Behavior

Marta E. Soden,[1,2] Samara M. Miller,[1,2] Lauren M. Burgeno,[1,2] Paul E.M. Phillips,[1,2] Thomas S. Hnasko,[3] and Larry S. Zweifel[1,2,*]
[1]Department of Pharmacology, University of Washington, Seattle, WA 98195, USA
[2]Department of Psychiatry and Behavioral Sciences, University of Washington, Seattle, WA 98195, USA
[3]Department of Neurosciences, University of California, San Diego, La Jolla, CA 92093, USA
*Correspondence: larryz@uw.edu

http://dx.doi.org/10.1016/j.celrep.2016.05.067

SUMMARY

Nearly all animals engage in a complex assortment of social behaviors that are essential for the survival of the species. In mammals, these behaviors are regulated by sub-nuclei within the hypothalamus, but the specific cell types within these nuclei responsible for coordinating behavior in distinct contexts are only beginning to be resolved. Here, we identify a population of neurons in the ventral premammillary nucleus of the hypothalamus (PM_V) that are strongly activated in male intruder mice in response to a larger resident male but that are not responsive to females. Using a combination of molecular and genetic approaches, we demonstrate that these PM_V neurons regulate intruder-specific male social behavior and social novelty recognition in a manner dependent on synaptic release of the excitatory neurotransmitter glutamate. These data provide direct evidence for a unique population of neurons that regulate social behaviors in specific contexts.

INTRODUCTION

In rodents, olfactory information is a major modality for social communication. Inputs from the main olfactory bulb and accessory olfactory bulb directly innervate sub-nuclei of the medial amygdala that transmit this information to the hypothalamus (Scalia and Winans, 1975; Kevetter and Winans, 1981; Choi et al., 2005; Sosulski et al., 2011). The hypothalamus also receives direct olfactory information relevant to social cues (Yoon et al., 2005). Hypothalamic sub-populations have been identified within the ventrolateral ventromedial hypothalamus (VMH_{VL}) that regulate key social behaviors, including aggressive responses to conspecific threats and mating (Lin et al., 2011; Lee et al., 2014). Neurons within the medial pre-optic area (MPO) of the hypothalamus have also been isolated in mice and shown to regulate parental care or agression toward pups, depending on the animals' sexual experience (Wu et al., 2014). Numerous other sub-nuclei within the hypothalamus have been implicated in the regulation of social behaviors (Swanson, 2000), but virtually nothing is known about the cell types within these regions that contribute to these behaviors.

The ventral premammillary nucleus of the hypothalamus (PM_V) is highly connected with the brain's social networks (Canteras et al., 1992; Swanson, 2000; Cavalcante et al., 2014), and mapping studies using the immediate early gene *Fos* demonstrated that PM_V neurons are activated in multiple social contexts (Cavalcante et al., 2006; Borelli et al., 2009; Motta et al., 2009; Donato et al., 2010, 2013). Early analysis of the catecholamine-producing neurons of the brain identified a non-canonical dopamine neuron population within the PM_V (Hedreen, 1978; Meister and Elde, 1993; Zoli et al., 1993), but whether these neurons regulate social behavior and whether they use dopamine as a neurotransmitter is not known. Based on previous evidence that neurons in the PM_V express mRNA for the dopamine transporter (DAT; Meister and Elde, 1993), we used DAT as a genetic marker to isolate this population. We demonstrate that PM_V-DAT neurons are connected to brain regions implicated in conspecific social behavior and are principally glutamatergic, but do not release detectable dopamine. We show that PM_V-DAT neurons are most highly activated when male mice are intruders into the residence of a larger male. Chemogenetic inhibition of PM_V-DAT neurons specifically reduces exploratory social behavior in intruder males and impairs social novelty recognition. Activation of PM_V-DAT neurons increases exploratory social investigation of familiar mice in a manner dependent on synaptic glutamate release. Our genetic isolation and characterization of this unique neuronal population provides direct evidence for a hypothalamic cell type that regulates male intruder behavior and social novelty in specific contexts.

RESULTS

Genetic Isolation of PM_V-DAT Neurons

Given the previous identification of neurons in the PM_V that express DAT (Hedreen, 1978; Meister and Elde, 1993), we sought to determine the function of these neurons within social contexts. To confirm the presence of these neurons, we performed a differential search of the Allen Institute for Brain Science Mouse Brain Atlas in situ hybridization data (Lein et al., 2007). Three of the top 30 genes enriched in the PM_V are canonical markers for the synthesis and release of the neurotransmitter dopamine: dopamine transporter (DAT; *Slc6a3*), dopa-decarboxylase (DDC; *Ddc*), and the vesicular monoamine transporter (VMAT; *Slc18a2*). Of these markers, DAT appeared to be the most selective for the PM_V within the hypothalamus. Based on this observation, along with previous data demonstrating DAT mRNA in the PM_V (Meister and Elde, 1993) and observations that mice expressing Cre recombinase from the DAT locus are the most reliable for isolating dopamine-producing neurons (Lammel et al., 2015), we genetically isolated PM_V-DAT neurons utilizing the DAT-Cre mouse line ($Slc6a3^{Cre/+}$; Zhuang et al., 2005). To validate the expression data from the Mouse Brain Atlas, we performed immuno-isolation of actively translating mRNA using RiboTag mice (Sanz et al., 2009). These mice allow for Cre-dependent expression of an affinity-tagged ribosomal protein, Rpl22-HA, and subsequent immunoprecipitation (IP) of polyribosomes. $Slc6a3^{Cre/+};Rpl22^{HA/+}$ mice were used to isolate mRNA from DAT neurons in the PM_V following microdissection (Figures 1A–1C). We observed a significant enrichment of *Slc6a3*, *Ddc*, *Slc18a2*, and tyrosine hydroxylase (*Th*) mRNA in the IP relative to the input, which contains mRNA from all cell types in the region (Figure 1D). Similar enrichment of these markers was observed when mRNA was isolated from canonical dopamine neurons in the ventral tegmental area (VTA) of $Slc6a3^{Cre/+};Rpl22^{HA/+}$ mice (Figures S1A and S1B).

Quantification of PM_V-DAT neurons revealed that these cells make up 25% of neurons in this region (952 YFP+ cells/3767 NeuN+ cells, alternate sections counted from three mice; Figure S1C). Connectivity mapping of PM_V-DAT neurons through conditional expression of an EGFP-tagged synaptic marker protein synaptophysin-EGFP (AAV1-FLEX-synapto-EGFP) revealed axonal projections to numerous downstream targets previously implicated in the regulation of social behavior (Canteras et al., 1992). These included the VMH_{VL}, the MPO, the medial postero-ventral subdivision of the medial amygdala (MeA_{PV}), the amygdalo-hippocampal region of the medial amygdala (MeA_{AHi}), the postero-dorsal and ventral subdivisions of bed nucleus of the stria-terminalis ($BNST_{PD}$ and $BNST_{PV}$), the anterior hypothalamic nucleus (AHN), and the periaqueductal gray (PAG; Figure 1E). Analysis of PM_V-DAT projections obtained from the Allen Brain Institute Mouse Connectivity Atlas (Oh et al., 2014) revealed similar results. To establish the density of PM_V-DAT projections to various targets, we generated an analysis program to quantify fluorescent pixels in target areas using data from the Connectivity Atlas. While the BNST received the most total input, the densest innervation was observed in the VMH_{VL} (Figure S1D).

PM_V-DAT Neurons Release Glutamate, but Not Detectable Dopamine

Although previous studies have demonstrated the presence of 6-OH-DA-sensitive neurons in the PM_V (Hedreen, 1978), these neurons were not immunoreactive for dopamine (Zoli et al., 1993). To test whether PM_V-DAT neurons release detectable dopamine, we expressed ChR2 in PM_V-DAT neurons and assayed for light-evoked dopamine release using fast-scan cyclic voltammetry (FSCV) in acute slices from multiple target regions (VMH_{VL}, MeA_{MPV}, and BNST). We did not detect any dopamine transients following optical stimulation of PM_V-DAT fibers using multiple stimulus parameters (5, 10, 20, and 30 Hz; Figures 2A and 2B). In contrast, expression of ChR2-mCherry in VTA dopamine neurons and optical stimulation of terminals in the nucleus accumbens resulted in robust dopamine release (Figures 2A and 2B). Pre-treating mice with L-DOPA, the dopamine precursor shown to be transported by PM_V neurons (Zoli et al., 1993), did not enhance detection of dopamine release (Figures 2A and 2B). These data do not rule out that dopamine may be released below the level of detection of FSCV. However, consistent with our FSCV results, we did not detect any protein for TH or DAT in the cell bodies or terminal fields of DAT-PM_V neurons using immunohistochemistry (Figures S2A and S2B).

Because we detected mRNA, but no protein, for dopamine markers in PM_V-DAT neurons, we compared the relative amounts of mRNA for these markers in PM_V-DAT neurons versus VTA-DAT neurons using the IP fractions from our RiboTag analysis. Although mRNA for dopaminergic markers is highly enriched in PM_V-DAT neurons relative to other neurons within the region, these cells contained only 5%–10% the amount of dopaminergic mRNAs compared to VTA-DAT neurons (Figure S2C). In contrast, PM_V-DAT neurons contained three times as much mRNA encoding the vesicular glutamate transporter vGluT2 (*Slc17a6*) as did VTA-DAT neurons. Additional electrophysiological analysis of PM_V-DAT neurons revealed that these neurons do not share common electrophysiological properties with VTA dopaminergic neurons (Figures 2C–2G and S2D–S2G), indicating that these neurons are a unique population.

Consistent with the detection of mRNA encoding vGluT2, analysis of synaptic connectivity in the VMH_{VL} (Figure 2H) revealed the majority of light-evoked responses from ChR2-expressing terminals of PM_V-DAT neurons are monosynaptic excitatory currents blocked by the glutamatergic antagonist CNQX (Figure 2I). Delayed and unsynchronized inhibitory and excitatory currents were also seen in a small number of cells (five inhibitory, seven excitatory; Figures 2J and 2K). A smaller sample of recordings from the BNST (n = 6/15 cells connected) and MPN (n = 9/14 cells connected) produced similar results.

PM_V-DAT Neurons Regulate Context-Specific Social Behavior

To confirm previous observations of *Fos* induction in the PM_V following exposure of animals to opposite- and same-sex odorants (Cavalcante et al., 2006; Leshan et al., 2009; Donato et al., 2010), we exposed male and female mice to clean bedding (control) or bedding soiled by a same-sex or opposite-sex mouse, followed by immunohistochemistry for Fos. We observed a modest increase in Fos protein in response to opposite-sex odorants in male and female mice. However, male, but not female, mice showed increased Fos in response to same-sex odorants that was significantly higher than Fos observed in response to opposite-sex odorants (Figures S3A–S3C).

To establish whether PM_V-DAT neurons are activated in male mice in a context-dependent manner, we quantified Fos expression in virally labeled cells. Mice were exposed to one of five social encounter conditions: no encounter (control), intruder into the cage of a male resident, intruder into the cage of a female resident, resident exposed to a male intruder, and resident

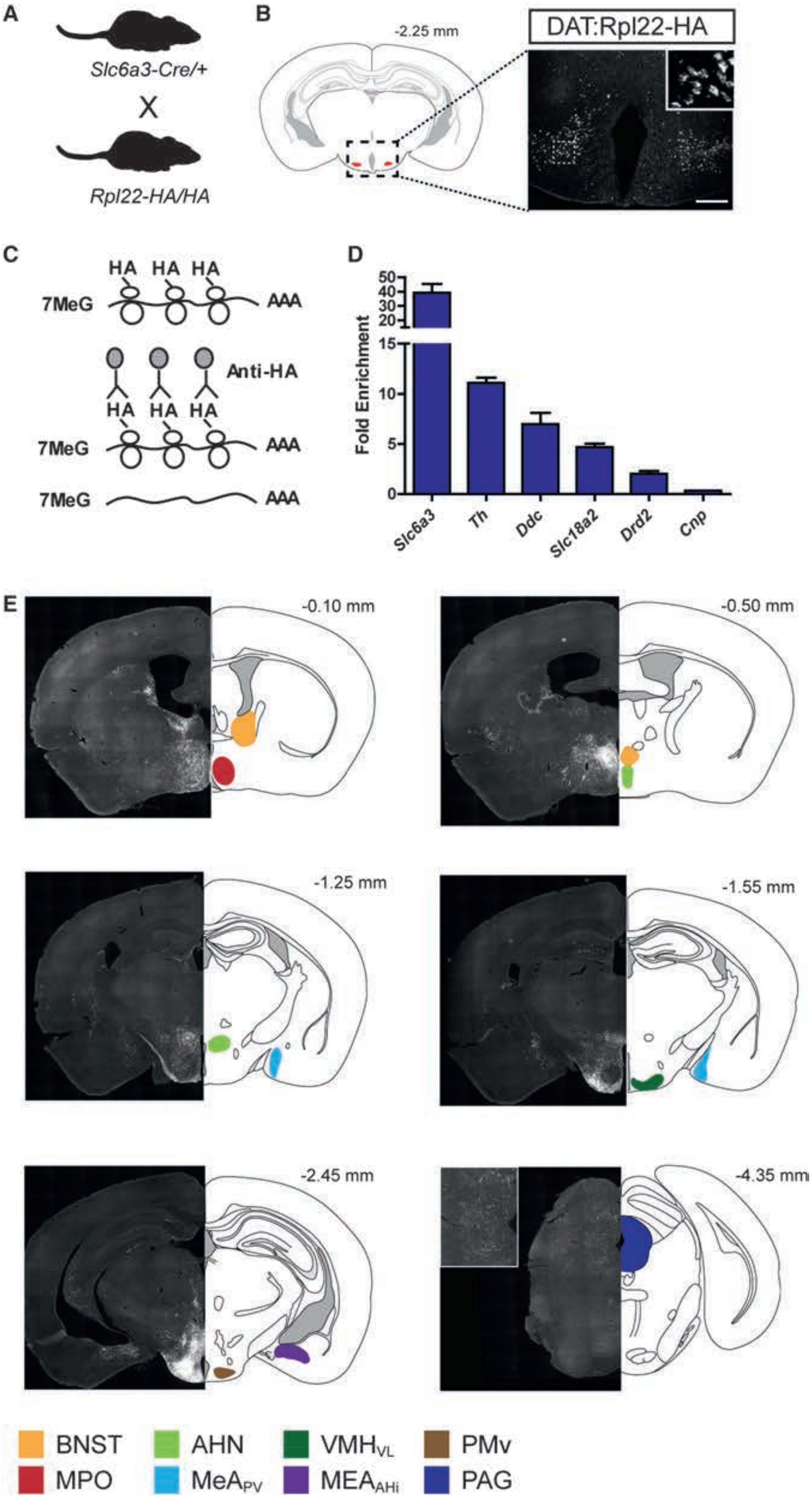


Figure 1. Molecular Profile and Connectivity of $\mathrm{PM_V}$-DAT Neurons

(A) Schematic generation of DAT-Cre-RiboTag mice.

(B) Atlas image depicting the $\mathrm{PM_V}$; boxed region is depicted in immunohistochemistry image of Rpl22-HA (right) and inset. Scale bar, 250 μm.

(C) Cartoon depicting RiboTag technique: in Cre-positive cells the ribosomal protein Rpl22 is labeled with an HA tag; immunoprecipitation of HA isolates ribosome-associated mRNAs.

(D) Fold enrichment (IP compared to input) of specific mRNAs isolated from $\mathrm{PM_V}$-DAT neurons (n = 3 pooled samples, each from 2–3 mice).

(E) Projections of $\mathrm{PM_V}$-DAT neurons are revealed by expression of synapto-EGFP in axon terminals. Atlas images show approximate distance from bregma; indicated brain regions are color-coded. See also Figure S1.

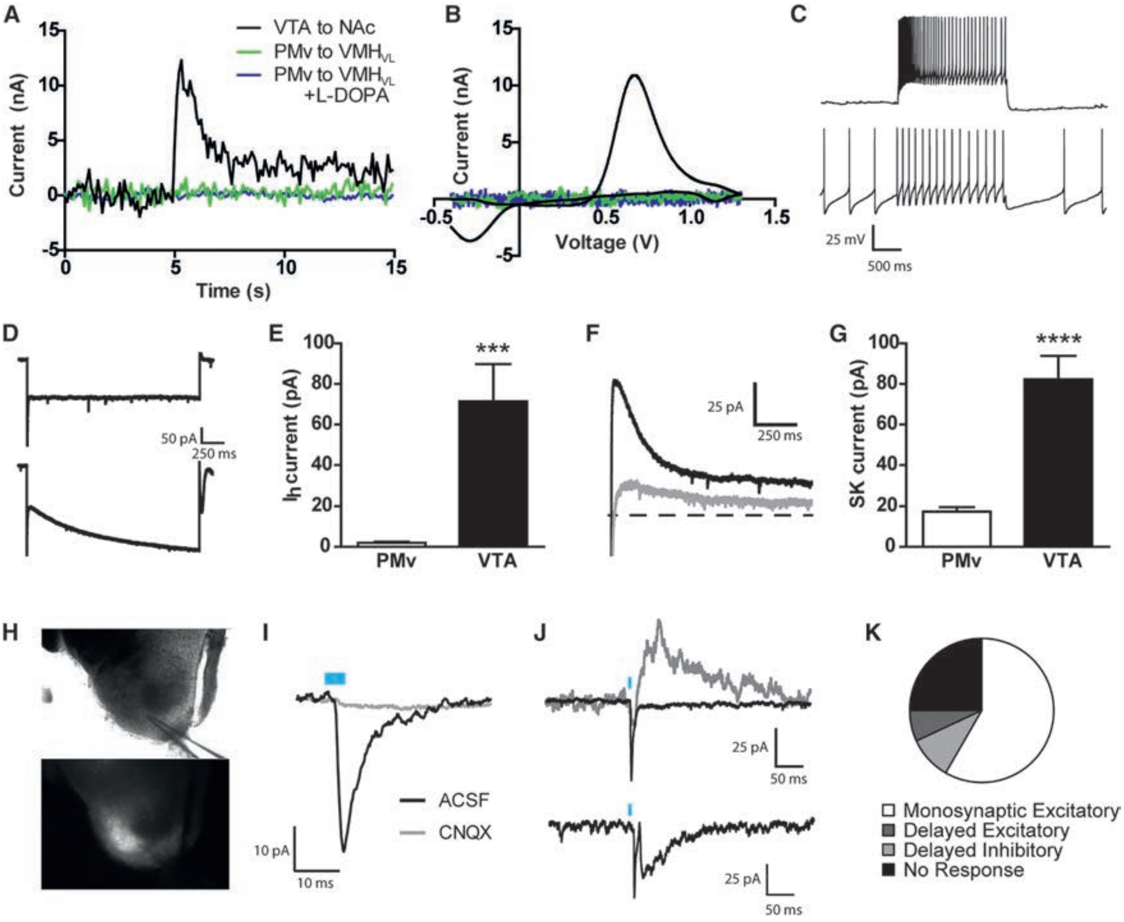


Figure 2. PM$_{\mathrm{V}}$-DAT Neurons Are Principally Glutamatergic and Do Not Release Dopamine

(A and B) Example traces (A) and current/voltage plots (B) showing light-evoked dopamine release detected using FSCV in control slices expressing ChR2 in VTA-DAT neurons, recording in the NAc. No signal was detected in the VMH$_{\mathrm{VL}}$ when ChR2 was expressed in PM$_{\mathrm{V}}$-DAT neurons, even with pre-loading of L-DOPA.

(C) Example traces showing that PM$_{\mathrm{V}}$-DAT neurons did not spontaneously fire in slice but did fire accommodating action potentials with current injection (top trace), while VTA-DAT neurons did fire spontaneously (bottom trace).

(D and E) Example traces (D) and quantification (E) of I_h current in PM$_{\mathrm{V}}$-DAT neurons (top) and VTA-DAT neurons (bottom).

(F and G) Example traces (F) and quantification (G) of SK currents in PM$_{\mathrm{V}}$-DAT neurons (gray) and VTA-DAT neurons (black).

(H) DIC (top) and fluorescent (bottom) images of an acute slice with fluorescent ChR2-mCherry fibers in the VMH$_{\mathrm{VL}}$ and patch electrode visible. Scale bar, 250 μm.

(I) Example trace of EPSC in the VMH$_{\mathrm{VL}}$ evoked by 5-ms blue light stimulation; the EPSC could be blocked by bath application of CNQX (10 μM).

(J) Top: example recordings showing a light-evoked monosynaptic EPSC (black trace, holding at -60 mV) and a delayed, unsynchronized IPSC (gray trace, holding at 0 mV). Bottom: example recording showing a light-evoked monosynaptic EPSC (initial fast depolarization) and a delayed, unsynchronized EPSC. All traces are averages of 15 sweeps.

(K) Proportion of VMH$_{\mathrm{VL}}$ neurons recorded that received each type of synaptic input.

See also Figure S2.

exposed to a female intruder. Fos levels in PM$_{\mathrm{V}}$-DAT neurons were significantly higher in male same-sex intruders relative to all other contexts (Figures 3A and 3B). PM$_{\mathrm{V}}$-DAT neurons were a subset of the total Fos+ PM$_{\mathrm{V}}$ neurons, but were comprised of a significantly larger subset in the same-sex intruder context compared to other contexts (Figure S3D). To test whether Fos activation in male intruder mice was primarily driven by the odorant context of the resident cage or by the resident mouse itself, we measured Fos in an "intruder" male following an encounter with a "resident-like" male (i.e., larger, singly housed, sexually experienced) in a neutral environment (clean cage). Fos in PM$_{\mathrm{V}}$-DAT neurons was only modestly activated in this context (Figures 3A and 3B), indicating that contextual odorants play a large role in driving PM$_{\mathrm{V}}$-DAT activation in intruder mice.

Our observation that PM$_{\mathrm{V}}$-DAT neurons are most strongly activated in male mice when these animals are intruders is consistent with previous reports (Borelli et al., 2009; Motta et al., 2009) and suggests that these neurons may preferentially function within this context. To test this hypothesis, we selectively attenuated the activity of PM$_{\mathrm{V}}$-DAT neurons through conditional expression of the inhibitory DREADD receptor hM4Di (AAV1-FLEX-hM4Di-YFP). Consistent with previous reports (Armbruster et al., 2007), activation of hM4Di by the selective agonist clozapine-N-oxide (CNO) reduced the excitability of PM$_{\mathrm{V}}$-DAT neurons (Figures S3E–S3H). Behavioral analysis of hM4Di and control mice (AAV1-FLEX-mCherry) revealed that reduced excitability of PM_V-DAT neurons was associated with a significant reduction in social investigation (anogenital sniffing, oral-facial sniffing, and grooming) by male intruder mice, but had no effect on resident male behavior or on male response to females. An encounter with a "resident-like" male in a neutral cage was also not affected (Figures 3C–3E and S3I).

A reduction in social exploratory behavior by intruder mice may represent a deficit in social recognition within this context. To test this hypothesis, we analyzed behavior in a three-chamber social preference and social recognition assay. Reduced excitability of PM_V-DAT neurons through hM4Di activation did not alter behavioral preference for a novel mouse versus a novel object, but did reduce social recognition of a novel versus familiar mouse (Figures 3F–3G and S3J–S3K).

PM_V-DAT Neurons Regulate Social Behavior through Glutamate Release

To further investigate the extent to which activation of PM_V-DAT neurons influences social recognition and social investigation, we developed a co-habitation assay to monitor social investigation of a familiar cage-mate, while PM_V-DAT neurons are selectively activated using the light-activated ion channel ChR2. Because PM_V-DAT neurons should be minimally activated in this context, this assay allowed us to establish the sufficiency of activation of PM_V-DAT neurons for social investigation behaviors.

To determine the optimal stimulation parameters for this assay, we analyzed the synaptic fidelity of light-evoked action potentials in postsynaptic VMH_{VL} neurons using a common stimulation paradigm (5-ms light stimuli at 20 Hz). Fidelity diminished with repetitive stimulation at this frequency (Figure S4A). This decrease was not observed when action potential firing was recorded in PM_V-DAT neuron cell bodies (Figure S4B). Quantitative analysis of light-evoked postsynaptic currents confirmed a significant synaptic rundown at frequencies of 5 Hz and above (Figures S4C and S4D), suggesting that low-frequency stimulation in these neurons is optimal for maintaining synaptic connectivity.

To monitor exploratory social behavior between familiar mice, we tested conspecific males co-housed from birth. One mouse was surgically injected with either AAV1-FLEX-ChR2-mCherry or AAV1-FLEX-mCherry and a fiber-optic was implanted above the PM_V (Figure 4A). Mice were assayed for investigatory behavior of their cage-mate with or without optical stimulation (Figure 4B). While 3-Hz light stimulation had no effect on social interaction in control mice, mice expressing ChR2 significantly increased social investigation (anogenital sniffing, oral-facial sniffing, and grooming) of their cage-mate during light stimulation (Figures 4C and S4E). Light stimulation had no effect on exploration of a familiar object in the home cage (Figure 4D) and did not affect distance traveled in the cage (Figure S4F). Light stimulation also had no effect in a real-time place preference (RTPP) assay, suggesting that activation of the PM_V is neither rewarding nor aversive (Figure 4E). Finally, light stimulation did not affect behavior in an open-field assay (Figures S4G and S4H), indicating no overt role for PM_V-DAT neurons in regulating anxiety.

To further establish the extent to which activation of PM_V-DAT neurons can enhance social investigation, we stimulated PM_V-DAT neurons in resident animals, a context in which PM_V-DAT neurons are not robustly activated, but in which animals are already actively engaged in social investigation. We observed an increase in investigation time in ChR2-expressing animals, indicating that artificially activating PM_V-DAT neurons can drive increased social behavior even when animals are already socially engaged (Figure S4I). In contrast, stimulation of PM_V-DAT neurons in intruder mice did not enhance social investigation (Figure S4J), indicating that these neurons are already optimally functioning in this context.

To confirm that glutamate is the critical neurotransmitter responsible for the effect of PM_V-DAT neurons on social behavior, we repeated our co-habitation experiments by expressing ChR2 in DAT-PM_V neurons in mice in which the gene encoding vGluT2, *Slc17a6*, is inactivated in dopamine neurons (*Slc6a3*$^{iCre/+}$; *Slc17a6*$^{lox/lox}$, or DAT-vGlut2 KO) (Hnasko et al., 2010). Consistent with PM_V-DAT neurons being glutamatergic, light-induced excitatory post-synaptic currents (EPSCs) were observed in control mice (*Slc6a3*$^{Cre/+}$; *Slc17a6*$^{lox/+}$), but not in DAT-vGlut2 KO mice (Figure 4F, inset). Optical stimulation of PM_V-DAT neurons significantly enhanced social investigation in control mice, but not in DAT-vGlut2 KO mice (Figure 4F). Stimulation of PM_V-DAT neurons in DAT-vGlut2 KO and control mice was not associated with increased investigation of a familiar object (Figure 4G) or with RTPP (Figure 4H).

Because DAT-vGlut2 KO mice are a loss of function, we next asked whether these mice have altered baseline social behavior in the resident-intruder assay in the absence of light stimulation. Consistent with hM4Di-mediated inhibition of PM_V-DAT neurons, DAT-vGlut2 KO mice displayed significantly reduced social exploratory behavior when they were intruders in the cage of a resident male (Figure 4I). This effect was not observed when the mice were residents in response to a male or female intruder (Figures 4J and 4K). A caveat to this approach is that genetic inactivation of vGlut2 in DAT-expressing neurons removes this protein from all such neurons, including those in the VTA. It has been demonstrated that inactivation of DAT-expressing neurons in the VTA disrupts social behavior (Gunaydin et al., 2014; Yu et al., 2014). Therefore, we inactivated DAT neurons in the VTA by expressing hM4Di in these cells (Figure S4K). Inhibition of DAT neurons in the VTA did not affect investigatory behavior in intruder mice, but did reduce this behavior in resident males toward male intruders (Figures S4L–S4N).

DISCUSSION

Our data identify a unique population of neurons in the mouse hypothalamus that regulate intruder-specific male behavior. Numerous studies have identified the hypothalamus as a key regulator of socially motivated behavior (Swanson, 2000), but, to date, a defined group of neurons specifically tuned to influence conspecific behavior exclusively in intruder males had not been demonstrated. Our observations are supported by previous findings of increased Fos in the PM_V of male intruder rats (Borelli et al., 2009; Motta et al., 2009). Motta et al. (2009) also have observed increased Fos protein in the

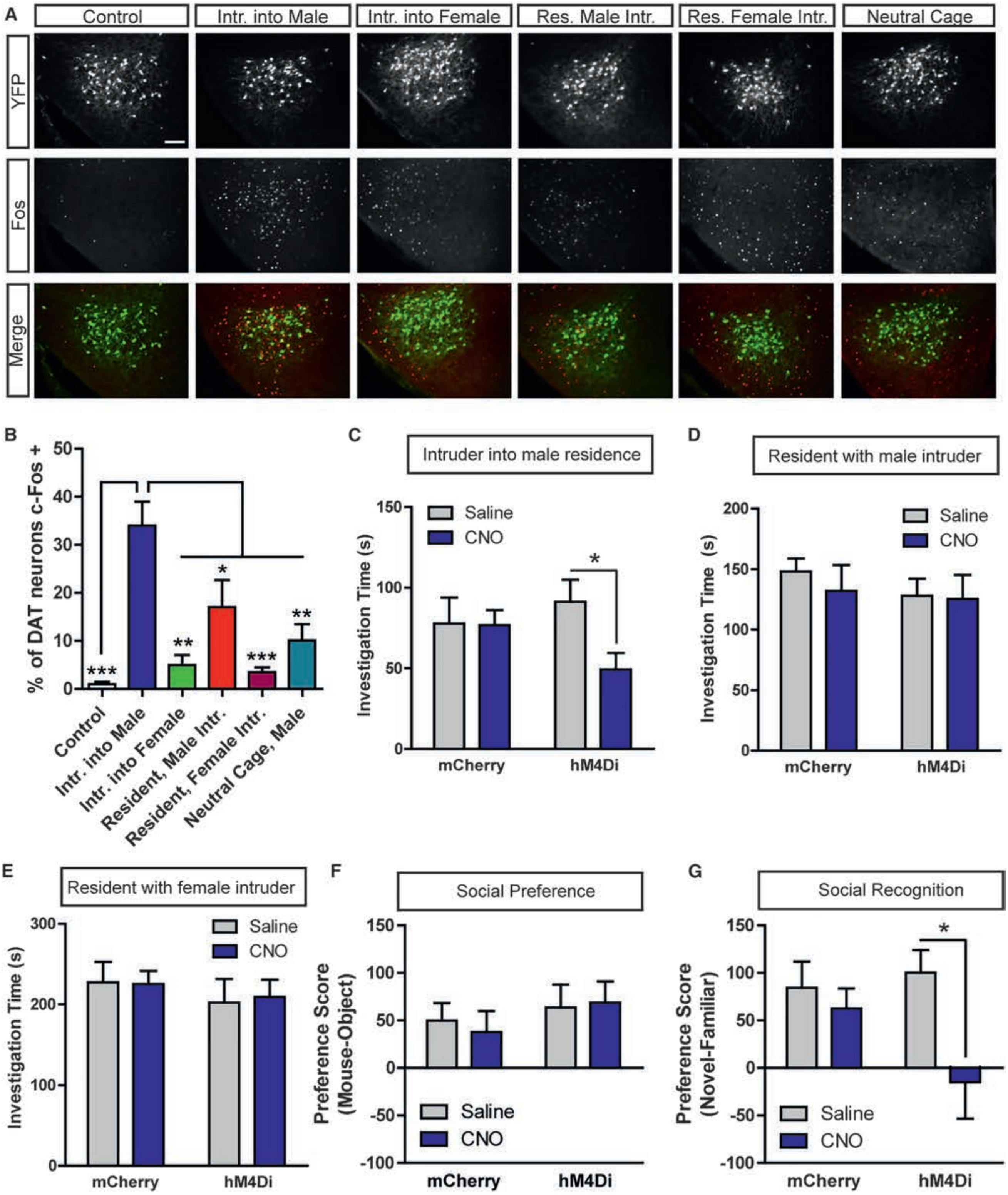


Figure 3. PM_V-DAT Neurons Regulate Intruder-Specific Behavior

(A and B) Images (A) and quantification (B) of Fos levels in PM_V-DAT neurons during social encounters (n = 3 mice/group; one-way ANOVA $F_{(5,12)} = 12.26$, $p < 0.001$; Tukey's multiple comparison $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$; scale bar, 100 μm).

(legend continued on next page)

dorsal premammillary nucleus (PM_D) and found that lesioning the PM_D reduced exploratory social behavior and defensive behavior by much smaller subordinate intruder rats. These data collectively point to specific populations within the PM_V and PM_D that regulate intruder-specific behavior.

It has been reported that lesions of the PM_V increase aggressive behavior in male rats (Vandenberg et al., 1983). Since PM_V-DAT neurons only constitute approximately 25% of the total neurons of the PM_V, and numerous Fos-positive PM_V neurons in the resident-intruder assays are not DAT-positive, PM_V-DAT neurons likely are not the only population involved in context-specific social behavior. Thus, lesioning of the entire PM_V may disrupt multiple cell types that when collectively destroyed lead to different behavioral outcomes.

We also have found that optogenetic stimulation of PM_V-DAT neurons engages social exploration of a familiar mouse. These findings, together with our observations that social recognition is impaired when PM_V-DAT neurons are inhibited, suggest that these neurons influence exploration and novelty detection specific for socially relevant stimuli. Similarly, it also has been demonstrated that dopamine neurons of the VTA selectively influence social exploration, but not novelty exploration (Gunaydin et al., 2014). Interestingly, we have found that inhibition of VTA-DAT neurons reduces social behavior by resident males, but not by intruder males, and this behavior is not sensitive to inactivation of vGlut2, further supporting the notion that VTA-DAT neurons and PM_V-DAT neurons operate through distinct neurotransmitter systems in different behavioral contexts.

Direct optogenetic stimulation of all neurons or a specific subpopulation of neurons within the VMH_{VL} results in robust aggressive posturing in resident male mice toward intruders (Lin et al., 2011; Lee et al., 2014) and toward inanimate objects (Lin et al., 2011). Optogenetic stimulation of excitatory neurons of the MeA, many of which project to the VMH_{VL}, also increases male aggressive behavior (Hong et al., 2014), and similar results have been reported for stimulation of a subpopulation of aromatase neurons in the MeA (Unger et al., 2015). We have found that PM_V-DAT inputs to target structures, including the VMH_{VL}, are principally excitatory and are most effective at high-fidelity synaptic transmission at low frequencies. Stimulation of PM_V-DAT neurons promotes social exploration, but not aggression, and does not evoke aggression toward inanimate objects. Thus, we propose that social behavior engaged by PM_V-DAT neurons is tightly controlled to promote social exploration without escalating to aggression.

Isolation of mRNA associated with polyribosomes from DAT-expressing neurons of the PM_V reveals that these cells are highly enriched for dopaminergic markers compared to cells in the surrounding tissue, making them the only known population to contain mRNA for all dopaminergic markers, but not to release detectable dopamine. A likely explanation for our inability to detect dopamine release from these neurons is that while these neurons do contain mRNA for the molecular machinery to synthesize and release this neurotransmitter, their mRNA levels are an order of magnitude less than those in conventional VTA-DAT neurons. Why neurons in the PM_V contain any mRNA at all for dopaminergic enzymes is not clear. One possibility is that these neurons are derived from a common lineage of other dopamine neurons in the hypothalamus and midbrain that express these markers, but the mRNA in PM_V-DAT neurons is not translated into protein or these proteins are rapidly degraded and thus are in quantities below the level of detection. A second possibility is that these neurons utilize dopamine as a neurotransmitter early during development and then switch to glutamatergic once circuit connectivity is established, while maintaining residual dopaminergic mRNA expression following this switch. Evidence for developmental neurotransmitter switching has been widely reported (Spitzer, 2015). A third possibility is that these neurons have the capacity to enhance dopamine production and release under specific environmental demands. Neurotransmitter switching has been previously reported in the adult hypothalamus (Dulcis et al., 2013). Future experiments designed to determine the relevance of dopamine marker mRNA expression in PM_V-DAT neurons will further inform the identity and function of the unique neuronal population.

EXPERIMENTAL PROCEDURES

See the Supplemental Experimental Procedures for additional experimental procedures.

Mice

All procedures were approved and conducted in accordance with the guidelines of the Institutional Animal Care and Use Committee of the University of Washington. Mice 8 weeks or older were used for all experiments, except for slice electrophysiology, where 5- to 8-week-old mice were used.

RiboTag

Immunoprecipitation was performed as described previously (Sanz et al., 2009). Briefly, 1 × 1 mm punches of PM_V and VTA were removed, homogenized, and incubated with anti-HA antibody (Covance), coupled to magnetic beads (Pierce) overnight at 4°C. Following elution from magnetic beads, RNA from both IP and input samples was obtained using the RNeasy micro kit (QIAGEN). cDNA was generated using oligo dT primers (Invitrogen). TaqMan (Applied Biosystems) primers were used for qRT-PCR analysis.

Electrophysiology and Voltammetry

Whole-cell recordings were made using an Axopatch 700B amplifier (Molecular Devices). Light-evoked synaptic transmission was induced with 5-ms light pulses delivered from an optic fiber placed in the bath directly. I_h currents were induced by 2-s hyperpolarizing voltage steps from −70 to −120 mV. SK currents were induced by depolarizing voltage steps from −70 to 0 mV. Capacitance measurements were calculated by Clampex software using 5-mV

(C) Social investigation of resident animal by intruder (experimental) was decreased following inhibition of PM_V-DAT neurons by CNO/hM4Di (n = 10–11 mice/group; two-way repeated-measures ANOVA virus × CNO $F_{(1,19)} = 4.43$, $p < 0.05$; Bonferroni multiple comparisons $^*p < 0.05$).

(D and E) Inhibition of PM_V-DAT neurons did not affect the investigation of a male (D) or female (E) intruder by a resident (experimental).

(F) Inhibition of PM_V-DAT neurons by CNO/hM4Di did not affect preference for a mouse over an object.

(G) Inhibition of PM_V-DAT neurons eliminated the preference for a novel mouse versus a familiar mouse; n = 10–11 mice/group; two-way repeated-measures ANOVA: $F_{(1,19)} = 7.38$, $p < 0.05$; Bonferroni multiple comparisons: $^*p < 0.05$). Data are represented as mean ± SEM.

See also Figure S3.

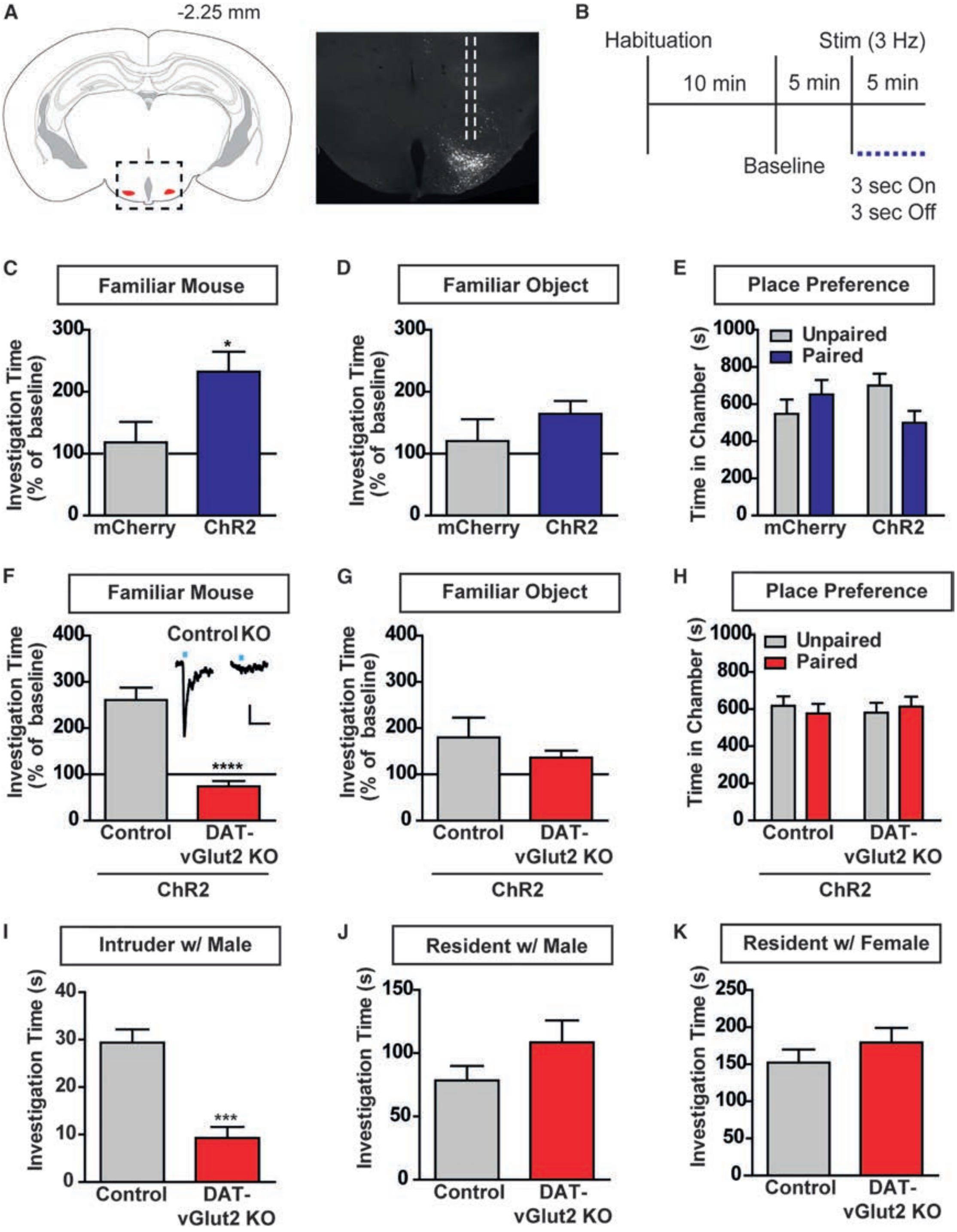


Figure 4. PM$_V$-DAT Neurons Regulate Social Behaviors through Glutamate Release

(A) Atlas and histology images showing ChR2-mCherry expression in the PM$_V$; dashed lines indicate track mark from fiber optic.

(B) Schematic of cohabitation behavioral assay.

(C) Activation of PM$_V$-DAT neurons increased social investigation of a familiar cage mate (n = 6–8 animals/group; Student's t test *$p < 0.05$).

(D) Activation did not increase investigation of a familiar object.

(E) Pairing one side of a two-chambered box with light stimulation did not lead to a significant preference for either side.

(F) Inactivation of vGlut2 in DAT-vGlut2 KO mice eliminates light-evoked excitatory currents driven by ChR2 expression in DAT-PM$_V$ neurons and recorded in VMH$_{VL}$ (inset). Light activation of PM$_V$-DAT neurons expressing ChR2 increased social investigation of a cage mate in DAT-vGlut2 heterozygous animals (control), but not in DAT-vGlut2 knockout animals (n = 6 animals/group; Student's t test ****$p < 0.0001$).

(legend continued on next page)

hyperpolarizing steps. Fast-scan cyclic voltammetry was performed using carbon-fiber microelectrodes as described previously (Clark et al., 2010). 5-ms light stimuli were delivered as described for electrophysiology.

Fos Induction

For social encounters, mice were assigned to the resident, intruder, or control condition. Resident animals were singly housed for at least 2 weeks, while intruder and control mice were group housed. Animals experienced a 20-min social encounter with an appropriately matched resident or intruder animal and were euthanized and perfused 90 min following the start of the encounter. Control animals remained in their home cage. Fos-positive neurons were identified and automatically counted using ImageJ software. Virally transduced DAT neurons and Fos-positive DAT neurons were counted by hand by an experienced investigator blind to the condition.

Behavior

For resident/intruder encounters, resident mice were singly housed for at least 2 weeks, were sexually experienced, and were 3–4 weeks older than intruder animals, which were group housed. Saline or CNO (1 mg/kg) was administered intraperitoneally 40 min prior to the start of the encounter. Each mouse received saline and CNO on subsequent days (order of administration was counterbalanced across groups) and encountered a different resident or intruder mouse on each day. Scored social behaviors included anogenital sniffing, oronasal sniffing, following, and grooming. For the three-chamber assay, mice were given 10 min to explore the empty arena and then were briefly returned to their home cage, while the novel object (empty wire pencil cup) was introduced to one chamber. The first mouse (contained in a wire pencil cup) was introduced to the opposite chamber, and the experimental animal was returned to the arena for a 10-min exploration and then briefly removed again, while the novel mouse was added, before a final 10-min exploration. The first 5 min of each exploration period was scored for the time spent in each chamber. For home-cage social encounters, mice implanted with fiber optics were housed with a single littermate. After the implanted mouse was connected to the fiber optic cable, they were allowed a 10-min habituation period, which was not scored, followed by a 5-min baseline period and 5 min of light stimulation (3 Hz, 5 ms, 3 s on, 3 s off).

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures and four figures and can be found with this article online at http://dx.doi.org/10.1016/j.celrep.2016.05.067.

AUTHOR CONTRIBUTIONS

M.E.S. performed the electrophysiology, behavior, and histology experiments. S.M.M. performed the RiboTag experiments and assisted with behavior. L.M.B. and M.E.S. performed the voltammetry experiments. P.E.M.P. and T.S.H. provided resources. M.E.S. and L.S.Z. designed the experiments and wrote the paper.

ACKNOWLEDGMENTS

The authors wish to thank Drs. Richard Palmiter and Stephanie Padilla for helpful discussion and comments on the manuscript, Dr. Christina Sanford for providing the Python script used for projection analysis, and Cerise Knakal for technical assistance. This work was supported by the NIH (R01-MH094536 to L.S.Z, P50-MH106428-5877 to P.E.M.P., and R01-DA036612 to T.S.H).

Received: January 25, 2016
Revised: April 22, 2016
Accepted: May 17, 2016
Published: June 23, 2016

REFERENCES

Armbruster, B.N., Li, X., Pausch, M.H., Herlitze, S., and Roth, B.L. (2007). Evolving the lock to fit the key to create a family of G protein-coupled receptors potently activated by an inert ligand. Proc. Natl. Acad. Sci. USA *104*, 5163–5168.

Borelli, K.G., Blanchard, D.C., Javier, L.K., Defensor, E.B., Brandão, M.L., and Blanchard, R.J. (2009). Neural correlates of scent marking behavior in C57BL/6J mice: detection and recognition of a social stimulus. Neuroscience *162*, 914–923.

Canteras, N.S., Simerly, R.B., and Swanson, L.W. (1992). Projections of the ventral premammillary nucleus. J. Comp. Neurol. *324*, 195–212.

Cavalcante, J.C., Bittencourt, J.C., and Elias, C.F. (2006). Female odors stimulate CART neurons in the ventral premammillary nucleus of male rats. Physiol. Behav. *88*, 160–166.

Cavalcante, J.C., Bittencourt, J.C., and Elias, C.F. (2014). Distribution of the neuronal inputs to the ventral premammillary nucleus of male and female rats. Brain Res. *1582*, 77–90.

Choi, G.B., Dong, H.W., Murphy, A.J., Valenzuela, D.M., Yancopoulos, G.D., Swanson, L.W., and Anderson, D.J. (2005). Lhx6 delineates a pathway mediating innate reproductive behaviors from the amygdala to the hypothalamus. Neuron *46*, 647–660.

Clark, J.J., Sandberg, S.G., Wanat, M.J., Gan, J.O., Horne, E.A., Hart, A.S., Akers, C.A., Parker, J.G., Willuhn, I., Martinez, V., et al. (2010). Chronic microsensors for longitudinal, subsecond dopamine detection in behaving animals. Nat. Methods *7*, 126–129.

Donato, J., Jr., Cavalcante, J.C., Silva, R.J., Teixeira, A.S., Bittencourt, J.C., and Elias, C.F. (2010). Male and female odors induce Fos expression in chemically defined neuronal population. Physiol. Behav. *99*, 67–77.

Donato, J., Jr., Lee, C., Ratra, D.V., Franci, C.R., Canteras, N.S., and Elias, C.F. (2013). Lesions of the ventral premammillary nucleus disrupt the dynamic changes in Kiss1 and GnRH expression characteristic of the proestrus-estrus transition. Neuroscience *241*, 67–79.

Dulcis, D., Jamshidi, P., Leutgeb, S., and Spitzer, N.C. (2013). Neurotransmitter switching in the adult brain regulates behavior. Science *340*, 449–453.

Gunaydin, L.A., Grosenick, L., Finkelstein, J.C., Kauvar, I.V., Fenno, L.E., Adhikari, A., Lammel, S., Mirzabekov, J.J., Airan, R.D., Zalocusky, K.A., et al. (2014). Natural neural projection dynamics underlying social behavior. Cell *157*, 1535–1551.

Hedreen, J. (1978). Neuronal degeneration caused by lateral ventricular injection of 6-hydroxy-dopamine. I. The ventral premammillary nucleus and other cell groups. Brain Res. *157*, 129–135.

Hnasko, T.S., Chuhma, N., Zhang, H., Goh, G.Y., Sulzer, D., Palmiter, R.D., Rayport, S., and Edwards, R.H. (2010). Vesicular glutamate transport promotes dopamine storage and glutamate corelease in vivo. Neuron *65*, 643–656.

Hong, W., Kim, D.W., and Anderson, D.J. (2014). Antagonistic control of social versus repetitive self-grooming behaviors by separable amygdala neuronal subsets. Cell *158*, 1348–1361.

Kevetter, G.A., and Winans, S.S. (1981). Connections of the corticomedial amygdala in the golden hamster. II. Efferents of the “olfactory amygdala”. J. Comp. Neurol. *197*, 99–111.

(G) Activation did not increase investigation of a familiar object in either group.
(H) Pairing one side of a two-chambered box with light stimulation did not cause a significant preference in either group.
(I) Social investigation of a resident male by an intruder male was reduced in DAT-vGlut2 knockout animals (n = 6 animals/group; Student’s t test ***$p < 0.001$).
(J and K) Social investigation of a male (J) or female (K) intruder by a resident was unaffected. Data are represented as mean ± SEM.
See also Figure S4.

Lammel, S., Steinberg, E.E., Földy, C., Wall, N.R., Beier, K., Luo, L., and Malenka, R.C. (2015). Diversity of transgenic mouse models for selective targeting of midbrain dopamine neurons. Neuron *85*, 429–438.

Lee, H., Kim, D.W., Remedios, R., Anthony, T.E., Chang, A., Madisen, L., Zeng, H., and Anderson, D.J. (2014). Scalable control of mounting and attack by Esr1+ neurons in the ventromedial hypothalamus. Nature *509*, 627–632.

Lein, E.S., Hawrylycz, M.J., Ao, N., Ayres, M., Bensinger, A., Bernard, A., Boe, A.F., Boguski, M.S., Brockway, K.S., Byrnes, E.J., et al. (2007). Genome-wide atlas of gene expression in the adult mouse brain. Nature *445*, 168–176.

Leshan, R.L., Louis, G.W., Jo, Y.H., Rhodes, C.J., Münzberg, H., and Myers, M.G., Jr. (2009). Direct innervation of GnRH neurons by metabolic- and sexual odorant-sensing leptin receptor neurons in the hypothalamic ventral premammillary nucleus. J. Neurosci. *29*, 3138–3147.

Lin, D., Boyle, M.P., Dollar, P., Lee, H., Lein, E.S., Perona, P., and Anderson, D.J. (2011). Functional identification of an aggression locus in the mouse hypothalamus. Nature *470*, 221–226.

Meister, B., and Elde, R. (1993). Dopamine transporter mRNA in neurons of the rat hypothalamus. Neuroendocrinology *58*, 388–395.

Motta, S.C., Goto, M., Gouveia, F.V., Baldo, M.V., Canteras, N.S., and Swanson, L.W. (2009). Dissecting the brain's fear system reveals the hypothalamus is critical for responding in subordinate conspecific intruders. Proc. Natl. Acad. Sci. USA *106*, 4870–4875.

Oh, S.W., Harris, J.A., Ng, L., Winslow, B., Cain, N., Mihalas, S., Wang, Q., Lau, C., Kuan, L., Henry, A.M., et al. (2014). A mesoscale connectome of the mouse brain. Nature *508*, 207–214.

Sanz, E., Yang, L., Su, T., Morris, D.R., McKnight, G.S., and Amieux, P.S. (2009). Cell-type-specific isolation of ribosome-associated mRNA from complex tissues. Proc. Natl. Acad. Sci. USA *106*, 13939–13944.

Scalia, F., and Winans, S.S. (1975). The differential projections of the olfactory bulb and accessory olfactory bulb in mammals. J. Comp. Neurol. *161*, 31–55.

Sosulski, D.L., Bloom, M.L., Cutforth, T., Axel, R., and Datta, S.R. (2011). Distinct representations of olfactory information in different cortical centres. Nature *472*, 213–216.

Spitzer, N.C. (2015). Neurotransmitter Switching? No Surprise. Neuron *86*, 1131–1144.

Swanson, L.W. (2000). Cerebral hemisphere regulation of motivated behavior. Brain Res. *886*, 113–164.

Unger, E.K., Burke, K.J., Jr., Yang, C.F., Bender, K.J., Fuller, P.M., and Shah, N.M. (2015). Medial amygdalar aromatase neurons regulate aggression in both sexes. Cell Rep. *10*, 453–462.

Vandenberg, M.J., Terhorst, G.J., and Koolhaas, J.M. (1983). The nucleus premammillaris ventralis (Pmv) and aggressive-behavior in the rat. Aggress. Behav. *9*, 41–47.

Wu, Z., Autry, A.E., Bergan, J.F., Watabe-Uchida, M., and Dulac, C.G. (2014). Galanin neurons in the medial preoptic area govern parental behaviour. Nature *509*, 325–330.

Yoon, H., Enquist, L.W., and Dulac, C. (2005). Olfactory inputs to hypothalamic neurons controlling reproduction and fertility. Cell *123*, 669–682.

Yu, Q., Teixeira, C.M., Mahadevia, D., Huang, Y., Balsam, D., Mann, J.J., Gingrich, J.A., and Ansorge, M.S. (2014). Dopamine and serotonin signaling during two sensitive developmental periods differentially impact adult aggressive and affective behaviors in mice. Mol. Psychiatry *19*, 688–698.

Zhuang, X., Masson, J., Gingrich, J.A., Rayport, S., and Hen, R. (2005). Targeted gene expression in dopamine and serotonin neurons of the mouse brain. J. Neurosci. Methods *143*, 27–32.

Zoli, M., Agnati, L.F., Tinner, B., Steinbusch, H.W., and Fuxe, K. (1993). Distribution of dopamine-immunoreactive neurons and their relationships to transmitter and hypothalamic hormone-immunoreactive neuronal systems in the rat mediobasal hypothalamus. A morphometric and microdensitometric analysis. J. Chem. Neuroanat. *6*, 293–310.


OUR
NETWORK
IS YOUR
NETWORK


Powered by
people in the know.
Like you.

Cell Reports
Report

OPEN ACCESS
CellPress

The Adipose Transcriptional Response to Insulin Is Determined by Obesity, Not Insulin Sensitivity

Mikael Rydén,[1] Olga Hrydziuszko,[2] Enrichetta Mileti,[3] Amitha Raman,[1] Jette Bornholdt,[4] Mette Boyd,[4] Eva Toft,[1,5] Veronica Qvist,[1,5] Erik Näslund,[6] Anders Thorell,[6,7] Daniel P. Andersson,[1] Ingrid Dahlman,[1,5] Hui Gao,[8] Albin Sandelin,[4] Carsten O. Daub,[3,*] and Peter Arner[1,9,*]
[1]Department of Medicine, Karolinska University Hospital Huddinge, Karolinska Institutet, 141 86 Stockholm, Sweden
[2]Bioinformatics Short-term Support and Infrastructure (BILS), Science for Life Laboratory, Tomtebodavägen 23A, 171 65 Solna, Sweden
[3]Department of Biosciences and Nutrition, Science for Life Laboratory, Karolinska Institutet, 141 83 Huddinge, Sweden
[4]The Bioinformatics Centre, Department of Biology, Biotech Research and Innovation Centre, University of Copenhagen, 2200 Copenhagen N, Denmark
[5]Department of Medicine, Ersta Hospital, Karolinska Institutet, 116 91 Stockholm, Sweden
[6]Department of Clinical Sciences, Danderyd Hospital, Karolinska Institutet, 182 88 Stockholm, Sweden
[7]Department of Surgery, Ersta Hospital, Karolinska Institutet, 116 91 Stockholm, Sweden
[8]Department of Biosciences and Nutrition, Karolinska Institutet, 141 86 Stockholm, Sweden
[9]Lead Contact
*Correspondence: carsten.daub@ki.se (C.O.D.), peter.arner@ki.se (P.A.)
http://dx.doi.org/10.1016/j.celrep.2016.07.070

SUMMARY

Metabolically healthy obese subjects display preserved insulin sensitivity and a beneficial white adipose tissue gene expression pattern. However, this observation stems from fasting studies when insulin levels are low. We investigated adipose gene expression by 5′Cap-mRNA sequencing in 17 healthy non-obese (NO), 21 insulin-sensitive severely obese (ISO), and 30 insulin-resistant severely obese (IRO) subjects, before and 2 hr into a hyperinsulinemic euglycemic clamp. ISO and IRO subjects displayed a clear but globally similar transcriptional response to insulin, which differed from the small effects observed in NO subjects. In the obese, 231 genes were altered; 71 were enriched in ISO subjects (e.g., phosphorylation processes), and 52 were enriched in IRO subjects (e.g., cellular stimuli). Common cardio-metabolic risk factors and gender do not influence these findings. This study demonstrates that differences in the acute transcriptional response to insulin are primarily driven by obesity per se, challenging the notion of healthy obese adipose tissue, at least in severe obesity.

INTRODUCTION

Up to 30% of obese subjects display normal fasting plasma glucose/lipid levels and normotension, a phenotype referred to as ‘‘metabolically healthy obesity,’’ which implies that a significant proportion of obese individuals may need less vigorous interventions to avoid metabolic/cardiovascular complications (Blüher, 2010; Karelis, 2008; Primeau et al., 2011; Samocha-Bonet et al., 2012; Sims, 2001). A hallmark characteristic among these individuals is high insulin sensitivity. Several studies have shown that insulin-sensitive obese (ISO) subjects have lower visceral fat accumulation, less ectopic fat and arterial atherosclerosis, higher plasma adiponectin levels, and a more favorable inflammation profile than insulin-resistant obese (IRO) individuals (Blüher, 2010; Karelis, 2008; Primeau et al., 2011; Samocha-Bonet et al., 2012; Xu et al., 2013). It is also well established that the two obesity phenotypes differ in the subcutaneous white adipose tissue (sWAT) itself (Xu et al., 2013). ISO individuals have smaller fat cells and less pronounced inflammation than IRO individuals, which is also reflected at the gene expression level (Elbein et al., 2011; Qatanani et al., 2013). However, the transcriptional profiles of sWAT have been investigated in the fasting state, when insulin levels are low (Elbein et al., 2011; Qatanani et al., 2013). As insulin is expected to induce profound alterations in gene expression, it is not clear how such changes relate to insulin sensitivity and clinical profiles. This has prompted some investigators to determine the transcriptional response to insulin in sWAT collected before and during hyperinsulinemic euglycemic clamp for 6 hr. Comparisons in limited numbers of lean ISO and IRO subjects have reported some between-group differences (Soronen et al., 2012; Westerbacka et al., 2006). Although relevant, these studies were not designed to address the transcriptional response to insulin in subjects matched for BMI. Thus, in order to fully evaluate the idea of a healthy obese state, insulin responses need to be determined in obese subjects discordant in insulin sensitivity and ideally compared with those in healthy non-obese (NO) subjects.

Although ISO and IRO individuals display different clinical phenotypes, it has been a matter of debate whether they also confer different risks for cardiovascular morbidity and/or mortality (Flint et al., 2010; Lu et al., 2014; Ortega et al., 2013; Song et al., 2007). In fact, several recent meta-analyses have refuted the notion that ‘‘healthy obesity‘‘or preserved insulin sensitivity protects against cardiometabolic complications (Fan et al., 2013; Kramer et al., 2013; Roberson et al., 2014). These controversies prompted us

Table 1. Clinical Characteristics of the Study Groups

				p Value			
Variable	ISO (n = 21; M/F, 0/21)	IRO (n = 30; M/F, 6/24)	NO (n = 17; M/F, 3/14)	Chi-Square Test	ISO versus IRO	NO versus ISO	NO versus IRO
Menopause, yes/no	6/15	11/13	4/10	0.40	–	–	–
Nicotine use, yes/no	2/19	4/26	1/16	0.70	–	–	–
Age (years)	41 ± 12	45 ± 11	42 ± 13		0.22	0.75	0.42
BMI (kg/m^2)	39 ± 3	39 ± 5	24 ± 3		0.81	<0.0001	<0.0001
Waist-to-hip ratio	0.93 ± 0.07	1.00 ± 0.07	0.88 ± 0.05		0.0007	0.01	<0.0001
Total fat (kg)	55 ± 7	53 ± 9	22 ± 8		0.37	<0.0001	<0.0001
fP-Glucose (mmol/l)	5.1 ± 0.4	5.9 ± 2.0	5.0 ± 0.4		0.041	0.85	0.035
fP-Insulin (mU/l)	9.7 ± 5.1	18.9 ± 9.5	6.0 ± 3.3		<0.0001	0.13	<0.0001
fP-Cholesterol (mmol/l)	4.9 ± 0.7	5.2 ± 1.2	4.5 ± 0.9		0.19	0.35	0.027
fP-HDL cholesterol (mmol/l)	1.4 ± 0.3	1.2 ± 0.3	1.5 ± 0.3		0.044	0.12	0.0006
fP-Triglycerides (mmol/l)	1.1 ± 0.5	1.5 ± 0.8	0.9 ± 0.3		0.010	0.24	0.0004
Systolic blood pressure (mmHg)	134 ± 14	142 ± 15	120 ± 13		0.044	0.0035	<0.0001
Diastolic blood pressure (mmHg)	82 ± 15	83 ± 11	76 ± 10		0.76	0.17	0.083
Resting pulse rate (beats per minute)	68 ± 12	74 ± 13	62 ± 11		0.074	0.14	0.0016
M value (mg/kg · min)	6.2 ± 0.9	3.1 ± 0.9	8.6 ± 2.0		<0.0001	<0.0001	<0.0001
Mean fP-insulin level during clamp (60–120 min, mU/l)	223 ± 43	258 ± 56	202 ± 32		0.011	0.18	0.0003

Values are given as actual numbers or means ± SD. They were compared by Student's t test or chi-square test. ISO, insulin-sensitive obese; IRO, insulin-resistant obese; M, males; F, females; fP, fasting plasma.

to investigate how gene expression in the sWAT of ISO individuals responds to insulin stimulation (hyperinsulinemia). Assuming that the ISO group displays a more "beneficial" gene expression profile in the fasting state compared to the IRO group, we hypothesized that the differences in transcriptional response profiles between the ISO and IRO groups would be more pronounced upon hyperinsulinemia and that the ISO group would be more similar to the NO group than the IRO group. Therefore, we assessed global transcriptional profiles in sWAT from healthy NO subjects and from obese subjects subdivided into ISO and IRO groups according to hyperinsulinemic euglycemic clamp measures. Subcutaneous WAT biopsies were taken before and at the end of the 2-hr clamp. We chose this short period of hyperinsulinemia in order to evaluate direct transcriptional effects of insulin, assuming that longer duration of stimulation may cause secondary effects on gene transcription. Samples were analyzed using global transcriptional profiling with the 5′cap analysis of gene expression (CAGE) (Takahashi et al., 2012). CAGE is based on sequencing the 5′ end of mRNA, thereby assessing the transcriptional start sites (TSS) and their usage (expression) with high resolution and reproducibility (Kawaji et al., 2014). Adjacent TSSs for presumably the same transcripts are collapsed into "tag clusters" corresponding to gene promoters (Frith et al., 2008). Individual genes typically have several tag clusters depending on tissue type (Carninci et al., 2006; Forrest et al., 2014). CAGE provides less biased results than conventional expression arrays, allows for more in-depth analysis, and also detects uncharacterized novel gene transcripts (Suzuki et al., 2009). However, in order to obtain biologically interpretable insights, the present work was focused on analyses of tag clusters corresponding to annotated genes.

RESULTS

Cohort Description

Clinical data are summarized in Table 1. As expected, there were major differences in the clinical profile between the NO group and the obese groups, in particular for the IRO group. Compared with the ISO group, the IRO group displayed significantly higher values for waist-to-hip ratio, systolic blood pressure, fasting plasma glucose, insulin, and triglycerides, as well as lower HDL (high-density lipoprotein)-cholesterol. Mean plasma insulin levels during clamp were also slightly higher in the IRO group. M values in the NO group were, on average, $\sim$40% higher than those in the ISO group, but there was a considerable overlap (Figure 1A).

Gene Expression Profiles Are Altered after Insulin Stimulation for 2 hr in All Subjects

Taking into account the expression data from all the subjects put together, there was a clear overall expression response to hyperinsulinemia, including several genes involved in insulin signaling. Using a false discovery rate (FDR) of <0.05, we found 786 CAGE tag clusters to be altered during the clamp (expressed as fasting divided by hyperinsulinemia; f/hi) (Figure 1B). Out of these, 493 corresponded to annotated genes, and there was a more pronounced transcriptional upregulation (641 out of 786 tag clusters; 82%) upon hyperinsulinemia (i.e., the f/hi quotient was decreased). Eight insulin responding tag clusters (Figure 1B) corresponded to eight genes in the insulin signaling pathway and included *PPP1R3B*, *PPP1R3C*, *PIK3R*, and *IRS2* (Figure 1C). A subset of these genes (*SREBF1*, *PIK3R3*, and *IRS2*) was validated by qPCR in the

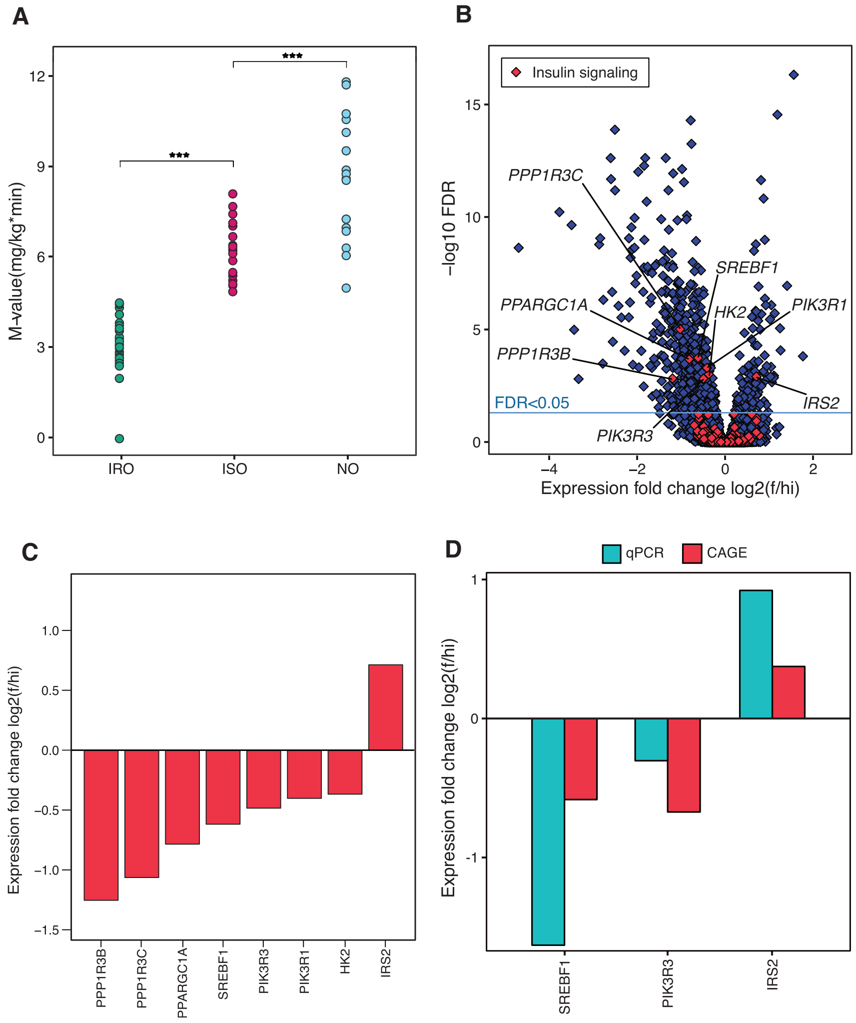


Figure 1. Clamp Validation

(A) Individual M values are detailed for the NO, ISO, and IRO groups. ***$p < 0.001$.

(B) Volcano plot of the 30,331 detected CAGE tag clusters, corresponding to 15,518 genes, expressed as log fold change fasting/hyperinsulinemia (f/hi) along the x axis and false discovery rate (FDR) on the y axis. In blue: 29,964 tag clusters (15,398 genes) responding to the clamp. In red: 367 tag clusters (122 genes) involved in the insulin signaling pathway. The eight tag clusters corresponding to eight genes involved in the insulin signaling pathway are highlighted. The horizontal line shows an FDR < 0.05.

(C) Expression fold changes (fasting/hyperinsulinemia; f/hi) of the eight insulin signaling pathway genes shown in (B).

(D) qPCR validation of selected genes from (C) in the NO group expressed as f/hi. Paired two-sided t tests gave p values between <0.0001 and 0.015.

NO group and confirmed the findings obtained by CAGE (Figure 1D).

Global Gene Expression Profiles in NO, ISO, and IRO Subjects in the Fasting and Hyperinsulinemic States

In the present dataset, it is possible to compare the groups in several different ways, as outlined in Figures S1A–S1C. Principal-component analysis (PCA) for all differentially expressed tag clusters between groups and conditions, corresponding to the comparisons in Figure S1B, showed a clear distinction between the NO group and both obese groups (Figures S2A and S2B). As expected, the ISO and IRO groups in the fasting state displayed a higher expression of genes in pro-inflammatory pathways compared with the NO group (data not shown). When focusing on the actual insulin response, i.e., the changes from fasting to hyperinsulinemia (f/hi; see comparison in Figure S1C), the majority of the differentially expressed tag clusters between the ISO group (246 out of 295; 83%) and the IRO group (190 out of 246; 77%) were upregulated (f/hi; Figure 2A). Surprisingly, after correction for multiple testing, insulin altered only the expression of four tag clusters in the NO group, two of which were not annotated, while two represented the genes *FRMD6-AS2* and *NPC1* (Figure 2A). The effect on *NPC1* expression was confirmed by qPCR (Figure S2C). PCA of data from the fasting and hyperinsulinemic states showed that the two obese groups were indistinguishable but that both were clearly separated from the NO group (Figure 2B). Principal component (PC) 1 primarily reflected the insulin effect (Figure 2C). In contrast, in the fasting state, PC2 separated the NO group from the obese groups but not the IRO group from the ISO group (Figure 2D). Upon hyperinsulinemia, there was no further change in the NO group, while the ISO and IRO groups were altered but remained different from the NO group (Figure 2D).

Analysis of Insulin-Induced Genes in ISO and IRO Subjects

As indicated for insulin responses in Figure 2 and further supported by the global comparisons in Figure S2, the predominant difference between the controls and the IRO and ISO groups appeared to be obesity per se. This made it difficult to assess the possible differences between the ISO and IRO groups when including all three groups. To further evaluate what drives the differences in insulin-induced transcriptional response in obesity, we performed a multiple regression analysis of the data from the obese groups from Figure 2A in relation to individual insulin sensitivity (i.e., M value) and other, possibly, contributing factors. This demonstrated that the M value and BMI contributed to the variations. However, the results were not influenced by gender or common risk factors such as waist-to-hip ratio, fasting lipid/insulin levels, pulse rate, or blood pressure (Table 2), indicating that obesity and insulin sensitivity, rather than associated cardio-metabolic risk factors, explains the differences in gene

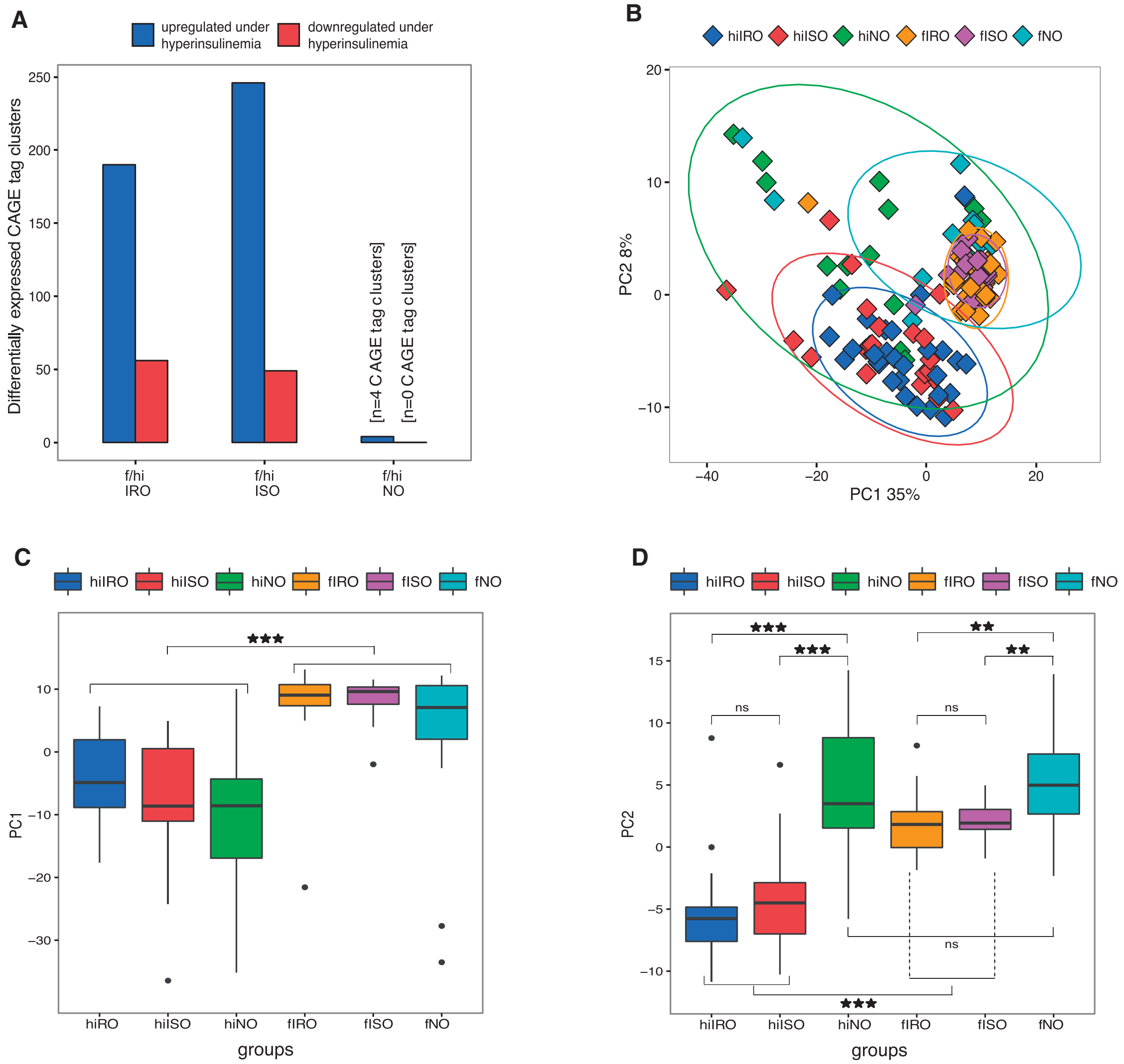


Figure 2. Overall Differences in Insulin-Induced Gene Expression among the NO, ISO, and IRO Groups

(A) The number of differentially expressed CAGE tag clusters altered by fasting/hyperinsulinemia (f/hi) according to the way of comparison in Figure S1C.

(B) Principal-component (PC) analysis plot summarizing the high-dimensional transcriptional data from (A) for NO, ISO, and IRO groups along PC1 and PC2, explaining the most variance in the data. Circles represent 95% confidence intervals.

(C and D) Statistical analyses (t test) of group differences in (B) along PC1 (C) and PC2 (D). **$p < 0.01$; ***$p < 0.001$; ns, not significant.

See also Figure S1C.

expression upon hyperinsulinemia. A comparison of the two obese groups showed that 380 tag clusters (FDR < 0.05), corresponding to 231 genes, were altered in hyperinsulinemia (Figure 3A; Table S1). Almost half were responding in both the ISO and IRO groups (161 tag clusters, 42%; 108 genes, 47%), and among these, the degree of upregulation was more pronounced in the ISO group (116 out of the 133 upregulated tag clusters, 87%; Figure 3B). The 231 genes mapped to KEGG (Kyoto Encyclopedia of Genes and Genomes) pathways involved in, e.g., metabolic function, insulin signaling/resistance, MAPK signaling, circadian rhythm, and cancer (Figure 3C). Genes in these pathways were found among common as well as ISO- and IRO-group-associated genes. In order to gain some further insight into the biological meaning of the findings, a gene ontology (GO) analysis focusing on biological processes was also performed. This revealed that the top five most significantly enriched GO terms ($p < 0.05$) were response to lipid, cellular response to lipid, response to organic cyclic compound, response to steroid hormone, and fat cell differentiation (Table S2). Because, in the present article, we focused on the effects of obesity and insulin resistance, we used these two MeSH (Medical Subject Headings) terms in a PubMatrix (http://pubmatrix.irp.nia.nih.gov/) search of the genes listed in Table S1. This revealed 34 ISO- and 27 IRO-enriched genes, as well as 48 common genes (Figure 3D). The genes (down- or upregulated) displaying the most pronounced differences between the

Table 2. Correlations between Clinical Parameters and Individual Changes in Overall Gene Expression during Hyperinsulinemia, PC1, in Obese Subjects

Variable	Estimate	SE	t Value	p Value
Intercept	−37.51	19.23	1.95	0.0583
M value	−3.01	1.22	−2.47	0.0181
BMI	1.33	0.46	2.91	0.0060
fP-Cholesterol	1.15	1.73	0.66	0.5122
fP-HDL cholesterol	6.24	5.60	1.11	0.2720
fP-Triglycerides	−3.48	2.50	−1.39	0.1722
fP-Insulin	0.03	0.28	0.10	0.9196
Systolic blood pressure	0.04	0.13	0.31	0.7609
Diastolic blood pressure	−0.16	0.11	−1.46	0.1516
Pulse rate	−0.03	0.13	−0.22	0.8263
Gender	−4.13	5.43	−0.76	0.45

Multiple regression was used. For the whole model, $r^2 = 0.45$ and $p = 0.0045$. PC1, principal component 1; fP, fasting plasma; HDL, high-density lipoprotein.

ISO and IRO groups are displayed in Figures 3E and 3F. Among those most prominent in ISO individuals were *RORC*, *RPGR*, *KLF9*, *IRS2*, and *DDIT4*. Those in the IRO group included *PPARGC1* and *TAGAP*. Expression changes of these seven genes were confirmed by qPCR (Figure S3). The five most significant and biologically relevant GO processes for the genes identified in Figure 3D are summarized in Table 3. The common ones belonged to biological pathways involved in lipid metabolism and cell differentiation (e.g., *NPC1*, *PPARGC1B*, and *CEBPD*). The ISO-enriched genes (e.g., *IRS2*, *TRIB1*, and *MIDN*) were primarily involved in phosphorylation processes, while the genes significantly altered in the IRO group (e.g., *PIK3R1*, *FOS*, and *PPARGC1A*) were associated with fat cell differentiation and cellular stimuli such as responses to endogenous/exogenous factors.

Influence of Gender

Most of the included subjects were female, but there were men in the NO and IRO groups. Exclusion of male subjects did not impact our findings in a major way, as exemplified by PCA plots (Figures S2D and S2E).

DISCUSSION

Herein, we report findings on global transcriptional profiles in the sWAT of NO and obese subjects, using a sequencing method that allows for broader analyses than conventional expression arrays (Kawaji et al., 2014). We found a clear transcriptional response to insulin in our global analysis. However, after correction for multiple testing, this was only significant among the obese individuals except for four tag clusters in the NO group. Thus, NO individuals may regulate their short-term insulin response in WAT in a different way, possibly via post-transcriptional mechanisms. The small effects in NO also exclude the possibility that we used an insulin infusion rate during clamp that was too high, potentially masking differences between the ISO and IRO groups. Because of the strong influence of obesity per se, it was necessary to omit the NO group in the analyses in order to allow identification of the differences between the ISO and IRO groups. Around half of the genes were common for the ISO and IRO groups, although the magnitude of the insulin response was more pronounced for the upregulated genes in the ISO group, most probably reflecting these subjects' higher level of insulin sensitivity.

Our analyses identified 108 insulin-responding genes that were common and 123 that were enriched in either the ISO or the IRO grooup. This shows that there are small but quantifiable differences between the ISO and IRO groups in the adipose response to insulin at the gene expression level. Notably, the obesity/insulin resistance-linked genes (i.e., with at least one publication in the literature) constituted less than half (109 out of 231) of all the insulin-responsive genes in obesity (individual genes exemplified in Results). The remaining 122 genes may be of interest in future studies of the pathophysiological mechanisms of insulin action in human WAT. Nevertheless, it should be stressed that, although statistically significant, the differences between the ISO and IRO groups were small and primarily of quantitative nature, supporting the conclusion that the ISO and IRO groups display a strong similarity, at least in the short-term response to insulin.

The observation that the two obese phenotypes displayed very similar insulin responses could provide a clue for why ISO subjects display an increased morbidity and cardiovascular risk, compared with NO individuals (Fan et al., 2013; Kramer et al., 2013; Roberson et al., 2014). The sWAT in ISO individuals may simply not be as metabolically normal as previously believed. Our present findings are in line with results following weight reduction induced by low-calorie diet (Viguerie et al., 2012). This study showed similar improvements in sWAT gene expression in obese subjects with or without the metabolic syndrome. Thus, the transcriptional control upon either hyperinsulinemia or weight loss appears to be very similar in ISO and IRO subjects.

We categorized the obese into ISO or IRO based solely on insulin sensitivity. Currently, there is no consensus on how to define a "metabolically healthy obese," phenotype and various scoring strategies have been used, as reviewed (Blüher, 2010; Karelis, 2008; Primeau et al., 2011; Samocha-Bonet et al., 2012; Xu et al., 2013). Nevertheless, as mentioned earlier, the cardio-metabolic risk profile had no influence on our findings.

The obese subjects were subdivided based on the 25th upper percentile of insulin sensitivity (Hoffstedt et al., 2010). However, this has no important bearing on the results, as we obtained similar findings using individual insulin sensitivity values as a continuous variable (in multiple regression analyses). There is no consensus on how to express clamp data. We used the most common way, i.e., M value/body weight. However, correcting M values for lean body mass yielded virtually the same results. Insulin levels during clamp were slightly higher in the IRO group than in the ISO group (Table 1). These small differences can hardly have influenced our findings on gene expression. Furthermore, as mentioned earlier, fasting insulin levels had no bearing on the results.

In the present work, we obtained sequencing-based data on 136 samples from 68 individuals. This can be regarded as very

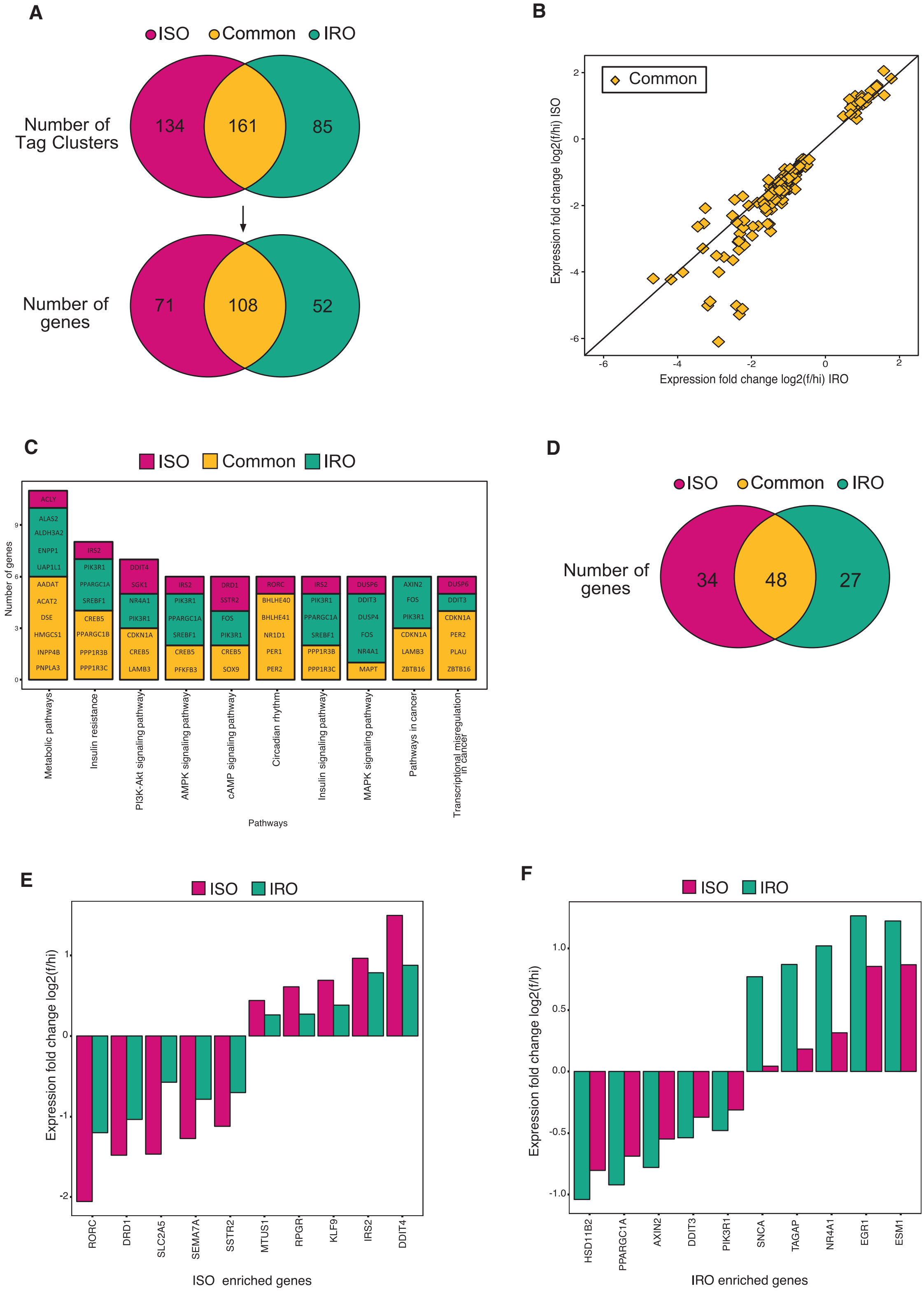


Figure 3. Identification of Genes Altered by Hyperinsulinemia in ISO and IRO Groups

(A) Venn diagram of tag clusters and genes significantly altered by hyperinsulinemia in the ISO or IRO group.

(B) Comparison between expression fold change (fasting/hyperinsulinemia; f/hi) of tag clusters common in ISO and IRO groups.

(legend continued on next page)

Table 3. Gene Ontology Analysis of Insulin-Induced Genes in ISO and IRO Groups

GO Term	Description	p Value	Genes	Type
GO:0071396	cellular response to lipid	1.8E-2	HDAC5, KLF9, RORC, SGK1, SSTR2, TRIB1, VDR	ISO
GO:0071383	cellular response to steroid hormone stimulus	1.8E-2	KLF9, RORC, SGK1, SSTR2, VDR	ISO
GO:0033673	negative regulation of kinase activity	1.7E-2	DRD1, DUSP6, IRS2, LRP5, MIDN, TRIB1	ISO
GO:0045963	negative regulation of phosphate metabolic process	1.5E-3	APOC1, DDIT4, DRD1, DUSP6, IRS2, LRP5, MIDN, MYO1D,TRIB1	ISO
GO:0010563	negative regulation of phosphorus metabolic process	1.5E-3	APOC1, DDIT4, DRD1, DUSP6, IRS2, LRP5, MIDN, MYO1D,TRIB1	ISO
GO:0010035	response to inorganic substance	7.0E-4	FABP4, FOS, NR4A1, PPARGC1A, SLC1A3, SLC30A1, TNFAIP3	IRO
GO:0042493	response to drug	4.0E-3	EGR1, FOS, HSD11B2, PPARGC1A, SLC1A3, SNCA, SREBF1	IRO
GO:005159	response to cAMP	3.0E-2	EGR1, FOS, PIK3R1, SREBF1	IRO
GO:0071869	response to catecholamine	3.0E-2	EGR1, SNCA, PPARGC1A	IRO
GO:0045444	fat cell differentiation	1.0E-3	DDIT3, ENPP1, FABP4, NR4A1, SREBF1, PPARGC1A	IRO
GO:0045444	fat cell differentiation	8.0E-6	AACS, CEBPD, CREB5, LGALS12, NR1D1, PER2, SNAI2, ZBTB16, ZC3H12A	common
GO:0032330	regulation of chondrocyte differentiation	3.0E-4	CTGF, RARG, SNAI2, SOX9, ZBTB16	common
GO:0048545	response to steroid hormone	3.0E-3	AACS, ABCC2, CDKN1A, CTGF, ERRFI1, NPC1, NR1D1, PPARGC1B, RARG	common
GO:0071396	cellular response to lipid	1.0E-6	AACS, ADAMTS1, ERRFI1, HMGCS1, NPC1, NR1D1, PDK4, PLAU, RARG, SNAI2, SOX9, ZC3H12A	common
GO:0001503	ossification	1.0E-3	BCOR, CEBPD, COL5A2, CTGF, PPARGC1B, SKI, SNAI2, SOX9, ZBTB16	common

The top five most significant and biologically relevant Gene Ontology (GO):Biological processes are listed for the genes identified in Figure 3D. The p values are Bonferroni corrected in the GO analysis. ISO, insulin-sensitive obese; IRO, insulin-resistant obese.

large numbers, using a genome-wide sequencing technique by today's standards. Moreover, based on results from a smaller study on 40 subjects using CAGE (Persson et al., 2015), our present cohort was sufficiently large to detect gene expression differences between the three groups. Despite this, we did not have sufficient statistical power to allow a subgroup analysis of, e.g., the influence of different age groups.

There are some caveats with the present study. Because the obese subjects were scheduled to undergo bariatric surgery, our results may only pertain to individuals with severe obesity. We investigated solely sWAT as it is, for obvious ethical reasons, virtually impossible to obtain visceral WAT biopsies during clamp. Still, there is no obvious reason to believe that regional differences in gene expression are important for the present findings. In fact, previous studies of subcutaneous and visceral WAT have found that the impact of obesity and metabolic status is very similar in the two regions in the fasting state (Klimcáková et al., 2011). Moreover, studies of visceral WAT would require intraoperative fat biopsies during general anesthesia. It has been convincingly demonstrated that this procedure induces an acute insulin-resistant state immediately following the incision of the intra-abdominal wall (Felländer et al., 1994). This makes valid assessments of acute insulin responses in any target tissue during general surgery uncertain. Another aspect is the possibility that the gene expression response during fasting and hyperinsulinemia may differ between sWAT and tissues such as skeletal muscle and liver. However, previous comparisons have shown that differences in gene expression between BMI-matched individuals discordant in insulin sensitivity are more pronounced in sWAT than in skeletal muscle (Elbein et al., 2011). Furthermore,

(C) KEGG pathways to which insulin-responding genes from (A) are mapping.

(D) Number of genes corresponding to genes in (A) after the PubMatrix filtering step described in Results.

(E and F) Top differentially altered genes in the ISO (D) and IRO (E) groups. The y axis shows the expression fold change induced by insulin expressed as f/hi.

with regard to influence of gender, the number of men was too small to allow a detailed comparison in this respect. Again, for lack of statistical power, we had to keep the male subjects in the analyses. Nevertheless, gender distribution did not influence our correlation analysis, and PCA of insulin gene responses in women was similar to that in all subjects (Figures S2D and S2E). Therefore, we do not believe that sex impacts on our results to any significant degree. Finally, because both pre- and postmenopausal women were included, we cannot exclude that menopausal state may influence our results. Nevertheless, as evident from Table 1, the three groups were well matched for this factor.

In summary, obese subjects with preserved insulin sensitivity have globally almost the same WAT gene expression response to insulin as BMI- and body-fat-matched insulin-resistant individuals. This is independent of cardio-metabolic risk factors, thereby questioning the notion of a "healthy obese state," at least in the sWAT of women with severe obesity. A comparison with NO subjects confirms that the major factor explaining the differences in short-term insulin response is obesity per se. Nevertheless, small but clear quantitative differences in the transcriptional response to insulin in the ISO and IRO groups are observed that are linked to specific biological pathways involving, e.g., phosphorylation processes, cellular stimuli, and fat cell differentiation. The clinical and pathophysiological relevance of these differences will be addressed in additional studies, which we encourage by making the full tag cluster dataset publically available. The dataset can also be used for future studies addressing other gene regulatory mechanisms, e.g., those facilitated by specific gene promoters, enhancer usage, and long non-coding RNAs.

EXPERIMENTAL PROCEDURES

Subjects

The 51 obese subjects were enrolled in a clinical trial studying the outcome of gastric bypass surgery (ClinicalTrials.gov ID NCT01727245). From one obese patient, sWAT RNA was of insufficient quality, preventing accurate analyses. Hence, only CAGE data from the remaining 50 subjects are reported herein. We also recruited 17 healthy never-obese subjects. In this group, RNA quality was adequate in samples from 15 subjects. Investigations were performed in the morning after an overnight fast. Height, weight, hip and waist circumference, resting pulse rate, blood pressure, and total body fat content by dual X-ray absorptiometry (DEXA) were determined (Arner et al., 2015). Venous blood was obtained and analyzed by The Karolinska University Hospital's accredited routine clinical chemistry laboratory. Abdominal sWAT biopsies were obtained from the paraumbilical region by needle aspiration under local anesthesia. Thereafter, a hyperinsulinemic, euglycemic clamp was performed as described previously (Hagström-Toft et al., 2001). An intravenous bolus dose of insulin (1.6 U/m^2 body surface area; Actrapid, Novo Nordisk) was given, followed by intravenous infusion of insulin (0.12 U/m^2 min) for 120 min. Plasma glucose values were measured in duplicate every fifth minute (HemoCue). Euglycemia was maintained between 4.5 and 5.5 mmol/l (81 and 99 mg/dl) by a variable intravenous infusion of glucose (200 mg/ml). The mean glucose infusion rate (glucose disposal) between 60 and 120 min was determined (M value, milligrams of glucose uptake per kilograms of body weight per minute). Mean plasma insulin at 60 and 120 min of clamp was calculated. The second sWAT biopsy was obtained from the contralateral paraumbilical side during the last 5 min of the clamp.

ISO was predefined based on data from a previously published study of obese women as the 25th upper percentile of the M value (Hoffstedt et al., 2010). This gave a cutoff value of 4.65 mg/kg/min. Based on this figure, the 50 subjects from which CAGE data could be obtained were divided into two subgroups: 21 as ISO and 29 as IRO. Expressing M values per lean body mass instead of total body weight resulted in an identical subdivision, except that one subject was reclassified from IRO to ISO. Three ISO and nine IRO patients were on pharmacotherapy against hypertension. One IRO patient had diet-/lifestyle-treated type 2 diabetes. The study was approved by the regional ethics board in Stockholm. Informed written consent was obtained before enrollment.

qPCR

RNA isolation, cDNA synthesis, real-time qPCR, and analysis of gene expression (using the ΔΔCt method) were performed as described previously (Gao et al., 2014). Details on TaqMan probes (Applied Biosystems) and SyBR primers (some of which were designed by us but provided by Sigma-Aldrich) are available upon request. Gene expression was normalized to *18S* rRNA or *LRP10*.

5′Cap Analysis of Gene Expression Profiling and Data Processing

Total adipose RNA was extracted using the RNeasy Lipid Tissue Mini Kit (QIAGEN), followed by RNA up-concentration, which was measured using a Nanodrop ND-1000 (Thermo Fisher Scientific). RNA quality was determined using the bioanalyzer RNA 6000 Pico Kit (Agilent Technologies). CAGE libraries were prepared as described previously (Takahashi et al., 2012), with an input of 1,000 ng total RNA. Samples were run individually. Four CAGE libraries with different barcodes were pooled prior to sequencing and applied to the same sequencing lane. Libraries were prepared in a random order to avoid systematic errors, as described previously (Takahashi et al., 2012), and sequenced using Illumina Hi-Seq 2500 or 2000. Sequenced reads were mapped to the human genome using Bowtie (Langmead et al., 2009). Nearby mapped reads on the same strand were merged into 35,639 tag clusters (genomic regions) using Paraclu (Frith et al., 2008). Tag per million (TPM) normalized read counts in these clusters from the 136 RNA samples formed the expression data matrix of interest. Tag clusters with an expression ≥ 0.5 TPM in at least 20% of the samples were maintained, resulting in 30,331 tag clusters. ENSMBL genome annotation (Kersey et al., 2014) was used to assign the tag clusters to a total of 15,518 genes. All raw data are uploaded at https://export.uppmax.uu.se/b2013047/CellReportsTables/.

Bioinformatic Analyses

Pathway and gene ontology analyses were performed using standard webtools, including KEGG Mapper (http://www.genome.jp/kegg/tool/map_pathway1.html) and ToppFun (https://toppgene.cchmc.org).

Statistical Methods

Unless otherwise stated, values are means ± SD and compared by an unpaired two-sided t test assuming unequal variances. One-sided tests were used in qPCR validation experiments comparing ISO and IRO subjects. Nominal parameters were compared by chi-square test. Specific for CAGE data, the PCA scores were tested using an unpaired t test assuming unequal variance. Multiple regression and differential expression analyses for CAGE data were performed using global linear models (GLMs) implemented in edgeR (Robinson et al., 2010), and significance was determined by Benjamini-Hochberg-corrected FDR.

SUPPLEMENTAL INFORMATION

Supplemental Information includes three figures and two tables and can be found with this article online at http://dx.doi.org/10.1016/j.celrep.2016.07.070.

AUTHOR CONTRIBUTIONS

Conception and Design, P.A., C.O.D., and M.R.; Analysis and Interpretation of the Data, P.A., C.O.D., O.H., E.M., M.R., and A.S.; Drafting of the Article, P.A., C.O.D., O.H., E.M., and M.R.; Critical Revision of the Article for Important Intellectual Content, P.A., C.O.D., E.M., and M.R.; Final Approval of the Article, all co-authors; Collection of Data, D.P.A., J.B., M.B., I.D., H.G., E.M., E.N., A.R., V.Q., A.T., and E.T.

ACKNOWLEDGMENTS

We thank nurses Yvonne Widlund, Britt-Marie Leijonhufvud, and Katarina Hertel, as well as laboratory engineers Eva Sjölin and Elisabeth Dungner for excellent technical assistance. The computations were performed on resources provided by SNIC through Uppsala Multidisciplinary Center for Advanced Computational Science (UPPMAX) under Project b2013047. We thank Science for Life Laboratory in Stockholm as well as the National High-throughput DNA Sequencing Centre in Copenhagen for help with CAGE sequencing.

The work in Sweden was supported by two grants from the Novo Nordisk Foundation (including the Tripartite Immuno-metabolism Consortium [TrIC], grant NNF15CC0018486), the Swedish Research Council, the EASD/Lilly Foundation, the Stockholm County Council, the Swedish Diabetes Association, the Erling Persson Family Foundation, CIMED, and the Diabetes Research Program at Karolinska Institutet. The work in Denmark was supported by the Lundbeck and Novo Nordisk Foundations.

Received: October 19, 2015
Revised: April 22, 2016
Accepted: July 26, 2016
Published: August 18, 2016

REFERENCES

Arner, P., Bäckdahl, J., Hemmingsson, P., Stenvinkel, P., Eriksson-Hogling, D., Näslund, E., Thorell, A., Andersson, D.P., Caidahl, K., and Rydén, M. (2015). Regional variations in the relationship between arterial stiffness and adipocyte volume or number in obese subjects. Int. J. Obes. *39*, 222–227.

Blüher, M. (2010). The distinction of metabolically ‘healthy’ from ‘unhealthy’ obese individuals. Curr. Opin. Lipidol. *21*, 38–43.

Carninci, P., Sandelin, A., Lenhard, B., Katayama, S., Shimokawa, K., Ponjavic, J., Semple, C.A., Taylor, M.S., Engström, P.G., Frith, M.C., et al. (2006). Genome-wide analysis of mammalian promoter architecture and evolution. Nat. Genet. *38*, 626–635.

Elbein, S.C., Kern, P.A., Rasouli, N., Yao-Borengasser, A., Sharma, N.K., and Das, S.K. (2011). Global gene expression profiles of subcutaneous adipose and muscle from glucose-tolerant, insulin-sensitive, and insulin-resistant individuals matched for BMI. Diabetes *60*, 1019–1029.

Fan, J., Song, Y., Chen, Y., Hui, R., and Zhang, W. (2013). Combined effect of obesity and cardio-metabolic abnormality on the risk of cardiovascular disease: a meta-analysis of prospective cohort studies. Int. J. Cardiol. *168*, 4761–4768.

Felländer, G., Nordenström, J., Tjäder, I., Bolinder, J., and Arner, P. (1994). Lipolysis during abdominal surgery. J. Clin. Endocrinol. Metab. *78*, 150–155.

Flint, A.J., Hu, F.B., Glynn, R.J., Caspard, H., Manson, J.E., Willett, W.C., and Rimm, E.B. (2010). Excess weight and the risk of incident coronary heart disease among men and women. Obesity (Silver Spring) *18*, 377–383.

Forrest, A.R., Kawaji, H., Rehli, M., Baillie, J.K., de Hoon, M.J., Haberle, V., Lassmann, T., Kulakovskiy, I.V., Lizio, M., Itoh, M., et al.; FANTOM Consortium and the RIKEN PMI and CLST (DGT) (2014). A promoter-level mammalian expression atlas. Nature *507*, 462–470.

Frith, M.C., Valen, E., Krogh, A., Hayashizaki, Y., Carninci, P., and Sandelin, A. (2008). A code for transcription initiation in mammalian genomes. Genome Res. *18*, 1–12.

Gao, H., Mejhert, N., Fretz, J.A., Arner, E., Lorente-Cebrián, S., Ehrlund, A., Dahlman-Wright, K., Gong, X., Strömblad, S., Douagi, I., et al. (2014). Early B cell factor 1 regulates adipocyte morphology and lipolysis in white adipose tissue. Cell Metab. *19*, 981–992.

Hagström-Toft, E., Thörne, A., Reynisdottir, S., Moberg, E., Rössner, S., Bolinder, J., and Arner, P. (2001). Evidence for a major role of skeletal muscle lipolysis in the regulation of lipid oxidation during caloric restriction in vivo. Diabetes *50*, 1604–1611.

Hoffstedt, J., Arner, E., Wahrenberg, H., Andersson, D.P., Qvisth, V., Löfgren, P., Rydén, M., Thörne, A., Wirén, M., Palmér, M., et al. (2010). Regional impact of adipose tissue morphology on the metabolic profile in morbid obesity. Diabetologia *53*, 2496–2503.

Karelis, A.D. (2008). Metabolically healthy but obese individuals. Lancet *372*, 1281–1283.

Kawaji, H., Lizio, M., Itoh, M., Kanamori-Katayama, M., Kaiho, A., Nishiyori-Sueki, H., Shin, J.W., Kojima-Ishiyama, M., Kawano, M., Murata, M., et al.; FANTOM Consortium (2014). Comparison of CAGE and RNA-seq transcriptome profiling using clonally amplified and single-molecule next-generation sequencing. Genome Res. *24*, 708–717.

Kersey, P.J., Allen, J.E., Christensen, M., Davis, P., Falin, L.J., Grabmueller, C., Hughes, D.S., Humphrey, J., Kerhornou, A., Khobova, J., et al. (2014). Ensembl Genomes 2013: scaling up access to genome-wide data. Nucleic Acids Res. *42*, D546–D552.

Klimcáková, E., Roussel, B., Márquez-Quiñones, A., Kovácová, Z., Kovácikova, M., Combes, M., Siklová-Vítková, M., Hejnová, J., Srámková, P., Bouloumié, A., et al. (2011). Worsening of obesity and metabolic status yields similar molecular adaptations in human subcutaneous and visceral adipose tissue: decreased metabolism and increased immune response. J. Clin. Endocrinol. Metab. *96*, E73–E82.

Kramer, C.K., Zinman, B., and Retnakaran, R. (2013). Are metabolically healthy overweight and obesity benign conditions?: A systematic review and meta-analysis. Ann. Intern. Med. *159*, 758–769.

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. *10*, R25.

Lu, J., Bi, Y., Wang, T., Wang, W., Mu, Y., Zhao, J., Liu, C., Chen, L., Shi, L., Li, Q., et al. (2014). The relationship between insulin-sensitive obesity and cardiovascular diseases in a Chinese population: results of the REACTION study. Int. J. Cardiol. *172*, 388–394.

Ortega, F.B., Lee, D.C., Katzmarzyk, P.T., Ruiz, J.R., Sui, X., Church, T.S., and Blair, S.N. (2013). The intriguing metabolically healthy but obese phenotype: cardiovascular prognosis and role of fitness. Eur. Heart J. *34*, 389–397.

Persson, H., Kwon, A.T., Ramilowski, J.A., Silberberg, G., Söderhäll, C., Orsmark-Pietras, C., Nordlund, B., Konradsen, J.R., de Hoon, M.J., Melén, E., et al. (2015). Transcriptome analysis of controlled and therapy-resistant childhood asthma reveals distinct gene expression profiles. J. Allergy Clin. Immunol. *136*, 638–648.

Primeau, V., Coderre, L., Karelis, A.D., Brochu, M., Lavoie, M.E., Messier, V., Sladek, R., and Rabasa-Lhoret, R. (2011). Characterizing the profile of obese patients who are metabolically healthy. Int. J. Obes. *35*, 971–981.

Qatanani, M., Tan, Y., Dobrin, R., Greenawalt, D.M., Hu, G., Zhao, W., Olefsky, J.M., Sears, D.D., Kaplan, L.M., and Kemp, D.M. (2013). Inverse regulation of inflammation and mitochondrial function in adipose tissue defines extreme insulin sensitivity in morbidly obese patients. Diabetes *62*, 855–863.

Roberson, L.L., Aneni, E.C., Maziak, W., Agatston, A., Feldman, T., Rouseff, M., Tran, T., Blaha, M.J., Santos, R.D., Sposito, A., et al. (2014). Beyond BMI: The “metabolically healthy obese” phenotype & its association with clinical/subclinical cardiovascular disease and all-cause mortality – a systematic review. BMC Public Health *14*, 14.

Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics *26*, 139–140.

Samocha-Bonet, D., Chisholm, D.J., Tonks, K., Campbell, L.V., and Greenfield, J.R. (2012). Insulin-sensitive obesity in humans – a ‘favorable fat’ phenotype? Trends Endocrinol. Metab. *23*, 116–124.

Sims, E.A. (2001). Are there persons who are obese, but metabolically healthy? Metabolism *50*, 1499–1504.

Song, Y., Manson, J.E., Meigs, J.B., Ridker, P.M., Buring, J.E., and Liu, S. (2007). Comparison of usefulness of body mass index versus metabolic risk factors in predicting 10-year risk of cardiovascular events in women. Am. J. Cardiol. *100*, 1654–1658.

Soronen, J., Laurila, P.P., Naukkarinen, J., Surakka, I., Ripatti, S., Jauhiainen, M., Olkkonen, V.M., and Yki-Järvinen, H. (2012). Adipose tissue gene expression analysis reveals changes in inflammatory, mitochondrial respiratory and lipid metabolic pathways in obese insulin-resistant subjects. BMC Med. Genomics *5*, 9.

Suzuki, H., Forrest, A.R., van Nimwegen, E., Daub, C.O., Balwierz, P.J., Irvine, K.M., Lassmann, T., Ravasi, T., Hasegawa, Y., de Hoon, M.J., et al.; FANTOM Consortium; Riken Omics Science Center (2009). The transcriptional network that controls growth arrest and differentiation in a human myeloid leukemia cell line. Nat. Genet. *41*, 553–562.

Takahashi, H., Lassmann, T., Murata, M., and Carninci, P. (2012). 5′ end-centered expression profiling using cap-analysis gene expression and next-generation sequencing. Nat. Protoc. *7*, 542–561.

Viguerie, N., Montastier, E., Maoret, J.J., Roussel, B., Combes, M., Valle, C., Villa-Vialaneix, N., Iacovoni, J.S., Martinez, J.A., Holst, C., et al. (2012). Determinants of human adipose tissue gene expression: impact of diet, sex, metabolic status, and cis genetic regulation. PLoS Genet. *8*, e1002959.

Westerbacka, J., Cornér, A., Kannisto, K., Kolak, M., Makkonen, J., Korsheninnikova, E., Nyman, T., Hamsten, A., Fisher, R.M., and Yki-Järvinen, H. (2006). Acute in vivo effects of insulin on gene expression in adipose tissue in insulin-resistant and insulin-sensitive subjects. Diabetologia *49*, 132–140.

Xu, X.J., Pories, W.J., Dohm, L.G., and Ruderman, N.B. (2013). What distinguishes adipose tissue of severely obese humans who are insulin sensitive and resistant? Curr. Opin. Lipidol. *24*, 49–56.

Cell Reports
Article

Parent-of-Origin DNA Methylation Dynamics during Mouse Development

Yonatan Stelzer,[1] Hao Wu,[1] Yuelin Song,[1,2] Chikdu S. Shivalila,[1,2] Styliani Markoulaki,[1] and Rudolf Jaenisch[1,2,3,*]
[1]Whitehead Institute for Biomedical Research, Cambridge, MA 02142, USA
[2]Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02142, USA
[3]Lead Contact
*Correspondence: jaenisch@wi.mit.edu

http://dx.doi.org/10.1016/j.celrep.2016.08.066

SUMMARY

Parent-specific differentially methylated regions (DMRs) are established during gametogenesis and regulate parent-specific expression of imprinted genes. Monoallelic expression of imprinted genes is essential for development, suggesting that imprints are faithfully maintained in embryos and adults. To test this hypothesis, we targeted a reporter for genomic methylation to the imprinted Dlk1-Dio3 intergenic DMR (IG-DMR) to assess the methylation of both parental alleles at single-cell resolution. Biallelic gain or loss of IG-DMR methylation occurred in a small fraction of mouse embryonic stem cells, significantly affecting developmental potency. Mice carrying the reporter in either parental allele showed striking parent-specific changes in IG-DMR methylation, causing substantial and consistent tissue- and cell-type-dependent signatures in embryos and postnatal animals. Furthermore, dynamics in DNA methylation persisted during adult neurogenesis, resulting in inter-individual diversity. This substantial cell-cell DNA methylation heterogeneity implies that dynamic DNA methylation variations in the adult may be of functional importance.

INTRODUCTION

Parental imprinting is a heritable epigenetic mechanism resulting in parent-specific monoallelic expression of subset of genes (Ferguson-Smith, 2011; Reik and Walter, 2001), and such imprinting is essential during early mammalian development (McGrath and Solter, 1984; Surani and Barton, 1983). While methylation imprints established during gametogenesis are thought to be stable in development, complex tissue-specific expression of imprinted genes can occur in the developing embryo (Barton et al., 1991; Thomson and Solter, 1988), with possible functional consequences in the animal (Davies et al., 2005; Frost and Moore, 2010; Wilkinson et al., 2007). Due to their monoallelic nature, imprinted genes are specifically susceptible to alterations that may be caused by loss-of-function mutations or by epimutations in regulatory elements. Indeed, loss of imprinting (LOI) correlates with mild to severe developmental abnormalities, organ malfunctions, behavior anomalies, and cancer (Avior et al., 2016; Peters, 2014; Robertson, 2005; Yamazawa et al., 2010).

DNA methylation is central for the regulation of parental imprinting, as gamete-specific differentially methylated regions (DMRs) act in *cis* to regulate the monoallelic parent-of-origin expression of multiple imprinted genes (Barlow and Bartolomei, 2014). Following fertilization, imprinted DMRs are protected from global de-methylation and de novo methylation in somatic cells, with the exception of primordial germ cells, where all methylation imprints are removed and re-established in a sex-dependent manner during gametogenesis (Lee et al., 2014; Reik, 2007). Recent advances in sequencing technologies facilitated single-base-resolution DNA methylation maps of multiple embryonic and adult tissues (Hon et al., 2013; Kundaje et al., 2015; Ziller et al., 2013), enabling insights into the stability of imprinted DMRs in adult tissues and the identification of novel imprinted DMRs in both humans (Court et al., 2014; Stelzer et al., 2013) and mice (Xie et al., 2012). It is believed that following fertilization, imprinted DMRs are mostly maintained by the activity of Dnmt1 (Li et al., 1993; Tucker et al., 1996) and that loss of parent-specific methylation is stochastic and may contribute to disease (Ferguson-Smith, 2011; Reik, 2007; Reik and Walter, 2001; Robertson, 2005). Nevertheless, because of the “snapshot” nature of sequencing data, the present understanding of imprint maintenance during embryonic development and in adult tissues is limited and precludes the assessment of tissues and cell-type heterogeneity at single-cell resolution.

The imprinted Dlk1-Dio3 locus on mouse chromosome 12 is characterized by the reciprocal expression of maternal non-coding transcripts and paternal protein coding genes regulated by both *cis*- (Lin et al., 2003) and *trans*-acting (Cockett et al., 1996; Seitz et al., 2003) mechanisms. The intergenic DMR (IG-DMR) serves as an imprinted control center regulating parent-specific expression of genes in this locus (da Rocha et al., 2008; Lin et al., 2003). Mice with uniparental disomy and genetic manipulations of the locus have substantiated that proper imprinting is essential for normal development, with LOI resulting in early embryonic lethality (Georgiades et al., 2000; Lin et al., 2003, 2007; Tevendale et al., 2006). Targeted deletions of individual genes in the Dlk1-Dio3 locus lead to complex abnormalities in the embryo and postnatal animal and include cartilage,

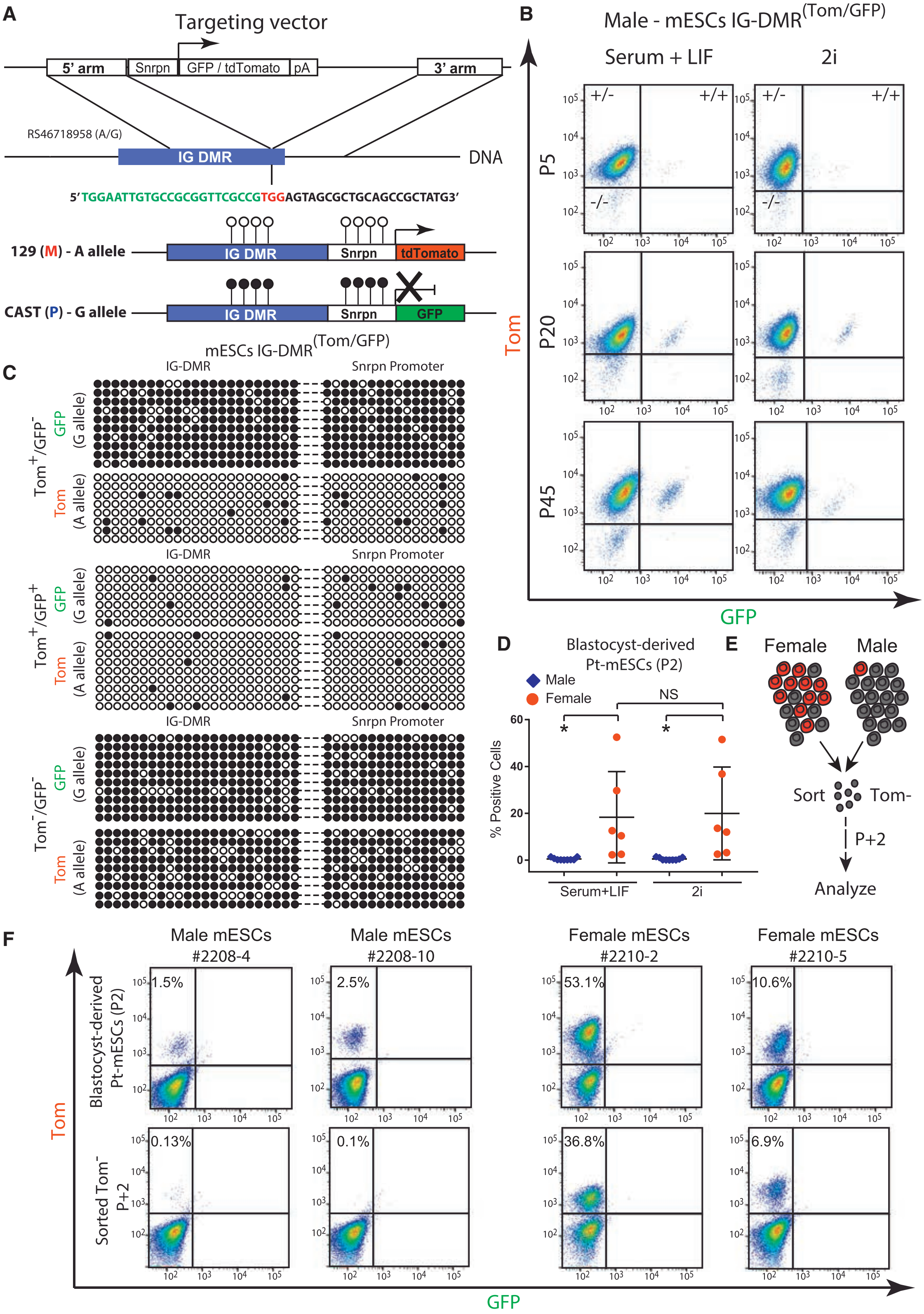


Figure 1. Allele-Specific Targeting of the Dlk1-Dio3 IG-DMR

(A) Schematic representation of CRISPR/Cas-mediated allele-specific targeting of Snrpn-GFP or Snrpn-Tom, adjacent to the IG-DMR region; green sequence, endogenous IG-DMR region; black sequence, targeting CRISPR; red sequence, protospacer adjacent motif (PAM) recognition site.

(legend continued on next page)

bone, muscle, and placenta defects (Andersen et al., 2013; Sekita et al., 2008; Takahashi et al., 2009), obesity (Moon et al., 2002), and metabolic and behavioral dysfunctions (Labialle et al., 2014; Qian et al., 2016; Sittig and Redei, 2014).

We have recently established a reporter of genomic methylation (RGM) that relies on an imprinted gene promoter (*Snrpn*) driving a fluorescent protein (Stelzer and Jaenisch, 2015; Stelzer et al., 2015). Here, we utilized RGM to facilitate a comprehensive study of the dynamics of imprinted DMRs in embryos and adult mice. RGM was targeted in mouse embryonic stem cells (mESCs) to each allele of the Dlk1-Dio3 IG-DMR. Aberrant methylation at the IG-DMR strongly affected developmental potency in chimera assays. Furthermore, we identify sex-dependent differences in the degree and kinetics of paternal allele demethylation, with blastocyst-derived female mESCs displaying rapid demethylation during early passages. Mice carrying the reporter in either allele were used to assess the maintenance of imprints in embryos and adult mice. Surprisingly, methylation changes at the Dlk1-Dio3 DMR were found to be dynamic in most tissues of the embryo and the postnatal animal. In particular, methylation imprints varied at the single-cell level during adult neurogenesis, resulting in inter-individual diversity and epigenetic variability.

RESULTS

Allele-Specific Targeting of the Dlk1-Dio3 IG-DMR

We utilized CRISPR/Cas9-mediated gene editing in F1 hybrid 129X*Castaneous* (CAST) male mESCs to target Snrpn-GFP or Snrpn-tdTomato (Tom) to each allele of the Dlk1-Dio3 IG-DMR (Figure 1A). The IG-DMR acquires paternal methylation during spermatogenesis, while the maternal allele is hypomethylated in the oocyte (da Rocha et al., 2008). Consistent with this notion, cells targeted with Snrpn-Tom to the maternal allele and Snrpn-GFP to the paternal allele (IG-DMR$^{Tom/GFP}$) expressed the Tom reporter, but not the GFP reporter (Figures 1B and S1A). Bisulfite sequencing of targeted cell lines demonstrated that while the maternally targeted Tom allele was hypomethylated in the IG-DMR and downstream Snrpn promoter regions, the paternally targeted GFP reporter allele exhibited high levels of DNA methylation which spread from the IG-DMR region into the Snrpn promoter (Figure 1C), resulting in its repression. During expansion of targeted cell lines, a small fraction of cells emerged that were either double positive or double negative for reporter expression (Figure 1B). These subpopulations slightly increased during consecutive passages suggesting that they do not confer significant growth advantage (Figures 1B and S1B). Bisulfite sequencing of sorted double-positive and double-negative cells indicated hypomethylation or hypermethylation, respectively, of both parental alleles as well as the Snrpn promoters (Figures 1C). Notably, during prolonged culturing of the sorted cell populations, a new population of cells emerged that had switched the allelic reporter activity repressing the maternal Tom allele and activating the paternal GFP (Figure S1C). Thus, all these data demonstrated that reporter activity faithfully reflects parent-specific gain or loss of DNA methylation at the Dlk1-Dio3 IG-DMR and that insertion of RGM does not affect the methylation levels of adjacent sequences. Recent studies have shown that culturing mESCs with inhibitors of MEK and GSK3 (2i) results in global hypomethylation (Ficz et al., 2013; Habibi et al., 2013). When cultured in standard 2i culture conditions, we observed no significant increase of double-positive cells compared with culturing in serum and leukemia inhibitory factor (LIF) (Figures 1B and S1B).

To investigate whether the in vitro loss of parent-specific methylation also occurs in newly derived mESCs, we isolated the inner cell mass (ICM) from blastocysts carrying the paternally transmitted (Pt) GFP or Tomato reporter (see Figure S1D and Experimental Procedures). As documented for targeted male cell lines (Figures 1B and S1B), newly isolated male mESCs exhibited rare and stable population of cells with aberrant paternal reporter activity (Figure 1D). Nevertheless, and in strike contrast, female mESCs were significantly more likely to activate paternal reporter activity, with some cell lines exhibiting >50% GFP- or Tom-positive cells (Figure 1D). To test whether the observed variation reflects intrinsic sex-specific differences, male and female mESCs harboring the Tomato reporter in the paternal allele of the IG-DMR were sorted for Tomato-negative cells and analyzed in subsequent passages (Figure 1E). Only two passages following sorting, female, but not male, cells showed robust reactivation of the reporter (Figure 1F). Culturing female cell lines in 2i showed no significant increase of positive cells compared with culturing in serum and LIF (Figure 1D), suggesting that X chromosome number (2 X in female versus 1 X in male cells), but not culture conditions, play a role in the rapid demethylation of the IG-DMR as was observed previously (Zvetkova et al., 2005). In summary, our data suggest that female mouse ESCs with two X chromosomes exhibit rapid demethylation of the paternal allele of the IG-DMR as revealed by reporter activity.

Allele-Specific Methylation, Gene Expression, and Reporter Activity

The Dlk1-Dio3 imprinted locus comprises multiple maternally expressed non-coding genes with unknown functions, including the long intergenic noncoding RNA (lincRNA) *Gtl2* and large

---

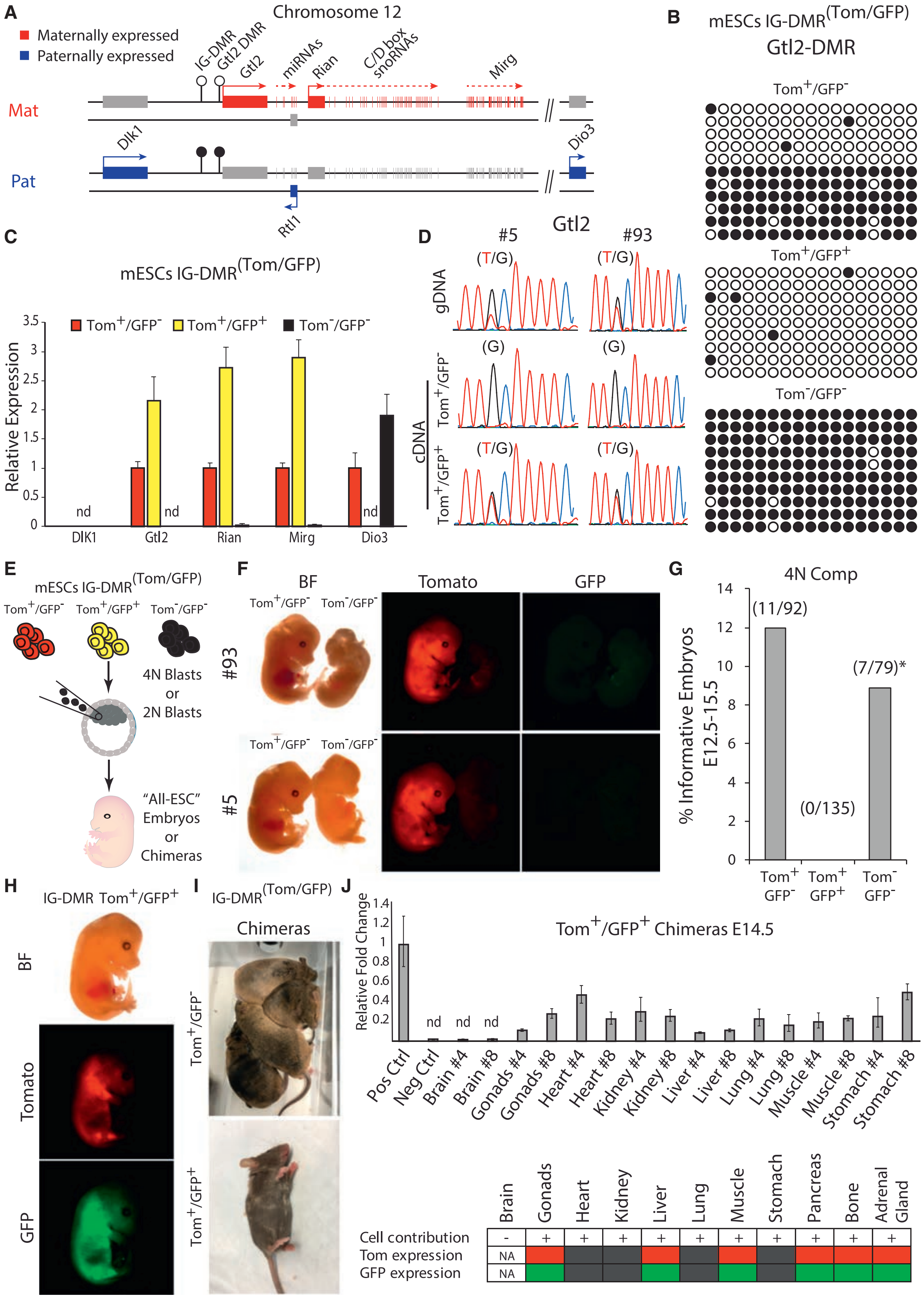
(B) Flow cytometric analysis of GFP/Tom reporter ESCs at different passages, cultured in serum plus LIF or 2i.

(C) Allele-specific bisulfite sequencing was performed on sorted IG-DMR Tom^{+}/GFP^{-}, IG-DMR Tom^{+}/GFP^{+} and IG-DMR Tom^{-}/GFP^{-} mESCs. Each row represents a distinct PCR amplicon (marked with dashed line) that includes the endogenous IG-DMR (left) and the downstream integrated Snrpn promoter region (right). Open circles represent unmethylated CpGs, and black circles represent methylated CpGs.

(D) Dot plot showing the percentage of GFP/Tom-positive cells in passage 2 (P2) male and female mESCs cultured in serum plus LIF or 2i, as measured by flow cytometry. Black lines indicate mean ± SD for each group. Statistical differences between genotypes were calculated using one-way ANOVA; $^{*}p < 0.05$; NS, not significant; Pt, paternally transmitted.

(E) Schematic diagram for sorting and analyzing paternally transmitted (Pt) Tomato-negative (Tom^{-}) male or female mESCs presented in (F).

(F) Flow cytometric analysis of the proportion of Tom-positive cells in passage 2 (P2) male and female mESCs (top) and in sorted Tom^{-} cells following two consecutive passages (bottom). Pt, paternally transmitted.



	Brain	Gonads	Heart	Kidney	Liver	Lung	Muscle	Stomach	Pancreas	Bone	Adrenal Gland
Cell contribution	-	+	+	+	+	+	+	+	+	+	+
Tom expression	NA	🟥	⬛	⬛	🟥	⬛	🟥	⬛	🟥	🟥	🟥
GFP expression	NA	🟩	⬛	⬛	🟩	⬛	🟩	⬛	🟩	🟩	🟩

(legend on next page)

clusters of C/D box small nucleolar RNAs (snoRNAs) and microRNAs (miRNAs). Additionally, three protein-coding genes being expressed exclusively from the paternal allele (Figure 2A). These include the atypical Notch ligand delta-like homolog 1 (*Dlk1*), a retrotransposon-like *Rtl1*, and the type 3 iodothyronine deiodinase (*Dio3*) (da Rocha et al., 2008). The IG-DMR serves as a *cis*-acting regulatory center that establishes post-zygotic ''secondary'' DMRs such as in the promoter of *Gtl2* (Figure 2A). We tested the methylation levels associated with the *Gtl2* promoter DMR in the three IG-DMR$^{Tom/GFP}$ cell populations. Figure 2B shows that the methylation state of the IG-DMR corresponded to that of the downstream Gtl2 promoter DMR, suggesting that the mechanism that mediates the establishment of *Gtl2* DMR is functional in mESCs. qPCR on representative genes in the locus demonstrated that IG-DMR methylation strictly correlated with the expression patterns of maternal and paternal genes (Figure 2C). Thus, hypomethylation of the paternal allele (Tom$^+$/GFP$^+$) resulted in a near 2-fold increase in maternal gene expression and loss of expression of the paternal gene *Dio3*. Conversely, hypermethylation of the maternal alleles (Tom$^-$/GFP$^-$) resulted in complete repression of all maternal genes and a 2-fold increase in the expression of Dio3 as compared to cells with intact parent-of-origin methylation levels (Tom$^+$/GFP$^-$). Furthermore, utilizing a heterozygous SNP in the Gtl2 coding region demonstrated maternal monoallelic expression of Gtl2 in control IG-DMR Tom$^+$/GFP$^-$ cells, while hypomethylated (Tom$^+$/GFP$^+$) cells exhibited biallelic Gtl2 expression (Figure 2D), consistent with 2-fold increase in expression (Figure 2C). We conclude that IG-DMR methylation reporter activity strictly correlates with parent-specific gene expression of multiple genes in the Dlk1-Dio3 region. Expression of other imprinted genes, such as H19, PEG3, and Snrpn, was not altered in cells with aberrant IG-DMR methylation (Figure S1E).

Dlk1-Dio3 Loss of Imprinting Affects Developmental Potency of ESCs

To assess whether Dlk1-Dio3 LOI would affect the developmental potential of ESCs, we utilized tetraploid complementation (Tam and Rossant, 2003), the most stringent assay for developmental potency (Figure 2E). Embryos were analyzed at embryonic days 12.5 to 15.5 (E12.5–15.5). Figure 2F shows that while control IG-DMR Tom$^+$/GFP$^-$ cells generated normal embryos with comparable frequencies to previous reports (Buganim et al., 2014), aberrantly hypermethylated IG-DMR Tom$^-$/GFP$^-$ embryos exhibited growth defects at midgestation that included severe brain malformations and muscle defects, whereas 4n embryos from biallelically hypomethylated IG-DMR Tom$^+$/GFP$^+$ ESCs died prior to gastrulation (Figure 2G). Biallelically hypomethylated (Tom$^+$/GFP$^+$) cells, when injected into 2n host blastocysts, contributed to chimeric embryos and postnatal animals, though with lower efficiency than control IG-DMR Tom$^+$/GFP$^-$ ESCs (Figures 2H and 2I). qPCR detected the presence of donor cells in all tissues of chimeric embryos, except in brain (Figure 2J). Notably, some tissues maintained the expression of both fluorescent markers, while other tissues, with evident contribution of donor cells (e.g., kidney, heart, and lung), appeared double-negative Tom$^-$/GFP$^-$, indicating biallelic hypermethylation of the IG-DMR (Figures 2J and S2). Our results demonstrate that incorporation of both IG-DMR Tom$^-$/GFP$^-$ and Tom$^+$/GFP$^+$ cells into chimeric embryos results in developmental defects, while IG-DMR Tom$^+$/GFP$^+$ cells display a more severe phenotype, with lack of contribution to the brain in chimeric embryos.

Parent-Specific Imprints Are Maintained Faithfully in Some, but Not Other, Tissues

In order to study parent-specific methylation dynamics in vivo, cells with a maternal Snrpn-GFP reporter were injected into blastocysts to generate chimeras, which were bred to obtain transgenic males (Figure 3A). Mice carrying the reporter allele were born at the expected Mendelian ratio, implying that the reporter had no adverse effect. Since the IG-DMR is methylated during spermatogenesis, first-generation (F1) males and females carrying the reporter allele were expected to be GFP negative. When maternally and paternally transmitted F2 embryos were analyzed (IG-DMR$^{Mat\text{-}GFP}$ and IG-DMR$^{Pat\text{-}GFP}$, respectively), IG-DMR$^{Mat\text{-}GFP}$ blastocysts were positive for GFP expression and IG-DMR$^{Pat\text{-}GFP}$ blastocysts were negative (Figures 3A, 3B, S3A, and S3B), indicating proper parent-specific reporter expression. All IG-DMR$^{Mat\text{-}GFP}$ embryos expressed the GFP reporter throughout development (Figures 3B, S3C, and S3D), but close examination revealed differential GFP activity between

Figure 2. Functional Consequences of Parent-Specific Loss of Methylation in the ID-DMR Region

(A) Organization of imprinted genes in the mouse *Dlk1-Dio3* locus; open lollipops represent unmethylated regions, and black lollipops represent methylated regions.

(B) Bisulfite sequencing was performed on Gtl2 promoter DMR in distinct IG-DMR$^{(Tom/GFP)}$ sorted mESC populations.

(C) Quantitative real-time PCR of the mean relative fold change ± SD of representative genes in the Dlk1-Dio3 region in three IG-DMR$^{(Tom/GFP)}$ sorted mESCs from two independently targeted cell lines. Data were normalized to Gapdh housekeeping control; nd, not detected.

(D) Sequencing of Gtl2 in two independent mESC IG-DMR$^{(Tom/GFP)}$ lines. Heterozygous SNP was identified in the genomic DNA (gDNA); monoallelic versus biallelic expression was evaluated in the cDNA.

(E) Schematic representation of the blastocyst-injection strategy.

(F) Representative images of E13.5 4n complementation embryos, obtained from two independent mESC IG-DMR Tom$^-$/GFP$^-$ and control mESC IG-DMR Tom$^+$/GFP$^-$ lines.

(G) Summary of 4n embryo injections; *, all embryos analyzed exhibited muscle and brain phenotypes.

(H) Representative images of IG-DMR Tom$^+$/GFP$^+$ mESC contributions to E14.5 chimeric embryos and (I) postnatal mice.

(J) Top: quantitative real-time PCR detection of IG-DMR Tom$^+$/GFP$^+$ mESC contributions to different organs in E14.5 chimeric embryos. Samples were normalized to ultra-conserved noncoding element in the mouse genome. Shown are mean relative fold change ± SD of GFP detection in two embryos (#4 and #8) compared with GFP-positive (Pos Ctrl) and WT cells (Neg Ctrl); nd, not detected. Bottom: summary of Tom and GFP expression in different organs of chimeric embryos.

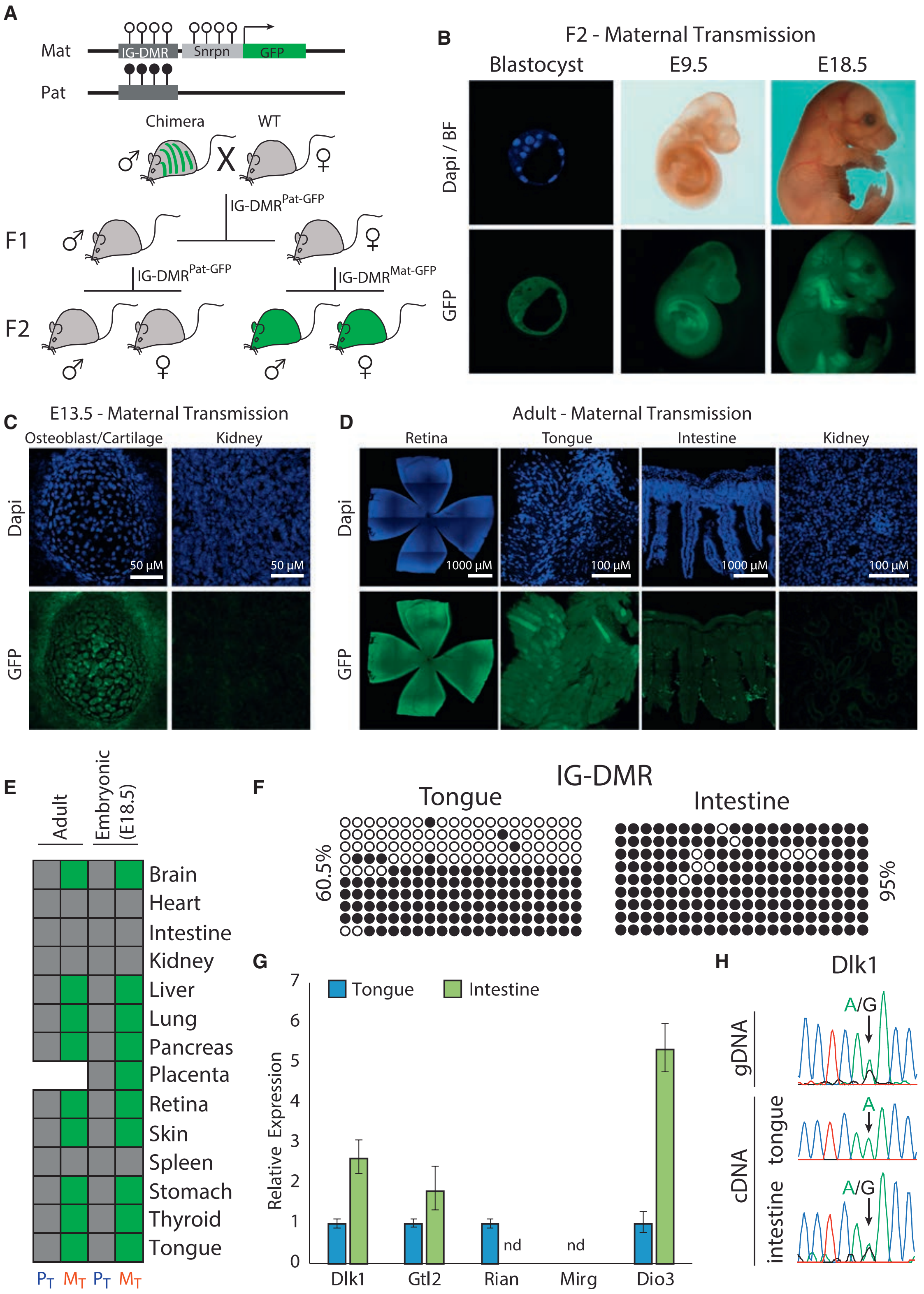


Figure 3. Generation of IG-DMR^{GFP} Reporter Mice Reveals Tissue-Specific Reporter Activity

(A) Mating scheme for generation of parent-specific IG-DMR^{GFP} reporter mice.

(B) Representative images of F2 IG-DMR$^{Mat-GFP}$ embryos at different developmental stages. DAPI and anti-GFP staining; (C) embryonic and (D) adult tissues obtained from 5- to 7-week-old mice.

(legend continued on next page)

some tissues (Figure 3C). Furthermore, tissues that repressed the GFP reporter in the developing embryo, such as kidney, heart, and intestine, persistently silenced the IG-DMR$^{\text{Mat-GFP}}$ allele in adult animals (Figures 3D and 3E).

We tested whether reporter activity faithfully reflects the methylation patterns in wild-type (WT) untargeted mice. The IG-DMR region was hypermethylated in intestine consistent with maternal reporter silencing, whereas the region was hemimethylated in tongue, consistent with maternal reporter expression (Figures 3E and 3F). Gene expression analysis demonstrated complete downregulation of the maternally expressed gene Rian in intestine, consistent with hypermethylation of the IG-DMR region. Another maternal gene, Gtl2, was found to be expressed in both intestine and tongue, corroborating independent regulation by its secondary promoter DMR (Figures 3G and 2A). Paternally expressed genes (*Dlk1* and *Dio3*) exhibited elevated expression levels in intestine as compared with tongue (Figure 3G), suggesting tissue-specific differences in regulation of gene expression. Consistent with methylation signatures, proper monoallelic expression in tongue and biallelic expression in intestine was identified using an informative SNP in the *Dlk1* coding region (Figure 3H).

Cell-Type-Dependent Imprinting

In addition to tissue-specific imprinting, adult tissues revealed cell-type-dependent reporter expression in IG-DMR$^{\text{Mat-GFP}}$ animals. Figure 4A shows selective expression of GFP in some, but not other, cells in the stomach, as well as regional heterogeneity. In the liver, reporter activity gradually decreased with age and was restricted to epithelial cells surrounding the liver bile duct that retained proper GFP expression (Figure 4B). To further analyze the differences between the two cell populations, we used fluorescence-activated cell sorting (FACS) to isolate GFP+ cells from adult mouse livers. Figures 4C–4E show a small fraction of GFP-positive cells with a similar expression level of endogenous Snrpn and Gtl2 in GFP-positive cells compared with GFP-negative cells that displayed a 2-fold increase of paternally expressed genes Dlk1 and Dio3 and silencing of the maternally expressed gene Rian (Figure 4E). We performed bisulfite sequencing of the Gtl2-associated promoter DMR in the two cell populations and identified intermediate methylation levels (Figure 4F). Thus, our results support regulation of Gtl2 by its promoter DMR independent of the IG-DMR methylation, a finding consistent with its methylation state in ESCs (Figure 2B). While it was previously speculated that Rian and Mirg might be further processed from a large non-coding transcript originating from the Gtl2 promoter (Royo and Cavaillé, 2008), our data suggest that these transcripts are independently regulated.

Imprinted genes in the *Dlk1-Dio3* region are highly enriched in the brain. In addition, we show that loss of parent-specific methylation in the IG-DMR region results in marked brain phenotypes. At the macroscopic level, the brain of 5- to 7-week-old mice expressed the RGM reporter in a parent-of-origin-specific pattern (Figure S4A). However, similar to the stomach and liver, close examination revealed consistent variations in GFP expression between different anatomical regions of the brain (Figure S4B), with overall variations being associated with cell-type-specific reporter activity. Thus, while some cell types such as dopaminergic neurons robustly expressed GFP, other cell types such as astrocytes were GFP negative (Figures 5A, 5B, and S5–S7). In addition to cell-type-dependent GFP expression, we also noticed considerable heterogeneity within some cell types. Figure 5C shows that while most calbindin-positive Purkinje cell were GFP negative, some cerebellum lobes contained groups of adjacent GFP-positive Purkinje cells (Figures S8A and S8B). Staining for the NeuN neuronal marker identified cellular variation in GFP expression associated with different cortical layers, with the external granular and pyramidal layers containing high numbers of NeuN^{+}GFP^{+} neurons and the internal granular and pyramidal layers exhibiting a high fraction of NeuN^{+}GFP^{-} cells (Figure 5D). Figure 5E summarizes the anatomical and cell-type-specific methylation differences. Given the high expression levels of multiple imprinted regulatory transcripts in the Dlk1-Dio3 region, these cell-type-dependent differences in IG-DMR methylation may result in substantial gene expression differences between cell types and anatomical regions (Figures S5–S8). To validate the reporter activity in untargeted WT cells, we isolated pre- and postnatal astrocytes and performed bisulfite sequencing. Figure 5F shows that, consistent with lack of IG-DMR$^{\text{Mat-GFP}}$ reporter expression, fetal and postnatal astrocytes exhibited hypermethylation of the IG-DMR region. The downstream Gtl2 promoter DMR identified hemimethylated levels, suggesting that Gtl2 maintains monoallelic regulation (Figure S9A).

Maintenance of Imprinting in the Adult Brain Is Variable at the Single-Cell Level

A recent report demonstrated biallelic expression of Dlk1 in neural stem cells (NSCs) and astrocytes in the post-natal neurogenic niche suggested to be mediated by hypermethylation of the IG-DMR region (Ferrón et al., 2011). In agreement with these findings, GFAP-positive cells residing in the subventricular zone (SVZ) ependymal wall were found to be GFP negative in adult IG-DMR$^{\text{Mat-GFP}}$ brains (Figures 6A and S9B). This was in contrast to overall high reporter expression in the E13.5 SVZ, suggesting that the imprinting status of the IG-DMR changes in pre- versus postnatal NSCs (Figure S9C). As NSCs migrate along the rostral migratory stream (RMS) to replenish the olfactory bulb (OB) neurons, we hypothesized that GFP negative NSCs may contribute to neuronal heterogeneity over time. Consistent with this notion, we identified NeuN^{+}GFP^{-} cells in the adult OB (Figures 5E and S9D). To study whether gain of maternal IG-DMR methylation

(E) Summary of tissue-specific reporter activity in maternal transmitted (M_T) and paternal transmitted (P_T) E18.5 embryos (n = 13) and 5- to 7-week-old adult mice (n = 6). Stitched pictures are shown for both retina and intestine.

(F) Bisulfite sequencing of the IG-DMR region in WT tongue and intestine. Shown are percentages of methylated CpGs.

(G) Quantitative real-time PCR of the mean relative fold change ± SD of representative genes in the *Dlk1-Dio3* region in tongue and intestine from two WT mice. Expression was normalized to Gapdh; nd, not detected.

(H) Sequencing analysis of heterozygous SNP identified in the Dlk1 coding region was performed on cDNAs obtained from WT tongue and intestine tissues.

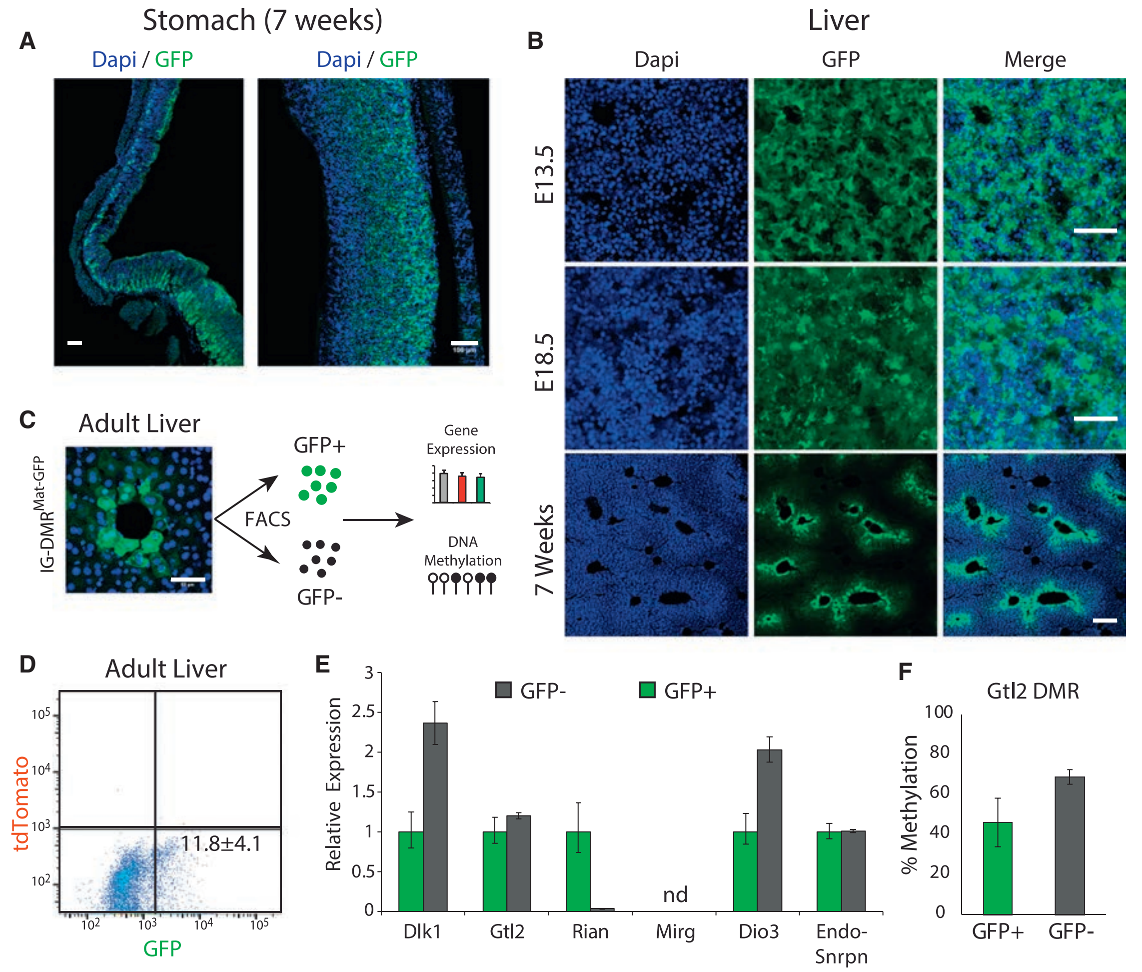


Figure 4. Heterogeneous Reporter Activity in Adult IG-DMR$^{GFP/Mat}$ Tissues

(A) Whole-mount stitching (left) and region specific (right) images of DAPI and anti-GFP staining in adult stomach sections; scale bar, 100μm.

(B) Representative images of DAPI and anti-GFP staining in liver sections of embryos and adults; scale bar, 50 μm.

(C) Single-cell suspension was established from 5-week-old IG-DMR$^{Mat-GFP}$ liver tissues following cell sorting and DNA/RNA extraction; scale bar, 50 μm.

(D) Flow cytometric of GFP-positive cells (mean ± SD of two independent livers).

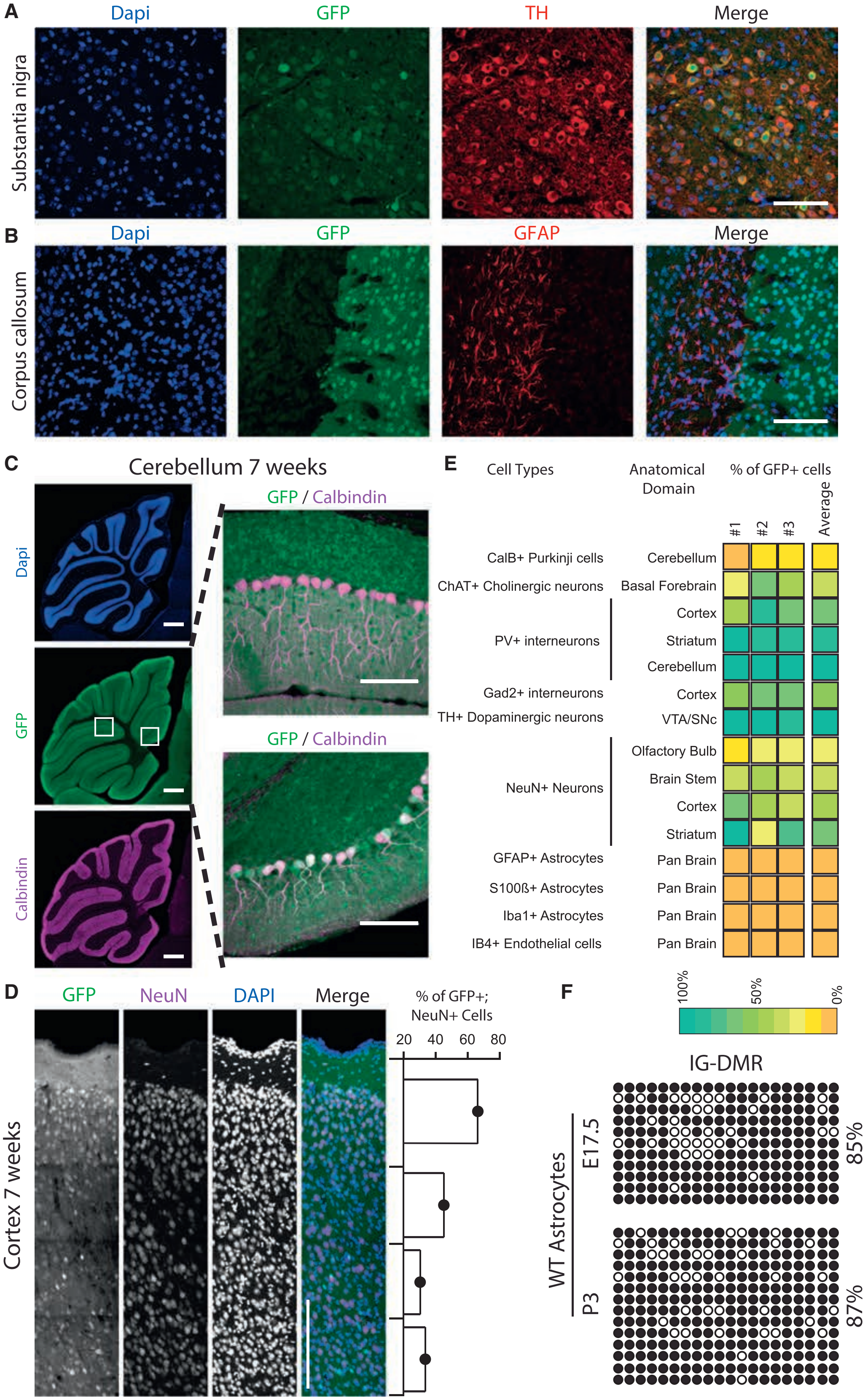
(E) Quantitative real-time PCR analysis of representative genes in the Dlk1-Dio3 region in sorted IG-DMR$^{Mat-GFP}$ liver cells. Shown is mean relative fold change ± SD of two biological replicates; nd, not detected.

(F) Bisulfite sequencing of *Gtl2* promoter DMR in sorted GFP-positive and GFP-negative IG-DMR$^{Mat-GFP}$ liver cells. Shown are mean methylation levels ± SD of two biological replicates. For each sample, more than ten amplicons were sequenced to calculate the percentage of methylated CpGs.

in adult NSCs is irreversible, 5-week-old mice were injected with EDU, a nucleoside analog of thymidine that allows marking of dividing cells and their post-mitotic daughters, and stained for EDU after 10 days (Figure 6B). Figure 6C shows that the vast majority of EDU$^+$-labeled cells repressed GFP expression, consistent with biallelic hypermethylation of the IG-DMR region. GFP repression was identified in EDU$^+$ NSCs located at the SVZ, in EDU$^+$ cells along the RMS, and in EDU$^+$ cells that migrated to the OB cortex and glomeruli. Furthermore, EDU$^+$ cells originating from the dentate gyrus in the subgranular zone (SGZ) were GFP negative, suggesting that adult neurogenesis in the SGZ may contribute to neuronal heterogeneity in the hippocampus over time (Figure 6C). These results suggest that hypermethylation of the IG-DMR in NSCs is irreversible, potentially contributing to neuronal epigenetic variability over time.

DISCUSSION

In this study, we utilized RGM (Stelzer et al., 2015) to report on parent-specific methylation changes of the Dlk1-Dio3 DMR. The RGM reporter was used to isolate subpopulations of ESCs that either had methylated the maternal allele or demethylated the paternal allele, which allowed assessing the consequence on developmental potency; when injected into 4n host embryos, the double-negative cells generated only abnormal E13.5 embryos, whereas injection of double-positive ESCs led to pre-gastrulation death, indicating that LOI at this locus in mESCs results in impaired developmental potency. The generation of transgenic mice carrying the reporter in the maternal and paternal allele identified striking parent- and tissue-specific changes in IG-DMR methylation during development resulting



(legend on next page)

in tissue- and cell-type-dependent methylation signatures in the embryo and adult. Significantly, methylation changes were dynamic in tissues of the postnatal animal. This was particularly evident during adult neurogenesis, resulting in inter-individual diversity and epigenetic variability at the single-cell level.

Current understanding of the status of imprinted DMRs during development and in adult tissues is based on extensive molecular studies and high-resolution sequencing maps. Recent advancements in single-cell sequencing (Smallwood et al., 2014) and allele-specific RNA fluorescence in situ hybridization (RNA-FISH) (Hansen and van Oudenaarden, 2013) technologies hold the promise for elucidating single-cell parent-of-origin methylation and expression. Such methodology was recently used to uncover allele-specific expression heterogeneity of H19/Igf2 in single cells in mutant animals (Ginart et al., 2016). However, a serious limitation of current methods is that they provide only "snapshots" of bulk cell populations, thus precluding the evaluation of DNA methylation dynamics at single-cell resolution. Here, we provide a systematic single-cell analysis of parent-specific methylation dynamics during mouse development. We show that unlike the deleterious effects of loss of parent-specific methylation in mESCs, during embryonic development, the IG-DMR region is subjected to dynamic methylation changes in a tissue- and cell-type-dependent manner. These methylation patterns persist in adult tissues, consistent with the notion that gain of parent-specific methylation is irreversible. Although the full impact of parent-specific methylation dynamics during development and in postnatal animals remains to be identified, the consistent tissue and cellular patterns documented here in multiple animals favor a rather regulated and non-stochastic process. In support of this notion, we show that methylation-mediated silencing of the IG-DMR reporter in tissues such as intestine or in cell types such as astrocytes does not simply silence the gene at this locus but rather regulates gene dosage, as revealed by the biallelic expression of Dlk1 in the intestine.

Maternal deletion of the mouse IG-DMR region is comparable to biallelic hypermethylation of that region and was shown to result in prenatal lethality (Lin et al., 2003, 2007). Furthermore, induced pluripotent stem cells (iPSCs) hypermethylated in both IG-DMR parental alleles failed to generate "all-iPSCs mice" using 4n complementation (Stadtfeld et al., 2010), in agreement with the prenatal death of biallelically hypermethylated embryos described here. The consequence of biallelic hypomethylation of the IG-DMR had not been assessed previously, as it had not been possible to generate such cells using classical genetics. Here, we show that biallelically hypomethylated IG-DMR ESCs displayed reciprocal upregulation of maternal genes and repression of paternal genes in the Dlk1-Dio3 region. These cumulative gene expression perturbations resulted in pre-gastrulation death of 4n embryos, affecting an earlier developmental window than biallelic hypermethylation of the locus. In 2n chimeric embryos, the biallelically hypomethylated cells contributed to many tissues with the notable exception of the brain. We also detected significant differences in the rate of acquiring and in the extent of aberrant paternal demethylation between male and female ESCs consistent with previous results that showed global demethylation in XX ESCs (Zvetkova et al., 2005).

Genes in the *Dlk1-Dio3* locus are highly expressed in the brain, an organ that was previously associated with complex parent-of-origin effects (Davies et al., 2005; Ferrón et al., 2015; Perez et al., 2015; Sittig and Redei, 2014; Wilkinson et al., 2007; Xie et al., 2012). A recent report suggested that gain or loss of DNA methylation in the IG-DMR region might be regulated in a dynamic manner in the adult neurogenic niche (Ferrón et al., 2011). Consistent with this notion, we show striking cell-type-dependent variation in IG-DMR methylation in the adult brain. Furthermore, our data suggest that loss of parent-specific methylation in adult NSCs actively shapes the brain epigenome over time. Given the potential dosage effects on dozens of regulatory genes in the Dlk1-Dio3 region, this epigenetic heterogeneity may account for substantial gene expression differences during aging. Future studies combining allele-specific expression in single cells and transgenic animals will allow us to elucidate the full impact of parent-specific methylation heterogeneity on gene dosage in vivo. Our results may provide a general framework for elucidating the contribution of dynamic changes in epigenetic state to gene dosage in normal developmental context, as well as in disease. The substantial cell-to-cell epigenetic heterogeneity illustrates the limitations of bulk approaches to the study of dynamic epigenetic variations.

EXPERIMENTAL PROCEDURES

Reporter Cell Lines

To generate IG-DMR reporter cell lines, targeting vectors and CRISPR/Cas9 were transfected into 129XCast F1 mESCs using Xfect mESC transfection reagent (Clontech Laboratories), according to the provider's protocol. 48 hr following transfection, cells were FACS sorted for either GFP or tdTomato expression and plated on mouse embryonic fibroblast (MEF) feeder plates. Single colonies were analyzed for proper allelic integration by Southern blot and PCR analysis. Clones carrying the Snrpn-tdTomato reporter targeted into the IG-DMR maternal 129 allele were re-transfected with Snrpn-GFP reporter vector to target the IG-DMR paternal Cast allele to

Figure 5. Cell-type-Specific Reporter Activity in the Adult Brain

(A–C) Representative images of brain sections from 7-week-old mice stained with DAPI (blue), anti-GFP (green), anti-tyrosine hydroxylase (TH, red), anti-glial fibrillary acidic protein (GFAP, red), and anti-calbindin (purple); scale bar, 100 μm. (A) Overlap between GFP and TH is shown in the substantia nigra region; (B) GFP and GFAP in the corpus callosum are mutually exclusive. (C) Whole-mount stitching (left) and region-specific (right) images demonstrate cellular mosaicism in the cerebellum Purkinje cells: most cells are calbindin$^+$GFP$^-$ (right top image), whereas some lobes contain double-positive calbindin$^+$GFP$^+$ cells (right bottom image); scale bar represents 500 μm (left images) and 100 μm (right images).

(D) Representative stitching images of 7-week-old IG-DMR$^{\text{Mat-GFP}}$ cortical layers stained with DAPI (blue), anti-GFP (green), and anti-NeuN (purple). Shown are percentages of GFP$^+$NeuN$^+$ neurons for each layer (right); scale bar, 250 μm.

(E) Heatmap summarizing the percentage of overlap between different cell type markers and GFP as measured in different brain anatomical regions. Shown are mean values of three independent IG-DMR$^{\text{Mat-GFP}}$ brains.

(F) Bisulfite sequencing performed on the IG-DMR region in WT astrocytes isolated from E17.5 and postnatal day 3 (P3) brains. Shown are percentages of methylated CpGs.

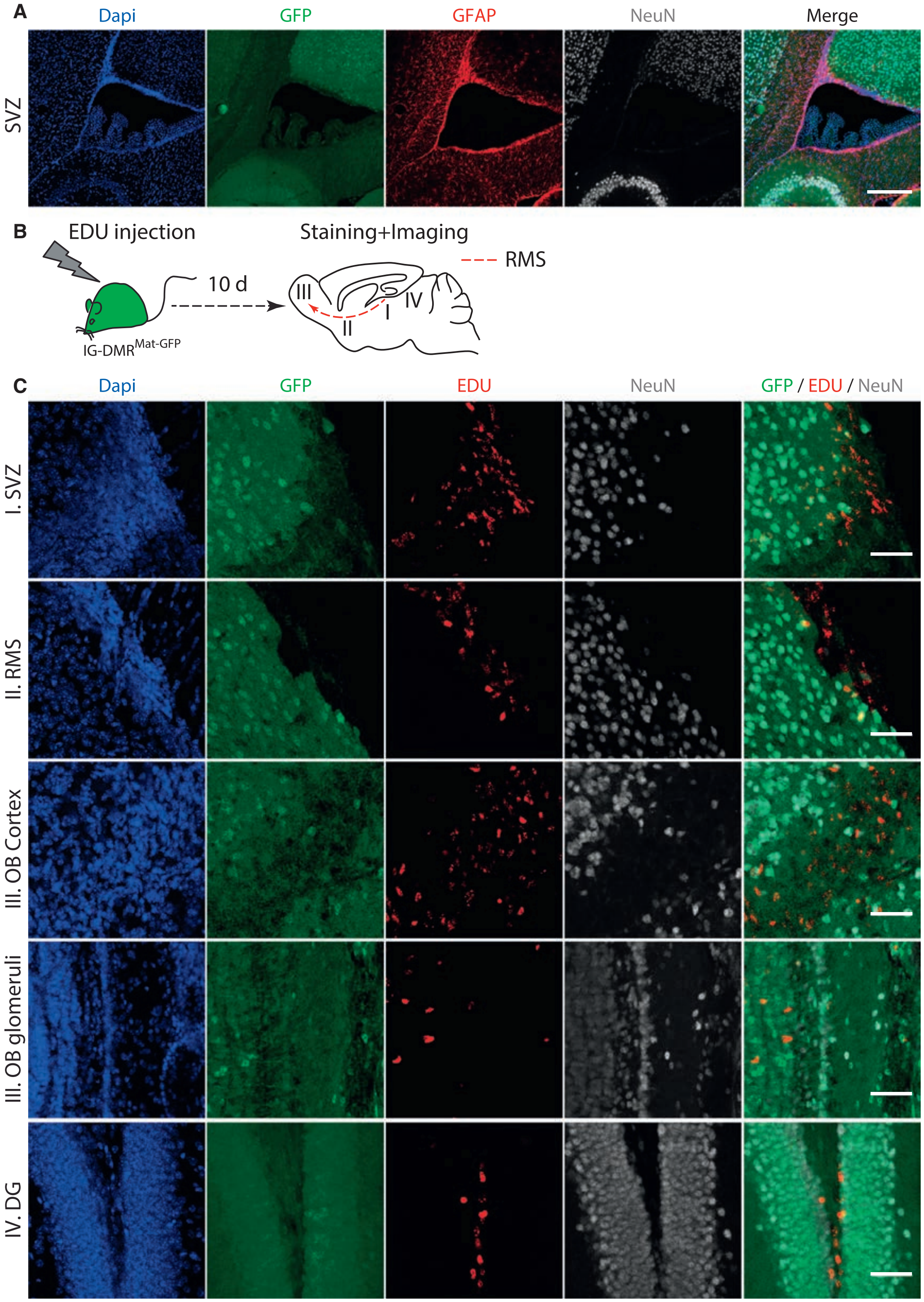


Figure 6. Parent-Specific DNA Methylation in Neural Progenitors of the Adult Brain

(A) Staining of the Sub Ventricular Zone (SVZ) in 7-week-old IG-DMR$^{Mat-GFP}$ brain with DAPI (blue), anti-GFP (green), anti-GFAP (red), and anti-NeuN (gray); scale bar, 250 μm. Shown are stitched images.

(legend continued on next page)

establish double-targeted cells (see complete list of primers in Table S1). To establish blastocyst-derived mESCs, males carrying the IG-DMR-Snrpn-GFP or IG-DMR-Snrpn-Tomato methylation reporter were crossed with BDF1 females following blastocyst isolation. ICM-derived mESCs were obtained according to previously established protocols (Markoulaki et al., 2008).

mESC Culture

Targeted mESCs were cultured on irradiated MEFs with standard ESC medium: (500 ml) DMEM supplemented with 10% fetal bovine serum (FBS) (HyClone), 10 μg recombinant LIF, 0.1 mM β-mercaptoethanol (Sigma-Aldrich), penicillin/streptomycin, 1 mM L-glutamine, and 1% nonessential amino acids (all from Invitrogen). For experiments in 2i culture conditions, mESCs were cultured on gelatin-coated plates with N2B27 + 2i + LIF medium containing (500 ml) 240 ml DMEM/F12 (Invitrogen; 11320), 240 ml Neurobasal media (Invitrogen; 21103), 5 ml N2 supplement (Invitrogen; 17502048), 10 ml B27 supplement (Invitrogen; 17504044), 10 μg recombinant LIF, 0.1 mM β-mercaptoethanol (Sigma-Aldrich), penicillin/streptomycin, 1 mM L-glutamine, and 1% nonessential amino acids (all from Invitrogen), as well as 50 μg/ml BSA (Sigma), PD0325901 (Stemgent, 1 μM), and CHIR99021 (Stemgent, 3 μM).

Tetraploid and Diploid Embryo Injections

All blastocyst injections were performed with B6D2F2 (C57Bl/6xDBA) host embryos. To obtain tetraploid (4n) blastocysts, electrofusion was performed at ~44–47 hr after human chorionic gonadotropin (HCG) injection using a BEX LF-301 cell fusion device (Protech International). Both 4n and 2n embryos were otherwise treated the same and cultured in Evolve KSOMaa (Zenith Biotech) until they formed blastocysts (94–98 hr after HCG injection), at which point they were placed in a drop of Evolve w/HEPES KSOMaa (Zenith) medium under mineral oil for injection. A flat-tip microinjection pipette with an internal diameter of 16 μm (Origio) was used to introduce 10–12 cells into the blastocoel cavity. Within 1–2 hr after injection, blastocysts were transferred to day 2.5 recipient CD1 Elite females (15–20 blastocysts per female).

Generation of Reporter Mice

Male chimeras carrying IG-DMR-Snrpn-GFP methylation reporter were crossed with BDF1 females. Male and female offspring carrying the paternally transmitted allele were bred to obtain offspring carrying a maternally or paternally transmitted allele in the F2 generation (Figure 2A). F2 offspring harboring the reporter allele were analyzed at different ages. Mice were handled in accordance with institutional guidelines and approved by the Committee on Animal Care (CAC) and Department of Comparative Medicine (DCM) of Massachusetts Institute of Technology.

Tissue Processing and Immunohistochemistry

Neonatal and adult mice were perfused via a transcardial route with 4% paraformaldehyde (PFA)/PBS. E9.5–E18.5 embryos were fixed by overnight immersion in 4% PFA/PBS at 4°C. Fixed tissues and embryos were dissected and either imaged intact or sectioned with a vibratome (Leica VT1100) at 100–150 μm or a cryostat (Leica Microsystems) at 15–50 μm thickness followed by immunohistochemical analysis. For vibratome sectioning, tissues were embedded with 3% agarose gel. For cryosectioning, tissues were equilibrated in 30% sucrose/PBS prior to embedding in optimal cutting temperature (OCT) compound. Immunostaining procedures for tissue sections were previously described (Wu et al., 2014). Briefly, sections were permeabilized with PBST (1 − PBS solution with 0.5% Triton X-100) for 1 hr at room temperature before blocking with 10% normal donkey serum (NDS) in PBST. Sections were then incubated with appropriately diluted primary antibodies in PBST with 5% NDS for 12–24 hr at 4°C, washed three times with PBST at room temperature, and then incubated with desired secondary antibodies in TBST with 5% NDS and DAPI to counterstain the nuclei. Sections were washed with PBST three times before they were mounted onto slides with Fluoromount G (SouthernBiotech).

The following antibodies were used in this study: chicken anti-GFP (1:1,000, Aves Labs), mouse anti-NeuN (1:1,000, EMD Millipore), mouse anti-GFAP (1:1,000, Sigma-Aldrich), rabbit anti-GFAP (1:1,000, Dako), rabbit anti-Iba1 (1:1,000, Wako Pure Chemicals Industries), rabbit anti-S100β (1:1,000, Dako), rabbit anti-Pax6 (1:1,000, EMD Millipore), rabbit anti-calbindin (1:1,000, Swant), rabbit anti-PV (1:1,000, Swant), rabbit anti-GAD2 (1:1,000, Sigma-Aldrich), rabbit anti-tyrosine hydroxylase (anti-TH; 1:1,000, EMD Millipore), goat anti-ChAT (1:1,000, EMD Millipore), and Alexa Fluor 647-conjugated GS-IB4 (1:500, Thermo Fisher Scientific).

Microscopy and Image Analysis

Images were captured on a Zeiss LSM710 confocal microscope and processed with Zen software, ImageJ/Fiji, and Adobe Photoshop. For imaging-based quantification, unless otherwise specified, three representative images from different mice were quantified manually and data were plotted with GraphPad.

Flow Cytometry

To assess the proportion of GFP and tdTomato in the established reporter cell lines, mESCs were treated with EDTA to obtain a single-cell suspension and assessed on the LSR II SORP, LSRFortessa SORP, or FACSCanto II.

Bisulfite Conversion, PCR, and Sequencing

Bisulfite conversion of DNA was established using the EpiTect Bisulfite Kit (QIAGEN) following the manufacturer's instructions. The resulting modified DNA was amplified by first round of nested PCR, following a second round using loci specific PCR primers (see complete list of primers in Table S1). The first round of nested PCR was done as follows: 94°C for 4 min, 55°C for 2 min, and 72°C for 2 min; repeat steps 1–3 one time; 94°C for 1 min, 55°C for 2 min, and 72°C for 2 min; repeat steps 5–7 35 times; 72°C for 5 min, and hold at 12°C. The second round of PCR was as follows: 95°C for 4 min, 94°C for 1 min, 55°C for 2 min, and 72°C for 2 min; repeat steps 2–4 35 times; 72°C for 5 min, and hold at 12°C. The resulting amplified products were gel-purified, subcloned into a pCR2.1-TOPO-TA cloning vector (Life Technologies), and sequenced.

Reverse Transcription of RNA and Quantitative Real-Time PCR

RNA was isolated using the RNeasy Mini Kit (QIAGEN) including on-column DNase digest to remove genomic DNA. Reverse transcription was performed on 0.5–1 μg total RNA using random hexamer primers and SuperScript III First-Strand Synthesis SuperMix (Life Technologies) according to the manufacturer's instructions. All PCR reactions were performed in a 96-well plate on an ABI 7900 HT Fast Real-Time PCR system (Applied Biosystems) using FAST SYBR Green Master Mix (Applied Biosystems). Relative quantification of gene expression was calculated using Gapdh primers or primers amplifying the ultraconserved mouse genomic region. See Table S1 for a complete list of primers.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, nine figures, and one table and can be found with this article online at http://dx.doi.org/10.1016/j.celrep.2016.08.066.

AUTHOR CONTRIBUTIONS

Y. Stelzer and R.J. conceived the idea for this project. Y. Stelzer and H.W. designed and conducted experiments and interpreted the data. Y. Song assisted

(B) Schematic representation of EDU labeling of 5-week-old IG-DMR$^{\text{Mat-GFP}}$ mice.

(C) Representative images of EDU positive cells in different anatomical regions of the brain. Shown are staining with DAPI (blue), anti-GFP (green), anti-EDU (red), and anti-NeuN (gray) in the SVZ, RMS, olfactory bulb (OB) cortex and glomeruli, and in dentate gyrus (DG), demonstrating that EDU$^+$ and GFP$^+$ cells are mutually exclusive; scale bar, 50 μm.

with bisulfite sequencing and SNP analysis. C.S.S. assisted with targeting of mESCs, and S.M. conducted 2N and 4N blastocyst injections. Y. Stelzer and R.J. wrote the manuscript with input from all other authors.

ACKNOWLEDGMENTS

We thank Patti Wisniewski and Colin Zollo for FACS analyses and cell sorting, Wendy Salmon for assistance with confocal imaging, and Maria Mihaylova for liver cell dissociation. The authors would like to thank Frank Soldner, Malkiel Cohen, Robert N. Plasschaert, and Shawn Liu for critically reading the manuscript. This study was supported by NIH grant HD 045022. Y. Stelzer is supported by a Human Frontier Postdoctoral Fellowship. H.W. is supported by a NARSAD Young Investigator Fellowship (grant 22950). R.J. is an advisor to Stemgent and a cofounder of Fate Therapeutics and Fulcrum Therapeutics.

Received: April 8, 2016
Revised: August 8, 2016
Accepted: August 19, 2016
Published: September 20, 2016

REFERENCES

Andersen, D.C., Laborda, J., Baladron, V., Kassem, M., Sheikh, S.P., and Jensen, C.H. (2013). Dual role of delta-like 1 homolog (DLK1) in skeletal muscle development and adult muscle regeneration. Development *140*, 3743–3753.

Avior, Y., Sagi, I., and Benvenisty, N. (2016). Pluripotent stem cells in disease modelling and drug discovery. Nat. Rev. Mol. Cell Biol. *17*, 170–182.

Barlow, D.P., and Bartolomei, M.S. (2014). Genomic imprinting in mammals. Cold Spring Harb. Perspect. Biol. *6*, a018382.

Barton, S.C., Ferguson-Smith, A.C., Fundele, R., and Surani, M.A. (1991). Influence of paternally imprinted genes on development. Development *113*, 679–687.

Buganim, Y., Markoulaki, S., van Wietmarschen, N., Hoke, H., Wu, T., Ganz, K., Akhtar-Zaidi, B., He, Y., Abraham, B.J., Porubsky, D., et al. (2014). The developmental potential of iPSCs is greatly influenced by reprogramming factor selection. Cell Stem Cell *15*, 295–309.

Cockett, N.E., Jackson, S.P., Shay, T.L., Farnir, F., Berghmans, S., Snowder, G.D., Nielsen, D.M., and Georges, M. (1996). Polar overdominance at the ovine callipyge locus. Science *273*, 236–238.

Court, F., Tayama, C., Romanelli, V., Martin-Trujillo, A., Iglesias-Platas, I., Okamura, K., Sugahara, N., Simón, C., Moore, H., Harness, J.V., et al. (2014). Genome-wide parent-of-origin DNA methylation analysis reveals the intricacies of human imprinting and suggests a germline methylation-independent mechanism of establishment. Genome Res. *24*, 554–569.

da Rocha, S.T., Edwards, C.A., Ito, M., Ogata, T., and Ferguson-Smith, A.C. (2008). Genomic imprinting at the mammalian Dlk1-Dio3 domain. Trends Genet. *24*, 306–316.

Davies, W., Isles, A.R., and Wilkinson, L.S. (2005). Imprinted gene expression in the brain. Neurosci. Biobehav. Rev. *29*, 421–430.

Ferguson-Smith, A.C. (2011). Genomic imprinting: the emergence of an epigenetic paradigm. Nat. Rev. Genet. *12*, 565–575.

Ferrón, S.R., Charalambous, M., Radford, E., McEwen, K., Wildner, H., Hind, E., Morante-Redolat, J.M., Laborda, J., Guillemot, F., Bauer, S.R., et al. (2011). Postnatal loss of Dlk1 imprinting in stem cells and niche astrocytes regulates neurogenesis. Nature *475*, 381–385.

Ferrón, S.R., Radford, E.J., Domingo-Muelas, A., Kleine, I., Ramme, A., Gray, D., Sandovici, I., Constancia, M., Ward, A., Menheniott, T.R., and Ferguson-Smith, A.C. (2015). Differential genomic imprinting regulates paracrine and autocrine roles of IGF2 in mouse adult neurogenesis. Nat. Commun. *6*, 8265.

Ficz, G., Hore, T.A., Santos, F., Lee, H.J., Dean, W., Arand, J., Krueger, F., Oxley, D., Paul, Y.L., Walter, J., et al. (2013). FGF signaling inhibition in ESCs drives rapid genome-wide demethylation to the epigenetic ground state of pluripotency. Cell Stem Cell *13*, 351–359.

Frost, J.M., and Moore, G.E. (2010). The importance of imprinting in the human placenta. PLoS Genet. *6*, e1001015.

Georgiades, P., Watkins, M., Surani, M.A., and Ferguson-Smith, A.C. (2000). Parental origin-specific developmental defects in mice with uniparental disomy for chromosome 12. Development *127*, 4719–4728.

Ginart, P., Kalish, J.M., Jiang, C.L., Yu, A.C., Bartolomei, M.S., and Raj, A. (2016). Visualizing allele-specific expression in single cells reveals epigenetic mosaicism in an H19 loss-of-imprinting mutant. Genes Dev. *30*, 567–578.

Habibi, E., Brinkman, A.B., Arand, J., Kroeze, L.I., Kerstens, H.H., Matarese, F., Lepikhov, K., Gut, M., Brun-Heath, I., Hubner, N.C., et al. (2013). Whole-genome bisulfite sequencing of two distinct interconvertible DNA methylomes of mouse embryonic stem cells. Cell Stem Cell *13*, 360–369.

Hansen, C.H., and van Oudenaarden, A. (2013). Allele-specific detection of single mRNA molecules in situ. Nat. Methods *10*, 869–871.

Hon, G.C., Rajagopal, N., Shen, Y., McCleary, D.F., Yue, F., Dang, M.D., and Ren, B. (2013). Epigenetic memory at embryonic enhancers identified in DNA methylation maps from adult mouse tissues. Nat. Genet. *45*, 1198–1206.

Kundaje, A., Meuleman, W., Ernst, J., Bilenky, M., Yen, A., Heravi-Moussavi, A., Kheradpour, P., Zhang, Z., Wang, J., Ziller, M.J., et al.; Roadmap Epigenomics Consortium (2015). Integrative analysis of 111 reference human epigenomes. Nature *518*, 317–330.

Labialle, S., Marty, V., Bortolin-Cavaillé, M.L., Hoareau-Osman, M., Pradère, J.P., Valet, P., Martin, P.G., and Cavaillé, J. (2014). The miR-379/miR-410 cluster at the imprinted Dlk1-Dio3 domain controls neonatal metabolic adaptation. EMBO J. *33*, 2216–2230.

Lee, H.J., Hore, T.A., and Reik, W. (2014). Reprogramming the methylome: erasing memory and creating diversity. Cell Stem Cell *14*, 710–719.

Li, E., Beard, C., and Jaenisch, R. (1993). Role for DNA methylation in genomic imprinting. Nature *366*, 362–365.

Lin, S.P., Youngson, N., Takada, S., Seitz, H., Reik, W., Paulsen, M., Cavaille, J., and Ferguson-Smith, A.C. (2003). Asymmetric regulation of imprinting on the maternal and paternal chromosomes at the Dlk1-Gtl2 imprinted cluster on mouse chromosome 12. Nat. Genet. *35*, 97–102.

Lin, S.P., Coan, P., da Rocha, S.T., Seitz, H., Cavaille, J., Teng, P.W., Takada, S., and Ferguson-Smith, A.C. (2007). Differential regulation of imprinting in the murine embryo and placenta by the Dlk1-Dio3 imprinting control region. Development *134*, 417–426.

Markoulaki, S., Meissner, A., and Jaenisch, R. (2008). Somatic cell nuclear transfer and derivation of embryonic stem cells in the mouse. Methods *45*, 101–114.

McGrath, J., and Solter, D. (1984). Completion of mouse embryogenesis requires both the maternal and paternal genomes. Cell *37*, 179–183.

Moon, Y.S., Smas, C.M., Lee, K., Villena, J.A., Kim, K.H., Yun, E.J., and Sul, H.S. (2002). Mice lacking paternally expressed Pref-1/Dlk1 display growth retardation and accelerated adiposity. Mol. Cell. Biol. *22*, 5585–5592.

Perez, J.D., Rubinstein, N.D., Fernandez, D.E., Santoro, S.W., Needleman, L.A., Ho-Shing, O., Choi, J.J., Zirlinger, M., Chen, S.K., Liu, J.S., and Dulac, C. (2015). Quantitative and functional interrogation of parent-of-origin allelic expression biases in the brain. eLife *4*, e07860.

Peters, J. (2014). The role of genomic imprinting in biology and disease: an expanding view. Nat. Rev. Genet. *15*, 517–530.

Qian, P., He, X.C., Paulson, A., Li, Z., Tao, F., Perry, J.M., Guo, F., Zhao, M., Zhi, L., Venkatraman, A., et al. (2016). The Dlk1-Gtl2 locus preserves LT-HSC function by inhibiting the PI3K-mTOR pathway to restrict mitochondrial metabolism. Cell Stem Cell *18*, 214–228.

Reik, W. (2007). Stability and flexibility of epigenetic gene regulation in mammalian development. Nature *447*, 425–432.

Reik, W., and Walter, J. (2001). Genomic imprinting: parental influence on the genome. Nat. Rev. Genet. *2*, 21–32.

Robertson, K.D. (2005). DNA methylation and human disease. Nat. Rev. Genet. *6*, 597–610.

Royo, H., and Cavaillé, J. (2008). Non-coding RNAs in imprinted gene clusters. Biol. Cell *100*, 149–166.

Seitz, H., Youngson, N., Lin, S.P., Dalbert, S., Paulsen, M., Bachellerie, J.P., Ferguson-Smith, A.C., and Cavaillé, J. (2003). Imprinted microRNA genes transcribed antisense to a reciprocally imprinted retrotransposon-like gene. Nat. Genet. *34*, 261–262.

Sekita, Y., Wagatsuma, H., Nakamura, K., Ono, R., Kagami, M., Wakisaka, N., Hino, T., Suzuki-Migishima, R., Kohda, T., Ogura, A., et al. (2008). Role of retrotransposon-derived imprinted gene, Rtl1, in the feto-maternal interface of mouse placenta. Nat. Genet. *40*, 243–248.

Sittig, L.J., and Redei, E.E. (2014). Fine-tuning notes in the behavioral symphony: parent-of-origin allelic gene expression in the brain. Adv. Genet. *86*, 93–106.

Smallwood, S.A., Lee, H.J., Angermueller, C., Krueger, F., Saadeh, H., Peat, J., Andrews, S.R., Stegle, O., Reik, W., and Kelsey, G. (2014). Single-cell genome-wide bisulfite sequencing for assessing epigenetic heterogeneity. Nat. Methods *11*, 817–820.

Stadtfeld, M., Apostolou, E., Akutsu, H., Fukuda, A., Follett, P., Natesan, S., Kono, T., Shioda, T., and Hochedlinger, K. (2010). Aberrant silencing of imprinted genes on chromosome 12qF1 in mouse induced pluripotent stem cells. Nature *465*, 175–181.

Stelzer, Y., and Jaenisch, R. (2015). Monitoring dynamics of DNA methylation at single-cell resolution during development and disease. Cold Spring Harb. Symp. Quant. Biol. *80*, 199–206.

Stelzer, Y., Ronen, D., Bock, C., Boyle, P., Meissner, A., and Benvenisty, N. (2013). Identification of novel imprinted differentially methylated regions by global analysis of human-parthenogenetic-induced pluripotent stem cells. Stem Cell Reports *1*, 79–89.

Stelzer, Y., Shivalila, C.S., Soldner, F., Markoulaki, S., and Jaenisch, R. (2015). Tracing dynamic changes of DNA methylation at single-cell resolution. Cell *163*, 218–229.

Surani, M.A., and Barton, S.C. (1983). Development of gynogenetic eggs in the mouse: implications for parthenogenetic embryos. Science *222*, 1034–1036.

Takahashi, N., Okamoto, A., Kobayashi, R., Shirai, M., Obata, Y., Ogawa, H., Sotomaru, Y., and Kono, T. (2009). Deletion of Gtl2, imprinted non-coding RNA, with its differentially methylated region induces lethal parent-origin-dependent defects in mice. Hum. Mol. Genet. *18*, 1879–1888.

Tam, P.P., and Rossant, J. (2003). Mouse embryonic chimeras: tools for studying mammalian development. Development *130*, 6155–6163.

Tevendale, M., Watkins, M., Rasberry, C., Cattanach, B., and Ferguson-Smith, A.C. (2006). Analysis of mouse conceptuses with uniparental duplication/deficiency for distal chromosome 12: comparison with chromosome 12 uniparental disomy and implications for genomic imprinting. Cytogenet. Genome Res. *113*, 215–222.

Thomson, J.A., and Solter, D. (1988). The developmental fate of androgenetic, parthenogenetic, and gynogenetic cells in chimeric gastrulating mouse embryos. Genes Dev. *2*, 1344–1351.

Tucker, K.L., Beard, C., Dausmann, J., Jackson-Grusby, L., Laird, P.W., Lei, H., Li, E., and Jaenisch, R. (1996). Germ-line passage is required for establishment of methylation and expression patterns of imprinted but not of nonimprinted genes. Genes Dev. *10*, 1008–1020.

Wilkinson, L.S., Davies, W., and Isles, A.R. (2007). Genomic imprinting effects on brain development and function. Nat. Rev. Neurosci. *8*, 832–843.

Wu, H., Luo, J., Yu, H., Rattner, A., Mo, A., Wang, Y., Smallwood, P.M., Erlanger, B., Wheelan, S.J., and Nathans, J. (2014). Cellular resolution maps of X chromosome inactivation: implications for neural development, function, and disease. Neuron *81*, 103–119.

Xie, W., Barr, C.L., Kim, A., Yue, F., Lee, A.Y., Eubanks, J., Dempster, E.L., and Ren, B. (2012). Base-resolution analyses of sequence and parent-of-origin dependent DNA methylation in the mouse genome. Cell *148*, 816–831.

Yamazawa, K., Ogata, T., and Ferguson-Smith, A.C. (2010). Uniparental disomy and human disease: an overview. Am. J. Med. Genet. C. Semin. Med. Genet. *154C*, 329–334.

Ziller, M.J., Gu, H., Müller, F., Donaghey, J., Tsai, L.T., Kohlbacher, O., De Jager, P.L., Rosen, E.D., Bennett, D.A., Bernstein, B.E., et al. (2013). Charting a dynamic DNA methylation landscape of the human genome. Nature *500*, 477–481.

Zvetkova, I., Apedaile, A., Ramsahoye, B., Mermoud, J.E., Crompton, L.A., John, R., Feil, R., and Brockdorff, N. (2005). Global hypomethylation of the genome in XX embryonic stem cells. Nat. Genet. *37*, 1274–1279.

Cell Reports
Article

The Gut Microbiota Modulates Energy Metabolism in the Hibernating Brown Bear *Ursus arctos*

Felix Sommer,[1,2] Marcus Ståhlman,[1] Olga Ilkayeva,[3] Jon M. Arnemo,[4,5] Jonas Kindberg,[5] Johan Josefsson,[6] Christopher B. Newgard,[3] Ole Fröbert,[6] and Fredrik Bäckhed[1,7,*]
[1]The Wallenberg Laboratory, Department of Molecular and Clinical Medicine, University of Gothenburg, 41345 Gothenburg, Sweden
[2]Institute for Clinical Molecular Biology, University of Kiel, Schittenhelmstraße 12, 24105 Kiel, Germany
[3]Sarah W. Stedman Nutrition and Metabolism Center and Duke Molecular Physiology Institute and Departments of Pharmacology and Cancer Biology and Medicine, Duke University Medical Center, Durham, NC 27701, USA
[4]Faculty of Forestry and Wildlife Management, Hedmark University College, 2418 Elverum, Norway
[5]Department of Wildlife, Fish, and Environmental Studies, Swedish University of Agricultural Sciences, 90183 Umeå, Sweden
[6]Department of Cardiology, Faculty of Health, Örebro University, 70185 Örebro, Sweden
[7]Novo Nordisk Foundation Center for Basic Metabolic Research, Section for Metabolic Receptology and Enteroendocrinology, Faculty of Health Sciences, University of Copenhagen, 2200 Copenhagen, Denmark
*Correspondence: fredrik.backhed@wlab.gu.se

http://dx.doi.org/10.1016/j.celrep.2016.01.026

SUMMARY

Hibernation is an adaptation that helps many animals to conserve energy during food shortage in winter. Brown bears double their fat depots during summer and use these stored lipids during hibernation. Although bears seasonally become obese, they remain metabolically healthy. We analyzed the microbiota of free-ranging brown bears during their active phase and hibernation. Compared to the active phase, hibernation microbiota had reduced diversity, reduced levels of Firmicutes and Actinobacteria, and increased levels of Bacteroidetes. Several metabolites involved in lipid metabolism, including triglycerides, cholesterol, and bile acids, were also affected by hibernation. Transplantation of the bear microbiota from summer and winter to germ-free mice transferred some of the seasonal metabolic features and demonstrated that the summer microbiota promoted adiposity without impairing glucose tolerance, suggesting that seasonal variation in the microbiota may contribute to host energy metabolism in the hibernating brown bear.

INTRODUCTION

Free-ranging brown bears (*Ursus arctos*) undergo cycles of intense eating and weight gain during the summer followed by prolonged dormant hypometabolic fasting for up to 6 months during the winter (Evans et al., 2012; Tøien et al., 2011). Despite the large fat accumulation before hibernation, bears remain metabolically healthy (Arinell et al., 2012; Nelson, 1973; Stenvinkel et al., 2013), which contrasts with the strong association between obesity and insulin resistance in humans. Thus, the brown bear may constitute a model for healthy obesity and studying hibernation might be a promising approach to develop novel therapies for obesity. The intestines of mammals harbor diverse microbial ecosystems that have profound effects on host physiology (Sommer and Bäckhed, 2013). The gut microbiota contributes to energy harvest from the diet (Bäckhed et al., 2004, 2007; Sommer and Bäckhed, 2013; Sommer et al., 2015) and is altered in obesity and type 2 diabetes (Khan et al., 2014). Furthermore, diet, which is seasonably variable in bears (Persson et al., 2001; Stenvinkel et al., 2013; Stofik et al., 2013), strongly affects the gut microbiota (Ley et al., 2008; Zoetendal and de Vos, 2014) and both fasting and hibernation alter the gut microbiota composition (Carey et al., 2013; Crawford et al., 2009; Dill-McFarland et al., 2014; Sonoyama et al., 2009).

Here, we investigated how hibernation in free-ranging brown bears affects the gut microbiota and plasma metabolites, and whether a seasonally altered microbiota contributes to the healthy obesity phenotype during summer. We used 16S rRNA profiling and next-generation sequencing to comprehensively analyze the fecal microbiota of free-ranging brown bears captured during hibernation (February) and during the active period (June) of the same year (Figure 1A). We showed that the winter microbiota comprised fewer bacterial taxa (Figure S1A) and was more homogenous than the summer microbiota (Figure S1B), which may reflect the varied diet among bears during the summer.

RESULTS AND DISCUSSION

Principal coordinate analysis of the overall composition of the bear fecal microbiota samples using unweighted UniFrac revealed a clear separation depending on the seasonal origin (Figures 1B and S2). We identified 24 bacterial phyla in the bear fecal microbiota (Table S1). The dominating bacterial phyla in the summer microbiota were Proteobacteria, Firmicutes, and Actinobacteria (Figure 1C). In the winter microbiota, Bacteroidetes increased in abundance, whereas Firmicutes and Actinobacteria were less abundant (Figure 1D). A number of low abundant phyla were only present in the summer microbiota (Figure S3A). At the species level, 199 of the 4,447 detected operational taxonomic units (OTUs) were significantly altered ($q < 0.05$; q = FDR

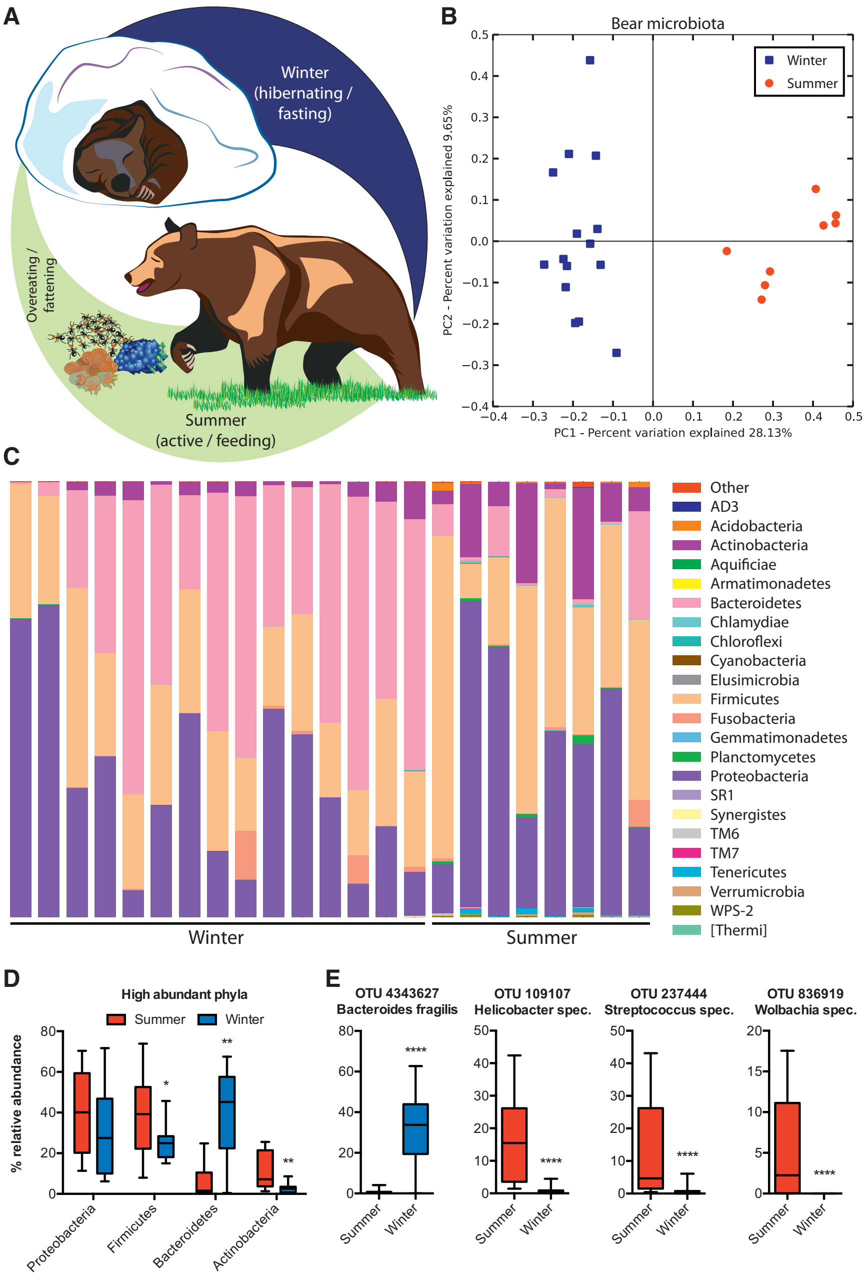


Figure 1. Seasonal Differences in the Bear Fecal Microbiota

(A) Seasonal cycle of the brown bear.

(B) Principal coordinate analysis of the bear fecal microbiota from summer and winter.

(C) Bacterial taxonomic representation in bear microbiota in summer and winter on phylum level.

(D) Significantly altered bacterial phyla between summer and winter in bear microbiota.

(E) Relative abundance (%) of high-abundant and season-dependent OTUs of the bear fecal microbiota.

Data are mean ± SEM n = 8 for summer and n = 15 for winter. *q < 0.05; **q < 0.01; ****q < 0.0001.

corrected p value) between winter and summer (Table S1). The most significant among these were OTU 4343627 (*Bacteroides fragilis*), which was enriched in the winter, and OTUs 109107 (*Helicobacter spec.*), 237444 (*Streptococcus spec.*), and 836919 (*Wolbachia spec.*), which were reduced in winter (Figure 1E). *Wolbachia* are symbionts of several insects (Hedges et al., 2008; Teixeira et al., 2008), and the increased abundance during summer presumably results from the intake of insects as part of the bear summer diet (Große et al., 2003). Furthermore, the winter microbiota had higher levels of several Enterobacteriaceae and lower levels of several Rhizobiales and Actinomycetales species (Figures S3B–S3D).

An enrichment of Bacteroidetes and lower relative abundance of Firmicutes has previously been observed in the microbiota of hibernating animals (Carey et al., 2013; Dill-McFarland et al., 2014; Sonoyama et al., 2009; Stevenson et al., 2014). The increase in Bacteroidetes may be explained by their capacity to switch their metabolism toward degradation of host glycans in the absence of dietary polysaccharides (Sonnenburg et al., 2005) or their capacity to metabolize protein and fat (Wu et al., 2011) putatively provided by the intestinal epithelium. In contrast, most Firmicutes taxa require dietary fiber. These changes in the microbiota phyla were accompanied by a loss of weight and body fat in the hibernating bear. Similar trends have been reported in studies comparing obese and lean subjects (Ley et al., 2005) or using calorie restriction in humans and mice (Crawford et al., 2009; Furet et al., 2010; Ley et al., 2006; Turnbaugh et al., 2009). Furthermore, two studies of calorie-restricted mice reported an increase in *Bacteroides fragilis* (Santacruz et al., 2009) but reduced Streptococcaceae and TM7 (Zhang et al., 2013). *Bacteroides fragilis* was the predominant bacterium in the microbiota from the hibernating bears, whereas both *Streptococcus* and TM7 were reduced during hibernation. Together these data indicate that many of the changes in the bear microbiota are associated with caloric restriction. In contrast to small hibernators, Verrucomicrobia including *Akkermansia muciniphila* were not increased during hibernation in free-ranging brown bears.

To identify metabolites that varied according to the activity status of the bears, we used targeted metabolomics to analyze blood samples taken from the jugular vein during winter and summer. Supporting previous publications (LeBlanc et al., 2001; Otis et al., 2011), we found that serum levels of cholesteryl esters, triglycerides, and free cholesterol were significantly

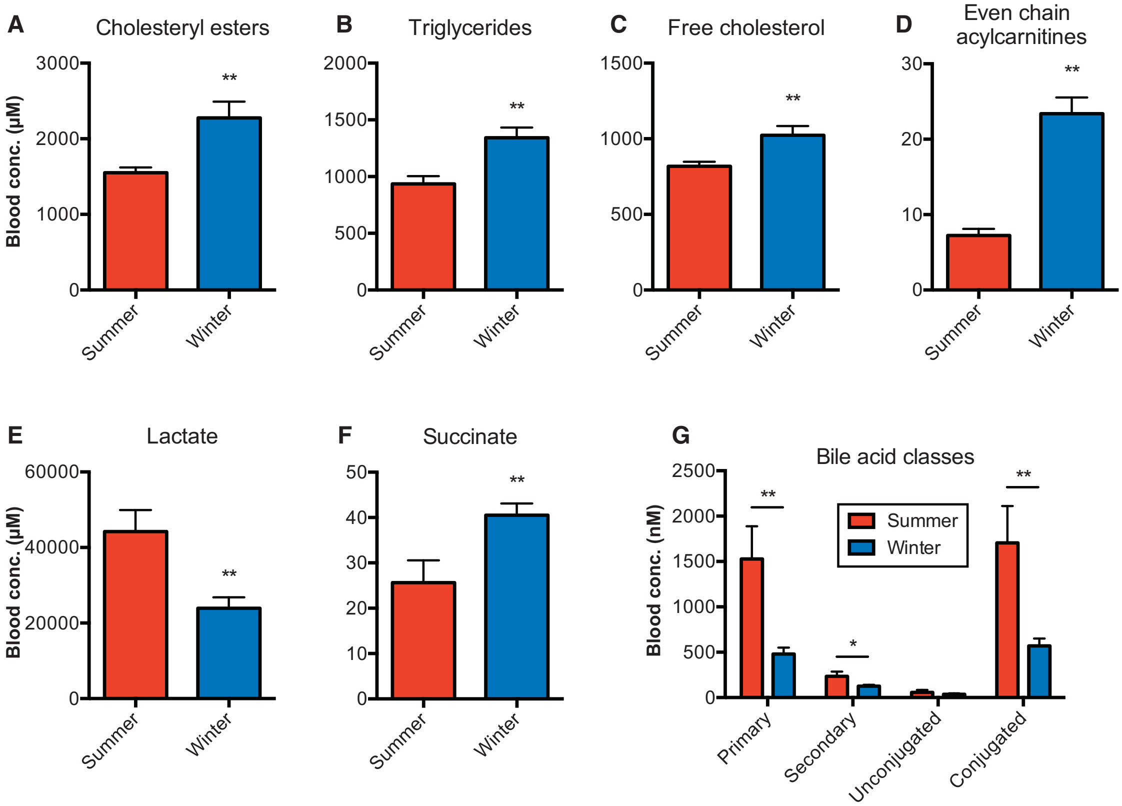


Figure 2. Seasonal Differences in Metabolites in Bear Blood

(A–G) Concentrations of the lipid classes (A) cholesteryl esters, (B) triglycerides, and (C) free cholesterol and of (D) even chain acylcarnitines, markers of FAO, and of the organic acids (E) lactate and (F) succinate, and of (G) bile acid classes in bear blood from summer and winter. Data are presented as mean ± SEM n = 12 for summer and n = 15 for winter. $^{*}q < 0.05$; $^{**}q < 0.01$.

higher in winter (Figures 2A–2C; Table S2). This finding is consistent with the fact that energy from fat stores is obtained by lipolysis during hibernation (Arinell et al., 2012; Nelson, 1973). Even chain acylcarnitines, which largely represent intermediates of mitochondrial fatty acid oxidation (FAO), were also higher in winter (Figure 2D), in agreement with high FAO activity during hibernation. In contrast, C3 and C5 acylcarnitines, markers of amino acid oxidation, were decreased during hibernation (Table S2). Finally lactate (Figure 2E) and the levels of several gluconeogenic amino acids (e.g., alanine, methionine, tyrosine) were reduced (Table S2) during winter compared to summer, whereas succinate was increased, which could indicate reduced glucose utilization and increased gluconeogenesis. Taken together, these data suggest that during hibernation bears mobilize and oxidize lipids as survival strategy, accompanied by reduced glucose utilization and increased utilization of amino acids for gluconeogenesis. The decreased lactate levels were also consistent with decreased abundance of lactate-producing bacteria during hibernation; e.g., the Firmicutes *Bacillus* or *Lactobacillus* and the Actinobacteria *Micrococcus* (Reddy et al., 2008). Similarly, the increased succinate levels during hibernation correlate with increased abundance of succinate-producing bacteria such as *Enterococcus* (Song and Lee, 2005) in the winter microbiota. However, the host can also produce lactate and succinate, and thus seasonal changes in host metabolism might also contribute to the differences in lactate and succinate.

We also observed that total bile acid levels in the serum were lower in the winter with large reductions in primary and conjugated bile acids (Figure 2G; Table S2). Notably, expression of the rate-limiting enzyme of bile acid production CYP7A1 is reduced in the liver of hibernating mammals (Fedorov et al., 2011; Otis et al., 2011), and the microbiota contributes to modifications of bile acids (Sayin et al., 2013). Bile acids promote lipid uptake and respond to food intake. Bears do not eat for up to 6 months during hibernation, which likely explains the reduced bile acid levels in the winter. Levels of deoxycholic acid and lithocholic acid, both of which are dependent on the microbiota, are known to have hemolytic activity (Oelberg et al., 1984; Schölmerich et al., 1984) and were reduced during hibernation (Table S2). Notably, several blood parameters that are linked to hemolysis and dehydration were also altered during hibernation (Table 1). For example, levels of red blood cells and hemoglobin were higher in the winter, whereas lactate dehydrogenase (marker of hemolysis) and bilirubin (used during recycling of hemoglobin) levels decreased during hibernation. Thus, microbiota-dependent changes in the bile acid profile might contribute to the reduced hemolysis during hibernation.

To test whether the seasonal differences in the bear microbiota affect host physiology, we colonized germ-free mice with a

Table 1. Signs of Dehydration and Reduced Hemolysis in Blood of Hibernating Brown Bears

Parameter	Unit	Summer Mean (Range)	Winter Mean (Range)	Ratio W/S	p Value
Total bile acids	nM	1,762 (137–4,379)	606 (171–1,177)	0.3	<0.01
White blood cells	10^9/l	7.7 (3.4–15.7)	6.2 (3.8–15.6)	0.8	ns
Red blood cells	10^{12}/l	6.6 (6.2–7)	8.6 (7.7–9.4)	1.3	<0.001
Hemoglobin	g/l	161 (132–176)	203 (183–223)	1.3	<0.001
Hematocrit	%	42.6 (36.6–46.1)	54.2 (48–60)	1.3	<0.001
Platelets	10^9/l	310 (251–359)	184 (65–265)	0.6	<0.001
Neutrophils	10^9/l	3.3 (2.4–4.4)	3.5 (2.1–4.4)	1.1	ns
Lymphocytes	10^9/l	1.2 (0.8–1.9)	1.5 (0.9–2.7)	1.2	ns
Monocytes	10^9/l	0.3 (0.2–0.5)	0.4 (0.3–0.7)	1.3	ns
Eosinophils	10^9/l	0.003 (0–0.01)	0 (0–0)	0.0	<0.05
Basophils	10^9/l	0.01 (0–0.04)	0.001 (0–0.01)	0.2	ns
Alkaline phosphatase	U/l	134 (100–174)	19.1 (13–27)	0.1	<0.001
Alanine transaminase	U/l	36 (23–60)	11.4 (9–14)	0.3	<0.001
Aspartate transaminase	U/l	90 (57–148)	53.1 (39–85)	0.6	<0.05
Bilirubin	µM	18 (9.9–30.9)	10.6 (5–23)	0.6	0.08
Lactate dehydrogenase	µkat/l	13.3 (13.3–13.3)	9.1 (7.2–11.2)	0.7	<0.001
Gamma glutamyltransferase	µkat/l	0.5 (0.3–0.7)	0.3 (0.2–0.5)	0.5	<0.001
C-reactive protein	mg/l	0.003 (0–0.01)	0.014 (0–0.04)	5.1	ns

Hematology analysis was performed on blood samples from brown bears during summer and winter, and marker enzymes were measured. Data show mean and range with n = 11–15 for summer and n = 7 for winter. ns, nonsignificant.

summer or winter bear microbiota (Figure 3A). 16S rRNA profiling of the colonized mice confirmed successful colonization (Figures 3B and S4). There was no seasonal difference in alpha diversity, possibly because all mice received the same food. Mice colonized with a summer bear microbiota trended toward a greater weight ($p = 0.09$) and showed a greater fat gain than mice colonized with a winter bear microbiota (Figures 3C and 3D) but did not display a significant difference in epididymal white adipose tissue weight (Figure 3E). In humans, adiposity is associated with reduced insulin sensitivity (Shulman, 2014). In contrast, brown bears seem to become only temporarily insulin resistant with mild hyperglycemia during hibernation but remain insulin sensitive during the rest of the seasonal cycle independent of fat accumulation (L. Nelson, personal communication; Stenvinkel et al., 2013). The increased weight and adiposity of the mice colonized with a summer bear microbiota were not due to higher bacterial abundance as tested by 16S rDNA qPCR (summer $2.27 \pm 1.16 \times 10^{11}$ and winter $2.55 \pm 1.03 \times 10^{11}$ 16S rDNA copies/g cecal content, $p = 0.6$). Despite their increased fat mass, mice colonized with summer bear microbiota showed no differences or even a slight improvement in glucose metabolism compared to mice colonized with a winter microbiota (Figure 3F). By performing targeted metabolomics, we showed that the seasonal metabolic phenotype of the bears could be partially transferred to germ-free mice by colonization with a bear microbiota. For example, mice colonized with a winter bear microbiota trended toward slightly higher serum levels of cholesteryl esters (Table S2) and triglycerides (Figure 3G) compared with mice that were colonized with summer microbiota.

Conclusions

In conclusion, our data show that the seasonal lifestyle of the brown bear with phases of severe hyperphagia and fat accumulation in the summer and prolonged fasting and inactivity during hibernation is accompanied by seasonal changes in metabolism and microbiota. Furthermore, colonization with the seasonal bear microbiota was sufficient to transfer some of the seasonal metabolic features to germ-free mice. Together this might indicate that the seasonal differences in the bear microbiota contribute to the seasonal metabolic changes, presumably due to the different physiologic demands of phases of severe hyperphagia and hibernation with prolonged fasting. Thus, the microbiota may be linked to the healthy obesity phenotype in brown bears and as such not only yields insights into the physiology of hibernating mammals, but also further supports targeting the microbiota as potential treatment of obesity. However, studying free-ranging animals also limits our information regarding, e.g., seasonal food intake and the experimental procedures that can be performed. Although informative, our findings from the colonization of germ-free mice cannot adequately reflect the physiologic state in a hibernating bear. Thus, further mechanistic studies using, e.g., bears in captivity in which calorie content, food consumption, microbiota composition, and the animal's physiology can be controlled, are required to functionally validate and elucidate which components of the microbiota contribute to the seasonal metabolic differences and the involved molecular pathways.

EXPERIMENTAL PROCEDURES

Blood and fecal samples were taken from 16 free-ranging Eurasian brown bears (*Ursus arctos*) during hibernation (February or March) and during the active period (June) of the same year.

Germ-free mice were colonized with a winter or summer bear fecal microbiota by oral gavage. Body composition was analyzed before and 14 days after colonization by MRI (EchoMRI) according to the manufacturer's instructions. Intraperitoneal glucose tolerance test was performed on day 15 post-colonization.

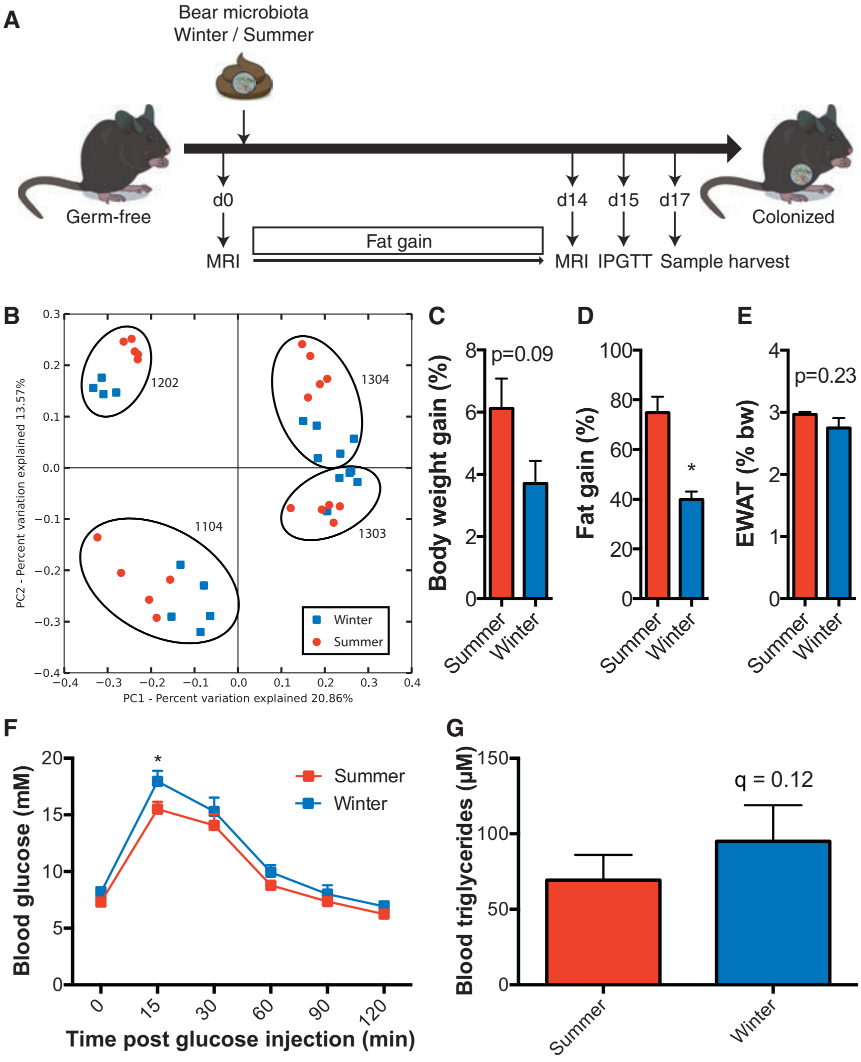


Figure 3. Metabolic Programming by the Seasonal Bear Microbiota

(A) Experimental scheme. Germ-free mice were colonized with a bear summer or winter microbiota and followed for 2 weeks.

(B) Principal coordinate analysis of the cecal microbiota of mice colonized with a bear fecal microbiota from summer or winter. 1104, 1202, 1303, and 1304 denote bear fecal donors.

(C–E) Weight gain (C), body-fat gain (D), and epididymal white adipose tissue (EWAT) (D) weight were determined.

(F) Glucose metabolism was assessed via intraperitoneal glucose tolerance test (IPGTT).

(G) Concentrations of triglycerides in blood of mice colonized with seasonal bear microbiota. Data are mean ± SEM of four experiments (n = 4) with each five animals per colonization. $^{*}p < 0.05$.

DNA was isolated from bear feces and ceca of colonized mice and 16S rRNA profiling performed as described previously (Sommer et al., 2014) by MiSeq sequencing.

Metabolites were analyzed by mass spectrometry as described.

Data were analyzed by Student's t test, and the statistical p values were further corrected for multiple testing using Benjamini Hochberg method in R program. Data are presented as mean ± SEM.

For detailed description of all experimental procedures, see the Supplemental Information.

ACCESSION NUMBERS

The Miseq sequences derived from the 16S profiling have been deposited to the European Nucleotide Archive and are available under accession numbers ERS1023032–ERS1023094.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, four figures, and two tables and can be found with this article online at http://dx.doi.org/10.1016/j.celrep.2016.01.026.

AUTHOR CONTRIBUTIONS

F.S. and F.B. conceived and designed the experiments; F.S., M.S., and O.I. performed the experiments: F.S., M.S., O.I., and F.B. analyzed the data; J.M.A., J.K., and J.J. contributed reagents/materials/analysis tools; F.S. and F.B. wrote the paper; and F.S., M.S., O.I., J.M.A., J.K., J.J., C.B.N., O.F., and F.B. commented on the manuscript.

ACKNOWLEDGMENTS

F.B. is founder and owns equity in Metabogen AB. We thank Valentina Tremaroli and Rozita Akrami for bioinformatics assistance and Valentina Tremaroli for performing 16S rRNA qPCR. The computations for pre-processing of 16S rRNA gene sequences were performed on resources provided by Swedish National Infrastructure for Computing (SNIC) through the Uppsala Multidisciplinary Center for Advanced Computational Science (UPPMAX). We thank Nina Sommer, Carina Arvidsson, and Anna Hallén for assistance with the colonization of germ-free mice and metabolic measurements, Rosie Perkins for editing the manuscript, Antonio Molinaro and Valentina Tremaroli for helpful discussions, Manuela Krämer for technical support with the microbiota sequencing, and Anna Hallén for help with the artwork. We thank Sven

Brunberg for bear handling and logistics. This study was supported by the Swedish Research Council, Torsten Söderberg and Ragnar Söderberg foundations, IngaBritt and Arne Lundberg's foundation, Novo Nordisk Foundation, Swedish Foundation for Strategic Research, Knut and Alice Wallenberg foundation, the regional agreement on medical training and clinical research (ALF) between Region Västra Götaland and Sahlgrenska University Hospital and the Lundbeck Foundation (R126-2012-12408). The Scandinavian Brown Bear Research Project is funded by the Swedish Environmental Protection Agency, the Norwegian Environment Agency, the Swedish Association for Hunting and Wildlife Management, WWF Sweden and the Research Council of Norway. F.B. is a recipient of ERC Consolidator Grant (European Research Council, Consolidator grant 615362, METABASE). This is paper no. 199 from the Scandinavian Brown Bear Research Project.

Received: September 30, 2015
Revised: November 20, 2015
Accepted: January 4, 2016
Published: February 4, 2016

REFERENCES

Arinell, K., Sahdo, B., Evans, A.L., Arnemo, J.M., Baandrup, U., and Fröbert, O. (2012). Brown bears (Ursus arctos) seem resistant to atherosclerosis despite highly elevated plasma lipids during hibernation and active state. Clin. Transl. Sci. *5*, 269–272.

Bäckhed, F., Ding, H., Wang, T., Hooper, L.V., Koh, G.Y., Nagy, A., Semenkovich, C.F., and Gordon, J.I. (2004). The gut microbiota as an environmental factor that regulates fat storage. Proc. Natl. Acad. Sci. USA *101*, 15718–15723.

Bäckhed, F., Manchester, J.K., Semenkovich, C.F., and Gordon, J.I. (2007). Mechanisms underlying the resistance to diet-induced obesity in germ-free mice. Proc. Natl. Acad. Sci. USA *104*, 979–984.

Carey, H.V., Walters, W.A., and Knight, R. (2013). Seasonal restructuring of the ground squirrel gut microbiota over the annual hibernation cycle. Am. J. Physiol. Regul. Integr. Comp. Physiol. *304*, R33–R42.

Crawford, P.A., Crowley, J.R., Sambandam, N., Muegge, B.D., Costello, E.K., Hamady, M., Knight, R., and Gordon, J.I. (2009). Regulation of myocardial ketone body metabolism by the gut microbiota during nutrient deprivation. Proc. Natl. Acad. Sci. USA *106*, 11276–11281.

Dill-McFarland, K.A., Neil, K.L., Zeng, A., Sprenger, R.J., Kurtz, C.C., Suen, G., and Carey, H.V. (2014). Hibernation alters the diversity and composition of mucosa-associated bacteria while enhancing antimicrobial defence in the gut of 13-lined ground squirrels. Mol. Ecol. *23*, 4658–4669.

Evans, A.L., Sahlén, V., Støen, O.G., Fahlman, Å., Brunberg, S., Madslien, K., Fröbert, O., Swenson, J.E., and Arnemo, J.M. (2012). Capture, anesthesia, and disturbance of free-ranging brown bears (Ursus arctos) during hibernation. PLoS ONE *7*, e40520.

Fedorov, V.B., Goropashnaya, A.V., Tøien, O., Stewart, N.C., Chang, C., Wang, H., Yan, J., Showe, L.C., Showe, M.K., and Barnes, B.M. (2011). Modulation of gene expression in heart and liver of hibernating black bears (Ursus americanus). BMC Genomics *12*, 171.

Furet, J.P., Kong, L.C., Tap, J., Poitou, C., Basdevant, A., Bouillot, J.L., Mariat, D., Corthier, G., Doré, J., Henegar, C., et al. (2010). Differential adaptation of human gut microbiota to bariatric surgery-induced weight loss: links with metabolic and low-grade inflammation markers. Diabetes *59*, 3049–3057.

Große, C., Kaczensky, P., and Knauer, F. (2003). Ants: A food source sought by Slovenian brown bears (Ursus arctos)? Can. J. Zool. *81*, 1996–2005.

Hedges, L.M., Brownlie, J.C., O'Neill, S.L., and Johnson, K.N. (2008). Wolbachia and virus protection in insects. Science *322*, 702.

Khan, M.T., Nieuwdorp, M., and Bäckhed, F. (2014). Microbial modulation of insulin sensitivity. Cell Metab. *20*, 753–760.

LeBlanc, P.J., Obbard, M., Battersby, B.J., Felskie, A.K., Brown, L., Wright, P.A., and Ballantyne, J.S. (2001). Correlations of plasma lipid metabolites with hibernation and lactation in wild black bears Ursus americanus. J. Comp. Physiol. B *171*, 327–334.

Ley, R.E., Bäckhed, F., Turnbaugh, P., Lozupone, C.A., Knight, R.D., and Gordon, J.I. (2005). Obesity alters gut microbial ecology. Proc. Natl. Acad. Sci. USA *102*, 11070–11075.

Ley, R.E., Turnbaugh, P.J., Klein, S., and Gordon, J.I. (2006). Microbial ecology: human gut microbes associated with obesity. Nature *444*, 1022–1023.

Ley, R.E., Hamady, M., Lozupone, C., Turnbaugh, P.J., Ramey, R.R., Bircher, J.S., Schlegel, M.L., Tucker, T.A., Schrenzel, M.D., Knight, R., and Gordon, J.I. (2008). Evolution of mammals and their gut microbes. Science *320*, 1647–1651.

Nelson, R.A. (1973). Winter sleep in the black bear. A physiologic and metabolic marvel. Mayo Clin. Proc. *48*, 733–737.

Oelberg, D.G., Sackman, J.W., Dubinsky, W.P., Adcock, E.W., and Lester, R. (1984). Mechanism of bile acid-induced hemolysis. Pediatr. Res. *18*, 207A–207A.

Otis, J.P., Sahoo, D., Drover, V.A., Yen, C.L., and Carey, H.V. (2011). Cholesterol and lipoprotein dynamics in a hibernating mammal. PLoS ONE *6*, e29111.

Persson, I., Wikan, S., Swenson, J.E., and Mysterud, I. (2001). The diet of the brown bear ursus arctos in the pasvik valley, Northeastern norway. Wildl. Biol. *7*, 27–37.

Reddy, G., Altaf, M., Naveena, B.J., Venkateshwar, M., and Kumar, E.V. (2008). Amylolytic bacterial lactic acid fermentation - a review. Biotechnol. Adv. *26*, 22–34.

Santacruz, A., Marcos, A., Wärnberg, J., Martí, A., Martin-Matillas, M., Campoy, C., Moreno, L.A., Veiga, O., Redondo-Figuero, C., Garagorri, J.M., et al.; EVASYON Study Group (2009). Interplay between weight loss and gut microbiota composition in overweight adolescents. Obesity *17*, 1906–1915.

Sayin, S.I., Wahlström, A., Felin, J., Jäntti, S., Marschall, H.U., Bamberg, K., Angelin, B., Hyötyläinen, T., Orešič, M., and Bäckhed, F. (2013). Gut microbiota regulates bile acid metabolism by reducing the levels of tauro-beta-muricholic acid, a naturally occurring FXR antagonist. Cell Metab. *17*, 225–235.

Schölmerich, J., Becher, M.S., Schmidt, K., Schubert, R., Kremer, B., Feldhaus, S., and Gerok, W. (1984). Influence of hydroxylation and conjugation of bile salts on their membrane-damaging properties–studies on isolated hepatocytes and lipid membrane vesicles. Hepatology *4*, 661–666.

Shulman, G.I. (2014). Ectopic fat in insulin resistance, dyslipidemia, and cardiometabolic disease. N. Engl. J. Med. *371*, 1131–1141.

Sommer, F., and Bäckhed, F. (2013). The gut microbiota–masters of host development and physiology. Nat. Rev. Microbiol. *11*, 227–238.

Sommer, F., Adam, N., Johansson, M.E.V., Xia, L., Hansson, G.C., and Bäckhed, F. (2014). Altered mucus glycosylation in core 1 O-glycan-deficient mice affects microbiota composition and intestinal architecture. PLoS ONE *9*, e85254.

Sommer, F., Nookaew, I., Sommer, N., Fogelstrand, P., and Bäckhed, F. (2015). Site-specific programming of the host epithelial transcriptome by the gut microbiota. Genome Biol. *16*, 62.

Song, H., and Lee, S.Y. (2005). Production of succinic acid by bacterial fermentation. Enzyme Microb. Technol. *39*, 352–361.

Sonnenburg, J.L., Xu, J., Leip, D.D., Chen, C.H., Westover, B.P., Weatherford, J., Buhler, J.D., and Gordon, J.I. (2005). Glycan foraging in vivo by an intestine-adapted bacterial symbiont. Science *307*, 1955–1959.

Sonoyama, K., Fujiwara, R., Takemura, N., Ogasawara, T., Watanabe, J., Ito, H., and Morita, T. (2009). Response of gut microbiota to fasting and hibernation in Syrian hamsters. Appl. Environ. Microbiol. *75*, 6451–6456.

Stenvinkel, P., Fröbert, O., Anderstam, B., Palm, F., Eriksson, M., Bragfors-Helin, A.C., Qureshi, A.R., Larsson, T., Friebe, A., Zedrosser, A., et al. (2013). Metabolic changes in summer active and anuric hibernating free-ranging brown bears (Ursus arctos). PLoS ONE *8*, e72934.

Stevenson, T.J., Duddleston, K.N., and Buck, C.L. (2014). Effects of season and host physiological state on the diversity, density, and activity of the arctic ground squirrel cecal microbiota. Appl. Environ. Microbiol. *80*, 5611–5622.

Stofik, J., Merganic, J., Merganicova, K., and Saniga, M. (2013). Seasonal changes in food composition of the brown bear (Ursus arctos) from the edge of its occurrence – Eastern Carpathians (Slovakia). Folia Zool. Brno *62*, 222–231.

Teixeira, L., Ferreira, A., and Ashburner, M. (2008). The bacterial symbiont Wolbachia induces resistance to RNA viral infections in Drosophila melanogaster. PLoS Biol. *6*, e2.

Tøien, Ø., Blake, J., Edgar, D.M., Grahn, D.A., Heller, H.C., and Barnes, B.M. (2011). Hibernation in black bears: independence of metabolic suppression from body temperature. Science *331*, 906–909.

Turnbaugh, P.J., Hamady, M., Yatsunenko, T., Cantarel, B.L., Duncan, A., Ley, R.E., Sogin, M.L., Jones, W.J., Roe, B.A., Affourtit, J.P., et al. (2009). A core gut microbiome in obese and lean twins. Nature *457*, 480–484.

Wu, G.D., Chen, J., Hoffmann, C., Bittinger, K., Chen, Y.Y., Keilbaugh, S.A., Bewtra, M., Knights, D., Walters, W.A., Knight, R., et al. (2011). Linking long-term dietary patterns with gut microbial enterotypes. Science *334*, 105–108.

Zhang, C., Li, S., Yang, L., Huang, P., Li, W., Wang, S., Zhao, G., Zhang, M., Pang, X., Yan, Z., et al. (2013). Structural modulation of gut microbiota in life-long calorie-restricted mice. Nat. Commun. *4*, 2163.

Zoetendal, E.G., and de Vos, W.M. (2014). Effect of diet on the intestinal microbiota and its activity. Curr. Opin. Gastroenterol. *30*, 189–195.

Cell Reports
Article

Zika Virus Infection in Mice Causes Panuveitis with Shedding of Virus in Tears

Jonathan J. Miner,[1,9,*] Abdoulaye Sene,[2,9] Justin M. Richner,[1] Amber M. Smith,[1] Andrea Santeford,[2] Norimitsu Ban,[2]
James Weger-Lucarelli,[3] Francesca Manzella,[4] Claudia Rückert,[3] Jennifer Govero,[1] Kevin K. Noguchi,[4] Gregory D. Ebel,[3]
Michael S. Diamond,[1,5,6,7,10,*] and Rajendra S. Apte[1,2,8,*]
[1]Department of Medicine
[2]Department of Ophthalmology and Visual Sciences
School of Medicine, Washington University, Saint Louis, MO 63110, US
[3]Department of Microbiology, Immunology and Pathology, College of Veterinary Medicine and Biomedical Sciences,
Colorado State University, Fort Collins, CO 80523, USA
[4]Department of Psychiatry
[5]Department of Molecular Microbiology
[6]Department of Pathology and Immunology
[7]The Center for Human Immunology and Immunotherapy Programs
[8]Department of Developmental Biology
School of Medicine, Washington University, Saint Louis, MO 63110, USA
[9]Co-first author
[10]Lead Contact
*Correspondence: miner@wustl.edu (J.J.M.), diamond@borcim.wustl.edu (M.S.D.), apte@vision.wustl.edu (R.S.A.)

http://dx.doi.org/10.1016/j.celrep.2016.08.079

SUMMARY

Zika virus (ZIKV) is an emerging flavivirus that causes congenital abnormalities and Guillain-Barré syndrome. ZIKV infection also results in severe eye disease characterized by optic neuritis, chorioretinal atrophy, and blindness in newborns and conjunctivitis and uveitis in adults. We evaluated ZIKV infection of the eye by using recently developed mouse models of pathogenesis. ZIKV-inoculated mice developed conjunctivitis, panuveitis, and infection of the cornea, iris, optic nerve, and ganglion and bipolar cells in the retina. This phenotype was independent of the entry receptors Axl or Mertk, given that *Axl*$^{-/-}$, *Mertk*$^{-/-}$, and *Axl*$^{-/-}$*Mertk*$^{-/-}$ double knockout mice sustained levels of infection similar to those of control animals. We also detected abundant viral RNA in tears, suggesting that virus might be secreted from lacrimal glands or shed from the cornea. This model provides a foundation for studying ZIKV-induced ocular disease, defining mechanisms of viral persistence, and developing therapeutic approaches for viral infections of the eye.

INTRODUCTION

ZIKV is a mosquito-transmitted flavivirus that is closely related to Dengue (DENV), West Nile (WNV), and yellow fever (YFV) viruses. Beyond the clinical syndromes of microcephaly, intrauterine growth restriction, and fetal demise, several clinical studies have observed eye malformations and pathology in neonates born to mothers infected with ZIKV during pregnancy (de Paula Freitas et al., 2016; McCarthy, 2016; Miranda et al., 2016; Sarno et al., 2016; Ventura et al., 2016a, 2016b). Manifestations of eye disease in newborns with ZIKV include chorioretinal atrophy, optic neuritis, bilateral iris colobomas, intraretinal hemorrhages, lens subluxation, and blindness (de Paula Freitas et al., 2016; Miranda et al., 2016; Ventura et al., 2016a, 2016b).

Viral infection in the eye can cause inflammation of uveal tissues (retina, choroid, iris, and ciliary body), also termed uveitis, which can lead to permanent vision loss if untreated (Niederkorn, 2006). In 2014, an Ebola virus (EBOV)-infected patient in the convalescent phase presented with uveitis. The aqueous humor of this patient's eye contained persistent EBOV RNA well after clearance of the virus in non-immune-privileged sites (Varkey et al., 2015). A subsequent study identified 57 EBOV survivors with uveitis, suggesting that infectious virus or viral RNA in the eye might have triggered this complication (Tiffany et al., 2016). Other animal models have suggested that some DNA viruses (e.g., ectromelia virus) might persist in the eye and recrudesce after immunosuppression (Sakala et al., 2015). ZIKV causes conjunctivitis in 10% to 15% of infected adults, and uveitis occurred in a patient several weeks after initial infection. Fluid sampled from the anterior chamber of this patient's eye contained viral RNA (Furtado et al., 2016), suggesting that ZIKV can replicate within the eye at some stage of its clinical syndrome.

ZIKV does not replicate efficiently in adult wild-type (WT) mice; this phenotype might be explained by the ability of ZIKV to antagonize human but not mouse STAT2, which is activated by signaling through the type I and type III interferon (IFN) receptors (Ashour et al., 2009; Grant et al., 2016). Accordingly, we and others have recently described models of ZIKV pathogenesis after congenital and adult infection in mice deficient in signaling through the type I IFN receptor (Aliota et al., 2016; Lazear et al., 2016; Miner et al., 2016; Rossi et al., 2016). Mice lacking the type I IFN receptor (*Ifnar1*$^{-/-}$ mice) developed neuroinvasive disease after ZIKV infection, which caused death in younger animals, although a fraction of older adult mice survived infection. Infection of *Ifnar1*$^{-/-}$ females with a French Polynesian strain of ZIKV resulted in fetal demise and intrauterine growth restriction in *Ifnar1*$^{+/-}$ fetuses (Miner et al., 2016), but ocular pathology was not assessed.

Some flaviviruses are thought to attach to or enter target cells by interacting with TAM receptors (Tyro3, Axl, or Mertk) via their ligands, Gas6 and protein S, which bind to phosphatidylserine displayed on the surface of the virion. In vitro studies have suggested that TAM receptors can facilitate infection with WNV, DENV, and ZIKV either by promoting binding (Hamel et al., 2015; Meertens et al., 2012) or by activating TAM receptors (Bhattacharyya et al., 2013), which negatively regulate signaling through Ifnar1, resulting in a more permissive environment for replication. Recent focus has been placed on ZIKV infection and Axl, because it is expressed highly on neuroprogenitor cells (Nowakowski et al., 2016) and subtypes of trophoblasts (Tabata et al., 2016), which are key cellular targets of ZIKV.

Here, we evaluated whether ZIKV infects and injures the eyes of mice. Although we did not detect eye pathology in congenitally infected mice, ZIKV infected the iris, retina, and optic nerve of adult mice and caused conjunctivitis, panuveitis, and neuroretinitis without global photoreceptor abnormalities. Acute uveitis was associated with high levels of ZIKV RNA and infectious virus in the eye and detectable viral RNA in the tears and lacrimal glands. ZIKV RNA persisted within the eye through the convalescent stage, while infectious virus was cleared within 28 days. Although Axl and Mertk are expressed in eye tissues (Prasad et al., 2006; Valverde et al., 2004), infection studies in *Axl*$^{-/-}$ and *Mertk*$^{-/-}$ mice revealed no difference in ocular or brain infection, suggesting that these TAM receptors lack an essential entry or signaling role in these tissues. Our experiments establish that ZIKV infects specific target cells in different regions of the eye and provide a model for the development and testing of treatments for acute and persistent viral infections in the eye.

RESULTS

ZIKV Infection of the Eye with Viral RNA Shedding into Tears

ZIKV does not replicate efficiently in WT C57BL/6 mice, in part because ZIKV NS5 antagonizes human but not mouse STAT2 (Grant et al., 2016), which transmits signals downstream of Ifnar1, a component of the type I IFN receptor. To overcome this limitation, we treated mice with an anti-Ifnar1 blocking antibody (Lazear et al., 2016; Miner et al., 2016) and subcutaneously inoculated them with low-passage ZIKV contemporary clinical isolates, including strains from French Polynesia (H/PF/2013) and Brazil (Paraíba 2015). WT adult mice treated with anti-Ifnar1 monoclonal antibody (mAb) and inoculated with these ZIKV isolates do not develop clinically apparent disease, although viremia and infection of multiple organs occurs, including in immune-privileged sites such as the testes (Lazear et al., 2016). In comparison, *Ifnar1*$^{-/-}$ mice develop neuroinvasive infection, causing some of these animals to succumb to infection (Aliota et al., 2016; Lazear et al., 2016; Rossi et al., 2016). In anti-Ifnar1 mAb-treated animals, we detected RNA of ZIKV H/PF/2013 or Paraíba 2015 in the eye at day 2 after infection, which increased at day 6 (Figure 1A). Similar results were obtained after infection of *Ifnar1*$^{-/-}$ mice with Paraíba 2015, with high intraocular levels of ZIKV RNA accumulating by day 7 (Figure 1B).

Given the data on persistence of eye infection in humans after infection with EBOV and ZIKV (Furtado et al., 2016; Varkey et al., 2015), we assessed infection in mice 28 days after inoculation. Notably, ZIKV RNA persisted in several tissues, including the eyes, brain, spleen, and other organs, long after the virus was cleared from serum (Figure 1C and data not shown). During the 2015 EBOV epidemic, there was concern for person-to-person spread through ocular secretions, including tears (Varkey et al., 2015). As such, we next tested whether ZIKV RNA was present in tear fluid after eye lavage of infected animals with 10 μl of PBS. We detected ZIKV RNA in tear fluid ($\sim 3 \times 10^2$ focus-forming units [FFUs] equivalents per mL) and in the lacrimal gland ($\sim 2.4 \times 10^3$ FFU equivalents per mL) on day 7 after infection (Figures 1D and 1E), suggesting that infectious virus, viral RNA, or ZIKV-infected cellular debris were present in ocular secretions.

To determine whether ZIKV RNA in eyes (day 7 and day 28) and tears (day 7) was infectious, we inoculated AG129 mice via an intraperitoneal route with ocular homogenates or tear fluid; these mice were utilized because they lack receptors for both type I and II IFN signaling and are highly vulnerable to ZIKV infection even after inoculation with 1 plaque forming unit (PFU) (Aliota et al., 2016), which we confirmed (data not shown). Inoculation with eye homogenates obtained from *Ifnar1*$^{-/-}$ mice infected with Paraíba 2015 at day 7 resulted in death of AG129 mice, which occurred uniformly by day 10 (Figure 2A). In contrast, mice inoculated with eye homogenates from day 28 or tears from day 7 did not develop signs of ZIKV infection (Figure 2A and data not shown). These data suggest that infectious virus was not produced in the eye after the acute phase of infection. We observed conjunctivitis in *Ifnar1*$^{-/-}$ mice infected with ZIKV Paraíba 2015, although this occurred in some but not all animals (Figure 2B and data not shown). In the AG129 mice inoculated with day 7 eye homogenates, we observed greater ocular pathology, including severe conjunctivitis with extraocular exudate in all animals, compared to milder disease in mice receiving a similar dose of the parental ZIKV Paraíba 2015 (Figures 2B and 2C).

Nonetheless, we observed similar titers in the spleen, brain, and eyes of AG129 mice inoculated with either parental or eye-derived virus (Figure 2D). We considered whether the severe pathology seen with the eye-derived virus might be due to a virus adaptation that enhances ocular tropism or injury. To evaluate this hypothesis, we performed next-generation sequencing of eye-, spleen-, and brain-derived virus from ZIKV-infected *Ifnar1*$^{-/-}$ animals (Figure S1 and Table S1). Although we did not identify any substitutions that were absent in the inoculating virus, we observed a large increase in the frequency of a single NS2A nucleotide mutation (C→T at genome position 3,895, from ~10% to ~80% in all biological replicates and in all tissues tested) that resulted in an alanine to valine change.

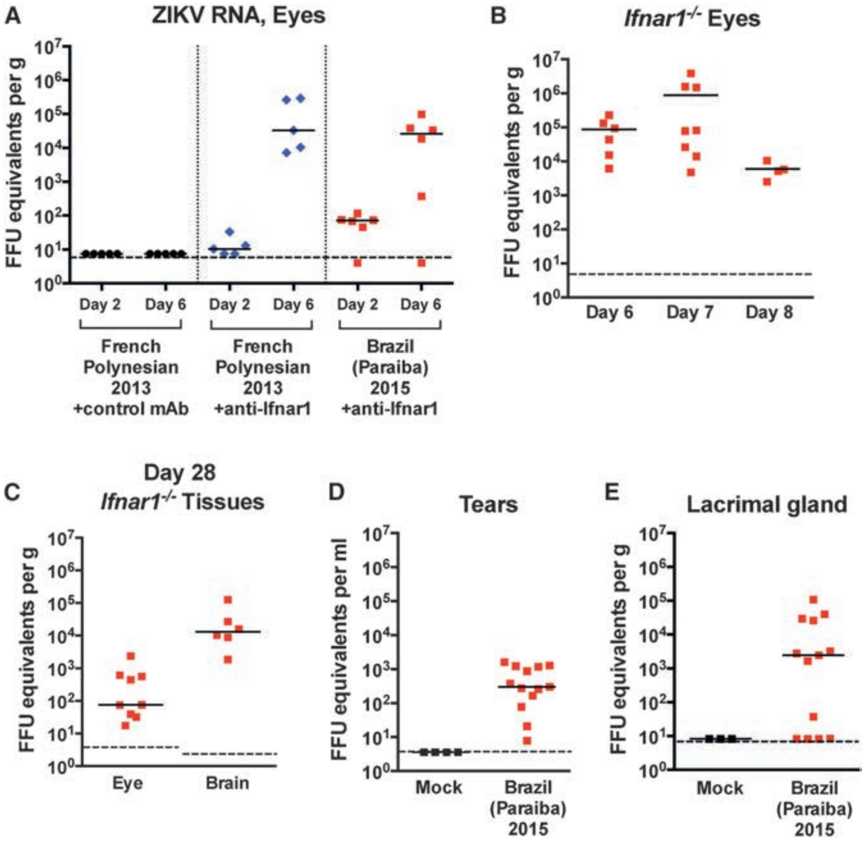


Figure 1. Viral Burden in the Eyes, Tears, and Lacrimal Gland of ZIKV-Infected Animals

4-week-old mice were infected with 10^3 FFUs of ZIKV Paraíba 2015. Viral burden was measured by qRT-PCR assay.

(A) Viral burden in the eyes of WT mice on days 2 and 6 after infection with ZIKV Paraíba 2015 or H/PF/2013. Mice were treated with 2 mg of an anti-Ifnar1 or control mAb 1 day prior to infection.

(B) Viral burden in the eyes of Ifnar1$^{-/-}$ mice on days 6, 7, and 8 after infection with ZIKV Paraíba 2015.

(C) Viral burden in the eyes and brain of Ifnar1$^{-/-}$ mice on day 28 after infection with ZIKV Paraíba 2015.

(D) Viral RNA in tears of mock- and ZIKV-infected Ifnar1$^{-/-}$ mice on day 7 after infection with ZIKV Paraíba 2015.

(E) Viral RNA in the lacrimal gland of mock- and ZIKV-infected Ifnar1$^{-/-}$ mice on day 7 after infection with ZIKV Paraíba 2015.

Symbols are derived from individual animals and pooled from two or three independent experiments. Bars indicate the mean of 5 to 13 mice per group. Dotted lines represent the limit of sensitivity of the assay.

Axl and Mertk Are Not Required for ZIKV Infection of the Eye and Brain In Vivo

The TAM receptors (Tyro3, Axl, and Mertk) are a family of receptor tyrosine kinases whose ligands, Gas6 and protein S, bind to phosphatidylserine on the surface of apoptotic cells and enveloped viruses (Meertens et al., 2012). Because prior studies have suggested that Axl might act as an attachment or entry receptor for ZIKV (Hamel et al., 2015; Nowakowski et al., 2016; Savidis et al., 2016) and TAM receptors are expressed in multiple cell types in the eye (Prasad et al., 2006; Valverde et al., 2004), we hypothesized Axl or its related TAM receptor, Mertk, might be required for efficient ZIKV replication in ocular tissues. Unexpectedly, *Axl*$^{-/-}$, *Mertk*$^{-/-}$, and *Axl*$^{-/-}$*Mertk*$^{-/-}$ double knockout (DKO) mice that were treated with anti-Ifnar1 mAb exhibited similar levels of ZIKV RNA in the serum, spleen, brain, and eyes on day 6 after infection as compared to similarly treated WT control animals (Figures 3A–3D). Thus, in mice deficient in IFN signaling, Axl and Mertk are not required for ZIKV infection in several tissues, including the eyes.

ZIKV Infects the Eyes and Brain of Young WT Mice

Congenital ZIKV infection in humans causes ocular pathology including optic neuritis, retinal mottling, and chorioretinal atrophy (de Paula Freitas et al., 2016; Miranda et al., 2016; Ventura et al., 2016a, 2016b). This could be a consequence of direct eye infection or it might be due to secondary brain defects that disrupt eye development. To test whether ocular pathology occurs in mice infected with ZIKV in utero, we modified a previously described congenital infection model with *Ifnar1*$^{-/-}$ dams and *Ifnar1*$^{+/+}$ sires such that *Ifnar1*$^{+/-}$ fetuses develop intrauterine growth restriction and brain injury but not isolated microcephaly (Miner et al., 2016). After infection with ZIKV Paraíba 2015 at embryonic day (E) 6.5 or E12.5 (equivalent to the late first and second trimesters, respectively), we confirmed the presence of ZIKV RNA in the heads of infected *Ifnar1*$^{+/-}$ fetuses 6 to 7 days later (E13.5 and E18.5) by qRT-PCR (Figure 4A). As reported previously (Miner et al., 2016), substantive demise of *Ifnar1*$^{+/-}$ fetuses was observed by E13.5 after ZIKV infection of *Ifnar1*$^{-/-}$ dams at E6.5, which precluded analysis of fetal ocular tissues. However, fetuses treated with an anti-Ifnar1 mAb survived ZIKV infection on E6.5 but did not show ocular abnormalities by histological analysis (Figure 4B). When *Ifnar1*$^{-/-}$ dams were inoculated later in gestation (E12.5), intrauterine growth restriction occurred without fetal demise (Figure 4C). Again, no histologically apparent pathology or developmental abnormality was observed at E18.5 in the eyes of *Ifnar1*$^{+/-}$ fetuses (Figure 4D).

Because we observed delayed clearance of ZIKV from the eyes of adult mice (see Figure 1C), we tested for viral persistence in the eyes of congenitally infected pups on postnatal day 8. After testing five different experimental conditions, including infection of anti-Ifnar1 mAb-treated pregnant WT mice at different gestational dates with ZIKV H/PF/2013 or Paraíba 2015, we detected viral RNA in eyes of only 2 of 41 congenitally infected animals (Figure 4E). These results suggest that ZIKV might not infect the eyes of fetuses efficiently in this model of in utero infection, even though it crosses the placenta (Miner et al., 2016). In contrast, postnatal infection of WT mice with ZIKV on day 8 after

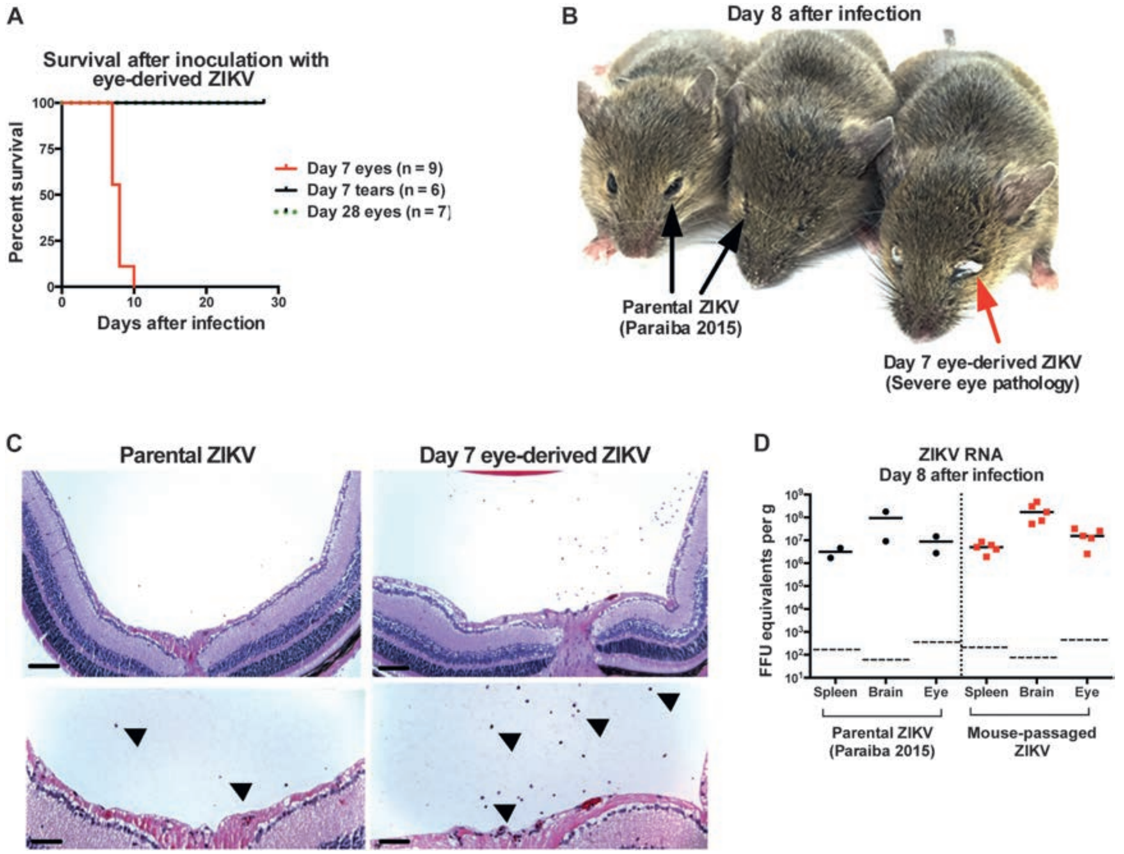


Figure 2. Survival and Viral Burden in AG129 Mice Inoculated with Tears and Eye Homogenate Derived from Infected Ifnar1$^{-/-}$ Mice

4-week-old AG129 mice were inoculated intraperitoneally with 10^4 FFUs of parental ZIKV Paraíba 2015 or eye-derived virus obtained from Ifnar1$^{-/-}$ mice.

(A) Kaplan-Meier survival curve in mice inoculated with day 7 or day 28 eye homogenates from Ifnar1$^{-/-}$ mice or 10 μL of tears obtained 7 days after ZIKV infection of Ifnar1$^{-/-}$ mice.

(B) A representative photograph demonstrating gross ocular pathology and exudate in mice inoculated with 10^4 FFUs of parental ZIKV Paraíba 2015 or eye-derived virus obtained from Ifnar1$^{-/-}$ mice.

(C) Representative H&E-stained eye sections from AG129 mice infected with parental and eye-derived ZIKV. Black arrowheads indicate inflammatory cell infiltrates in the posterior region of the eye. Scale bars represent 100 μm for upper panels and 75 μm for lower panels.

(D) Viral burden assessed by qRT-PCR assay in the spleen, brain, and eyes of AG129 mice inoculated with 10^4 FFUs of parental ZIKV Paraíba 2015 or eye-derived virus obtained from Ifnar1$^{-/-}$ mice.

Symbols are derived from individual animals and pooled from two or three independent experiments. Bars indicate the mean of two to five mice per group. Dotted lines represent the limit of sensitivity of the assay. See also Figure S1.

birth caused a subset of animals to become moribund, and ZIKV RNA was detected readily in the spleen, brain, and eyes 8 days later (Figure 4F). These data establish that ZIKV infection of the eye occurs in young mice even with intact type I IFN signaling. Examination of the brains of these postnatally infected animals revealed apoptosis (as detected by activated caspase-3 staining) in several CNS regions, including the optic tract, lateral geniculate nucleus, and the visual cortex, all components of the visual processing pathway (Figures 4G and 4H).

Anatomically, the eye is divided into the anterior (cornea, iris, ciliary body, and lens) and posterior (vitreous, retina, retinal pigment epithelium, choroid, and optic nerve) compartments, each with specialized cells and functions (Figure 5A). We evaluated the extent of eye injury in *Ifnar1*$^{-/-}$ or anti-Ifnar1 mAb-treated adult animals infected with ZIKV. Histopathological analysis revealed TUNEL-positive cells in the neurosensory retina (Figure 5B, white punctate staining). H&E staining showed anterior uveitis with infiltration of inflammatory cells in the anterior chamber (Figure 5C, upper right panel). The posterior eye also exhibited evidence of uveitis with inflammatory cell infiltration, exudate, and debris in the vitreous humor (Figure 5C, lower right panel). We next assessed the fundus of the eye for gross structural damage by using biomicroscopic and fundoscopic examination; no differences were observed between mock- and ZIKV-infected *Ifnar1*$^{-/-}$ or anti-Ifnar1 mAb-treated mice on day 7 (Figure S2 and data not shown). These experiments suggest that ZIKV infection does not induce significant pan-retinal abnormalities. Fluorescein angiography in ZIKV-infected *Ifnar1*$^{-/-}$ mice also revealed no evidence of vascular leakage or intraretinal hemorrhages (Figure S2). Thus, ZIKV infection in adult mice causes panuveitis without major structural damage or effects on vascular permeability.

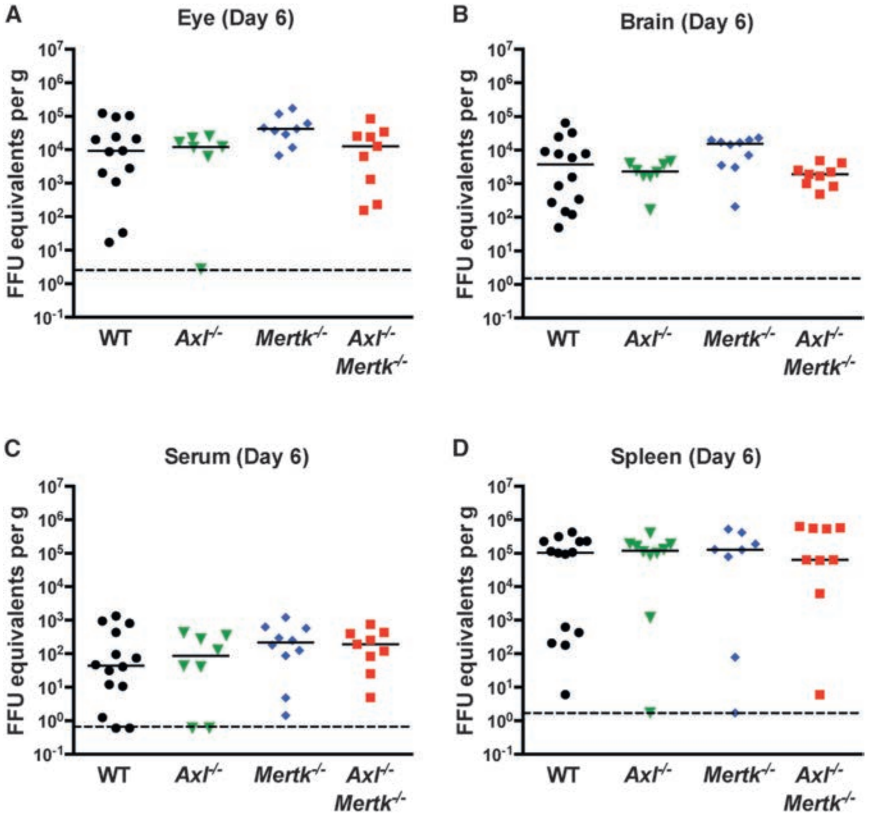


Figure 3. Viral Burden in WT, Axl$^{-/-}$, Mertk$^{-/-}$, or Axl$^{-/-}$Mertk$^{-/-}$ DKO Mice

4-to-6-week-old WT, Axl$^{-/-}$, Mertk$^{-/-}$, or Axl$^{-/-}$Mertk$^{-/-}$ DKO mice were treated with 2 mg of anti-Ifnar1 mAb 1 day prior to subcutaneous infection with 10^3 FFUs of ZIKV Paraíba 2015. On day 6 after infection, viral burden was measured by qRT-PCR in the eye (A), brain (B), serum (C), and spleen (D).

Symbols are derived from individual animals and pooled from two independent experiments. Bars indicate the mean of 8 to 14 mice per group. Dotted lines represent the limit of sensitivity of the assay. Data were analyzed by Kruskal-Wallis one-way ANOVA with a Dunn's multiple comparison test ($p > 0.1$ for comparison of all genotypes to the WT in all tissues).

To determine whether ZIKV infection of the eye caused functional deficits, we performed ex vivo electroretinography (ERG), which measures the transmission of light by photoreceptor cells. ERG testing of dissected eyes from *Ifnar1*$^{-/-}$ mice on day 7 after ZIKV infection revealed no apparent defects in photoreceptor function (Figures S3A– S3D). Similar results were obtained on days 7, 14, and 28 in ZIKV-infected WT mice treated with anti-Ifnar1 mAb (Figures S3E–S3H and data not shown). Thus, ZIKV infection and persistence in the eye does not cause global photoreceptor dysfunction. This ERG evaluation, however, does not rule out mild or focal defects in photoreceptor function or whether neuronal dysfunction of the inner retina and damage to the optic nerve, optic tract, or visual cortex occurs, any of which could result in blindness or selective visual field defects.

We evaluated the cellular tropism of ZIKV in the eye. Initially, we performed microdissection and measured ZIKV RNA levels in different compartments of the eye on day 7 after infection (Figure 6A). We included analysis of the anterior (cornea, iris, and lens) and posterior (neurosensory retina, retinal pigment epithelium/choroid complex, and optic nerve) chambers (see Figure 3B). ZIKV RNA was detected in all eye regions, with ~100-fold higher levels in the retinal pigment epithelium/choroid complex than in the optic nerve ($p < 0.05$, Figure 6A). To confirm these findings, eyes were collected at 6 to 8 days after inoculation and assessed for viral RNA by in situ hybridization (ISH). Mock-infected animals stained for ZIKV RNA and infected animals stained with a negative control probe against a bacterial gene had no staining in the cornea, optic nerve, or retina (Figure 6B and data not shown). In comparison, abundant ZIKV RNA was apparent in the bipolar and ganglion cell neurons of the neurosensory retina (Figure 6C, upper panel), the optic nerve (Figure 6C, middle panel), and the cornea of infected animals (Figure 6C, lower panel).

DISCUSSION

Understanding the consequences of ocular infection by viruses has become of greater concern because of the potential implications related to viral persistence and spread. Case reports of human patients infected with herpes simplex virus-1 (HSV-1), EBOV, ZIKV, and WNV suggest that some viruses have the capacity to cause significant acute ocular pathology, including uveitis and keratitis (Chong et al., 2004; Furtado et al., 2016; Kuchtey et al., 2003; Varkey et al., 2015). Since ZIKV does not efficiently infect WT mice, we tested whether it targets the eyes of mice by using a published model of ZIKV pathogenesis in adult animals deficient in type I IFN immunity as well as in immunocompetent neonatal mice and fetuses. Although we did not observe evidence of congenital ocular disease, which has been described in humans (de Paula Freitas et al., 2016; McCarthy, 2016; Ventura et al., 2016a, 2016b), ZIKV infected the eye and caused conjunctivitis and severe uveitis in immunodeficient adult mice. RNA ISH and virological analysis revealed that ZIKV infects the cornea, vascularized choroid, bipolar and ganglion layers of the retina, and optic nerve. Although the mechanism by which ZIKV causes eye disease remains undefined, our studies provide an animal model that can be used to define host factors that restrict pathogenesis and test candidate antiviral therapies directly within the eye.

The most common form of ZIKV-induced ocular disease is conjunctivitis, which occurs in 10% to 15% of patients, but whether conjunctivitis is a direct consequence of ZIKV infection of the eye is not known. In contrast, ZIKV-induced uveitis is less common, although it has been described in humans (Furtado et al., 2016). How ZIKV triggers uveitis remains uncertain,

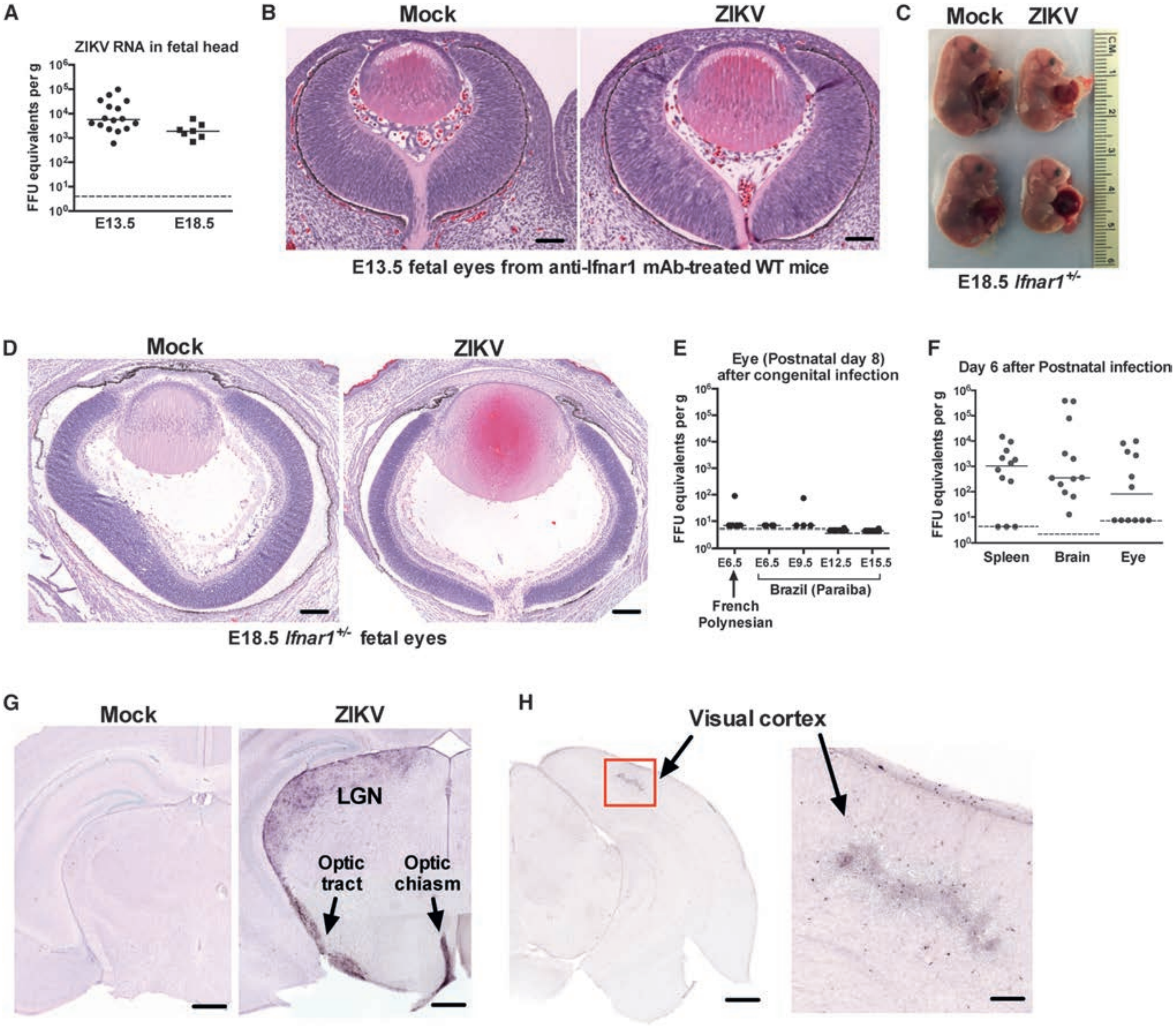


Figure 4. Histology and Viral Burden in the Eye after Congenital and Early Postnatal Infection with ZIKV

(A–E) Pregnant WT mice treated with 2 mg of an anti-Ifnar1 mAb or Ifnar1$^{-/-}$ mice were infected subcutaneously with 10^3 FFUs of ZIKV Paraíba 2015 WT, except where the French Polynesian strain is indicated in (E).

(A) Viral burden was assessed by qRT-PCR on E13.5 in WT fetuses or E18.5 in Ifnar1$^{+/-}$ fetuses after E6.5 and E12.5 infection, respectively.

(B) Representative H&E-stained eye sections from mock- and ZIKV-infected WT fetuses on E13.5.

(C) A representative photograph demonstrating the average size of mock- and ZIKV-infected Ifnar1$^{+/-}$ fetuses on E18.5.

(D) Representative H&E-stained eye sections from mock- and ZIKV-infected Ifnar1$^{+/-}$ fetuses on E18.5. The retinal detachment in the mock-infected sample is a commonly seen artifact of sectioning.

(E) Viral burden in the eyes of WT pups on postnatal day 8 after in utero ZIKV infection of anti-Ifnar1 mAb-treated WT pregnant dams at the indicated gestational date with the indicated strain.

(F–H) WT pups were inoculated subcutaneously on postnatal day 8 with 10^3 FFUs of ZIKV Paraíba 2015 WT and harvested 6 days later.

(F) Viral burden was assessed by qRT-PCR in the spleen, brain, and eye.

(G) Activated caspase-3 staining of mock- (left panel) and ZIKV-infected (right panel) WT pup brains, including the visual pathway (optic chiasm, optic tract, and lateral geniculate nucleus [LGN]). The image was generated by combining images of several coronal sections in the same animal into a single merged figure.

(H) Activated caspase-3 staining of a ZIKV-infected WT pup brain highlighting apoptosis in the visual cortex (red box) at higher magnification (right panel).

Results are from at least two or three independent experiments with 7 to 16 animals for viral burden analysis and 2 to 4 mice for histological analysis. Scale bars represent 25 μm for (B), 60 μm for (D), 500 μm for (G), and 1 mm for (H). Dotted lines represent the limit of sensitivity of the assay.

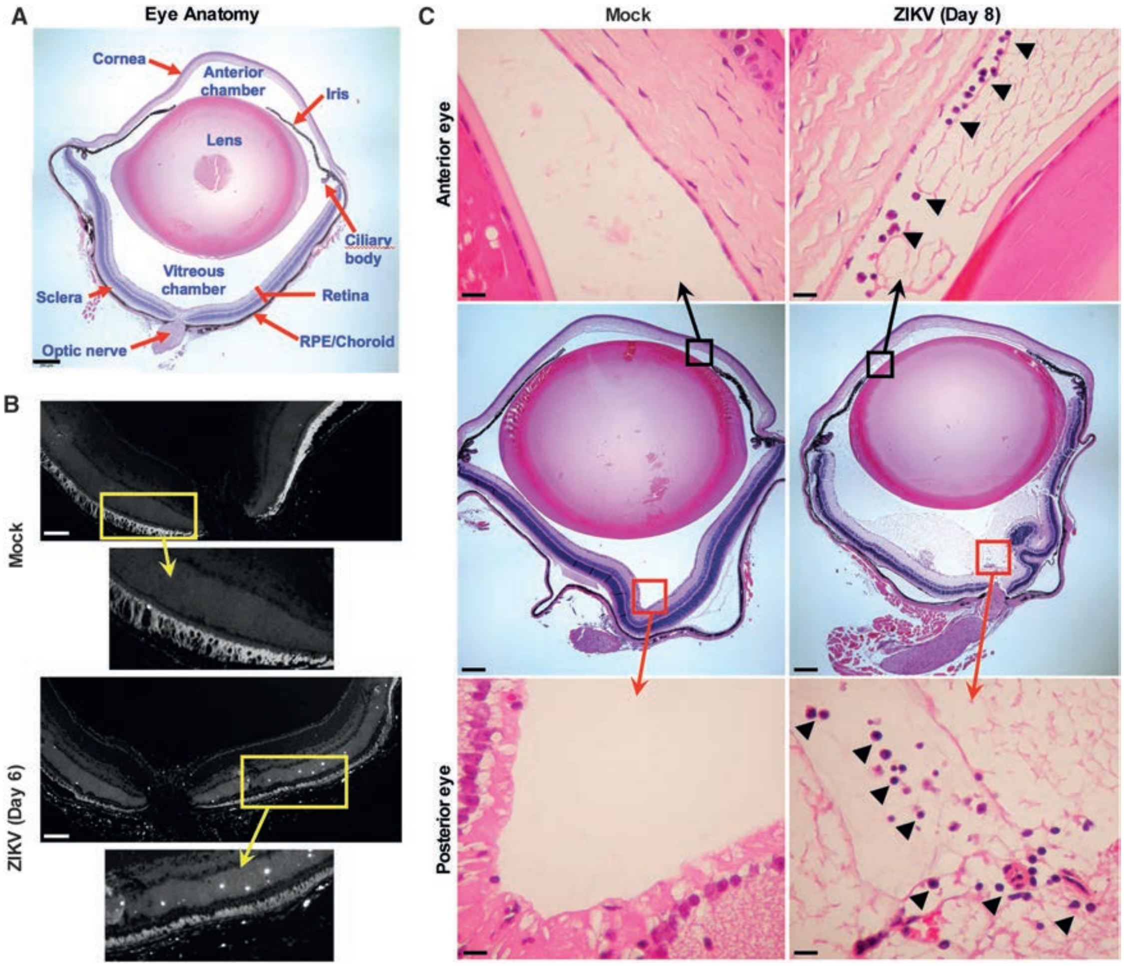


Figure 5. Histology of the Eye in Adult Ifnar1$^{-/-}$ Mice after Infection with ZIKV

4-to-8-week-old Ifnar1$^{-/-}$ mice were inoculated subcutaneously with 10^3 FFUs of ZIKV Paraíba 2015. On days 6 or 8 after infection, eyes were harvested.

(A) Anatomy of an uninfected eye, demonstrating anterior (cornea, lens, iris, and ciliary body) and posterior (sclera, retina, retinal pigment epithelium/choroid, and optic nerve) structures.

(B) TUNEL staining of the neurosensory retina of a mock (upper panel) or ZIKV-infected animal on day 6 (lower panel). Regions shown in higher magnification are indicated by a yellow box and arrow. Scale bars represent 100 μm.

(C) H&E-stained eye sections from mock- (left panels) and ZIKV-infected animals on day 8 (right panels). Regions shown in higher magnification are indicated by a box and displayed in the upper and lower panels. Black arrowheads indicate inflammatory cell infiltrates in the anterior (upper panels) and posterior (lower panels) chambers of the eye. Scale bars represent 250 μm (middle panels) and 25 μm (upper and lower panels).

The data and images are representative of two independent experiments with two to four animals per group. See also Figures S2 and S3.

although inflammatory cell recruitment might contribute to this, according to our histological findings. Direct infection of the eye might cause inflammation as a consequence of virus-induced cell death, cytokine production, and/or leukocyte recruitment. Alternatively, intracellular or extracellular viral RNA can act as a pathogen-associated molecular pattern (PAMP) and activate inflammatory responses via stimulating pattern recognition receptors (e.g., Toll-like or Rig-I-like receptors), even in the absence of infectious virus. In humans with EBOV- and ZIKV-induced uveitis, viral RNA was detectable within the intraocular fluid. We found that mice infected with ZIKV had no infectious virus in the eye by day 28, even though high levels of ZIKV RNA were present. Consistent with this observation, there is extensive literature describing WNV-induced ocular disease in humans, including uveitis, chorioretinitis, and retinal artery occlusion (Hershberger et al., 2003; Kaiser et al., 2003; Kuchtey et al., 2003). Neurotropic flaviviruses might first invade the brain, then infect the optic tract, and later transit in a retrograde direction into the eye along the optic nerve. Alternatively, ocular infection might result from hematogenous spread of virus across the blood-retinal barrier. Additional studies are needed to define the precise mechanisms of ocular invasion by ZIKV.

ZIKV infects both the eyes and testes in humans and in animal models (Lazear et al., 2016; Musso et al., 2015), suggesting that immune-privileged organs might support replication even weeks after resolution of viremia and clinical symptoms. Hepatitis C virus, a related *Flaviviridae* family member, can infect the human cornea (Lee et al., 2001) and is transmitted by corneal

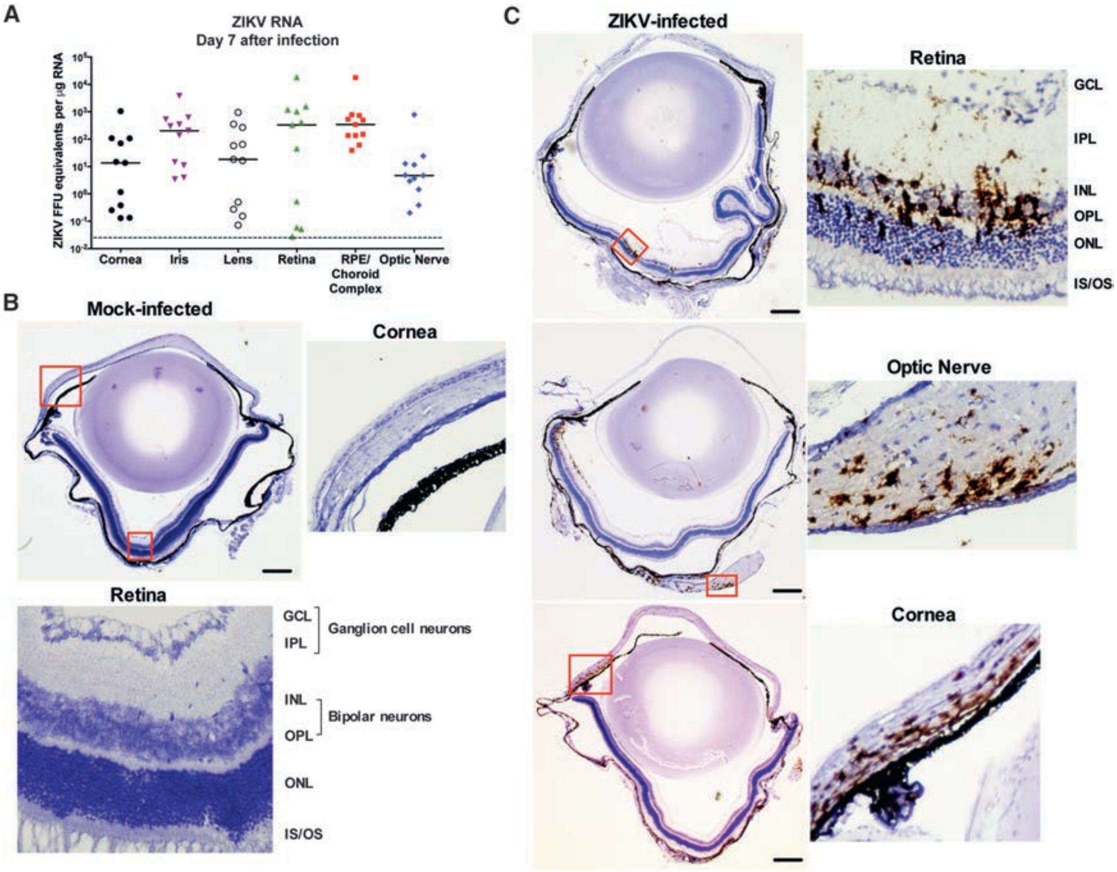


Figure 6. Viral Burden and Localization in the Anterior and Posterior Regions of the Eye

4-to-8-week-old Ifnar1$^{-/-}$ mice were inoculated subcutaneously with 10^3 FFUs of ZIKV Paraíba 2015. Eyes were harvested for microdissection and quantitation of ZIKV RNA by qRT-PCR (day 7) or for ZIKV RNA detection by ISH (day 6 or 8).

(A) Viral burden in the microdissected cornea, iris, lens, retina, retinal pigment epithelium/choroid, and optic nerve on day 7 after infection. Symbols are derived from individual animals and pooled from two independent experiments. Bars indicate the mean of 8 to 14 mice per group. Dotted lines represent the limit of sensitivity of the assay. Data were analyzed by Kruskal-Wallis one-way ANOVA with a Dunn's multiple comparison test ($p > 0.05$ only for samples from optic nerve compared to RPE/choroid complex).

(B and C) RNA ISH with a ZIKV-specific probe to stain eye sections of mock- (B) or ZIKV-infected (C) Ifnar1$^{-/-}$ mice. Red boxes indicate regions shown at higher magnification in adjacent panels. High-magnification views in (B) indicate regions of the cornea (right panel) and retina (bottom panel). High-magnification views in (C) show regions of the retina (right panel), optic nerve (middle panel), and cornea (lower panel). ISH data are representative of two independent experiments with at least two animals per group. Scale bar represents 200 μm.

transplantation (Tugwell et al., 2005). Given that ZIKV has capacity to infect the cornea of mice, human studies might be needed to confirm whether ZIKV analogously infects the human cornea. If corneal infection by ZIKV were established in humans, then widespread ocular infection during epidemics could necessitate testing by eye banks to ensure that ZIKV is not present in the corneas of infected donors.

Congenital ocular disease caused by ZIKV might be due to direct targeting of cells in the fetal eye. Alternatively, it could occur as part of a neurodevelopmental defect, which is seen in microcephaly cases even in the absence of an infectious cause (Derbent et al., 2004). Although a recent study suggested ocular pathology was present in SJL mouse pups that were infected with ZIKV in utero, histological characterization of their eyes was not demonstrated (Cugola et al., 2016). We did not observe histological abnormalities of the eyes of congenitally infected *Ifnar1*$^{+/-}$ fetuses from C57BL/6 *Ifnar1*$^{-/-}$ dams. Studies with other mouse and non-human primate models are needed to clarify the requirements and basis for ZIKV-induced congenital ocular disease.

TAM receptors might enhance attachment and entry of flaviviruses, including ZIKV (Hamel et al., 2015; Meertens et al., 2012; Savidis et al., 2016). This hypothesis has been strengthened by a correlation between levels of Axl expression and vulnerability of certain neuronal subtypes to ZIKV infection (Nowakowski et al., 2016). However, our studies in *Axl*$^{-/-}$, *Mertk*$^{-/-}$, and *Axl*$^{-/-}$*Mertk*$^{-/-}$ mice revealed no effect of a loss of expression of these TAM receptors on ZIKV replication, suggesting that Axl and Mertk are not required for CNS or ocular infection in mice. These results are analogous to prior studies with WNV, in which an absence of Axl and/or Mertk paradoxically resulted in enhanced infection in the brain, which was due in part to alterations in the permeability of endothelial cells lining the blood-brain barrier (Miner et al., 2015). Our data do not exclude the possibility that Axl might still act as an entry factor for ZIKV in specific cells in other tissue compartments (e.g., trophoblasts in the placenta or subsets of neurons in the CNS).

In summary, we have described a mouse model of ocular disease that demonstrates ZIKV tropism in specific regions of the eye, panuveitis, shedding of viral RNA into tears, and persistence in immunodeficient adult mice. We have also confirmed that ZIKV infects the eye of immunocompetent neonatal WT mice. Studies are planned to define the cellular mechanisms by which ZIKV invades and infects the eye and results in inflammation. Further analysis of the host and virus factors that facilitate ocular infection and mechanisms of immune-mediated clearance could lead to interventions that enhance elimination of viruses from immune-privileged sites, including the eye.

EXPERIMENTAL PROCEDURES

Ethics Statement

This study was carried out in accordance with the recommendations in the Guide for the Care and Use of Laboratory Animals of the NIH. The protocols were approved by the institutional animal care and use committee at the Washington University School of Medicine (assurance number A3381-01). Inoculations were performed under anesthesia induced and maintained with ketamine hydrochloride and xylazine, and all efforts were made to minimize animal suffering.

Viruses and Titration

The ZIKV strain H/PF/2013 (French Polynesia, 2013) was provided by the Arbovirus Branch of the Centers for Disease Control and Prevention with permission (X. de Lamballerie, Aix Marseille Université) (Baronti et al., 2014). The ZIKV strain Paraíba 2015 (Brazil, 2015) was provided by Steve Whitehead (NIH). ZIKV stocks were propagated in Vero cells and titrated by focus-forming assay (FFA) as previously described (Lazear et al., 2016; Miner et al., 2016).

Mouse Experiments

Ifnar1$^{-/-}$ mice (Hwang et al., 1995) were backcrossed onto a C57BL/6 background. *Axl*$^{-/-}$, *Mertk*$^{-/-}$, and *Axl*$^{-/-}$*Mertk*$^{-/-}$ mice have been published (Lu and Lemke, 2001) and were backcrossed for ten generations. Mice were bred in a specific-pathogen-free facility at Washington University or purchased (WT animals) from Jackson Laboratories. In some experiments, WT mice were treated with 2 mg of an anti-Ifnar1 blocking mouse mAb (MAR1-5A3) or isotype control mouse mAb (GIR-208; Leinco Technologies) (Sheehan et al., 2006, 2015) by intraperitoneal injection prior to infection with ZIKV. 4- to 8-week-old mice were inoculated with ZIKV by subcutaneous (footpad) route with 10^3 FFU in 50 μL of PBS. Mice were sacrificed at day 2, 6, 7, 8, or 28, depending on the experimental design, and organs were harvested. To study congenital infection, anti-Ifnar1 mAb-treated WT pregnant mice were infected with the indicated strains of ZIKV at E6.5, E9.5, E12.5, or E15.5, and tissues were harvested at postnatal day 8. For some experiments, pregnant *Ifnar1*$^{-/-}$ females carrying *Ifnar1*$^{+/-}$ fetuses were infected on E12.5. *Ifnar1*$^{+/-}$ fetal tissues were harvested on E18.5 as previously described (Miner et al., 2016). For studies in neonatal WT animals, mice were inoculated subcutaneously on postnatal day 7 or 8 and then tissues were harvested 6 days later and analyzed for the presence of viral RNA. For comparison of parental (Paraíba 2015) and mouse-passaged ZIKV, AG129 mice (van den Broek et al., 1995) were inoculated via an intraperitoneal route with 10^4 FFU of parental ZIKV or an equivalent dose (based on viral RNA) from eye homogenates collected from *Ifnar1*$^{-/-}$ mice on day 7 or day 28 after ZIKV infection. Alternatively, AG129 mice were inoculated via an intraperitoneal route with 10 μL of tears (diluted in 150 μL of PBS) collected from *Ifnar1*$^{-/-}$ mice on day 7 after infection with ZIKV (Paraíba 2015).

Measurement of Viral Burden

Tissues were weighed and homogenized with zirconia beads in a MagNA Lyser instrument (Roche Life Science) in 250 μL or 500 μL of PBS. Tissue samples and serum from ZIKV-infected mice were extracted with the RNeasy Mini Kit (tissues) or the Viral RNA Mini Kit (serum) (QIAGEN). ZIKV RNA levels were determined by one-step qRT-PCR on an ABI 7500 Fast Instrument using standard cycling conditions. Viral burden was expressed on a $\log_{10}$ scale as viral RNA equivalents per gram or per mL after comparison with a standard curve produced using serial 10-fold dilutions of ZIKV RNA. A published primer set was used to detect ZIKV RNA (Lanciotti et al., 2008; Miner et al., 2016).

Tears and Lacrimal Gland Collection

ZIKV-infected mice were euthanized on day 7 after infection. Tear fluid from both eyes was collected after gentle lavage with 10 μL of PBS using FP plus multiflex tips (Fisher Scientific). Intact lacrimal gland lobules were isolated by dissection.

TUNEL Staining

Cell death in the neurosensory retina of *Ifnar1*$^{-/-}$ mice was assessed 7 days after ZIKV infection by TUNEL staining. In brief, paraffin sections of eyes from ZIKV-infected mice were pretreated with proteinase K and stained with an ApopTag Fluorescein In Situ Apoptosis Detection Kit (Millipore) according to the manufacturer's instructions. Images of the retina were collected and merged into a composite image.

Brain Sectioning and Immunostaining for Activated Caspase-3

After ZIKV infection, animals were anesthetized and perfused via intracardiac injection with 4% paraformaldehyde in Tris-HCl buffer. Brains were then sectioned at 70 μm with a vibratome and stained for activated caspase-3 as described previously (Miner et al., 2016).

RNA ISH

RNA ISH was performed with an RNAscope 2.5 (Advanced Cell Diagnostics) according to the manufacturer's instructions. In brief, formalin-fixed paraffin-embedded tissue sections were deparaffinized by incubation for 60 min at 60°C. Endogenous peroxidases were quenched with hydrogen peroxide for 10 min at room temperature. Slides were then boiled for 15 min in RNAscope Target Retrieval Reagents and incubated for 30 min in RNAscope Protease Plus before probe hybridization. The probe targeting ZIKV RNA was designed and synthesized by Advanced Cell Diagnostics (catalog no. 467771). Positive (targeting *plr2a* gene) and negative (targeting bacterial gene *dapB*) control probes also were obtained from Advanced Cell Diagnostics (312471 and 310043, respectively). Tissues were counterstained with Gill's hematoxylin and visualized with standard bright-field microscopy.

Data Analysis

All data were analyzed with GraphPad Prism software. For viral burden analysis, the $\log_{10}$ transformed titers and levels of viral RNA were analyzed by the Mann-Whitney test or Kruskal-Wallis one-way ANOVA. A p value of < 0.05 indicated statistically significant differences.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, three figures, and one table and can be found with this article online at http://dx.doi.org/10.1016/j.celrep.2016.08.079.

AUTHOR CONTRIBUTIONS

J.J.M. and A.M.S. performed infections, tissue harvests, RNA isolation, qRT-PCR analysis, and some of the histological analysis. A.Sene performed ERG, fundoscopy, fluorescein angiography, TUNEL staining, and some of the histological analysis. A.Sene and N.B. dissected lacrimal glands and eye regions and collected tears. J.M.R. and A.M.S. performed all RNA ISH experiments. J.G. performed the ZIKV French Polynesian strain infection and harvest. A.Santeford performed tissue sectioning and histological analysis. F.M. and K.N. performed histological analysis of the brain. J.W., C.R., and G.D.E. performed the next-generation sequencing and analysis. A.Sene wrote parts of the initial draft. J.J.M. wrote the initial complete draft of the manuscript. J.J.M., M.S.D., and R.S.A. edited the manuscript. J.J.M, M.S.D, and R.S.A are corresponding authors. As a group, the corresponding authors shared responsibility for all data, figures, and text, determination of authorship, approval of content by authors, adherence to editorial policies, declaration of conflict of interests, and ensuring figures are representative of the original data. M.S.D. is the lead contact for journal communication and is the primary contact for reagent and resource sharing.

ACKNOWLEDGMENTS

This study was supported by grants from the NIH (R01 AI073755, R01 AI101400, and R01 AI104972 [to M.S.D.], R01 EY019287 [to R.S.A.], P30 EY02687 [Vision Core Grant], and R01 AI067380 [to G.D.E.]). R.S.A. also was supported by a Physician-Scientist Award from Research to Prevent Blindness, the Schulak Family Gift Fund for Retinal Research, and the Jeffrey Fort Innovation Fund. J.J.M. was supported by a Rheumatology Research Foundation Scientist Development Award. K.N. was supported by the Intellectual and Developmental Disabilities Research Center at Washington University (NIH/NICHD U54 HD087011). Additional funding comes from an unrestricted grant to the Department of Ophthalmology and Visual Sciences of Washington University School of Medicine from Research to Prevent Blindness. M.S.D. is a consultant for Inbios, Visterra, Sanofi, and Takeda Pharmaceuticals, is on the Scientific Advisory Boards of Moderna and OraGene, and is a recipient of research grants from Moderna, Sanofi, and Visterra.

Received: July 27, 2016
Revised: August 12, 2016
Accepted: August 25, 2016
Published: September 6, 2016

REFERENCES

Aliota, M.T., Caine, E.A., Walker, E.C., Larkin, K.E., Camacho, E., and Osorio, J.E. (2016). Characterization of lethal Zika virus infection in AG129 mice. PLoS Negl. Trop. Dis. *10*, e0004682.

Ashour, J., Laurent-Rolle, M., Shi, P.-Y., and García-Sastre, A. (2009). NS5 of dengue virus mediates STAT2 binding and degradation. J. Virol. *83*, 5408–5418.

Baronti, C., Piorkowski, G., Charrel, R.N., Boubis, L., Leparc-Goffart, I., and de Lamballerie, X. (2014). Complete coding sequence of zika virus from a French polynesia outbreak in 2013. Genome Announc. *2*, e00500–e00514.

Bhattacharyya, S., Zagórska, A., Lew, E.D., Shrestha, B., Rothlin, C.V., Naughton, J., Diamond, M.S., Lemke, G., and Young, J.A. (2013). Enveloped viruses disable innate immune responses in dendritic cells by direct activation of TAM receptors. Cell Host Microbe *14*, 136–147.

Chong, E.-M., Wilhelmus, K.R., Matoba, A.Y., Jones, D.B., Coats, D.K., and Paysse, E.A. (2004). Herpes simplex virus keratitis in children. Am. J. Ophthalmol. *138*, 474–475.

Cugola, F.R., Fernandes, I.R., Russo, F.B., Freitas, B.C., Dias, J.L.M., Guimarães, K.P., Benazzato, C., Almeida, N., Pignatari, G.C., Romero, S., et al. (2016). The Brazilian Zika virus strain causes birth defects in experimental models. Nature *534*, 267–271.

de Paula Freitas, B., de Oliveira Dias, J.R., Prazeres, J., Sacramento, G.A., Ko, A.I., Maia, M., and Belfort, R., Jr. (2016). Ocular findings in infants with microcephaly associated with presumed zika virus congenital infection in salvador, brazil. JAMA Ophthalmol. *134*, 529–535.

Derbent, M., Agras, P.I., Gedik, S., Oto, S., Alehan, F., and Saatçi, U. (2004). Congenital cataract, microphthalmia, hypoplasia of corpus callosum and hypogenitalism: report and review of Micro syndrome. Am. J. Med. Genet. A. *128A*, 232–234.

Furtado, J.M., Espósito, D.L., Klein, T.M., Teixeira-Pinto, T., and da Fonseca, B.A. (2016). Uveitis associated with zika virus infection. N. Engl. J. Med. *375*, 394–396.

Grant, A., Ponia, S.S., Tripathi, S., Balasubramaniam, V., Miorin, L., Sourisseau, M., Schwarz, M.C., Sánchez-Seco, M.P., Evans, M.J., Best, S.M., and García-Sastre, A. (2016). Zika virus targets human STAT2 to inhibit type I interferon signaling. Cell Host Microbe *19*, 882–890.

Hamel, R., Dejarnac, O., Wichit, S., Ekchariyawat, P., Neyret, A., Luplertlop, N., Perera-Lecoin, M., Surasombatpattana, P., Talignani, L., Thomas, F., et al. (2015). Biology of zika virus infection in human skin cells. J. Virol. *89*, 8880–8896.

Hershberger, V.S., Augsburger, J.J., Hutchins, R.K., Miller, S.A., Horwitz, J.A., and Bergmann, M. (2003). Chorioretinal lesions in nonfatal cases of West Nile virus infection. Ophthalmology *110*, 1732–1736.

Hwang, S.Y., Hertzog, P.J., Holland, K.A., Sumarsono, S.H., Tymms, M.J., Hamilton, J.A., Whitty, G., Bertoncello, I., and Kola, I. (1995). A null mutation in the gene encoding a type I interferon receptor component eliminates antiproliferative and antiviral responses to interferons alpha and beta and alters macrophage responses. Proc. Natl. Acad. Sci. USA *92*, 11284–11288.

Kaiser, P.K., Lee, M.S., and Martin, D.A. (2003). Occlusive vasculitis in a patient with concomitant West Nile virus infection. Am. J. Ophthalmol. *136*, 928–930.

Kuchtey, R.W., Kosmorsky, G.S., Martin, D., and Lee, M.S. (2003). Uveitis associated with West Nile virus infection. Arch. Ophthalmol. *121*, 1648–1649.

Lanciotti, R.S., Kosoy, O.L., Laven, J.J., Velez, J.O., Lambert, A.J., Johnson, A.J., Stanfield, S.M., and Duffy, M.R. (2008). Genetic and serologic properties of Zika virus associated with an epidemic, Yap State, Micronesia, 2007. Emerg. Infect. Dis. *14*, 1232–1239.

Lazear, H.M., Govero, J., Smith, A.M., Platt, D.J., Fernandez, E., Miner, J.J., and Diamond, M.S. (2016). A Mouse Model of Zika Virus Pathogenesis. Cell Host Microbe *19*, 720–730.

Lee, H.M., Naor, J., Alhindi, R., Chinfook, T., Krajden, M., Mazzulli, T., and Rootman, D.S. (2001). Detection of hepatitis C virus in the corneas of seropositive donors. Cornea *20*, 37–40.

Lu, Q., and Lemke, G. (2001). Homeostatic regulation of the immune system by receptor tyrosine kinases of the Tyro 3 family. Science *293*, 306–311.

McCarthy, M. (2016). Severe eye damage in infants with microcephaly is presumed to be due to Zika virus. BMJ *352*, i855.

Meertens, L., Carnec, X., Lecoin, M.P., Ramdasi, R., Guivel-Benhassine, F., Lew, E., Lemke, G., Schwartz, O., and Amara, A. (2012). The TIM and TAM families of phosphatidylserine receptors mediate dengue virus entry. Cell Host Microbe *12*, 544–557.

Miner, J.J., Daniels, B.P., Shrestha, B., Proenca-Modena, J.L., Lew, E.D., Lazear, H.M., Gorman, M.J., Lemke, G., Klein, R.S., and Diamond, M.S. (2015). The TAM receptor Mertk protects against neuroinvasive viral infection by maintaining blood-brain barrier integrity. Nat. Med. *21*, 1464–1472.

Miner, J.J., Cao, B., Govero, J., Smith, A.M., Fernandez, E., Cabrera, O.H., Garber, C., Noll, M., Klein, R.S., Noguchi, K.K., et al. (2016). Zika Virus Infection during Pregnancy in Mice Causes Placental Damage and Fetal Demise. Cell *165*, 1081–1091.

Miranda, H.A., 2nd, Costa, M.C., Frazão, M.A.M., Simão, N., Franchischini, S., and Moshfeghi, D.M. (2016). Expanded Spectrum of Congenital Ocular Findings in Microcephaly with Presumed Zika Infection. Ophthalmology *123*, 1788–1794.

Musso, D., Roche, C., Robin, E., Nhan, T., Teissier, A., and Cao-Lormeau, V.M. (2015). Potential sexual transmission of Zika virus. Emerg. Infect. Dis. *21*, 359–361.

Niederkorn, J.Y. (2006). See no evil, hear no evil, do no evil: the lessons of immune privilege. Nat. Immunol. *7*, 354–359.

Nowakowski, T.J., Pollen, A.A., Di Lullo, E., Sandoval-Espinosa, C., Bershteyn, M., and Kriegstein, A.R. (2016). Expression analysis highlights AXL as a candidate Zika virus entry receptor in neural stem cells. Cell Stem Cell *18*, 591–596.

Prasad, D., Rothlin, C.V., Burrola, P., Burstyn-Cohen, T., Lu, Q., Garcia de Frutos, P., and Lemke, G. (2006). TAM receptor function in the retinal pigment epithelium. Mol. Cell. Neurosci. *33*, 96–108.

Rossi, S.L., Tesh, R.B., Azar, S.R., Muruato, A.E., Hanley, K.A., Auguste, A.J., Langsjoen, R.M., Paessler, S., Vasilakis, N., and Weaver, S.C. (2016). Characterization of a novel murine model to study Zika virus. Am. J. Trop. Med. Hyg. *94*, 1362–1369.

Sakala, I.G., Chaudhri, G., Scalzo, A.A., Eldi, P., Newsome, T.P., Buller, R.M., and Karupiah, G. (2015). Evidence for persistence of ectromelia virus in inbred mice, recrudescence following immunosuppression and transmission to naïve mice. PLoS Pathog. *11*, e1005342.

Sarno, M., Sacramento, G.A., Khouri, R., do Rosário, M.S., Costa, F., Archanjo, G., Santos, L.A., Nery, N., Jr., Vasilakis, N., Ko, A.I., and de Almeida, A.R. (2016). Zika virus infection and stillbirths: a case of hydrops fetalis, hydranencephaly and fetal demise. PLoS Negl. Trop. Dis. *10*, e0004517.

Savidis, G., McDougall, W.M., Meraner, P., Perreira, J.M., Portmann, J.M., Trincucci, G., John, S.P., Aker, A.M., Renzette, N., Robbins, D.R., et al. (2016). Identification of Zika virus and Dengue virus dependency factors using functional genomics. Cell Rep. *16*, 232–246.

Sheehan, K.C., Lai, K.S., Dunn, G.P., Bruce, A.T., Diamond, M.S., Heutel, J.D., Dungo-Arthur, C., Carrero, J.A., White, J.M., Hertzog, P.J., et al. (2006). Blocking monoclonal antibodies specific for mouse IFN-alpha/beta receptor subunit 1 (IFNAR-1) from mice immunized by in vivo hydrodynamic transfection. J. Interferon Cytokine Res *26*, 804–819.

Sheehan, K.C., Lazear, H.M., Diamond, M.S., and Schreiber, R.D. (2015). Selective blockade of interferon-α and -β reveals their non-redundant functions in a mouse model of West Nile virus infection. PLoS ONE *10*, e0128636.

Tabata, T., Petitt, M., Puerta-Guardo, H., Michlmayr, D., Wang, C., Fang-Hoover, J., Harris, E., and Pereira, L. (2016). Zika virus targets different primary human placental cells, suggesting two routes for vertical transmission. Cell Host Microbe *20*, 155–166.

Tiffany, A., Vetter, P., Mattia, J., Dayer, J.-A., Bartsch, M., Kasztura, M., Sterk, E., Tijerino, A.M., Kaiser, L., and Ciglenecki, I. (2016). Ebola virus disease complications as experienced by survivors in Sierra Leone. Clin. Infect. Dis. *62*, 1360–1366.

Tugwell, B.D., Patel, P.R., Williams, I.T., Hedberg, K., Chai, F., Nainan, O.V., Thomas, A.R., Woll, J.E., Bell, B.P., and Cieslak, P.R. (2005). Transmission of hepatitis C virus to several organ and tissue recipients from an antibody-negative donor. Ann. Intern. Med. *143*, 648–654.

Valverde, P., Obin, M.S., and Taylor, A. (2004). Role of Gas6/Axl signaling in lens epithelial cell proliferation and survival. Exp. Eye Res. *78*, 27–37.

van den Broek, M.F., Müller, U., Huang, S., Aguet, M., and Zinkernagel, R.M. (1995). Antiviral defense in mice lacking both alpha/beta and gamma interferon receptors. J. Virol. *69*, 4792–4796.

Varkey, J.B., Shantha, J.G., Crozier, I., Kraft, C.S., Lyon, G.M., Mehta, A.K., Kumar, G., Smith, J.R., Kainulainen, M.H., Whitmer, S., et al. (2015). Persistence of Ebola virus in ocular fluid during convalescence. N. Engl. J. Med. *372*, 2423–2427.

Ventura, C.V., Maia, M., Bravo-Filho, V., Góis, A.L., and Belfort, R., Jr. (2016a). Zika virus in Brazil and macular atrophy in a child with microcephaly. Lancet *387*, 228.

Ventura, C.V., Maia, M., Travassos, S.B., Martins, T.T., Patriota, F., Nunes, M.E., Agra, C., Torres, V.L., van der Linden, V., Ramos, R.C., et al. (2016b). Risk factors associated with the ophthalmoscopic findings identified in infants with presumed zika virus congenital infection. JAMA Ophthalmol. *134*, 912–918.

Cell Reports
Article

Comprehensive RNA Polymerase II Interactomes Reveal Distinct and Varied Roles for Each Phospho-CTD Residue

Kevin M. Harlen,[1] Kristine L. Trotta,[1] Erin E. Smith,[1] Mohammad M. Mosaheb,[2] Stephen M. Fuchs,[2] and L. Stirling Churchman[1],*
[1]Department of Genetics, Harvard Medical School, Boston, MA 02115, USA
[2]Department of Biology, Tufts University, Medford, MA 02155, USA
*Correspondence: churchman@genetics.med.harvard.edu
http://dx.doi.org/10.1016/j.celrep.2016.05.010

SUMMARY

Transcription controls splicing and other gene regulatory processes, yet mechanisms remain obscure due to our fragmented knowledge of the molecular connections between the dynamically phosphorylated RNA polymerase II (Pol II) C-terminal domain (CTD) and regulatory factors. By systematically isolating phosphorylation states of the CTD heptapeptide repeat ($Y_1S_2P_3T_4S_5P_6S_7$), we identify hundreds of protein factors that are differentially enriched, revealing unappreciated connections between the Pol II CTD and co-transcriptional processes. These data uncover a role for threonine-4 in 3′ end processing through control of the transition between cleavage and termination. Furthermore, serine-5 phosphorylation seeds spliceosomal assembly immediately downstream of 3′ splice sites through a direct interaction with spliceosomal subcomplex U1. Strikingly, threonine-4 phosphorylation also impacts splicing by serving as a mark of co-transcriptional spliceosome release and ensuring efficient post-transcriptional splicing genome-wide. Thus, comprehensive Pol II interactomes identify the complex and functional connections between transcription machinery and other gene regulatory complexes.

INTRODUCTION

Transcription regulation controls both RNA transcript levels and the final sequence of the RNA transcript through regulated coordination with co-transcriptional RNA processing. Capping, splicing, and termination are coupled to transcription through phosphorylation of the C-terminal domain (CTD) of the largest RNA polymerase II (Pol II) subunit. The CTD extends from the Pol II body and consists of a conserved heptapeptide, Tyr1-Ser2-Pro3-Thr4-Ser5-Pro6-Ser7, repeated 26 times in yeast and 52 times in humans. Five of the seven residues are dynamically phosphorylated during the transcription cycle in a characteristic pattern across gene bodies. (Fuchs et al., 2009; Hintermair et al., 2012; Jeronimo et al., 2013; Komarnitsky et al., 2000; Mayer et al., 2010, 2012). Phosphorylation of particular CTD residues contributes to the recruitment of key regulatory factors at the right place and time, facilitating connections to co-transcriptional processes. Yet, the extent of connections between CTD phosphorylation marks and RNA-processing machinery are not clear. A focus on identifying direct phospho-CTD binding proteins has slowed experimental throughput and overlooked mechanisms that occur through phosphorylation marks blocking regulatory interactions or indirectly connecting transcription to other processes.

The best-characterized CTD residues are Ser2 and Ser5. Pol II complexes containing high levels of phospho-Ser2 (Ser2P) are thought to be involved in the latter stages of transcription as Ser2P rises slowly across gene bodies and peaks at the 3′ ends of genes (Barillà et al., 2001; Kim et al., 2004). A termination factor, Rtt103, directly binds Ser2P, yet is only recruited after the polyadenylation (polyA) site. This specificity is controlled in part by phospho-Tyr1 (Tyr1P), presumably through blocking Rtt103 binding before the polyA site (Mayer et al., 2012). It is unclear whether additional residues contribute to the precise recruitment of Rtt103 and, more generally, how multiple residues might work together in the coupling of other RNA-processing events to transcription.

Pol II complexes containing high levels of phospho-Ser5 (Ser5P) are enriched near the transcription start site (TSS) of genes, consistent with the essential role of Ser5P in the earliest mRNA-processing event, capping of the nascent RNA (Cho et al., 1997; Komarnitsky et al., 2000; McCracken et al., 1997; Schroeder et al., 2000; Schwer and Shuman, 2011). Recently, a connection between Ser5P and splicing has been suggested. In yeast, Ser5P increases transiently during a splicing-related transcription checkpoint at the 3′ splice site, where Pol II pauses and does not resume transcription until splicing-related events occur (Alexander et al., 2010; Chathoth et al., 2014). However it remains unclear whether there are direct connections between splicing and Ser5P or the CTD in general (Licatalosi et al., 2002). Initial studies indicated a component of the spliceosome (Prp40) could directly bind the CTD (Morris and Greenleaf, 2000). However, other work showed no defect on the co-transcriptional recruitment of the spliceosome in strains where the CTD-binding domain of Prp40 was removed (Görnemann et al., 2011). Thus, despite the observation of co-transcriptional splicing and spliceosomal assembly (Bentley, 2014; Kotovic et al., 2003; Lacadie and Rosbash, 2005; Lacadie et al., 2006; Moore et al., 2006; Tardiff et al., 2006), the degree of coupling between the transcription and splicing machineries, and what role Ser5P may play, remains unclear.

Not all CTD residues are as well characterized as Ser2 and Ser5. For example, Thr4 regulates elongation in human cells, but phospho-Thr4 (Thr4P) levels peak after polyA sites (Hintermair et al., 2012). Substitution of valine for each CTD threonine yielded a defect only in 3′ end processing of histone mRNAs in chicken DT40 cells (Hsin et al., 2011) and had minimal impact on the transcriptome in yeast (Rosonina et al., 2014). Yet, recent mass spectrometry analysis of the Pol II CTD showed that Thr4P is possibly as abundant as Ser2P, suggesting that key roles for this residue remain to be discovered (Schüller et al., 2016).

To determine unappreciated connections between CTD phosphorylation and RNA processing, we performed an unbiased and comprehensive analysis of phospho-specific Pol II complexes. We developed a sequential purification strategy to purify Pol II complexes enriched for each CTD phosphoisoform and used quantitative label-free mass spectrometry to identify factors that interact with phospho-specific Pol II complexes compared to all Pol II complexes. Application of this approach to *Saccharomyces cerevisiae* identified unique phospho-specific Pol II interactomes for the five phosphorylatable CTD residues. The factors enriched in each phospho-CTD interactome allow prediction and discovery of new roles for individual factors and phospho-CTD residues. By combining the interactome data with multiple high-resolution genomic techniques, we describe a previously uncharacterized role for Thr4 in transcription termination in yeast, through the recruitment of termination factors and regulation of Pol II dynamics.

Our proteomic analysis reveals a striking differential enrichment of the spliceosome as a function of CTD phosphorylation. The Ser5P interactome is strongly enriched with U1 spliceosomal components, and we demonstrate that Ser5P contributes directly to the co-transcriptional recruitment of the spliceosome. The Thr4P interactome has a complete lack of enriched spliceosome subunits, and we observe that this mark peaks in the terminal exon after dissociation of splicing machinery. Thr4 is required in the post-transcriptional completion of splicing, demonstrating that CTD modifications impact both co-transcriptional and post-transcriptional splicing through separate residues. Together, these data comprehensively define the protein factors that associate with Pol II and uncover roles for CTD modifications in splicing and termination regulation.

RESULTS

Proteomic Analysis of Phospho-specific Pol II Complexes

To determine the compendium of factors that associate to actively transcribing Pol II, we first analyzed purifications of total Pol II by label-free quantitative mass spectrometry. We adapted the immunoprecipitation approach developed for native elongating transcript sequencing (NET-seq), whereby engaged Pol II complexes are solubilized from chromatin and immunoprecipitated (Churchman and Weissman, 2011). Analysis of enriched factors was consistent with previous Pol II proteomics studies and in identifying many known transcription and RNA-processing factors (Figure S1A) (Mosley et al., 2011, 2013; Tardiff et al., 2007).

In order to investigate direct and indirect Pol II interactors specific to each CTD phosphoisoform, we developed a native sequential immunoprecipitation (IP) strategy to purify Pol II complexes enriched for different phosphorylated CTD residues (Figure 1A). Our strategy is designed to purify the host of proteins that interact both directly and indirectly with Pol II when a particular residue is phosphorylated. Thus, we obtain a comprehensive picture of the co-occurring processes marked by each phosphorylated residue. To this end, all Pol II complexes are first purified via a 3 × FLAG epitope tag to isolate the full complement of Pol II phosphoisoforms (five phosphorylatable residues across 26 repeats). From this pool, highly efficient IPs enrich for sets of phosphoisoforms with well-characterized and specific antisera targeting each phosphorylatable residue (Tyr1P, Ser2P, Ser5P, Thr4P, and Ser7P) or with a mock antisera (mCherry) (Figure S1B) (Hintermair et al., 2012; Mayer et al., 2012). Native IP conditions were chosen instead of crosslinking in order to reduce the possibility for false positives. Importantly, the repetitive nature of the CTD permits the purification of factors that directly bind the targeted phospho-CTD residue even in the case that their association is displaced by the antibody interaction. Nevertheless, for direct interactors, our purification strategy has the potential to be less quantitative and may yield false negatives. Each phospho-specific IP purifies a subset of Pol II complexes that are highly enriched for the selected phosphorylated residue, yet are also phosphorylated on other residues across the 26 repeats, thus allowing for the enrichment of indirectly associating factors recruited by frequently co-occurring patterns of Pol II modifications (Figure S1C). Furthermore, our strategy will isolate any proteins that are bound to the nascent RNA. In sum, this approach comprehensively purifies proteins that interact directly and indirectly with Pol II coincident with each phosphorylation mark.

To identify enriched factors in a quantitative manner, samples from the Pol II IP, each phospho-specific IP, and the mock IP were analyzed by mass spectrometry in biological triplicate. Significantly enriched factors in each phospho-specific IP were determined using quantitative label-free proteomic analysis (Hubner and Mann, 2011; Hubner et al., 2010). Specifically, the phospho-specific:mock IP fold increase of spectral intensities are compared to p values determined by t tests, generating a volcano plot (Figures 1B, 1C, S1A, and S1D–S1F). Only factors present in all three replicates are analyzed and are identified as significantly enriched using a false discovery rate of 0.05. As the mock IP is done after the first IP, factors enriched with each phosphoisoform represent proteins that are enriched over a background of all Pol II interactors. The stringent nature of this strategy identifies the most highly enriched factors associated with each phosphoisoform. Together, the five Pol II phospho-CTD interactomes contain nearly 200 enriched factors, including Pol II subunits, RNA-processing factors, transcription

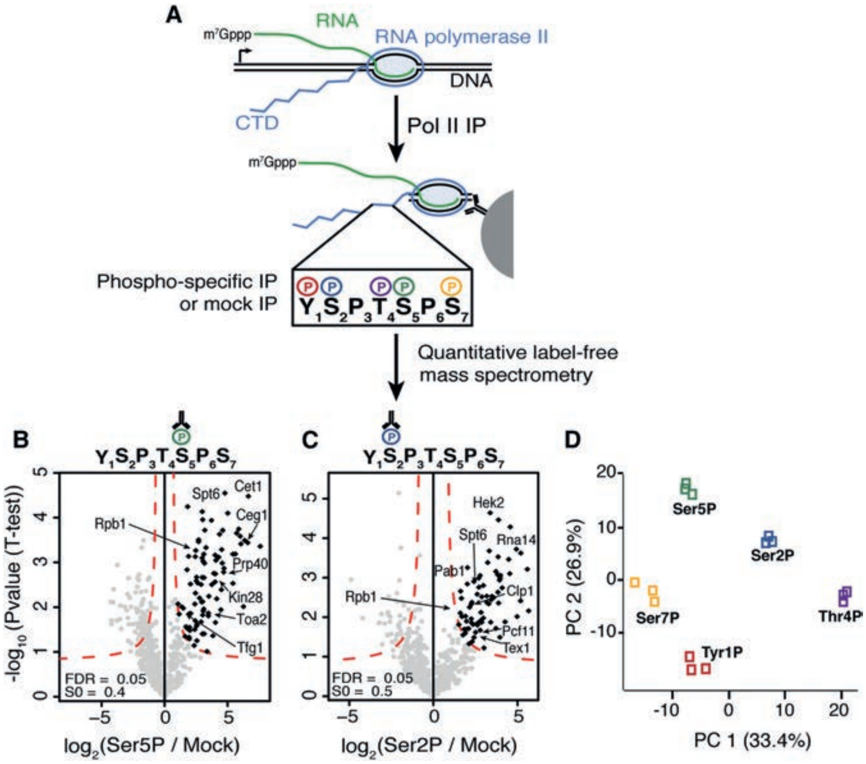


Figure 1. Purification and Analysis of Native Pol II Complexes

(A) Diagram of native Pol II purification followed by phospho-specific immunoprecipitations (IPs) and quantitative label-free mass spectrometry analysis for each CTD phospho-modification; m7Gppp, RNA cap.

(B and C) Volcano plots comparing Ser5P (B) and Ser2P (C) IPs to mock IPs. Specifically enriched factors determined using an FDR of 0.05 and an S0 value (curve bend, see Supplemental Experimental Procedures for details) of 0.4 for Ser5P and 0.5 for Ser2P are highlighted in black. Factors mentioned in the text are labeled.

(D) Principal component analysis of significantly enriched interactors from each phospho-CTD interactome. PC, principal component.

See also Figure S1.

factors, and chromatin-related factors. Our analysis successfully identified factors previously shown to directly associate with different phospho-CTD isoforms, such as enrichment of the mRNA capping complex (Cet1 and Ceg1) and the Ser5 kinase TFIIH in the Ser5P interactome (Figure 1B) (Komarnitsky et al., 2000; Schroeder et al., 2000) and enrichment of Pcf11 and other cleavage, polyadenylation, and export factors (Clp1, Rna14, and Tex1) in the Ser2P interactome (Figure 1C) (Barillà et al., 2001; Buratowski, 2005). Although our focus is not entirely on identifying direct phospho-CTD binders, these results demonstrate that our approach robustly purifies direct Pol II CTD interactors, presumably facilitated by the repetitive nature of the CTD.

Pol II Phospho-specific Interactomes Are Distinct

Utilizing the stringent nature of the interactome data, we asked whether the proteins that are highly enriched in each Pol II CTD interactome constitute distinct groups of factors, which would suggest that each Pol II phospho-CTD residue connects to different Pol II regulation modalities and/or co-transcriptional processes. Principal component analysis of all five phospho-CTD interactomes demonstrates that each interactome contains a unique set of factors that are reproducibly identified (Figure 1D) (Tables S1 and S2). Consistently, we observe minimal overlap of factors across all interactomes (Figure 2A). Gene ontology (GO) analysis of all co-purifying factors displays enrichment of many transcription-related GO terms as well as RNA processing and chromatin organization (Figure 2B). Additionally, the phospho-CTD interactomes show enrichment for protein domains involved in chromatin regulation, RNA processing, and in promoting protein-protein interactions (Figure 2C). Together these data indicate our approach isolates factors with functions in multiple transcriptional and co-transcriptional processes.

We reasoned that proteins with similar phospho-CTD interaction profiles should have related roles in transcription. Pairwise correlation analysis between the phospho-CTD interaction profiles of all enriched factors followed by hierarchical clustering identifies distinct groups of factors (Figure 2D). We find that factors regulating transcription initiation and early transcriptional processes, such as mRNA capping, are highly correlated and are anti-correlated with factors enriched for 3′ end processing of nascent RNA. Also, a majority of transcription elongation factors cluster together, revealing that proteins associated with the three major stages of transcription interact with Pol II in distinct modes with specific phospho-CTD states. These data demonstrate that the interactome of a phospho-CTD residue can be used to predict its function in transcription.

CTD Thr4 and Rtt103 Transition Pol II to Termination

As many functions have been attributed to the well-characterized CTD phosphoisoforms, Ser5P and Ser2P, we utilized the phospho-CTD interactome dataset to explore the lesser-studied Thr4P phosphoisoform. Moreover, mass spectrometry analysis of the Pol II CTD demonstrated that Thr4P is a prevalent modification, possibly as abundant as Ser2P (Schüller et al., 2016; Suh et al., 2016). Purification of Pol II CTD Thr4P complexes was reproducibly efficient (95% average IP efficiency, Figure S1B), and mass spectrometry analysis revealed significant associations between Thr4P and a number of canonical transcription elongation factors (Figure 3A). Interestingly, the transcription termination factor Rtt103 was also enriched with Thr4P. Rtt103 is a CTD binding protein that functions with the exonuclease complex Rat1/Rai1 for termination of polyadenylated transcripts (Buratowski, 2005; Kim et al., 2004) and was recently demonstrated to interact with a CTD containing phosphorylated Thr4 (Suh et al., 2016). To decipher whether interaction of Rtt103 with Pol II complexes is a direct result of Thr4 phosphorylation,

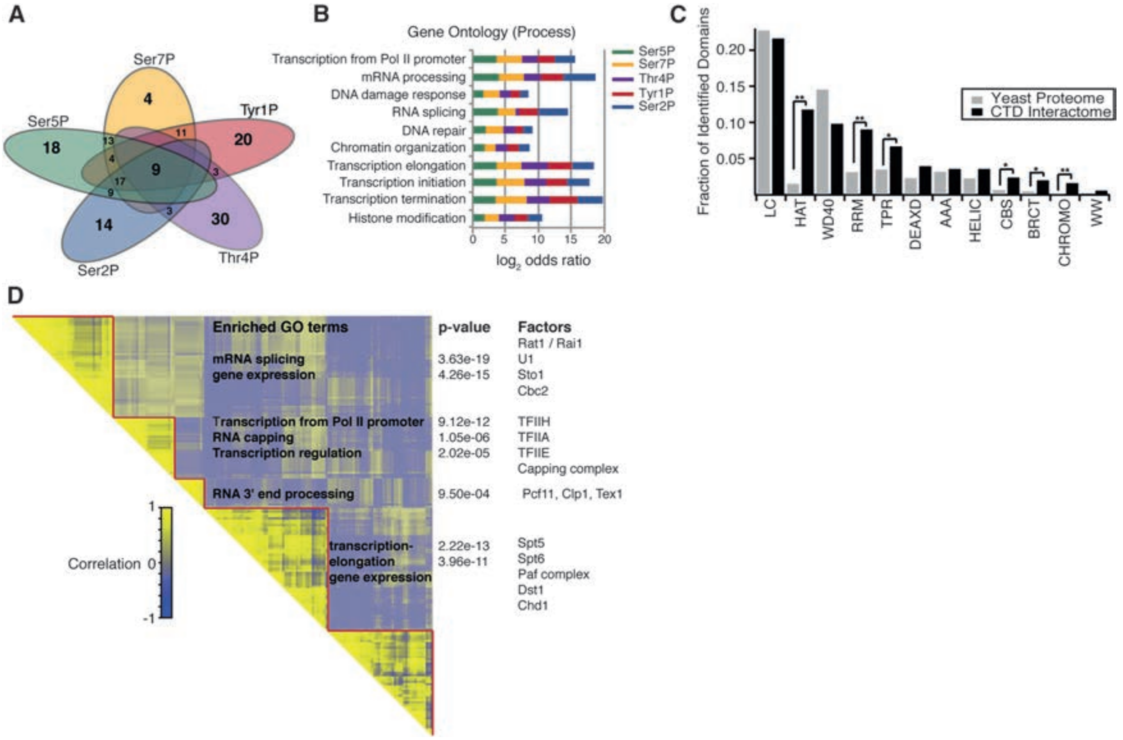


Figure 2. Pol II Phospho-CTD Interactomes Contain Unique Sets of Factors

(A) Venn diagram of enriched factors from the interactome of each CTD phosphoisoform.

(B) Gene ontology terms enriched in the phospho-CTD interactome data are plotted for each phosphoisoform.

(C) Protein domains enriched in the phospho-CTD interactomes. The abundance of the top 12 protein domains identified in the phospho-CTD interactome compared to their abundance in the yeast proteome. Fisher exact test, $^*p < 0.05$, $^{**}p < 0.01$. LC, low complexity; HAT, half-A-TPR repeat; WD40, repeat ending in tryptophan-aspatic acid repeat; RRM, RNA recognition motif; TPR, tetratricopeptide repeat; DEAXD, DEAD-like helicase; AAA, ATPases; HELIC, helicases; CBS, cystathionine beta synthase domain; BRCT, breast cancer c-terminal; CHROMO, chromatin organization modifier; WW, two conversed Trp resides.

(D) Clustered pairwise correlation matrix comparing phospho-CTD interaction profiles of all enriched factors. GO terms enriched in separate clusters, p values for the significance of enrichment of each GO term, and representative factors in each cluster are indicated.

we tested the ability of the Thr4P antisera to compete with Rtt103 for CTD binding. After purifying Pol II complexes using a FLAG epitope tag on the Rpb3 subunit, antisera directed against Thr4P, Ser2P, Ser5P, no antisera, or FLAG peptide (as a positive elution control) was incubated in vast excess (>13-fold higher concentration compared to the IP reactions) with the purified complexes on beads. Supernatant and bead fractions were assayed by western blot (Figure 3B). As expected, the Ser2P antisera were able to compete with Rtt103 for CTD binding (Kim et al., 2004), and the Ser5P antisera were not effective competition. Consistent with the mass spectrometry data, Thr4P antisera were also able to compete with Rtt103 for CTD binding, indicating both residues likely regulate the interaction between Rtt103 and the Pol II CTD.

Surprisingly, the Rat1/Rai1 exonuclease complex was not enriched in the Thr4P interactome, leading us to postulate that Rtt103 and Rat1/Rai1 may regulate separate steps in the termination process. To test this hypothesis, we used NET-seq, which maps actively transcribing Pol II complexes at nucleotide resolution genome-wide (Churchman and Weissman, 2011) and analyzed changes in Pol II density in *rtt103Δ* and *rai1Δ* deletion mutants after polyA sites. Aggregate analysis of wild-type NET-seq data around polyA sites reveals an increase in Pol II density just after polyA sites where 3′ end cleavage occurs co-transcriptionally (Figure 3C, black line). Subsequent termination leads to a decrease in Pol II density before neighboring genes increase the signal again. Individual deletion of RAI1 and RTT103 caused widespread termination defects, albeit in very different ways (Figures 3C and S2A). Loss of RAI1 (Figure 3C, green line) leads to a steady rise in Pol II density after polyA sites, indicating that transcription proceeds for much longer. In contrast, loss of RTT103 (Figures 3C and 3D, blue line) results in a pronounced peak of Pol II density directly after polyA sites precisely where Rtt103 occupancy peaks (Figure S2B) (Mayer et al., 2012). Furthermore, termination is not affected as Pol II density decreases normally thereafter, suggesting that instead Rtt103 binding regulates the onset of termination.

As Rtt103 is enriched in the Thr4P interactome, we postulated that Thr4 aids Rtt103 in regulating Pol II. We mapped Thr4P genome-wide at near nucleotide resolution using exonuclease-based chromatin immunoprecipitation (ChIP)-nexus (He et al., 2015), which revealed a peak of threonine-4 phosphorylation downstream of the polyA site, where the peak in Pol II density is observed in the *rtt103Δ* strain (Figure 3D purple line, and

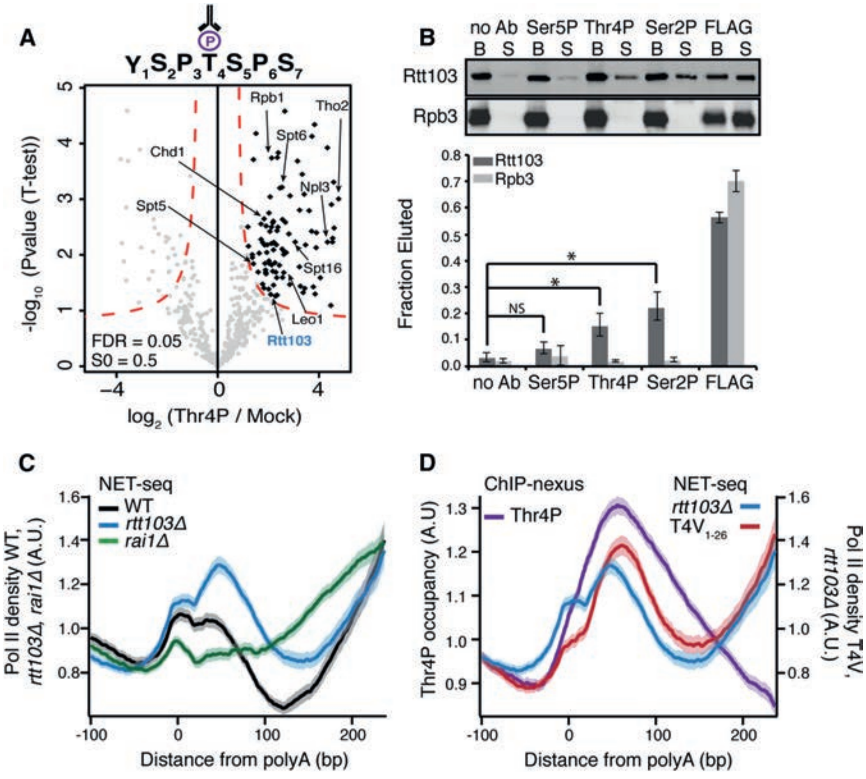


Figure 3. Thr4P Regulates Pol II Dynamics during Transcription Termination

(A) Volcano plot comparing Thr4P IPs to mock IPs. Specifically enriched factors are determined using an FDR of 0.05 and S0 = 0.5 and are highlighted in black. Factors mentioned in the text are labeled; Rtt103 is labeled in blue.

(B) Western blot (top) and quantification (bottom) of the amount of Rtt103 or Rpb3 eluted from purified Pol II complexes by phospho-CTD antisera (no Ab, Ser5P, Ser2P, Thr4P) or FLAG peptide. Competition assays were conducted in biological triplicate. Plotted are mean fraction eluted values; error bars, SD. NS, not significant; $^*p < 0.01$; Ab, antibody; B, beads; S, supernatant.

(C) Normalized average NET-seq profiles of WT, *rtt103Δ*, and *rai1Δ* cells at polyadenylation sites (polyA) of protein coding genes.

(D) Normalized average NET-seq profiles of *rtt103Δ* and T4V$_{1–26}$ mutant CTD cells and normalized average ChIP-nexus profile of Thr4P at the polyA site of protein coding genes. In (C) and (D), NET-seq and ChIP-nexus reads for each gene are normalized by total reads for each gene in the analyzed region; shaded areas represent the 95% confidence interval. n = 2,738. A.U., arbitrary units. See also Figures S2 and S3.

Figures S2B and S2D–S2G). Consistent with our detection of Rtt103 in the Thr4P interactome, the peak in Thr4P density overlaps with Rtt103 density (Figure S2B). To determine whether Thr4 controls 3′ end cleavage and transcription termination, we replaced each Thr4 residue in the endogenous 26 CTD heptapeptide repeats with a valine residue (T4V$_{1–26}$), representing a non-phosphorylatable state, and with a glutamic acid residue (T4E$_{1–26}$), mimicking the constitutively phosphorylated state. The T4E$_{1–26}$ mutant is lethal (Figure S3A), but the T4V$_{1–26}$ mutant is viable, as observed previously (Rosonina et al., 2014), with a modest growth defect (15% increase in doubling time). NET-seq analysis at the 3′ ends of genes demonstrates that the T4V$_{1–26}$ strain has a similar post-polyA transcription defect as the *rtt103Δ* strain with a strong increase in Pol II density after polyA sites, at exactly the location where Thr4P peaks in occupancy (Figures 3D and S2A, red line). A less stringent and inducible T4V mutant was created by generating an allele of RPB1 that encodes a CTD where Thr4 of the first eight repeats proximal to the Pol II core are mutated to valine (T4V$_{1–8}$). By placing the endogenous RPB1 under control of a Tet-off promoter, introduction of doxycycline induces incorporation of the mutant CTD into the Pol II complex (Malagon et al., 2006; Morrill et al., 2016). NET-seq analysis in this T4V$_{1–8}$ mutant shows defects similar to the T4V$_{1–26}$ mutant in post-polyA transcription (Figure S3B). RNA sequencing (RNA-seq) analysis of the T4V$_{1–26}$ strain shows a global downstream shift in the locations of polyA sites, consistent with a possibility of altered cleavage in this strain (Figures S3C–S3E). Importantly, no downregulation of transcription elongation or RNA-processing-related gene classes was detected in either the T4V$_{1–26}$ or *rtt103Δ* mutants (Table S3; Figures S2C and S6E). Together these data lead to a model by which both Ser2P and Thr4P recruit Rtt103 after polyA sites and indicate the presence of a post-polyA checkpoint controlled by Thr4 that ensures robust 3′ end cleavage followed by an efficient transition to transcription termination. Moreover, these data illustrate how analysis of Pol II CTD interactomes reveals distinct roles for CTD modifications and their interacting factors.

CTD Ser5P Is Enriched for Spliceosomal Components while Thr4P Is Depleted

To search for additional connections between Thr4P and RNA-processing events, we compared the interactomes by complete-linkage hierarchical clustering. Interestingly, CTD phosphorylation marks with similar ChIP occupancy (Ser5P/Ser7P and Tyr1P/Ser2P) have interactomes that cluster together (Figure 4A), and the Thr4P interactome does not cluster with any other phospho-CTD interactome. A major feature separating the Thr4P interactome is the absence of enriched spliceosomal subunits (Figure 4B) while other phospho-CTD interactomes enrich for eight to 18 spliceosomal subunits across five subcomplexes. In contrast, the Ser5P interactome is highly enriched for many components of the U1 spliceosomal complex, purifying nearly 70% of U1-associated proteins. To determine whether U1 is associating through a protein-protein interaction with Ser5P, we tested the ability of Ser5P antisera to displace U1 from purified Pol II complexes. Ser5P antisera were able to elute nearly half of the associated U1 from the CTD, while Thr4P antisera were unable to significantly compete with U1 binding (Figure 4C).

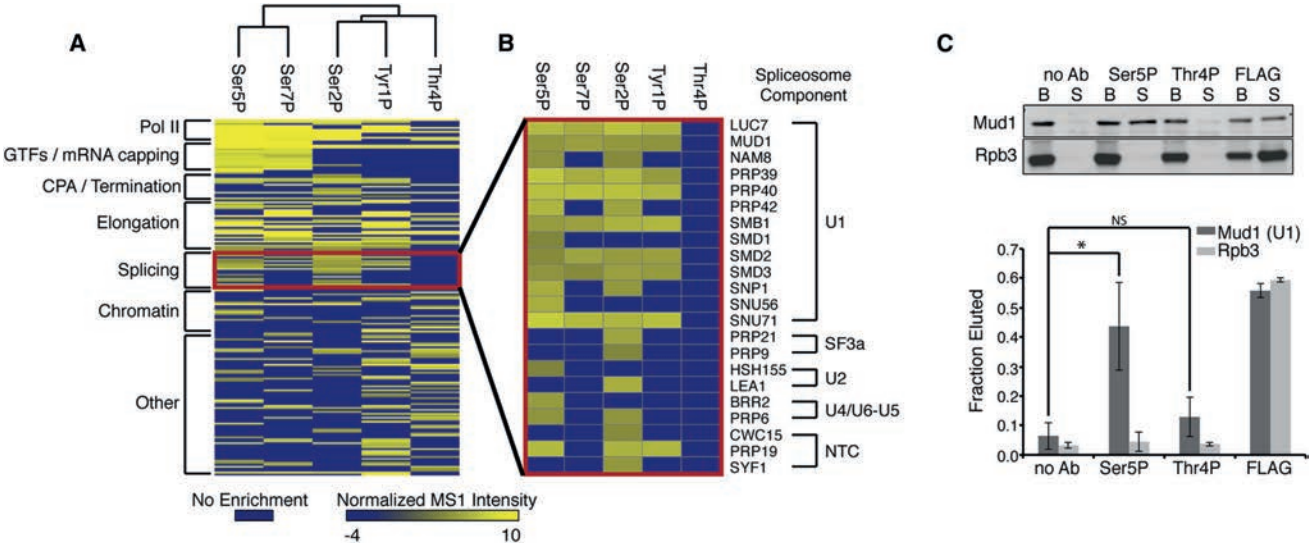


Figure 4. Phospho-CTD Interactomes Reveal Connection between Pol II CTD Ser5P, Thr4P, and Splicing

(A) Complete linkage clustering of each Pol II phospho-CTD interactome.

(B) Heatmap of splicing factors specifically enriched in each phospho-CTD Pol II interactome.

(C) Western blot (top) and quantification (bottom) of the amount of Mud1 (U1) or Rpb3 eluted from purified Pol II complexes by phospho-CTD antisera (no Ab, Ser5P, Thr4P) or FLAG peptide. Competition assays were conducted in biological triplicate. Plotted are mean fraction eluted values, error bars, SD; NS, not significant; *p < 0.05; Ab, antibody; B, beads; S, supernatant.

These data suggest that U1 directly associates with Pol II in a Ser5P-dependent manner. The difference in enrichment and interaction with the spliceosome between Ser5P and Thr4P suggest each phospho-modification may regulate different stages of splicing.

Co-transcriptional Spliceosome Recruitment Is Facilitated by Ser5P

To examine the roles of both Ser5P and Thr4P in coordinating splicing, we first analyzed the co-transcriptional recruitment of the spliceosome through mapping obligatory components of the U1 and U2 spliceosomal subcomplexes, Mud1 and Lea1, respectively. Analysis of U1 and U2 occupancy by ChIP-nexus displays enrichment for spliced genes and matches lower-resolution profiles around the 3′ splice site from U1 and U2 ChIP-chip data (Tardiff et al., 2006) (Figures S4A–S4E). We observe high levels of U1 and U2 enrichment over spliced genes, and, consistent with other studies, we observe low signal over unspliced genes (Figure S4D) (Kotovic et al., 2003; Moore et al., 2006; Tardiff et al., 2006). Hierarchical clustering of the data based on U1 occupancy breaks spliceosome occupancy into two clear categories of high- and low-occupancy genes, which had been observed previously (Figure 5A; Table S4) (Tardiff et al., 2006). The high-occupancy genes tend to be highly co-transcriptionally spliced and well expressed, to have longer introns and shorter terminal exons, and to encode ribosomal proteins (Figure S4F) (Carrillo Oesterreich et al., 2010). We focused our analysis on the high-occupancy genes. The ChIP-nexus data resolve a rapid and sharp increase in U1 occupancy 10–30 bp downstream of the 3′ splice site that remains high for 200 bp before decreasing (Figures 5B and 5C). Extending for almost 400 bp, U2 occupancy also increases sharply at the 3′ splice site, albeit more slowly than U1, consistent with ordered spliceosomal assembly (Figures 5C and S4G) (Hoskins et al., 2011).

From our Pol II CTD interactomes, we postulated that Ser5P occupancy would mirror U1. ChIP-nexus analysis of Ser5P occupancy revealed a characteristic peak of enrichment near transcription start sites of coding genes (Figures S5A–S5C). Interestingly, analysis around 3′ splice sites of high U1 occupancy genes displays a sharp increase in Ser5P 10–30 bp downstream of the 3′ splice site that persists for 200 bp, overlaying precisely with the U1 peak (Figures 5B and 5C). In contrast, Thr4P peaks late into the terminal exon, concomitant with low levels of Ser5P, U1, and U2 (Figures 5B and 5C). Interestingly, the peak of Thr4P in the terminal exon occurs just prior to polyA sites and is distinct from the peak of Thr4P after polyA sites (Figures 5B, 5C, and S5D). The bimodal Thr4P profile around the polyA sites is not observed at unspliced genes where only a single peak occurs after polyA sites (Figure S2G). These trends can be observed through comparison of raw ChIP-nexus signal between U1 and Ser5P (Pearson correlation, r = 0.61) or Thr4P (Pearson correlation, r = −0.02) across terminal exons (Figures 5D and S5F). These data suggest that Ser5P recruits the spliceosome at the 3′ splice site and that when the spliceosome dissociates, high levels of Thr4P are present. Importantly, U1 and U2 occupancy profiles are not altered in the $T4V_{1\text{–}26}$ mutant (Figure S5E), suggesting that Thr4 does not cause the dissociation of the spliceosome, and that instead high levels of Thr4P act as a mark of spliceosome release.

The increase in Ser5P and rapid recruitment of U1 at the 3′ splice site are coincident with a previously proposed splicing checkpoint involving Ser5P and Pol II pausing around the 3′ splice site (Alexander et al., 2010; Chathoth et al., 2014). Furthermore, pausing has been observed elsewhere at spliced

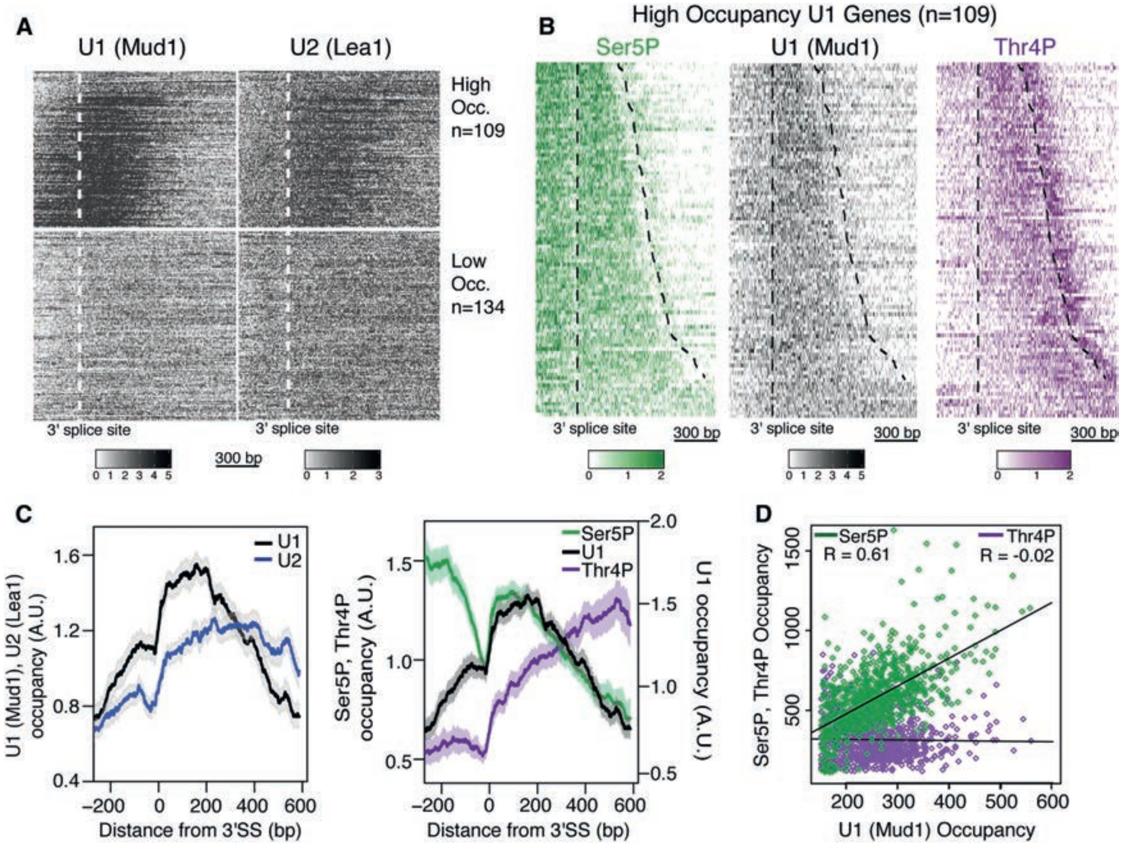


Figure 5. Pol II CTD Ser5P and Thr4P Regulate Spliceosome Occupancy and Splicing Dynamics

(A) Heatmap of normalized U1 (Mud1) and U2 (Lea1) ChIP occupancies around the 3′ splice site (300 bp upstream and 1,000 bp downstream) of non-overlapping intron-containing genes (see Supplemental Experimental Procedures). Genes are classified as high (High Occ.) or low occupancy (Low Occ.) based on U1 recruitment.

(B) Heatmap of normalized Ser5P, U1, and Thr4P ChIP occupancies around the 3′ splice site (300 bp upstream and 1,000 bp downstream) of high U1 occupancy genes. Dashed lines represent the 3′ splice site and polyadenylation site, respectively.

(C) Left: normalized average ChIP-nexus profiles of U1 (Mud1) and U2 (Lea1) around the 3′ splice site (3′SS) of high U1 occupancy spliced genes. Normalization of ChIP data to internal spike-ins allows direct comparison of U1 and U2 profiles. Right: normalized average ChIP-nexus profiles of Ser5P and Thr4P and U1 around the 3′ splice site of high U1 occupancy spliced genes. n = 109, Shaded areas represent 95% confidence intervals.

(D) Comparison of U1 (Mud1) and Ser5P (green) or U1 and Thr4P (purple) ChIP-nexus reads from the 3′ splice site to 100 bp past the polyA site of high U1 occupancy spliced genes (n = 109). R, Pearson correlation coefficient.

In (A)–(C), ChIP-nexus reads for each gene are normalized by total reads for each gene in the analyzed region.

See also Figures S4 and S5.

genes in yeast (Carrillo Oesterreich et al., 2010). However, these studies have been limited by low resolution or indirect methods to detect transcriptional pausing. We analyzed NET-seq profiles of high-occupancy genes around the 3′ splice site (3′ SS) in an effort to resolve putative splicing checkpoints at nucleotide resolution. Analysis by NET-seq revealed highly reproducible patterns of Pol II pausing at intron-exon junctions as well as increased Pol II density over exons similar to earlier observations in yeast and in human cells (Figures 6A, 6B, and S5G) (Alexander et al., 2010; Jonkers et al., 2014; Mayer et al., 2015; Veloso et al., 2014). Splicing intermediates, such as lariats, are known contaminants in NET-seq data, so the precise nucleotide position where their 3′ ends align are removed from analysis, resulting in a single nucleotide gap. In order to determine whether the remaining signal upstream and downstream of the 3′ SS are due to lariats from missplicing events, we sequenced total nascent RNA in order to analyze the abundance of alternative splicing junctions. Greater than 99% of transcripts are correctly spliced (Figure S5H), strongly indicating that the NET-seq signal around the 3′ SS is due to Pol II pausing. Moreover, pausing at the 3′ SS is observed uniformly across spliced genes with high U1 recruitment (Figure 6A). Together these data suggest coordination between Pol II dynamics, Ser5 phosphorylation, and recruitment of the spliceosome during a splicing checkpoint at the 3′ splice site.

We reasoned that if Ser5P is responsible for the rapid and precise recruitment of U1 directly following 3′ splice sites, then a loss of Ser5 would disrupt splicing. As complete loss of Ser5 is lethal, we created an allele of the RPB1 gene that encodes the CTD where Ser5 of the first eight repeats proximal to the Pol II core are mutated to alanine (S5A$_{1-8}$), similar to the T4V$_{1-8}$ strain.

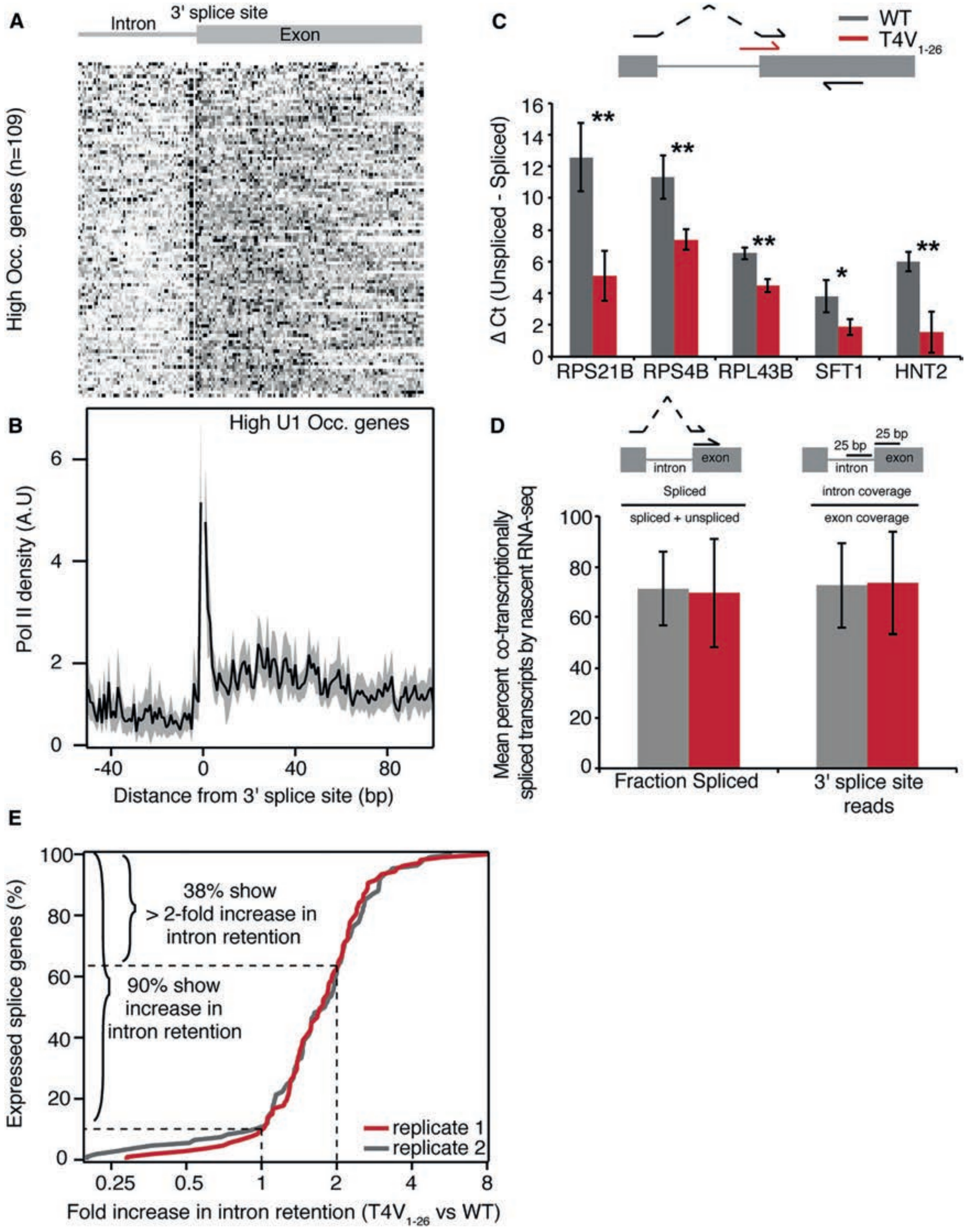


Figure 6. Pol II CTD Thr4 Regulates Post-transcriptional Splicing

(A) Heatmap of normalized NET-seq reads around the 3′ splice site of high U1 occupancy genes.

(B) Normalized average NET-seq reads of around the 3′ splice site of high U1 occupancy genes. NET-seq reads for each gene are normalized by total reads for each gene in the analyzed region. Shaded areas represent 95% confidence intervals. Splicing intermediates are contaminants in NET-seq data, so the precise 1-bp positions where they align are removed from analysis and lead to a gap in the Pol II density profiles. A.U., arbitrarty units, n = 109.

(C) Quantification of the mean difference in unspliced versus spliced transcripts in WT and T4V$_{1–26}$ CTD mutant cells as determined by qPCR of splice junctions at five spliced genes. Error bars, SD of three biological replicates. *$p < 0.05$, **$p < 0.01$. Top: diagram depicting primer pairs used to determine spliced (black forward arrow) and unspliced (red forward arrow) transcript abundance.

(legend continued on next page)

We also constructed similar alleles for Ser2 ($S2A_{1–8}$) and for Ser7 ($S7A_{1–8}$). Here, the endogenous RPB1 is again controlled by a Tet-off promoter, and we investigated the splicing phenotype in total RNA of these strains as well as the $T4V_{1–8}$ strain after exposing the cells to doxycycline. Real-time qPCR analysis of splice junctions revealed that, in contrast to $S2A_{1–8}$ and $S7A_{1–8}$, both $S5A_{1–8}$ and $T4V_{1–8}$ lead to intron retention at the RPL34B gene (Figure S6A), demonstrating the functional importance of Ser5P for efficient splicing, presumably through the recruitment of U1.

Thr4 Is Responsible for Efficient Post-transcriptional Splicing Genome-wide

The pre-polyA peak of Thr4P specifically at spliced genes, a lack of enriched spliceosomal subunits in the Thr4P interactome, and the splicing defect observed in the $T4V_{1–8}$ strain lead us to postulate that Thr4 has a role in splicing. The splicing defect caused by the $T4V_{1–8}$ mutation is likely to be a modest phenotype as there are still 18 wild-type repeats. Indeed, real-time qPCR analysis of the $T4V_{1–26}$ strain revealed a much stronger splicing defect at five genes (Figure 6C). To ask whether the splicing defect arises co-transcriptionally, we generated nascent RNA-seq data for wild-type (WT) and $T4V_{1–26}$ cells. Using two different metrics, we show that 73% of transcripts are co-transcriptionally spliced, consistent with previous reports, with no difference between WT and $T4V_{1–26}$ cells (Carrillo Oesterreich et al., 2010) (Figure 6D). Consistently, we observe no defect in the co-transcriptional recruitment of U1 or U2 in the $T4V_{1–26}$ strain (Figure S5E). We next considered whether post-transcriptional splicing is globally impacted by a replacement of Thr4 with valine by performing total RNA-seq. The inclusion of RNA standards allowed for an accurate and quantitative measure of intron retention genome-wide (Figures S6B–S6E). Comparison of $T4V_{1–26}$ and wild-type strains reveal 90% of spliced genes have increased intron retention in the $T4V_{1–26}$ strain and that 38% display a greater than 2-fold increase (Figures 6E and S6F–S6H). Importantly, no downregulation of transcription elongation, splicing, RNA processing, or RNA-turnover-related gene classes was detected (Table S3). Furthermore, a re-analysis of RNA-seq data from a T4V mutant reported in Rosonina et al. also reveals a splicing defect (Rosonina et al., 2014) (Figure S6I). Thus, Thr4 does not impact co-transcriptional splicing and is instead a critical residue for ensuring efficient splicing post-transcriptionally.

DISCUSSION

Transcription elongation proceeds through multiple stages, marked by distinct CTD phosphorylation patterns and by the set of factors regulating Pol II to coordinate co-transcriptional processes. Here, we combined phospho-specific CTD Pol II purification with quantitative mass spectrometry to identify the interactomes of each phospho-CTD residue. Importantly, we observe minimal overlap of factors across phospho-CTD interactomes, emphasizing the unique roles of each phosphorylation mark. The interactomes reveal the complement of factors interacting with Pol II throughout transcription elongation and will serve as a resource for future studies on the regulation of transcription and co-transcriptional processes.

The Ser5P and Ser2P modifications have been studied extensively, but recent studies have demonstrated that all the phospho-modifications of the CTD can regulate the transcription process (Descostes et al., 2014; Egloff et al., 2007; Hintermair et al., 2012; Hsin et al., 2011, 2014a, 2014b; Mayer et al., 2012; Rosonina et al., 2014). Furthermore, our work and that of others suggest that factors interact with Pol II in a highly regulated manner requiring multiple inputs (presence or absence of particular CTD phosphorylation marks, *cis*-acting elements in the nascent RNA, etc.) (Kim et al., 2004; Komarnitsky et al., 2000; Lacadie and Rosbash, 2005; Lacadie et al., 2006; Licatalosi et al., 2002; McCracken et al., 1997; Mayer et al., 2010, 2012; Moore et al., 2006; Schwer and Shuman 2011; Suh et al., 2016). For example, Ser2P and the lack of Tyr1P are proposed to facilitate the localization of Rtt103 to Pol II (Mayer et al., 2012). Here, we describe Thr4P as an additional input to this process. Our data are consistent with a model whereby Thr4 phosphorylation aids the recruitment of Rtt103 to the transcription elongation complex, followed by a Rtt103/Thr4-regulated release of Pol II from a region of pausing after polyA sites (Figure 7). These results raise the possibility that Pol II slows transcription until the arrival of Rtt103 and the onset of termination. Thus, Thr4 joins Tyr1 and Ser2 in facilitating the precise and timely association of Rtt103 with the transcription elongation complex, indicating that the "CTD code" includes intricate CTD phosphorylation combinations to orchestrate the connections between transcription and co-transcriptional processes. Interestingly, the recruitment of Rtt103 may be even more complicated as we also observe a strong enrichment for Rtt103 with Pol II CTD Ser7P (Table S1).

Our data reveal that the coupling of transcription to splicing is also controlled by multiple inputs, revealing CTD Ser5 and Thr4 as being critical links between transcription and splicing. Our analyses suggest a model in which after Pol II releases from a 3′ splice site pause, the U1 spliceosomal subcomplex is rapidly recruited to Pol II through physical interaction with Ser5P. U2 subsequently associates, which presumably leads to co-transcriptional splicing (Figure 7). Interestingly, Mud1 (U1) does not

(D) Mean fraction of co-transcriptionally spliced transcripts as determined by nascent RNA-seq in WT (black) or the $T4V_{1–26}$ CTD mutant (red) in high U1 occupancy genes. Two metrics were used to determine co-transcriptionally spliced reads: fraction spliced, which calculates the fraction of spliced reads by dividing the number of spliced reads by the total number of reads (spliced + unspliced) spanning the 3′ splice site and 3′ splice site coverage, which uses the ratio of intron to exon coverage in a region spanning 25 bp up- or downstream of the 3′ splice site and subtracting this value from 1. Top: diagrams depicting the fraction spliced and 3′ splice site metrics. Error bars, SD.

(E) Distribution of the increase in intron retention in the $T4V_{1–26}$ CTD mutant versus WT cells for two biological replicates (n = 107 and 108 genes). Genes included in the analysis contained reliably quantifiable intron coverage as determined by RNA standards in both samples (Figure S6B and Supplemental Experimental Procedures).

See also Figures S5 and S6.

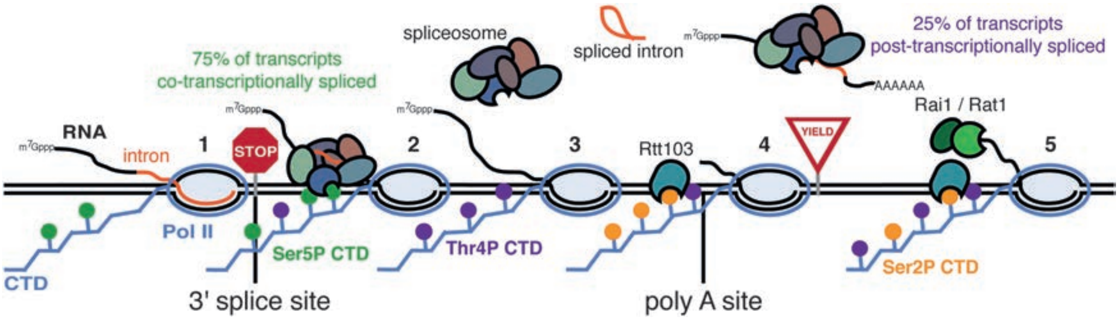


Figure 7. Coupling of Transcription to RNA Processing by Ser5P and Thr4P

Model displaying how Ser5 and Thr4 phosphorylation coordinates the pausing of Pol II the at the 3′ splice site (1), recruitment of the splicing machinery (2), completion of splicing and release of the spliceosome (3), pausing for 3′ end processing and recruitment of Rtt103 (4), and the transition from 3′ end processing to termination (5).

co-localize with the canonical Ser5P peak at the 5′ end of all genes, indicating that Ser5P is not sufficient for U1 recruitment. Certainly known interactions between U1 and the nascent RNA likely serve as additional inputs to ensure proper localization (Lacadie and Rosbash, 2005). After 3′ splice sites, Thr4P levels increase steadily and peak after U1 and U2. These occupancy data and the lack of spliceosomal subunits in the Thr4P interactome suggest that high levels of Thr4P mark spliceosomal release. Interestingly, we observe that Thr4 is strongly connected to post-transcriptional splicing, but how is this mediated? A modest peak of spliceosome recruitment occurs after polyA sites where both U1 and U2 are simultaneously re-recruited (Figure S5D). As links between splicing, 3′ end processing and termination have been described (Albulescu et al., 2012; Martins et al., 2011) an intriguing possibility is that Thr4 coordinates the commitment to post-transcriptional splicing during 3′ end processing and termination. In sum, the phospho-specific Pol II interactomes transform our understanding of critical co-transcriptional gene regulatory mechanisms through disentangling the complex interactions of CTD phosphorylation and regulatory factors.

EXPERIMENTAL PROCEDURES

CTD Immunoprecipitation

Yeast cultures grown to mid-log phase in YPD were flash frozen in liquid nitrogen and pulverized by a ball-bearing mixer mill (Retsch MM400). 1 g of grindate was resuspended in 5 ml of 1 × lysis buffer (20 mM HEPES [pH 7.4], 110 mM KOAc, 0.5 Triton X-100, 0.1% Tween 20, 10 mM MnCl$_2$, 8 U ml^{-1} RNasin [Promega], 1 × cOmplete EDTA free protease inhibitor cocktail [Roche], and 1 × PhosSTOP [Roche]). 660 U of RQ1 DNase (Promega) was added, and the lysate was incubated for 20 min on ice and centrifuged at 20,000 × *g* for 10 min at 4°C. Supernatant was added to 500 μl of anti-FLAG M2 affinity gel (Sigma-Aldrich) and rotated at 4°C for 1 hr. IPs were washed three times for 5 min at 4°C with 10 ml of 1 × wash buffer (20 mM HEPES [pH 7.4], 110 mM KOAc, 0.5 Triton X-100, 0.1% Tween 20, 8 U ml^{-1} RNasin, 1 mM EDTA) and eluted (two times) in 300 μl of 2 mg/ml 3 × FLAG peptide (Sigma-Aldrich). Elutions were combined, and 100 μl of eluate was used for mass spectrometry analysis. For phospho-CTD IPs, 250 μl of the combined elution was added to magnetic beads covalently coupled to 33 μg of phospho-specific CTD antibodies (Active Motif; Ser5P, 3E8 [61085]; Ser7P, 4E12 [61087]; Ser2P, 3E10 [61083]; Tyr1P, 3D12 [61383]; Thr4P, 6D7 [61361]) or 33 μg of mCherry rat monoclonal antibody (Life Technologies, M11217) for mock IPs and incubated at 4°C for 30 min. Samples were then washed three times by pipetting and one time by rotating for 5 min in 1 ml of wash buffer at 4°C and eluted using 0.1 M Gly (pH 2).

NET-Seq and Nascent RNA-Seq

NET-seq growth conditions, IPs, and isolation of nascent RNA and library construction were carried out as described in Churchman and Weissman (2012). NET-seq linker ligation was done directly to the 3′ end of isolated nascent RNA. For nascent RNA-seq, the RNA was fragmented followed by dephosphorylation and linker ligation. Libraries were then subjected to 3′ end sequencing.

RNA-Seq

Total RNA from mid-log yeast cultures was harvested using standard hot phenol-chloroform extraction methods. ERCC RNA standard mix 1 (Life Technologies) was added and rRNA was depleted using the Ribo-Zero Gold rRNA removal kit for yeast (Illumina). Library generation was carried out according to the procedure in Churchman and Weissman (2012) followed by 3′ end sequencing.

ChIP-Nexus

Methods to adapt ChIP-nexus to yeast were obtained through personal communication from Stephen Doris and Fred Winston and will be published elsewhere. Library generation was carried out as described in He et al. (2015).

Mass Spectrometry Analysis

Trichloroacetic acid (TCA) precipitated samples were separated on 12% acrylamide gels, extracted, and submitted to the Taplin Mass Spectrometry Facility at Harvard Medical School for analysis. All mass spectrometry data analysis was done using the Perseus software (Hubner and Mann, 2011; Hubner et al., 2010). Summed MS1 intensities for the triplicate IPs were loaded into Perseus along with the respective mock control dataset. Datasets were log$_2$ transformed and filtered for proteins present in all three specific or mock IPs. Missing values were imputed, and data were normalized by median subtraction. Enriched proteins were defined using a false discovery rate (FDR) of 0.05.

Processing and Alignment of Sequencing Data

NET-seq and RNA-seq reads were trimmed and aligned using TopHat2 and library generation artifacts removed. For NET-seq, the 5′ end of the sequencing, which corresponds to the 3′ end of the nascent RNA fragment, is recorded with a custom Python script using HTSeq package. For RNA-seq and nascent RNA-seq the same script was used, this time recording the entire read. NET-seq and nascent RNA-seq data are normalized by million mapped reads. Positions of splicing intermediates in NET-seq data are removed from analysis. ChIP-nexus reads are selected for the fixed barcode, trimmed, and aligned using Bowtie. Plus and minus strand coverage are combined and the 5′ end of the sequencing read is recorded and normalized by the spike-in, corresponding to the number of mapped *S. pombe* reads.

Gene Expression and Intron Retention Analysis

All RNA-seq reads were normalized using ERCC RNA standards mix 1 (Life Technologies) by the number of 10^4 uniquely mapping ERCC reads. For gene expression analysis the reads per gene were calculated for genes that were expressed in at least one replicate in either the WT or T4V strains. For intron retention analysis, the normalized reads per length in Kb (RPK) was calculated for each intron and exon in non-overlapping spliced genes as well as spliced genes where overlap did not occur on the same strand. To determine which RPK values could be reliably quantified, the concentration of ERCC standards was plotted against the RPK scores for the ERCC standards in each library (Figure S5A). For intron retention analysis, only those genes with intron and exon RPK values greater than 75 in both WT and T4V samples were analyzed.

Metagene Analysis

NET-seq and ChIP-nexus reads from wild-type or mutant strains were scored around the TSS, polyA, and the 3′ splice sites of non-overlapping genes in 1-bp bins using the deepTools program. The TSS and polyA average profiles were calculated using a sliding window of 25 bp for non-overlapping protein coding genes with a Reads per kilobase per million maped reads (RPKM) >10 in the WT NET-seq data and are at least 500 bp long. For average profiles around the 3′ splice site, high-occupancy non-overlapping spliced genes as well as spliced genes where overlap did not occur on the same strand were analyzed.

Real-Time qPCR

Yeast was cultured and RNA extracted as described in RNA-seq methods. For CTD_{1-8} mutants cells were grown in SC-leu media containing 2% glucose and 50 μg/ml doxycycline (Clontech Laboratories). Total RNA was harvested using standard hot phenol-chloroform extraction methods, ethanol precipitated, and DNase treated using RQ1 DNase (Promega). cDNA was then generated from gene-specific RT primers using SuperScript III (Life Sciences). qPCR was done using SsoFast Supermix (BioRad) and carried out on a Bio-Rad CFX384 Real-Time System with a C1000 Theromycylcer.

Antibody Competition Assay

Pol II was immunopurified as described above. Samples were split into five aliquots of 200 μl each. To each sample, 100 μl of wash buffer containing 0 or 20 μg of phospho-CTD antibody or 100 μl of 2 mg/ml FLAG peptide was added. Samples were incubated for 30 min at 4°C followed by centrifugation to collect the beads. The beads (B) and supernatant (S) fractions were then analyzed by western blot.

Statistical Methods

p values for mass spectrometry analysis, antibody competition assays, and qPCR data were calculated using two-sample, two-tailed t tests. FDRs in volcano plot analysis were calculated using a permutation-based FDR for each dataset. p value in Figure S3E was calculated using a paired two-sample t test. p values for GO term enrichment were obtained from GO term finder at the *Saccharomyces* Genome Database website. p values comparing protein domain enrichment were calculated using Fisher's exact test.

ACCESSION NUMBERS

The accession number for the NET-seq, RNA-seq, and ChIP-nexus datasets reported in this paper is GEO: GSE68484.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, six figures, and four tables and can be found with this article online at http://dx.doi.org/10.1016/j.celrep.2016.05.010.

AUTHOR CONTRIBUTIONS

K.M.H. and L.S.C. designed experiments and wrote the manuscript. K.M.H. performed IP experiments, analyzed mass spectrometry data, generated mutant strains, and performed and analyzed ChIP-nexus and qPCR experiments. M.M.M., S.M.F., and K.M.H. designed and synthesized the CTD mutant plasmids. K.M.H., K.L.T., and E.E.S. generated NET-seq libraries; K.L.T. generated RNA-seq libraries and qPCR data. K.M.H. analyzed sequencing data.

ACKNOWLEDGMENTS

We thank S. Buratowski, C. Guthrie, S. Gygi, R. Kingston, F. Winston, and L.S.C.'s laboratory for advice and discussions; S. Buratowski, M. Couvillion, A. Mayer, and F. Winston for their critical reading of the manuscript; U. Eser and J. di Iulio for help with data analysis. We thank the Taplin Mass Spectrometry Facility and The Bauer Core Facility at Harvard University for mass spectrometry assistance and next-generation sequencing. This work was supported by U.S. National Institutes of Health NHGRI grants R01HG007173 to L.S.C., a Damon Runyon Dale F. Frey Award for Breakthrough Scientists (to L.S.C.), and a Burroughs Wellcome Fund Career Award at the Scientific Interface (to L.S.C.). K.M.H. was supported by U.S. National Science Foundation Graduate Research Fellowship DGE1144152.

Received: December 1, 2015
Revised: March 22, 2016
Accepted: April 26, 2016
Published: May 26, 2016

REFERENCES

Albulescu, L.-O.O., Sabet, N., Gudipati, M., Stepankiw, N., Bergman, Z.J., Huffaker, T.C., and Pleiss, J.A. (2012). A quantitative, high-throughput reverse genetic screen reveals novel connections between Pre-mRNA splicing and 5′ and 3′ end transcript determinants. PLoS Genet. *8*, e1002530.

Alexander, R.D., Innocente, S.A., Barrass, J.D., and Beggs, J.D. (2010). Splicing-dependent RNA polymerase pausing in yeast. Mol. Cell *40*, 582–593.

Barillà, D., Lee, B.A., and Proudfoot, N.J. (2001). Cleavage/polyadenylation factor IA associates with the carboxyl-terminal domain of RNA polymerase II in Saccharomyces cerevisiae. Proc. Natl. Acad. Sci. USA *98*, 445–450.

Bentley, D.L. (2014). Coupling mRNA processing with transcription in time and space. Nat. Rev. Genet. *15*, 163–175.

Buratowski, S. (2005). Connections between mRNA 3′ end processing and transcription termination. Curr. Opin. Cell Biol. *17*, 257–261.

Carrillo Oesterreich, F., Preibisch, S., and Neugebauer, K.M. (2010). Global analysis of nascent RNA reveals transcriptional pausing in terminal exons. Mol. Cell *40*, 571–581.

Chathoth, K.T., Barrass, J.D., Webb, S., and Beggs, J.D. (2014). A splicing-dependent transcriptional checkpoint associated with prespliceosome formation. Mol. Cell *53*, 779–790.

Cho, E.J., Takagi, T., Moore, C.R., and Buratowski, S. (1997). mRNA capping enzyme is recruited to the transcription complex by phosphorylation of the RNA polymerase II carboxy-terminal domain. Genes Dev. *11*, 3319–3326.

Churchman, L.S., and Weissman, J.S. (2011). Nascent transcript sequencing visualizes transcription at nucleotide resolution. Nature *469*, 368–373.

Churchman, L.S., and Weissman, J.S. (2012). Native elongating transcript sequencing (NET-seq). Curr. Protoc. Mol. Biol. *4*, 14.1–17.

Descostes, N., Heidemann, M., Spinelli, L., Schüller, R., Maqbool, M.A., Fenouil, R., Koch, F., Innocenti, C., Gut, M., Gut, I., et al. (2014). Tyrosine phosphorylation of RNA polymerase II CTD is associated with antisense promoter transcription and active enhancers in mammalian cells. eLife *3*, e02105.

Egloff, S., O'Reilly, D., Chapman, R.D., Taylor, A., Tanzhaus, K., Pitts, L., Eick, D., and Murphy, S. (2007). Serine-7 of the RNA polymerase II CTD is specifically required for snRNA gene expression. Science *318*, 1777–1779.

Fuchs, S.M., Laribee, R.N., and Strahl, B.D. (2009). Protein modifications in transcription elongation. Biochim. Biophys. Acta *1789*, 26–36.

Görnemann, J., Barrandon, C., Hujer, K., Rutz, B., Rigaut, G., Kotovic, K.M., Faux, C., Neugebauer, K.M., and Séraphin, B. (2011). Cotranscriptional spliceosome assembly and splicing are independent of the Prp40p WW domain. RNA *17*, 2119–2129.

He, Q., Johnston, J., and Zeitlinger, J. (2015). ChIP-nexus enables improved detection of in vivo transcription factor binding footprints. Nat. Biotechnol. *33*, 395–401.

Hintermair, C., Heidemann, M., Koch, F., Descostes, N., Gut, M., Gut, I., Fenouil, R., Ferrier, P., Flatley, A., Kremmer, E., et al. (2012). Threonine-4 of mammalian RNA polymerase II CTD is targeted by Polo-like kinase 3 and required for transcriptional elongation. EMBO J. *31*, 2784–2797.

Hoskins, A.A., Friedman, L.J., Gallagher, S.S., Crawford, D.J., Anderson, E.G., Wombacher, R., Ramirez, N., Cornish, V.W., Gelles, J., and Moore, M.J. (2011). Ordered and dynamic assembly of single spliceosomes. Science *331*, 1289–1295.

Hsin, J.-P.P., Sheth, A., and Manley, J.L. (2011). RNAP II CTD phosphorylated on threonine-4 is required for histone mRNA 3′ end processing. Science *334*, 683–686.

Hsin, J.-P.P., Li, W., Hoque, M., Tian, B., and Manley, J.L. (2014a). RNAP II CTD tyrosine 1 performs diverse functions in vertebrate cells. eLife *3*, e02112.

Hsin, J.-P.P., Xiang, K., and Manley, J.L. (2014b). Function and control of RNA polymerase II C-terminal domain phosphorylation in vertebrate transcription and RNA processing. Mol. Cell. Biol. *34*, 2488–2498.

Hubner, N.C., and Mann, M. (2011). Extracting gene function from protein-protein interactions using Quantitative BAC InteraCtomics (QUBIC). Methods *53*, 453–459.

Hubner, N.C., Bird, A.W., Cox, J., Splettstoesser, B., Bandilla, P., Poser, I., Hyman, A., and Mann, M. (2010). Quantitative proteomics combined with BAC TransgeneOmics reveals in vivo protein interactions. J. Cell Biol. *189*, 739–754.

Jeronimo, C., Bataille, A.R., and Robert, F. (2013). The writers, readers, and functions of the RNA polymerase II C-terminal domain code. Chem. Rev. *113*, 8491–8522.

Jonkers, I., Kwak, H., and Lis, J.T. (2014). Genome-wide dynamics of Pol II elongation and its interplay with promoter proximal pausing, chromatin, and exons. eLife *3*, e02407.

Kim, M., Krogan, N.J., Vasiljeva, L., Rando, O.J., Nedea, E., Greenblatt, J.F., and Buratowski, S. (2004). The yeast Rat1 exonuclease promotes transcription termination by RNA polymerase II. Nature *432*, 517–522.

Komarnitsky, P., Cho, E.-J., and Buratowski, S. (2000). Different phosphorylated forms of RNA polymerase II and associated mRNA processing factors during transcription. Genes Dev. *14*, 2452–2460.

Kotovic, K.M., Lockshon, D., Boric, L., and Neugebauer, K.M. (2003). Cotranscriptional recruitment of the U1 snRNP to intron-containing genes in yeast. Mol. Cell. Biol. *23*, 5768–5779.

Lacadie, S.A., and Rosbash, M. (2005). Cotranscriptional spliceosome assembly dynamics and the role of U1 snRNA:5’ss base pairing in yeast. Mol. Cell *19*, 65–75.

Lacadie, S.A., Tardiff, D.F., Kadener, S., and Rosbash, M. (2006). In vivo commitment to yeast cotranscriptional splicing is sensitive to transcription elongation mutants. Genes Dev. *20*, 2055–2066.

Licatalosi, D.D., Geiger, G., Minet, M., Schroeder, S., Cilli, K., McNeil, J.B., and Bentley, D.L. (2002). Functional interaction of yeast pre-mRNA 3′ end processing factors with RNA polymerase II. Mol. Cell *9*, 1101–1111.

Malagon, F., Kireeva, M.L., Shafer, B.K., Lubkowska, L., Kashlev, M., and Strathern, J.N. (2006). Mutations in the Saccharomyces cerevisiae RPB1 gene conferring hypersensitivity to 6-azauracil. Genetics *172*, 2201–2209.

Martins, S.B., Rino, J., Carvalho, T., Carvalho, C., Yoshida, M., Klose, J.M., de Almeida, S.F., and Carmo-Fonseca, M. (2011). Spliceosome assembly is coupled to RNA polymerase II dynamics at the 3′ end of human genes. Nat. Struct. Mol. Biol. *18*, 1115–1123.

Mayer, A., Lidschreiber, M., Siebert, M., Leike, K., Söding, J., and Cramer, P. (2010). Uniform transitions of the general RNA polymerase II transcription complex. Nat. Struct. Mol. Biol. *17*, 1272–1278.

Mayer, A., Heidemann, M., Lidschreiber, M., Schreieck, A., Sun, M., Hintermair, C., Kremmer, E., Eick, D., and Cramer, P. (2012). CTD tyrosine phosphorylation impairs termination factor recruitment to RNA polymerase II. Science *336*, 1723–1725.

Mayer, A., di Iulio, J., Maleri, S., Eser, U., Vierstra, J., Reynolds, A., Sandstrom, R., Stamatoyannopoulos, J.A., and Churchman, L.S. (2015). Native elongating transcript sequencing reveals human transcriptional activity at nucleotide resolution. Cell *161*, 541–554.

McCracken, S., Fong, N., Rosonina, E., Yankulov, K., Brothers, G., Siderovski, D., Hessel, A., Foster, S., Shuman, S., and Bentley, D.L. (1997). 5′-Capping enzymes are targeted to pre-mRNA by binding to the phosphorylated carboxy-terminal domain of RNA polymerase II. Genes Dev. *11*, 3306–3318.

Moore, M.J., Schwartzfarb, E.M., Silver, P.A., and Yu, M.C. (2006). Differential recruitment of the splicing machinery during transcription predicts genome-wide patterns of mRNA splicing. Mol. Cell *24*, 903–915.

Morrill, S.A., Exner, A.E., Babokhov, M., Reinfeld, B.I., and Fuchs, S.M. (2016). DNA instability maintains the repeat length of the yeast RNA polymerase II C-terminal domain. J. Biol. Chem., Published online March 29, 2016.

Morris, D.P., and Greenleaf, A.L. (2000). The splicing factor, Prp40, binds the phosphorylated carboxyl-terminal domain of RNA polymerase II. J. Biol. Chem. *275*, 39935–39943.

Mosley, A.L., Sardiu, M.E., Pattenden, S.G., Workman, J.L., Florens, L., and Washburn, M.P. (2011). Highly reproducible label free quantitative proteomic analysis of RNA polymerase complexes. Mol. Cell Proteomics *10*, M110.000687.

Mosley, A.L., Hunter, G.O., Sardiu, M.E., Smolle, M., Workman, J.L., Florens, L., and Washburn, M.P. (2013). Quantitative proteomics demonstrates that the RNA polymerase II subunits Rpb4 and Rpb7 dissociate during transcriptional elongation. Mol. Cell. Proteomics *12*, 1530–1538.

Rosonina, E., Yurko, N., Li, W., Hoque, M., Tian, B., and Manley, J.L. (2014). Threonine-4 of the budding yeast RNAP II CTD couples transcription with Htz1-mediated chromatin remodeling. Proc. Natl. Acad. Sci. USA *111*, 11924–11931.

Schroeder, S.C., Schwer, B., Shuman, S., and Bentley, D. (2000). Dynamic association of capping enzymes with transcribing RNA polymerase II. Genes Dev. *14*, 2435–2440.

Schüller, R., Forné, I., Straub, T., Schreieck, A., Texier, Y., Shah, N., Decker, T.M., Cramer, P., Imhof, A., and Eick, D. (2016). Heptad-specific phosphorylation of RNA polymerase II CTD. Mol. Cell *61*, 305–314.

Schwer, B., and Shuman, S. (2011). Deciphering the RNA polymerase II CTD code in fission yeast. Mol. Cell *43*, 311–318.

Suh, H., Ficarro, S.B., Kang, U.B., Chun, Y., Marto, J.A., and Buratowski, S. (2016). Direct analysis of phosphorylation sites on the Rpb1 C-terminal domain of RNA polymerase II. Mol. Cell *61*, 297–304.

Tardiff, D.F., Lacadie, S.A., and Rosbash, M. (2006). A genome-wide analysis indicates that yeast pre-mRNA splicing is predominantly posttranscriptional. Mol. Cell *24*, 917–929.

Tardiff, D.F., Abruzzi, K.C., and Rosbash, M. (2007). Protein characterization of Saccharomyces cerevisiae RNA polymerase II after in vivo cross-linking. Proc. Natl. Acad. Sci. USA *104*, 19948–19953.

Veloso, A., Kirkconnell, K.S., Magnuson, B., Biewen, B., Paulsen, M.T., Wilson, T.E., and Ljungman, M. (2014). Rate of elongation by RNA polymerase II is associated with specific gene features and epigenetic modifications. Genome Res. *24*, 896–905.

Cell Reports
Resource

OPEN
ACCESS
CellPress

Macrophage Ontogeny Underlies Differences in Tumor-Specific Education in Brain Malignancies

Robert L. Bowman,[1,2] Florian Klemm,[1,3,4] Leila Akkari,[1,3,4] Stephanie M. Pyonteck,[1] Lisa Sevenich,[1] Daniela F. Quail,[1] Surajit Dhara,[5] Kenishana Simpson,[1] Eric E. Gardner,[6] Christine A. Iacobuzio-Donahue,[5,7] Cameron W. Brennan,[8] Viviane Tabar,[8] Philip H. Gutin,[8] and Johanna A. Joyce[1,3,4,9,*]
[1]Cancer Biology and Genetics Program
[2]Gerstner Sloan Kettering Graduate School of Biomedical Science
Memorial Sloan Kettering Cancer Center, New York, NY 10065, USA
[3]Department of Oncology
[4]Ludwig Institute for Cancer Research
University of Lausanne, 1066 Lausanne, Switzerland
[5]Human Oncology and Pathogenesis Program
[6]Molecular Pharmacology Program
[7]Department of Pathology
[8]Department of Neurosurgery
Memorial Sloan Kettering Cancer Center, New York, NY 10065, USA
[9]Lead Contact
*Correspondence: johanna@joycelab.org
http://dx.doi.org/10.1016/j.celrep.2016.10.052

SUMMARY

Extensive transcriptional and ontogenetic diversity exists among normal tissue-resident macrophages, with unique transcriptional profiles endowing the cells with tissue-specific functions. However, it is unknown whether the origins of different macrophage populations affect their roles in malignancy. Given potential artifacts associated with irradiation-based lineage tracing, it remains unclear if bone-marrow-derived macrophages (BMDMs) are present in tumors of the brain, a tissue with no homeostatic involvement of BMDMs. Here, we employed multiple models of murine brain malignancy and genetic lineage tracing to demonstrate that BMDMs are abundant in primary and metastatic brain tumors. Our data indicate that distinct transcriptional networks in brain-resident microglia and recruited BMDMs are associated with tumor-mediated education yet are also influenced by chromatin landscapes established before tumor initiation. Furthermore, we demonstrate that microglia specifically repress *Itga4* (CD49D), enabling its utility as a discriminatory marker between microglia and BMDMs in primary and metastatic disease in mouse and human.

INTRODUCTION

Macrophages are terminally differentiated cells of the myeloid lineage, with critical functions in tissue development and homeostasis (Okabe and Medzhitov, 2016). These cells serve as a nexus between adaptive and innate immunity, regulating responses to inflammation and wound healing (Mosser and Edwards, 2008). To facilitate these diverse functions, macrophages employ considerable plasticity in response to a range of cytokines. These responses fall within a spectrum of different phenotypes ranging from classically activated pro-inflammatory macrophages to alternatively activated anti-inflammatory macrophages (Xue et al., 2014). Macrophages also possess substantial diversity and plasticity, with recent studies revealing important insights into the developmental origins of tissue-resident macrophages and uncovering tissue-specific gene expression patterns and enhancer landscapes (Gautier et al., 2012; Ginhoux et al., 2010; Gomez Perdiguero et al., 2015; Lavin et al., 2014; Mass et al., 2016).

While the local tissue environment sculpts macrophage transcriptional profiles and epigenetic states in homeostasis (Lavin et al., 2014), it is unknown whether an inflammatory tissue environment may promote differences between macrophage populations of distinct ontogenies. This is particularly relevant in cancer, where tumor-associated macrophages (TAMs) are derived from monocytes and also potentially from tissue-resident macrophages (Du et al., 2008; Pyonteck et al., 2013; Solga et al., 2015).

Brain-resident macrophages, microglia (MG), develop from erythromyeloid precursors in the yolk sac (Gomez Perdiguero et al., 2015; Kierdorf et al., 2013a; Schulz et al., 2012). Unlike other tissue-resident macrophages, during homeostasis, MG undergo self-renewal and their pool is not replenished by monocytes (Ajami et al., 2007). Microglia are also resistant to myeloablative irradiation (Kennedy and Abkowitz, 1997). Indeed, this property has been used extensively in bone marrow transplantation (BMT) models to distinguish radio-resistant MG from BM-derived macrophages (BMDMs) (Huang et al., 2014; Sedgwick et al., 1991). However, only under conditions of blood-brain

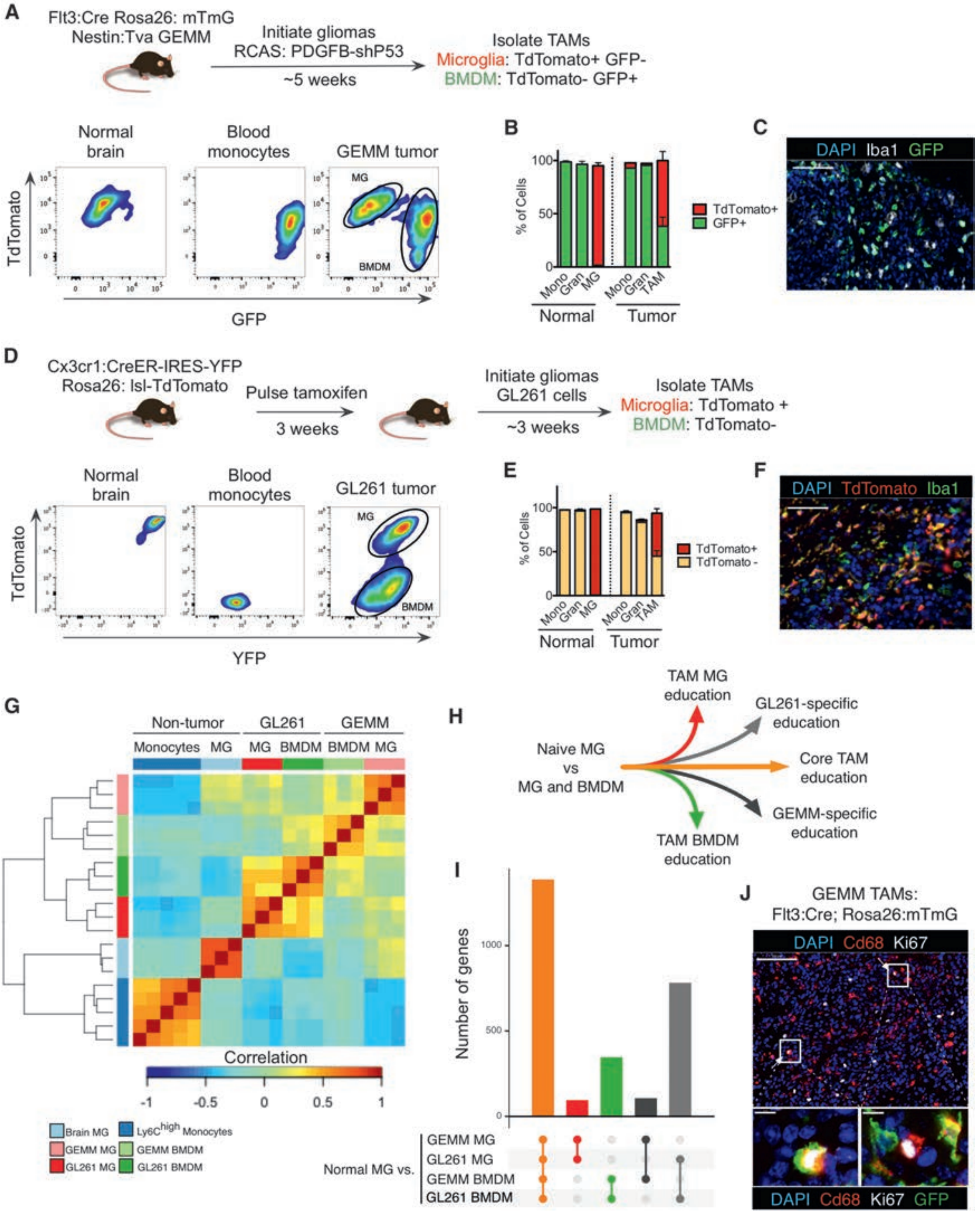


Figure 1. Lineage Tracing Systems Demonstrate Heterogeneity in TAM Ontogeny in Multiple Models of Glioma

(A) Experimental scheme for the GEMM-shP53 model (see Supplemental Experimental Procedures for details). Representative flow cytometry panels for TdTomato and GFP are shown for Cd45$^+$Cd11b$^+$Ly6G$^-$Ly6C$^-$ microglia (MG), Cd45$^+$Cd11b$^+$Ly6G$^-$Ly6C$^+$ monocytes, and Cd45$^+$Cd11b$^+$Ly6G$^-$Ly6C$^-$ TAMs from GEMM-shP53 gliomas.

(B) Quantitation of TdTomato$^+$ and GFP$^+$ monocytes (Mono) and granulocytes (Gran) in peripheral blood, MG in non-tumor-bearing brain, and monocytes, granulocytes, and TAMs in GEMM-shP53 gliomas as depicted in (A). Bars represent mean and SEM (n = 3–5 for each group).

(C) Representative immunofluorescence (IF) staining of Iba1 (white), GFP (green), and DAPI (blue) in a GEMM-shP53 tumor as depicted in (A). Scale bar, 50 μm. Data are representative of n = 5 tumors.

(D) Experimental design for Cx3cr1 lineage-tracing model (see Experimental Procedures for details). Monocytes, MG, and TAMs were isolated as described in (A) and evaluated for TdTomato and YFP reporter expression. Data are representative of n = 3 mice.

(legend continued on next page)

barrier (BBB) disruption (e.g., via irradiation [IR] or chemical manipulation) does there appear to be a significant contribution of BMDMs to the brain macrophage pool in a non-pathological context (Bruttger et al., 2015; Mildner et al., 2007). This is relevant to brain tumors such as gliomas, where there is also disruption of the BBB with disease progression (Dubois et al., 2014). IR-BMT has shown BMDM abundance in murine CNS cancers (Biffi et al., 2004; De Palma et al., 2005; Huang et al., 2014; Müller et al., 2015; Pyonteck et al., 2013); however, given the current lack of markers definitively distinguishing MG and BMDMs, it remains unclear if BMDM recruitment indeed occurs in brain tumors in the absence of irradiation. The need for markers distinguishing these cells is especially critical in human disease, where lineage tracing is not possible.

Here, we utilize multiple genetic lineage tracing models to demonstrate that BMDMs are indeed present in murine brain tumors. Gene expression profiling showed that while BMDMs and MG share features of tumor education, they also exhibit distinct activation modes. Our data suggest these faculties are a result of inherent transcriptional networks poised before the onset of tumorigenesis, where ontogeny pre-biases cells to engage in distinct macrophage activation states. Lastly, we identify markers that distinguish MG and peripherally derived macrophages under homeostasis, as well as in glioma and brain metastasis in both mice and humans.

RESULTS

Tumor-Associated BMDMs Are Present in Mouse Glioma Models

To track the ontogeny of myeloid cells in murine gliomas, we utilized a hematopoietic lineage tracing system, Flt3:Cre; Rosa26:mTmG, which has been used to show that peripheral myeloid cells develop from Flt3^{+} short-term hematopoietic stem cells (ST-HSCs) and are GFP^{+}, while parenchymal MG develop independently of ST-HSC precursors and are thus negative for the GFP reporter, remaining TdTomato^{+} (Boyer et al., 2011; Gomez Perdiguero et al., 2015). In non-tumor-bearing mice, >98% of blood monocytes were GFP^{+}, and <1% of MG showed recombination for the mTmG reporter (Figure 1A). The spleen was composed of GFP^{+} lymphocyte-rich follicles, surrounded by TdTomato^{+} stromal cells, while the brain parenchyma did not contain any detectable GFP^{+} cells (Figure S1A).

We next bred this line to the nestin:Tva (nTva) line to trace myeloid cell ontogeny in a genetically engineered mouse model (GEMM) of glioma. We induced gliomas by intracranial injection of DF1 cells transfected with RCAS vectors encoding platelet-derived growth factor β (PDGFB) and a short hairpin against P53 (Ozawa et al., 2014) (Figure 1A), termed GEMM-shP53 herein. Flow cytometry of end-stage gliomas demonstrated that all monocytes (Cd45^{+}Cd11b^{+}Ly6C^{high}Ly6G^{-}) and granulocytes (Cd45^{+}Cd11b^{+}Ly6C^{low}Ly6G^{high}) in the tumor were GFP^{+} (Figure 1B), while the bulk TAM compartment (Cd45^{+}Cd11b^{+}Ly6C^{-}Ly6G^{-}) was composed of both GFP^{+} TAM BMDMs and GFP^{-} TAM MG (Figures 1A, 1B, and S1B), confirmed by tissue immunofluorescence (IF) co-staining with the pan-macrophage marker Iba1 (Figure 1C). By contrast, the contralateral, non-malignant brain contained only GFP^{-} MG, demonstrating the specific abundance of TAM BMDMs only within the tumor mass (Figure 1B).

We and others have utilized IR-BMT to show that TAM BMDMs are recruited to murine gliomas (Huang et al., 2014; Pyonteck et al., 2013). However, IR can lead to ectopic recruitment of BMDMs to the brain and thereby increase their relative abundance (Müller et al., 2015). We verified these findings in the orthotopic, syngeneic GL261 glioma model and found the TAM compartment was composed of both TAM MG and TAM BMDMs using both IR-BMT lineage tracing and IR-independent Flt3:Cre lineage tracing (Figures S1C and S1D). TAM BMDM abundance was significantly increased in the IR-BMT model compared to the Flt3:Cre model (Figure S1D), reinforcing previous reports that IR-BMT can skew the ratio of MG and BMDMs. Critically, however, using Flt3:Cre lineage tracing, we found that BMDMs composed >35% of the bulk TAM population in gliomas without IR preconditioning, demonstrating that BMDM infiltration into tumors is not solely an artifact of IR (Figure S1D).

To exclude the possibility that this finding was due to a subset of TAM MG spontaneously upregulating *Flt3* expression, we utilized a complementary lineage-tracing approach previously indicated to be specific for MG in the normal brain: Cx3cr1:CreER-IRES YFP; Rosa26:lsl-TdTomato (see Supplemental Experimental Procedures for details) (Parkhurst et al., 2013). 3 days after tamoxifen-induced labeling, >99% of MG and circulating monocytes were TdTomato^{+} (Figure S1E). However, after 3 weeks, blood monocytes no longer retained the TdTomato^{+} reporter, indicating their turnover and replenishment by tamoxifen-“naive” monocytes (Figure S1E). By contrast, >99% of MG remained TdTomato^{+} (Figure S1E). We induced GL261 tumors in these mice, at 7 weeks of age, and observed both TdTomato^{+} TAM MG and TdTomato^{-} TAM BMDMs (Figures 1D and 1E). Meanwhile, all monocytes and granulocytes were TdTomato^{-} in the tumor and periphery (Figure 1E). These findings were substantiated by IF co-staining of tissue

---

(E) Flow cytometry quantitation of TdTomato^{+} and TdTomato^{-} monocytes and granulocytes in peripheral blood, MG in non-tumor-bearing brain, and monocytes, granulocytes, and TAMs in GL261 gliomas as depicted in (D). Bars represent mean and SEM (n = 3 for each group).

(F) Representative IF staining for Iba1 (green), TdTomato (red), and DAPI (blue) in a GL261 tumor. Scale bar, 50 μm.

(G) Pairwise correlation matrix of normalized RNA-seq counts from monocytes (n = 5), normal MG (n = 3), and the four TAM populations from the different models: GEMM-shP53 TAM MG, GEMM-shP53 TAM BMDMs, GL261 TAM MG, and GL261 TAM BMDMs (n = 3 for each group).

(H) Diagram depicting different modules of TAM education compared to normal MG.

(I) Differentially expressed genes between normal MG and the four TAM populations were tabulated. Bar chart depicts the number of differentially expressed genes shared between the different groups.

(J) Representative IF staining of Ki67^{+} TAM BMDMs and TAM MG in the GEMM-shP53 model as depicted in (A) (Ki67, white; CD68, red; DAPI, blue; and GFP, green; omitted from top panel). Scale bars represent 100 μm (top panel) and 10 μm (lower panels). Data are representative of n = 5 tumors.

sections with Iba1 (Figure 1F). Importantly, there was a gradient of eYFP reporter expression levels, with highest expression in $TdTomato^+$ TAM MG, slightly lower levels in $TdTomato^-$ TAM BMDMs, and lowest levels in monocytes (Figure S1F), demonstrating the capacity of TAM BMDMs to express *Cx3cr1* in brain tumors. Thus, *Cx3cr1* expression alone cannot be used to strictly identify MG in gliomas. Together, these complementary genetic lineage-tracing models show that BMDMs contribute to the TAM pool in several murine models of glioma, in the absence of IR.

RNA Sequencing Reveals Multimodal Patterns of TAM Education

We next analyzed the transcriptional profiles of TAM MG and TAM BMDMs in gliomas. We performed RNA-sequencing (RNA-seq) on sorted populations of TAM MG and TAM BMDM from GEMM-shP53 and GL261 tumors using the Flt3-based and Cx3cr1-based lineage tracing systems, respectively. We also collected MG and $Ly6C^{high}$ blood monocytes from non-tumor-bearing Flt3:Cre Rosa26:mTmG mice. Global correlation analyses revealed distinct clustering of all TAM populations from normal MG and monocytes, with further cell-type-specific and tumor-specific clustering (Figure 1G). As expected, monocytes were enriched for *Ly6c2* expression, while both TAM MG and TAM BMDMs expressed higher levels of macrophage differentiation markers (e.g., *Aif1* and *Mertk)* than monocytes (Figure S1G). Normal MG and TAM MG expressed higher levels of MG-enriched genes *(*e.g., *Cx3cr1*, *P2ry12*, and *Tmem119*) than monocytes and TAM BMDMs (Figure S1G).

We next delineated cell-type-specific, tumor-specific, and conserved patterns of tumor education among TAMs (Figure 1H). We identified differentially expressed genes between each TAM population from GEMM-shP53 and GL261 tumors compared to normal MG and monocytes (Figures 1I and S1H; Table S1A). Using normal MG as the reference, we found 91 genes specifically upregulated in TAM MG from both GEMM-shP53 and GL261 models (Figure 1I, red bar), and 342 genes upregulated in TAM BMDMs from both GEMM-shP53 and GL261 models (Figure 1I, green bar). We also identified genes that were specifically upregulated in TAM MG and TAM BMDMs from either the GEMM-shP53 ($n = 102$) or GL261 ($n = 778$) models. The largest gene set ($n = 1383$) was significantly upregulated in all TAM populations compared to normal MG (Figure 1I, orange bar). Similar patterns of expression were observed when monocytes were used as the reference population (Figure S1H; Table S1B).

Many cell-cycle-related genes were upregulated, suggesting increased TAM proliferation compared to normal MG and monocytes (Tables S1A and S1B). Indeed, we found $Ki67^+$ cells in both $Iba1^+GFP^+$ TAM BMDM and $Iba1^+GFP^-$ TAM MG in the Flt3-based lineage-tracing model (Figure 1J). Conserved upregulation of complement-related factors, extracellular matrix components, proteases, lipid metabolism mediators, and clotting factors were also evident in both TAM populations (Table S1A). In addition to these programmatic changes, compared to normal MG, there was upregulation of growth factors (*Igf1*, *Areg*, and *Osm)*, chemokines and cytokines (*Spp1*, *Ccl5*, *Cxcl9*, and *Cxcl10)*, and other immune modulators, including *Cd274*/PD-L1 and major histocompatibility complex (MHC class) I molecules (*H2-K1*, *H2-D1*, and *B2m*) (Table S1A). A similar distribution of differentially expressed genes was evident in comparing the TAM populations from both glioma models to blood monocytes (Figure S1H; Table S1B). Interestingly, we found several MG-enriched genes (e.g., *Tmem119*, *Olfml3*, *Lag3*, *Jam2*, and *Sparc*) (Gautier et al., 2012) enriched in TAM BMDMs in both GL261 and GEMM-shP53 models compared to monocytes (Table S1B). Despite this difference, there was still higher expression of MG-related genes in normal MG and TAM MG than in TAM BMDMs. Meanwhile, other MG-enriched genes showed no such induction in TAM BMDMs (*P2ry12, Sall1*, and *Mef2c*). Collectively, these data are consistent with *Cx3cr1* upregulation specifically in gliomas (Figure S1F) and the notion that macrophages acquire tissue-resident gene expression upon infiltration into a foreign tissue (Gosselin et al., 2014; Lavin et al., 2014).

TAM BMDMs and TAM MG Possess Distinct Education Patterns

We investigated transcriptional differences between TAMs derived from BMDMs versus MG and identified 378 differentially expressed genes enriched in TAM MG compared to TAM BMDMs in both GEMM-shP53 and GL261 models and 485 genes enriched in TAM BMDMs compared to TAM MG (Figure 2A; Table S2). As expected, among the 378 TAM MG genes, we found markers previously shown to be enriched in MG compared to other macrophage populations, including *P2ry12*, *Tmem119*, *Slc2a5*, *Pros1*, and *Sall1* (Figure 2A) (Gautier et al., 2012). Consistent with their tissue-specific functions, we found that normal MG and TAM MG were enriched for *Jam2* and *Ocln* (Figure S2A; Table S2), integral components of the blood-brain barrier (Liu et al., 2012). Similarly, TAM MG expressed higher levels of classical complement factors *C4b*, *C2*, and *Cfh* (Figure S2A), a pathway important for MG function in synaptic pruning and host defense (Stephan et al., 2012).

Meanwhile, TAM BMDMs expressed high levels of alternative complement cascade components *Cfb* and *Cfp* (Figure S2A) and enrichment of many immune effectors, including *Cd40*, *Jak2*, *Ifitm1*, *Ifitm2*, *Tlr11*, *Tlr5*, *Tlr8*, *Mefv*, and *Fas* (Figure S2A). In the GEMM-shP53 model, interleukin 1 (IL-1) pathway ligands were differentially expressed, with *Il1a* enriched in TAM MG and *Il1b* in TAM BMDMs, and similar trends were observed in the GL261 model (Figure S2B). While *Il1r1* levels did not significantly differ, TAM BMDMs expressed higher levels of the IL-1 signaling antagonist *Il1rn*, and the IL-1 decoy receptor *Il1r2* (Figure S2B). These results complement reports in non-cancer contexts demonstrating *Il1a* enrichment in MG compared to BMDMs, where IL-1 signaling played a critical role in MG repopulation and maintenance (Bruttger et al., 2015).

We next interrogated chemokines, growth factors, and immune modulators associated with different macrophage activation states. In addition to model-specific gene expression changes (Figure S2C; Table S3), we found in both GEMM and GL261 models that TAM BMDMs were enriched for chemokines involved in wound healing, including *Ccl22*, *Ccl17*, *Cxcl2*, *Cxcl3*, and *Cxcl16* (Figure 2B) (Xue et al., 2014). Interestingly, TAM MG were enriched for expression of *Ccl4* and *Tnf*, chemokines associated with a pro-inflammatory response (Xue et al., 2014). This difference in activation states was supported by a programmatic increase in antigen presentation centered on increased expression

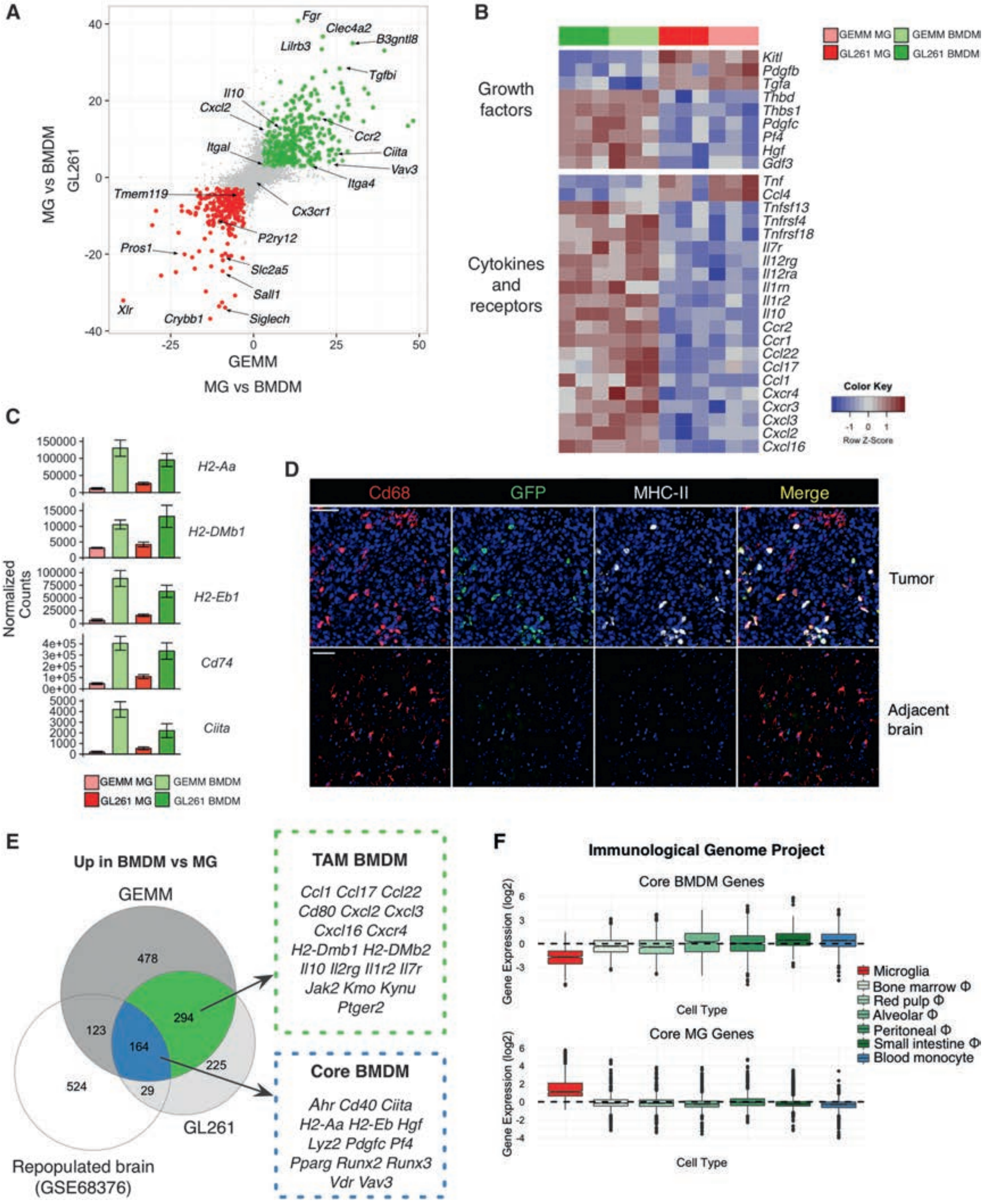


Figure 2. TAM BMDMs and TAM MG Possess Distinct Gene Expression Patterns

(A) Scatterplot depicting −log10 (p value) × sign (fold change) between TAM BMDMs and TAM MG in GEMM-shP53 gliomas (x axis) and GL261 gliomas (y axis). Significantly upregulated genes (log2 fold change of more than ±1 and FDR < 1%) are in green for BMDM and red for MG.

(B) Heatmap depicting row-normalized log2 gene expression values for indicated genes in GL261 TAM BMDMs (dark green), GEMM-shP53 TAM BMDMs (light green), GL261 TAM MG (dark red), and GEMM-shP53 TAM MG (light red).

(C) Bar plots depicting normalized gene expression values for indicated genes in these four different TAM populations. Bars represent mean ± SEM.

(D) Representative IF staining in GEMM-shP53 Flt3:Cre Rosa26:mTmG gliomas and adjacent normal brain for Cd68 (red, Alexa Fluor 594), GFP (green), and MHC II (white). DAPI is shown in blue, and TdTomato fluorescence is not shown. Scale bar, 100 μm. Data are representative of n = 5 tumors.

(E) Venn diagram depicting significantly upregulated genes in BMDMs versus MG in GL261 model, GEMM-shP53, and non-malignant brain (GSE68376 dataset). Select genes are listed.

(F) Boxplot of core BMDM genes (Figure 2E) and core MG genes (Figure S2D), where each data point represents the *Z* scored expression of a gene across the indicated cell populations using available datasets from the Immunological Genome Project.

of the MHC II master regulator *Ciita* (Reith et al., 2005) and its transcriptional targets *H2-Aa*, *H2-DMb1*, *H2-Eb1*, and *Cd74* in TAM BMDMs (Figure 2C). IF staining in Flt3:Cre; GEMM-shP53 tumors revealed a marked increase in MHC II in tumors, compared to adjacent brain, restricted to GFP$^+$ TAM BMDMs (Figure 2D). In addition to this antigen-presentation program, costimulatory molecules such as *Cd80*, *Cd40*, and *Cd200r4* were increased (Figure S2A). These findings were further complemented by TAM BMDM-enriched expression of the Aryl-hydrocarbon receptor (*Ahr*), a transcription factor previously shown to mediate immune suppression (Murray et al., 2014; Opitz et al., 2011) (Figure S2A). Critically, we also found that the immunosuppressive cytokine *Il10* was enriched in TAM BMDMs compared to TAM MG (Figure 2B). Collectively, these results suggest that TAM BMDMs engage in a chronic wound-healing-like state reminiscent of an alternatively activated macrophage (Mosser and Edwards, 2008). Similar phenotypes have been shown in models of oligodendrocyte cell death, where, despite high MHC II expression, myeloid cells did not activate a robust T cell response (Locatelli et al., 2012), suggestive of a tolerogenic program.

We next asked if the differences in inflammatory mediators were an inherent feature of BMDMs upon entry into the brain or rather a consequence of tumor education. Previous studies demonstrated that when MG are depleted and the brain preconditioned by IR, BMDMs can seed the brain and contribute significantly to the brain macrophage pool (hereafter termed “ectopic BMDM” in a normal “repopulated brain”) (Bruttger et al., 2015). We used this dataset for comparative analyses with our TAM BMDMs and TAM MG RNA-seq data to discriminate tumor education differences from ontogenetic, non-tumor-associated differences. This juxtaposition allowed us to identify genes enriched in TAM MG versus TAM BMDMs as well as normal MG versus “ectopic” BMDMs. These “core” MG genes included not only known MG markers such as *Jam2*, *Siglech*, and *P2ry12* but also complement factors *C2*, *C4b*, and *Cfh* as well as the pro-inflammatory cytokines *Ccl4* and *Tnf* (Figure S2D, n = 245 genes; Table S4). We identified genes enriched in TAM BMDMs (n = 294) specifically in the context of a tumor, including *Il10*, *Cxcl2*, *Cxcl3*, *Ccl17*, *Ccl22*, and *H2-Dmb1* (Figure 2E). In contrast to this tumor-specific expression profile, there were also 164 core BMDM genes enriched in BMDM compared to MG, regardless of the presence or absence of a tumor, including *Ciita*, *Ahr*, *Runx2*, *Runx3*, *Vav3*, and *Vdr* (Figure 2E; Table S4). These data indicate some features distinguishing TAM BMDMs and TAM MG are inherent to their differential ontogenies, while others are only acquired upon interaction with, and education by, the tumor microenvironment.

As many of the core BMDM genes are central players in innate immunity, we queried the immunological genome project database to determine if these genes were over-represented in any particular myeloid cell population. Interestingly, we found that these genes were actually repressed in MG compared to tissue-resident macrophages of the BM, spleen, lung, peritoneum, small intestine, and monocyte progenitors (Figure 2F). Meanwhile, core MG genes were indeed enriched in MG compared to other myeloid cells (Figure 2F). These data suggest that core BMDM genes are not specifically enriched in TAM BMDMs or macrophages in general but are specifically repressed in MG.

Recent studies have highlighted extensive epigenetic diversity among tissue-resident macrophages (Lavin et al., 2014); thus, we hypothesized that the MG-repressed genes may be epigenetically altered in MG compared to even the distantly related monocytes. Indeed, when we analyzed these published datasets, we observed increased H3K27 acetylation in the promoters of normal monocytes compared to normal MG for the core BMDM genes (Figure S2E). Similarly, there was increased H3K27 acetylation in the promoters of core MG genes in MG compared to monocytes (Figure S2E). Enhancer specification and epigenetic states in MG and other macrophage populations have been associated with differential PU.1 occupancy. Interrogating previously published data (Gosselin et al., 2014), we observed that several macrophage subsets (including BMDM) all showed increased PU.1 binding at the promoters of our core BMDM genes compared to normal MG (Figure S2F). Meanwhile, variability in PU.1 occupancy was minimal at the promoters of core MG genes (Figure S2F). Similar binding dynamics were evident in enhancer elements, where PU.1 occupancy in enhancer regions of core BMDM genes was higher in BMDMs than MG, with less pronounced differences present in core MG genes (Figure S2G). Thus, epigenetic landscapes established before the development of a tumor may play a role in regulating differential activation patterns subsequently observed in malignancy.

Identification of Transcription Factor Networks Underlying TAM Activation

Given the epigenetic differences in the non-malignant setting, we next determined if chromatin states also differed between TAM BMDMs and TAM MG. We performed assay for transposase-accessible chromatin sequencing (ATAC-seq) (Buenrostro et al., 2013) to assess chromatin accessibility in TAM BMDMs and TAM MG sorted from the GL261 model (Figure 1D). We found the ATAC-seq signal was associated with cell-type-specific gene expression. In TAM BMDMs, the promoters of core BMDM and TAM BMDM genes had higher ATAC-seq signal than core MG and TAM MG genes, while TAM MG promoters of core MG and TAM MG genes had a higher ATAC-seq signal than core BMDM and TAM BMDM genes (Figure S3A). Within enhancers and intronic elements of these gene sets, we identified 120 BMDM-specific peaks in TAM BMDM genes, including *Vav3*, and 704 MG-specific peaks in TAM MG genes, including *P2ry12* and *Sall1* (Figures 3A, S3B, and S3C; Table S5A).

We analyzed the transcription factor (TF) landscape underlying these different peaks and performed de novo motif analysis (motifs are shown in all capital letters). Motif analysis of these peaks revealed an enrichment of FOS/JUN and PU.1 binding sites in both TAM BMDM and TAM MG peaks (Figure 3B; Table S5B), reinforcing previous analyses demonstrating the critical role of PU.1 in establishing specific enhancer landscapes in tissue-resident macrophages (Gosselin et al., 2014). Besides these shared enrichments, we found TAM BMDM peaks enriched for RUNX and CREB/bZIP motifs, while TAM MG peaks were enriched for SMAD3 and MEF2A motifs (Figure 3B).

To determine if these motifs reflected pathway activation of particular TFs we modeled the expression of their predicted downstream targets (see Supplemental Experimental Procedures). We identified TF families with enriched activity in TAM

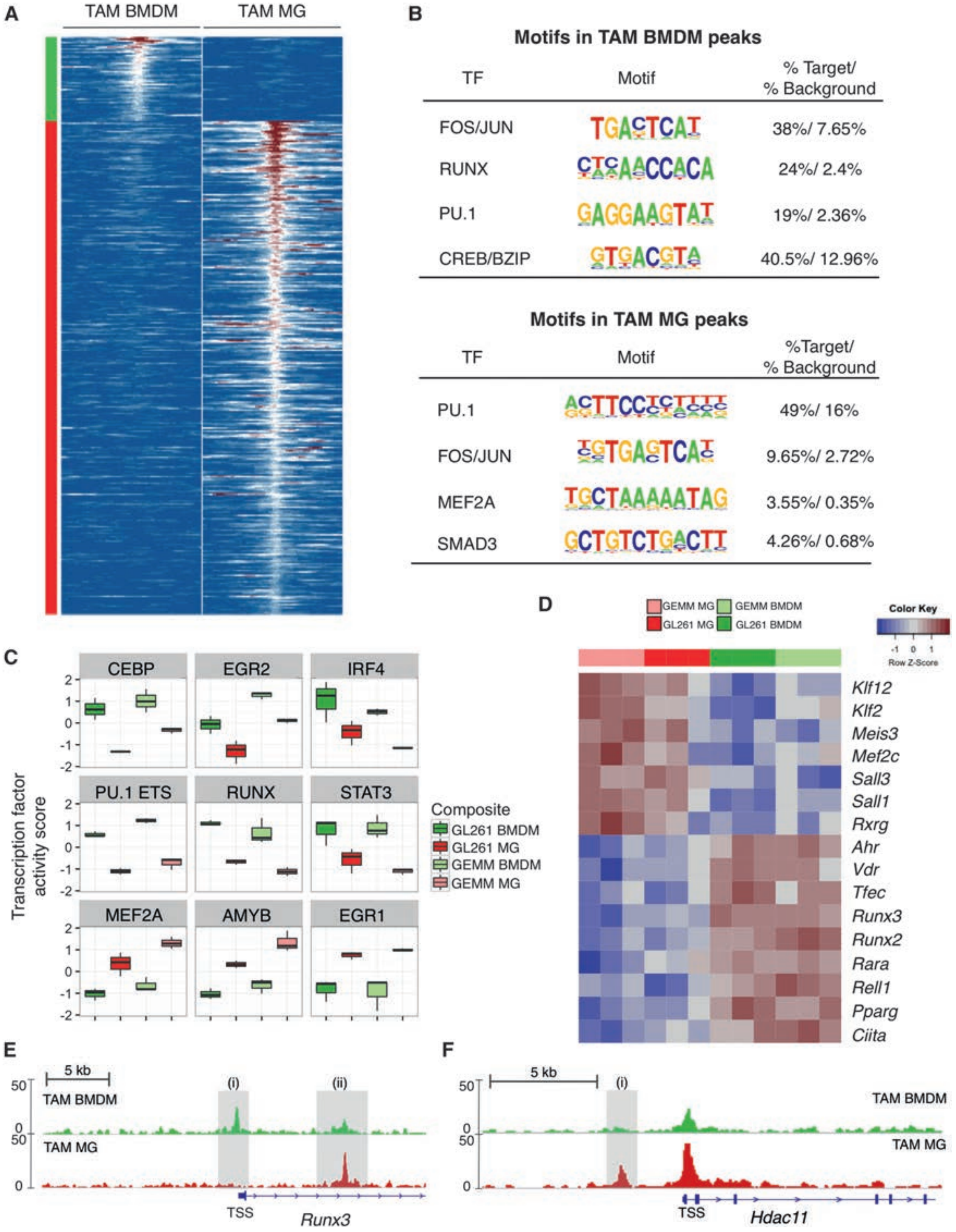


Figure 3. Cell-Specific Transcription Factor Activities Underlie Differences between TAM BMDMs and TAM MG

(A) Heatmap depicting ATAC-seq signal 1 kb upstream and downstream of peaks specifically enriched in GL261 TAM BMDMs (left) and GL261 TAM MG (right). Peaks were selected based on association with differentially expressed genes between TAM BMDMs (top, green) and TAM MG (bottom, red).

(B) Motifs identified by HOMER to be enriched in TAM BMDM and TAM MG peaks shown in (A).

(C) Boxplots depicting normalized TF activity scores for indicated motifs across TAM BMDMs and TAM MG from GL261 and GEMM-shP53 gliomas.

(D) Heatmap depicting row-normalized log2 gene expression values for indicated genes in four different TAM populations.

(E and F) ATAC-sequencing tracks from TAM BMDMs (top, green) and TAM MG (bottom, red) from GL261 gliomas for (E) *Runx3* and (F) *Hdac11*. Shaded gray regions indicate peaks specifically referenced in text. The y axis values indicate tags per 10,000,000 with a range of 0–50. TSS denotes transcription start site.

BMDMs relative to TAM MG in the GEMM-shP53 and GL261 models (and vice versa) (Figures 3C and S3D). Among a panel of different TFs, EGR1 and MEF2A were enriched in TAM MG (Figures 3C and S3D; Table S5C). Interestingly, MEF2 is associated with MG identity (Lavin et al., 2014). In TAM BMDMs, TF motifs involved in monocyte to macrophage differentiation were enriched, including RUNX, CEBP, and PU.1 (Figures 3C and S3D) (Alder et al., 2008). STAT3 and IRF4 were also enriched (Figure 3C), both of which have been associated with differential functions in macrophage activation (Mosser and Edwards, 2008; Ostuni and Natoli, 2011). We complemented these genome-wide TF activity analyses with motif enrichment analysis on the promoters of TAM BMDM-specific and TAM MG-specific genes using HOMER (Heinz et al., 2010) (Figure S3E). This also revealed an enrichment of MEF2 motifs in TAM MG, demonstrating the consistent role of tissue-specific transcriptional programs in TAM MG education. Meanwhile, TAM BMDM-specific genes were again enriched in PU.1, RUNX, and CEBP motifs (Figure S3E). These findings were further corroborated by increased expression of brain-specific TFs (*Mef2c*, *Sall1*, and *Sall3)* in TAM MG, while TAM BMDMs were enriched for *Ciita*, *Vdr*, *Ahr*, and *Runx* family members (Figure 3D; Table S2A).

Given the consistent enrichment of RUNX activity in TAM BMDMs, we next focused on examining the expression and chromatin state of *Runx* family members. *Runx2* and *Runx3* were enriched in TAM BMDMs compared to TAM MG (Figure 3D). While no differences were found in the chromatin state of *Runx1* or *Runx2*, in the first intron of *Runx3* (Figure 3E, ii), we observed a peak present in TAM MG but reduced in TAM BMDMs (Figure 3E). Meanwhile, the *Runx3* promoter showed little open chromatin in TAM MG and a distinct peak in TAM BMDMs near the transcription start site (Figure 3E, i). Interestingly, both peaks have been shown to be transforming growth factor β (TGF-β)-responsive PU.1 binding sites associated with *Runx3* expression (Chopin et al., 2013), indicating the same signal transduction pathway can produce distinct outputs in TAM BMDMs and TAM MG.

We also identified enrichment of the epigenetic modifiers *Hdac7* and *Hdac9* in TAM BMDMs, while *Hdac11* was enriched in TAM MG (Figure S3F), the latter of which has been shown to repress *Il10* expression in macrophages (Villagra et al., 2009). Interestingly, an upstream enhancer element in *Hdac11* was significantly enriched in TAM MG compared to TAM BMDMs, a peak that contained a SMAD-responsive element (Figure 3F-i). Collectively, these results suggest that differential genomic PU.1 occupancy underlies distinct open chromatin states in BMDMs and MG, whereupon additional factors such as TGF-β/SMAD signaling and RUNX family members cooperate with PU.1 to enforce distinct transcriptional networks. Subsequent regulation of TFs and chromatin modifying factors, such as *Hdac11*, may explain the distinct cytokine expression patterns observed, such as TAM BMDM expression of *Il10* and TAM MG expression of *Tnf.*

Itga4/Cd49d Distinguishes Microglia and Peripherally Derived Macrophages in Murine Models of Brain Malignancy

We next sought to identify tools capable of distinguishing TAM BMDMs and TAM MG in human disease, where genetic lineage tracing is not possible. Given that TAM BMDMs in gliomas upregulated *Cx3cr1* (Figure S1F), a proposed MG marker, we sought to identify TAM BMDM-specific markers that instead remained silent in TAM MG. From the 164 core BMDM genes, we identified 40 candidate transmembrane proteins that might serve as useful markers for flow cytometry. Among these, the integrin subunit alpha 4, *Itga4*/Cd49d, emerged as a promising candidate, particularly given previous reports that it, along with the integrin subunit alpha L, *Itgal*/Cd11a, is regulated by RUNX family members, including *Runx1* and *Runx3* (Domínguez-Soto et al., 2005). Consistently, we found that *Itga4* and *Itgal* were specifically repressed in MG compared to other macrophage populations (Figure S4A). This was confirmed by flow cytometry, where Cd49d expression in MG was negligible or absent compared to macrophages of the spleen, liver, lung, bone marrow, and blood Ly6C^{+} monocytes (Figure 4A). Ly6G^{+} granulocytes were also Cd49d^{-}, which, along with Cd49d^{+} lymphocytes in the Cd45^{+}Cd11b^{-} gate, served as useful gating controls in subsequent experiments (Figure 4A).

We examined Cd49d and Cd11a expression in TAM BMDMs and TAM MG using Flt3:Cre-based lineage tracing in the GEMM-shP53 model. After gating on Cd45^{+}Cd11b^{+}Ly6C^{-} Ly6G^{-} cells, the normal brain only contained Cd45^{low}Cd49d^{-} cells, and all peripheral monocytes were Cd45^{high}Cd49d^{+} (Figure 4B). In tumors, we found two cell populations, Cd45^{low}Cd49d^{-} and Cd45^{high}Cd49d^{+}, which contained GFP^{-} TdTomato^{+} MG and GFP^{+} TdTomato^{-} BMDMs, respectively (Figure 4B). Similar results were found for Cd11a (Figure 4B) and were replicated in the GL261 model using both Cx3cr1-based and Flt3:Cre lineage-tracing strategies (Figures S4B and S4C). Lastly, we evaluated Cd49d expression in a Pten loss-of-function PDGFB-driven glioma model (GEMM-*Pten^{Flox}*) where *Pten$^{Flox/Flox}$*; nTva^{+} mice were injected with RCAS vectors encoding PDGFB and Cre (Huse et al., 2009). Using IR-BMT for lineage tracing, we found that Cd49d distinguishes donor and host-derived cells, including in glioma models with extended latency (~12 weeks for the GEMM-*Pten^{Flox}* model) (Figure S4D).

To evaluate other models of brain malignancy, we utilized an intracardiac injection model of brain metastasis (BrM) colonization using a tumor cell line (99LN-BrM). 99LN-BrM cells were originally derived from the lymph node of a MMTV:PyMT breast cancer GEMM and subjected to in vivo selection. We used this syngeneic, immunocompetent BrM model in conjunction with Cx3cr1-based lineage tracing and found that BrM lesions contained both TdTomato^{+} Iba1^{+} and TdTomato^{-} Iba1^{+} cells, indicating recruitment of both TAM MG and TAM BMDMs, respectively (Figures 4C and 4D). We validated these findings by flow cytometry, where Cd49d and Cd11a served as reliable markers of BMDMs as in the glioma models described above (Figure 4E). We again found that eYFP levels, a direct readout of *Cx3cr1* expression, were similar between TAM BMDMs and TAM MG in BrM, reinforcing the necessity of the Cx3cr1:CreER lineage tracing approach over that of the Cx3cr1 reporter (Figure S4E). Lastly, we confirmed these data in a well-established xenograft BrM model using brain homing MDA-MB-231 cells (Bos et al., 2009), in conjunction with IR-BMT lineage tracing using mRFP^{+} donor cells. In this model, we identified two cell populations,

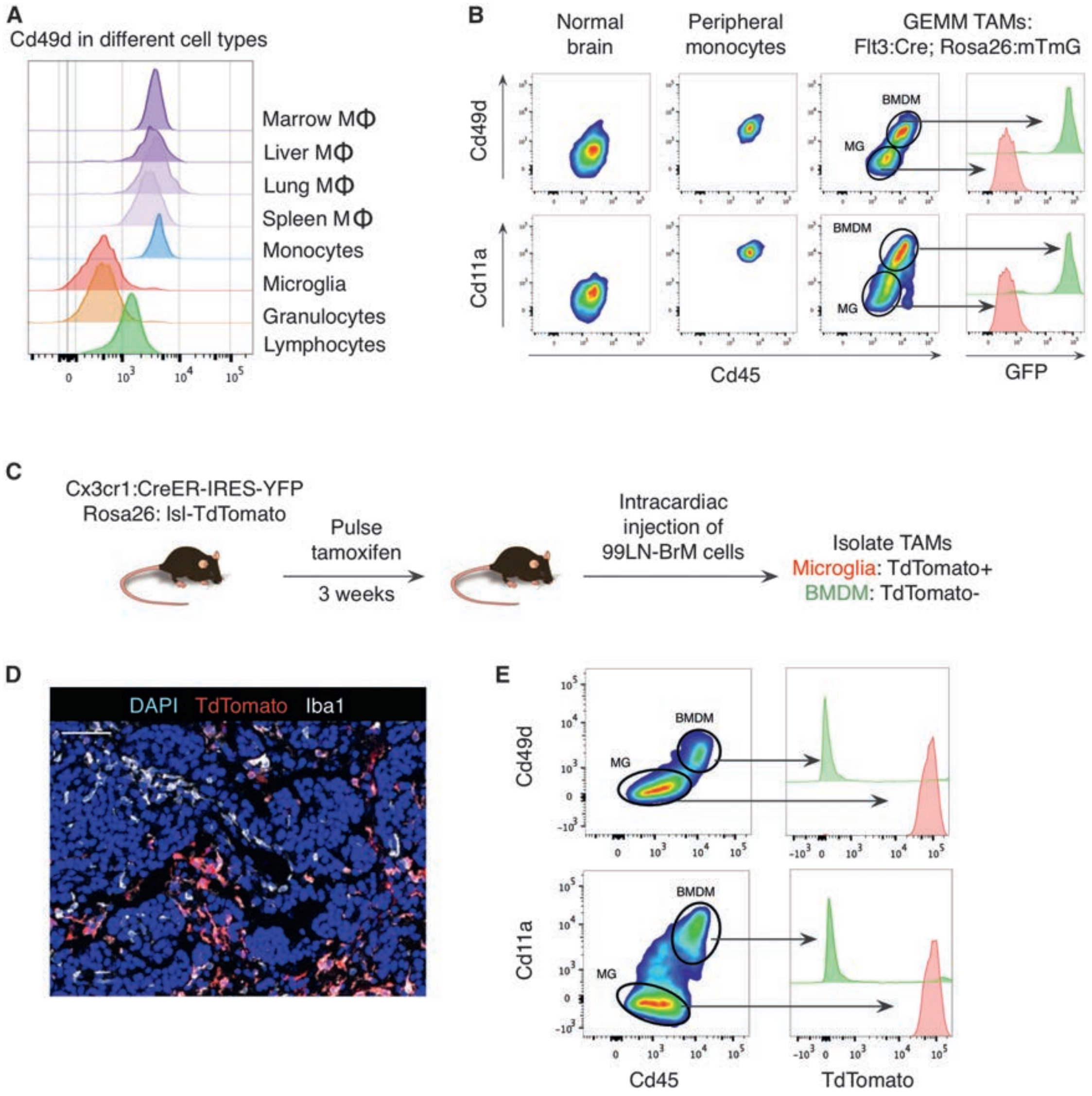


Figure 4. *Itga4*/Cd49d Distinguishes TAM BMDMs and TAM MG in Murine Brain Malignancy

(A) Histogram of Cd49d expression for indicated populations from non-tumor-bearing mice.

(B) Flow cytometry for Cd45 and either Cd49d (top) or Cd11a (bottom) in normal blood monocytes, normal MG (from adjacent normal brain), or TAMs isolated from Flt3:Cre Rosa26:mTmG mice with GEMM-shP53 tumors. Adjacent histograms indicate GFP expression in indicated populations.

(C) Experimental schematic for the 99LN-BrM model in Cx3cr1-lineage tracing mice.

(D) Representative IF staining of TdTomato (red), Iba1 (white), and DAPI (blue) 99LN-BrM tumors as depicted in (C). Scale bar, 50 μm.

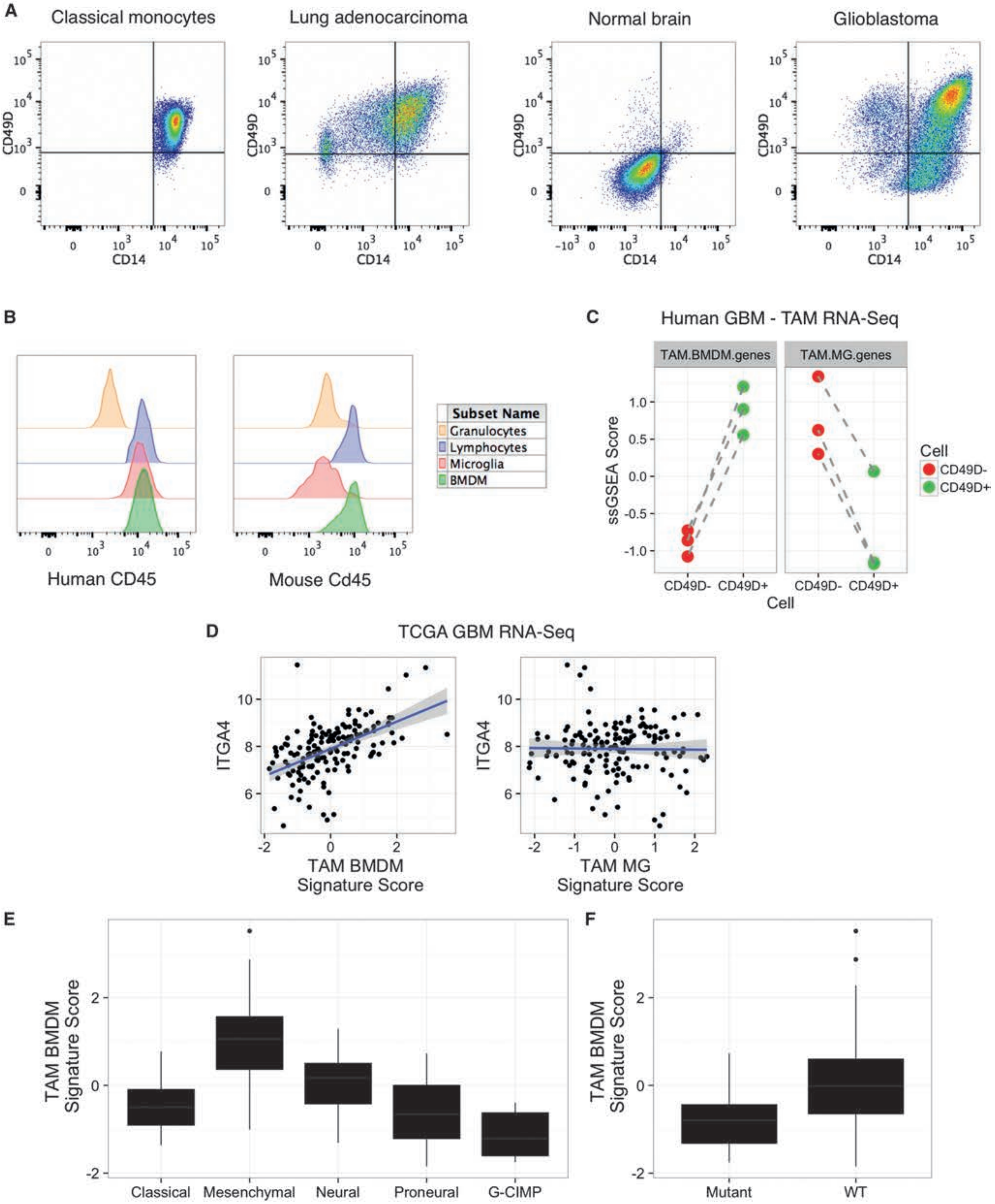
(E) Flow cytometry as in (B) for the 99LN-BrM model, with TdTomato expression indicated in the adjacent histogram. Flow plots are representative of n = 5–8 mice.

Cd45^{low}Cd49d^{-} MG and Cd45^{high}Cd49d^{+} BMDM. The mRFP^{+} donor cells were exclusively found within the Cd45^{high}Cd49d^{+} BMDM gate (Figure S4F).

Together, our results obtained in multiple models of brain malignancy with distinct lineage-tracing approaches demonstrate that TAM BMDM accumulation is independent of BBB preconditioning by IR or intracranial injection. These data also thoroughly establish Cd49d as an efficient marker to distinguish resident MG and peripherally derived macrophages in homeostasis as well as in primary and metastatic brain malignancies.

CD49D Identifies Microglia and Macrophages in Human Brain Malignancies

We next investigated whether CD49D could be used to discriminate MG and peripherally derived macrophages in human brain tumors. We assessed CD49D expression by flow cytometry across a panel of surgical samples composed of non-malignant normal brain (n = 3), untreated high-grade glioma (GBM) (n = 3), lung adenocarcinomas (n = 6), and peripheral blood mononuclear cells (PBMCs) (n = 6). Consistent with our data in mice, granulocytes (CD45^{+}CD11B^{+}CD66B^{+}CD14^{low}CD16^{+}) did not



(legend on next page)

express CD49D, and were used as a reference guide for gating $CD49D^+$ and $CD49D^-$ TAMs (Figure S5A). Importantly, we never identified $CD49D^-$ TAMs in primary lung tumors or $CD49D^-$ monocytes in healthy donor PBMCs, indicating that, as predicted, low expression of CD49D is restricted to MG and is not a general phenotype of tissue-resident macrophages (Figure 5A). By contrast, the $CD45^+CD11B^+CD66B^-CD14^+CD16^-$ compartment in non-malignant brain was predominantly composed of $CD49D^-$ MG (Figure 5A). Critically, in each GBM sample we identified both $CD49D^+$ and $CD49D^-$ TAMs, presumably representing BMDM and brain-resident MG, respectively (Figure 5A).

Interestingly, in human samples, we found no difference in CD45 expression between $CD49D^-$ and $CD49D^+$ TAMs (Figure 5B), a marker previously suggested to be informative for distinguishing BMDMs and MG in brain malignancy (Hussain et al., 2006; Parney et al., 2009; Sedgwick et al., 1991). Indeed, CD45 expression differed most prominently between granulocytes and TAMs, as opposed to MG and BMDMs (Figure 5B). However, this lack of differential CD45 expression is not the case in mouse, where Cd45 adequately discriminates MG and BMDM in the models tested (Figure 5B). We next sorted paired $CD49D^-$ and $CD49D^+$ TAMs from GBM patients to verify these populations indeed reflected TAM MG and TAM BMDMs, respectively. Using genes specific for TAM MG and TAM BMDMs from our mouse models (Figure 2A), we found that $CD49D^-$ TAMs were indeed enriched for TAM MG genes ($p \leq 7.78 \times 10^{-3}$), while $CD49D^+$ TAMs were enriched for TAM BMDM genes ($p \leq 5.01 \times 10^{-3}$) (Figure 5C).

Previous analyses of TAM expression in human gliomas have utilized bulk $CD11B^+$ cells, a population likely composed of both TAM BMDMs and TAM MG, as well as other myeloid populations. We queried one available RNA-seq dataset from bulk $CD11B^+$ cells (Szulzewsky et al., 2016), which showed increased *ITGA4*/CD49D expression in purified $CD11B^+$ cells in GBM compared to normal MG from either post-mortem samples or resections from epileptic patients (Figure S5B). This was complemented by a relative decrease in the MG-enriched transcript *P2RY12* in GBM compared to non-malignant brain (Figure S5B). In querying an additional microarray-based dataset of purified $CD11B^+$ cells (Gabrusiewicz et al., 2016), we observed that peripheral blood $CD11B^+$ cells from GBM patients expressed similar levels of *ITGA4* compared to GBM tumor samples, while there was higher *P2RY12* expression in GBM samples than in peripheral blood (Figure S5C), as we would have expected. We extended these analyses to whole-tissue RNA-seq data from the TCGA-GBM cohort (Brennan et al., 2013) and observed that *ITGA4* expression was significantly increased in GBM compared to normal brain (Figure S5D). Collectively, these analyses suggest that TAMs in GBM represent a heterogeneous population composed of both BMDMs and MG, reinforcing the necessity of refined sorting strategies for accurate discrimination between these cells and highlighting the utility of a CD49D-based gating approach.

We next assessed TAM BMDM and TAM MG gene set expression in the TCGA cohort as a whole. TAM BMDM genes and TAM MG genes showed high intra-gene set correlation, where TAM BMDM genes such as *RUNX2*, *IL10*, *RUNX3*, *ITGA4*, and *VDR* showed significant pairwise correlations and TAM MG genes such as *MEF2C*, *P2RY12*, *RXRG*, *SALL1*, *KLF12* and *SALL3* similarly showed significant pairwise correlations (Figure S5E). Moreover, *ITGA4* showed a high correlation with a TAM BMDM gene signature score ($p \leq 2.2 \times 10^{-16}$), but not with a TAM MG signature score (Figure 5D), showing increased *ITGA4* expression is specific to TAM BMDM abundance and not TAMs as a whole.

Previous transcriptional and epigenetic analyses have identified distinct GBM subtypes (Noushmehr et al., 2010; Verhaak et al., 2010), where the mesenchymal subtype was enriched for tumor stroma and inflammatory molecules. Here, we find TAM BMDM signature scores are significantly different among molecular subtypes of GBM ($p \leq 2.2 \times 10^{-16}$), with the highest scores in the mesenchymal GBM subtype and the lowest scores in G-CIMP patients (Figure 5E). Correspondingly, TAM BMDM signature scores were lowest in patients with *IDH1* mutations (Figure 5F; $p \leq 5.93 \times 10^{-3}$). By comparison, TAM MG signature scores displayed a blunted association with tumor subtype ($p \leq 0.041$) and no association with *IDH1* mutation status ($p \leq 0.153$) (Figures S5F and S5G). These analyses reinforce our findings that TAM BMDMs and TAM MG are distinguishable immune cell populations with distinct abundance and characteristics in specific subtypes of human GBM.

DISCUSSION

IR-BMT has been used widely in animal models to perform lineage tracing of TAMs in brain malignancy (Ajami et al., 2007; De

Figure 5. CD49D Discriminates TAM BMDMs and TAM MG in Human Brain Malignancy

(A) Classical monocytes, MG, and TAMs were defined as $CD45^+CD11B^+CD66B^-CD14^+CD16^-$. Gated cells are then shown for CD14 and CD49D in representative samples of human classical monocytes from peripheral blood (n = 6), TAMs from a lung adenocarcinoma patient (n = 6), MG from a non-malignant brain (n = 3), and TAMs from a GBM patient (n = 3).

(B) Histogram of CD45 expression by flow cytometry in human (left) and mouse (right) samples. In human GBM, CD45 expression is shown for granulocytes ($CD45^+CD11B^+CD66B^+CD16^+CD14^{low}$), lymphocytes ($CD45^+CD11B^-$), TAM MG ($CD45^+CD11B^+CD66B^-CD16^-CD14^+CD49D^-$), and TAM BMDMs ($CD45^+CD11B^+$ $CD66B^-CD16^-CD14^+CD49D^+$). In mouse, Cd45 expression is shown for granulocytes ($Cd45^+Cd11b^+Ly6C^{low}Ly6G^+$), lymphocytes ($Cd45^+Cd11b^-$), TAM MG ($Cd45^+Cd11b^+Ly6C^-Ly6G^-Tomato^+GFP^-$), and TAM BMDMs ($Cd45^+Cd11b^+Ly6C^-Ly6G^-Tomato^-GFP^+$) from a Flt3:Cre Rosa26:mTmG GEMM-shP53 glioma. Data are representative of n = 3 patients and n = 5 mice.

(C) *Z* scored single sample gene set enrichment analysis (ssGSEA) scores for TAM BMDM genes (left, paired t test, $p \leq 5.01 \times 10^{-3}$) and TAM MG genes (right, paired t test, $p \leq 7.78 \times 10^{-3}$) in matched $CD49D^-$ and $CD49D^+$ TAMs from GBM patients. Dashed lines indicate matched samples (n = 3 patients).

(D) Scatterplot of TAM BMDM signature score (x axis, left, Spearman rho = 0.564, $p \leq 2.2 \times 10^{-16}$) and TAM MG signature score (x axis, right, Spearman rho = 0.067, $p \leq 0.411$) and *ITGA4* expression (y axis) from TCGA-GBM RNA-seq data. Solid blue line indicates line of best fit, with shaded areas depicting SD confidence intervals.

(E and F) *Z* scored TAM BMDM signature scores across (E) GBM subtype (ANOVA $p \leq 2.2 \times 10^{-16}$) and (F) *IDH1* mutation status (Student's t test $p \leq 5.93 \times 10^{-3}$).

Palma et al., 2005; Huang et al., 2014; Mildner et al., 2007; Müller et al., 2015), albeit with concerns regarding potential artifacts due to effects of IR on BBB disruption. Alternative chemical BMT approaches have been suggested, though similar effects on BBB permeability cannot be ignored (Alder et al., 2008; Kierdorf et al., 2013b). Here, we confirm that IR-BMT leads to increased TAM BMDM content in the GL261 glioma model, a finding that has been recently reported by juxtaposing IR-BMT with and without head-shielding (Müller et al., 2015). While IR-BMT may confound lineage-tracing studies, it remains to be seen if IR preconditioning before the onset of tumorigenesis significantly alters TAM activity in tumor development or if the inflammatory environment of the tumor supersedes any antecedent effects of the IR-BMT protocol.

Other than IR-BMT, the most widely employed approach to discriminate MG and peripherally derived macrophages relies upon Cd45 expression, with Cd45^{high} cells considered BMDMs and Cd45^{low} cells considered MG (Gabrusiewicz et al., 2011; Sedgwick et al., 1991). While this marker seems adequate in the murine models we have employed here, cell-type-specific CD45 expression appears to be different between mouse and human. Our data indicate that CD45 does not accurately discriminate MG and BMDMs in patient samples, emphasizing the need for extensive flow cytometry panels to clearly distinguish these cells in both species. Additionally, our genetic lineage tracing models also show that expression of *Cx3cr1*, which is commonly used to trace normal MG, is subject to upregulation in BMDMs upon tumor education (Figures 2A, S1F, and S4E) and thus cannot be used to discriminate MG and BMDMs in brain tumors.

Instead, we present *Itga4* (Cd49d) as an effective, consistent marker that works in both mice and humans to distinguish MG and peripherally derived macrophages in multiple brain malignancies. Cd49d may also prove a useful tool in determining the precise origin and kinetics of peripherally derived macrophages in brain tumors. Recent efforts to understand the heterogeneity and origins of non-parenchymal myeloid cells in the brain (including perivascular, meningeal, and choroid plexus macrophages) revealed that a subset of these cells are labeled using similar Flt3-Cre and Cx3cr1-CreER based lineage tracing systems as employed here (Goldmann et al., 2016). Thus it will be of interest to determine if any of these populations, in addition to monocytes, contribute to the TAM pool.

Our data support the hypothesis that epigenetic states influence stimulus-dependent transcriptional induction, thus leading to differential TAM education between MG and BMDMs. Differential genomic occupancy of PU.1 between MG and other macrophage populations in non-cancer contexts has been shown to dictate differential enhancer selection (Gosselin et al., 2014). Indeed, within this dataset, we found that PU.1 binding sites at enhancers and promoters were already different between MG and BMDM for the genes we identified to be specific to their respective TAM populations. This suggests that TAM BMDMs and TAM MG are poised to engage in different transcriptional networks based on initial enhancer selection. It is likely that differential expression of binding partners influences PU.1 genomic occupation. Cooperative binding is evident between PU.1 and CEBPβ to promote macrophage differentiation and in B cell development, where PU.1 occupancy is influenced by E2A expression (Heinz et al., 2010). Such a hypothesis has also been shown to account for MG-specific PU.1 binding in cooperation with TGF-β-induced SMAD activity (Gosselin et al., 2014). Similar dynamics may be at play in brain tumors, where binding partners that are absent in MG and expressed in BMDMs can sculpt genomic PU.1 occupancy. For example, the RUNX family member *Runx3* is one such candidate, which is enriched in TAM BMDMs versus TAM MG and shows motif enrichment in promoters where PU.1 binds in BMDMs, but not MG.

While our studies here focus predominantly on identifying recurrent signatures distinguishing TAM MG and TAM BMDMs across multiple mouse models and patient samples, there were also tumor-specific gene expression patterns in TAM education (Figures 2E, S2C, and S2D), which may provide insights into how tumor-derived signals can generate inter-tumoral heterogeneity in TAM activation profiles. In addition, analysis of TCGA data showed that gene signatures associated with TAM BMDMs were differentially enriched in the distinct tumor subtypes of GBM. Recent reports have identified mixed activation states in bulk TAM populations in glioma patients (Gabrusiewicz et al., 2016; Szulzewsky et al., 2016), and our data now show that TAM MG and TAM BMDMs possess distinct activation states, potentially resolving this mixed phenotype. Importantly, the identification of CD49D as a cell-surface marker to discriminate between TAM MG and TAM BMDMs in human disease will permit extensive interrogation of these cell populations in patient samples.

Collectively, the studies presented here definitively demonstrate that peripherally derived macrophages are indeed present in multiple mouse and human brain malignancies and have distinct transcriptional profiles from their brain-resident counterparts. We posit that while macrophages can acquire tissue-resident macrophage-like traits upon entry into a tissue (Lavin et al., 2014), an inflammatory microenvironment, such as in the context of cancer or neuroinflammation, may further amplify differences between the cells, leading to diverse functional outcomes for tissue-resident and peripherally derived macrophage populations.

EXPERIMENTAL PROCEDURES

Tumor and Lineage Tracing Models

Mouse models of gliomagenesis and brain metastasis, cell line generation, and the use of lineage tracing models have been previously reported (Boyer et al., 2011; Parkhurst et al., 2013; Quail et al., 2016; Sevenich et al., 2014) and are described in full in Supplemental Experimental Procedures.

Institutional Review Board Approval and Patient Information

All human specimens were collected from patients consented to Memorial Sloan Kettering Cancer Center (MSKCC) institutional review board (IRB) protocols #06-107, #14-230. Glioma patients that presented with contrast-enhancing brain lesions and no prior history of brain malignancy or therapy were included. Tumor specimens were collected from the operating room and processed as described below. Pathological analyses confirmed grade IV GBM. Non-malignant normal brain samples were collected from two sources: non-malignant sites distant from low-grade disease and post-mortem samples with no history of brain malignancy. Pathological analysis confirmed the absence of tumor. Samples from patients with primary lung tumors were included based on pathological analysis of lung adenocarcinoma, with no screening based on prior malignancy or therapy.

Flow Cytometry and Cell Sorting

For blood analysis, mice were bled via either retro-orbital or submandibular routes under isoflurane anesthesia. For all other tissue analyses, mice were anesthetized with 1.25% avertin and transcardially perfused with PBS. Single-cell suspensions from spleen and bone marrow were isolated by macrodissection and mechanical tissue dissociation. Liver, kidney, and lung were macrodissected and dissociated using the Mouse Tumor Dissociation Kit (mTDK; Miltenyi) and the OctoMACS dissociator. Mouse and human brain specimens were macrodissected and dissociated using the Brain Tumor Dissociation Kit (BTDK; Miltenyi) and a single-cell suspension generated using the OctoMACS dissociator. Human lung tumors were dissociated with the Human Tumor Dissociation Kit (hTDK; Miltenyi.) All tissue suspensions were filtered through a 40-μM mesh filter and underwent red blood cell lysis (Pharm-Lyse BD). Normal brain and brain tumor tissues were incubated with Myelin Removal Beads (Miltenyi). Single-cell suspensions were FC blocked (BD #553141) for 15 min at 4°C and then incubated with directly conjugated antibody panels for 15 min at 4°C. Cell suspensions were washed (PBS + 2% fetal bovine serum) and resuspended in a DAPI solution. All flow cytometry analysis was completed on a BD Fortessa device, and all sorting was performed on an Aria III. Cells were sorted directly into TRIzol LS and snap frozen in liquid nitrogen. Antibodies and methods for immunohistochemistry can be found in Supplemental Experimental Procedures.

Statistical Methods

RNA Sequencing, ATAC Sequencing, and Bioinformatics

RNA was isolated by chloroform extraction and isopropanol precipitation. RNA-sequencing libraries were generated with the SMART-Seq preparation kit (CloneTech). Single-end, 100-bp sequencing was performed by GeneWiz on an Illumina HiSeq 2500. FASTQ files were mapped to the mouse genome (mm10) or the human genome (hg19) using STAR (version 2.5.0e) with default parameters (Dobin et al., 2013). Transcript abundance was quantified using STAR with a GTF file from iGenomes (Illumina). A count matrix was produced in R and differential gene expression was assessed with DESeq2 using a fold change cutoff of ±2 and a false discovery rate of 5% (Love et al., 2014). Gene Ontology analysis was performed using DAVID with default parameters (Dennis et al., 2003). ATAC-sequencing was performed as previously described (Buenrostro et al., 2013). Paired-end, 50-bp sequencing was performed on an Illumina HiSeq 2500 with an average read depth of ∼35,000,000 reads per sample. Reads were mapped to mm10 using STAR (version 2.5.0e) using (–alignIntronMax 1–alignEndsType EndToEnd). Peak calling, annotation, and differential peak identification was performed using HOMER.

Methods for analyzing external datasets, TF activity analysis, and additional statistical methods are described in Supplemental Experimental Procedures.

ACCESSION NUMBERS

The accession number for the gene expression and ATAC-sequencing data generated in this study is GEO: GSE86573.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, five figures, and five tables and can be found with this article online at http://dx.doi.org/10.1016/j.celrep.2016.10.052.

AUTHOR CONTRIBUTIONS

R.L.B. and J.A.J. conceived the study, designed and interpreted experiments, and wrote the manuscript. R.L.B., F.K., L.A., S.M.P., L.S., D.F.Q., S.D., and K.S. performed experiments and analyzed results. R.L.B. performed all computational analyses. E.E.G., C.A.I.-D., C.W.B., V.T., and P.H.G. provided patient samples. J.A.J. supervised the study. All authors commented on the manuscript.

ACKNOWLEDGMENTS

We thank X. Chen for excellent technical support, Drs. D. Yan and N. Ben-Chetrit for experimental assistance, and members of the J.A.J. lab for insightful comments and discussion. We are grateful to Dr. B. Santomasso, N. Distefano, and Dr. L. DeAngelis (MSKCC Neurosurgery) and Dr. C. Rudin (MSKCC Thoracic Oncology) for help in acquiring patient samples. We thank Drs. E. Holland and T. Ozawa for generously providing the RCAS vectors and Nestin-Tva mice, Dr. C. Forsberg for the *Flt3:Cre Rosa26:mTmG* mice, Drs. C. Sawyers and B. Carver for the *Pten*$^{Flox/Flox}$ mice, Drs. S. Coniglio and J. Segall for the GL261 cell line, and Dr. J. Massagué for the brain-homing MDA-MB-231 cell line. This research was supported by grant R01CA181355 (J.A.J.), the Ludwig Institute for Cancer Research (J.A.J.), the MSKCC Center for Metastasis Research (J.A.J.), MSK Cancer Center Support grant P30 CA008748 from the National Cancer Institute, the Gerstner Sloan Kettering Graduate School (R.L.B.), National Cancer Institute fellowship 5F31CA167863 (R.L.B.), Deutsche Forschungsgemeinschaft fellowships KL 2491/1-1 (F.K.) and SE2234/1-1 (L.S.), the American Brain Tumor Association in honor of Joel A. Gingras (L.A.), and the Canadian Institutes of Health Research (D.F.Q.).

Received: September 18, 2016
Revised: October 12, 2016
Accepted: October 19, 2016
Published: November 10, 2016

REFERENCES

Ajami, B., Bennett, J.L., Krieger, C., Tetzlaff, W., and Rossi, F.M. (2007). Local self-renewal can sustain CNS microglia maintenance and function throughout adult life. Nat. Neurosci. *10*, 1538–1543.

Alder, J.K., Georgantas, R.W., 3rd, Hildreth, R.L., Kaplan, I.M., Morisot, S., Yu, X., McDevitt, M., and Civin, C.I. (2008). Kruppel-like factor 4 is essential for inflammatory monocyte differentiation in vivo. J. Immunol. *180*, 5645–5652.

Biffi, A., De Palma, M., Quattrini, A., Del Carro, U., Amadio, S., Visigalli, I., Sessa, M., Fasano, S., Brambilla, R., Marchesini, S., et al. (2004). Correction of metachromatic leukodystrophy in the mouse model by transplantation of genetically modified hematopoietic stem cells. J. Clin. Invest. *113*, 1118–1129.

Bos, P.D., Zhang, X.H., Nadal, C., Shu, W., Gomis, R.R., Nguyen, D.X., Minn, A.J., van de Vijver, M.J., Gerald, W.L., Foekens, J.A., and Massagué, J. (2009). Genes that mediate breast cancer metastasis to the brain. Nature *459*, 1005–1009.

Boyer, S.W., Schroeder, A.V., Smith-Berdan, S., and Forsberg, E.C. (2011). All hematopoietic cells develop from hematopoietic stem cells through Flk2/Flt3-positive progenitor cells. Cell Stem Cell *9*, 64–73.

Brennan, C.W., Verhaak, R.G., McKenna, A., Campos, B., Noushmehr, H., Salama, S.R., Zheng, S., Chakravarty, D., Sanborn, J.Z., Berman, S.H., et al.; TCGA Research Network (2013). The somatic genomic landscape of glioblastoma. Cell *155*, 462–477.

Bruttger, J., Karram, K., Wörtge, S., Regen, T., Marini, F., Hoppmann, N., Klein, M., Blank, T., Yona, S., Wolf, Y., et al. (2015). Genetic cell ablation reveals clusters of local self-renewing microglia in the mammalian central nervous system. Immunity *43*, 92–106.

Buenrostro, J.D., Giresi, P.G., Zaba, L.C., Chang, H.Y., and Greenleaf, W.J. (2013). Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nat. Methods *10*, 1213–1218.

Chopin, M., Seillet, C., Chevrier, S., Wu, L., Wang, H., Morse, H.C., 3rd, Belz, G.T., and Nutt, S.L. (2013). Langerhans cells are generated by two distinct PU.1-dependent transcriptional networks. J. Exp. Med. *210*, 2967–2980.

De Palma, M., Venneri, M.A., Galli, R., Sergi Sergi, L., Politi, L.S., Sampaolesi, M., and Naldini, L. (2005). Tie2 identifies a hematopoietic lineage of proangiogenic monocytes required for tumor vessel formation and a mesenchymal population of pericyte progenitors. Cancer Cell *8*, 211–226.

Dennis, G., Jr., Sherman, B.T., Hosack, D.A., Yang, J., Gao, W., Lane, H.C., and Lempicki, R.A. (2003). DAVID: Database for annotation, visualization, and integrated discovery. Genome Biol. *4*, 3.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15–21.

Domínguez-Soto, A., Relloso, M., Vega, M.A., Corbí, A.L., and Puig-Kröger, A. (2005). RUNX3 regulates the activity of the CD11a and CD49d integrin gene promoters. Immunobiology *210*, 133–139.

Du, R., Lu, K.V., Petritsch, C., Liu, P., Ganss, R., Passegué, E., Song, H., Vandenberg, S., Johnson, R.S., Werb, Z., and Bergers, G. (2008). HIF1alpha induces the recruitment of bone marrow-derived vascular modulatory cells to regulate tumor angiogenesis and invasion. Cancer Cell *13*, 206–220.

Dubois, L.G., Campanati, L., Righy, C., D'Andrea-Meira, I., Spohr, T.C., Porto-Carreiro, I., Pereira, C.M., Balça-Silva, J., Kahn, S.A., DosSantos, M.F., et al. (2014). Gliomas and the vascular fragility of the blood brain barrier. Front. Cell. Neurosci. *8*, 418.

Gabrusiewicz, K., Ellert-Miklaszewska, A., Lipko, M., Sielska, M., Frankowska, M., and Kaminska, B. (2011). Characteristics of the alternative phenotype of microglia/macrophages and its modulation in experimental gliomas. PLoS ONE *6*, e23902.

Gabrusiewicz, K., Rodriguez, B., Wei, J., Hashimoto, Y., Healy, L.M., Maiti, S.N., Thomas, G., Zhou, S., Wang, Q., Elakkad, A., et al. (2016). Glioblastoma-infiltrated innate immune cells resemble M0 macrophage phenotype. JCI Insight *1*, e85841.

Gautier, E.L., Shay, T., Miller, J., Greter, M., Jakubzick, C., Ivanov, S., Helft, J., Chow, A., Elpek, K.G., Gordonov, S., et al.; Immunological Genome Consortium (2012). Gene-expression profiles and transcriptional regulatory pathways that underlie the identity and diversity of mouse tissue macrophages. Nat. Immunol. *13*, 1118–1128.

Ginhoux, F., Greter, M., Leboeuf, M., Nandi, S., See, P., Gokhan, S., Mehler, M.F., Conway, S.J., Ng, L.G., Stanley, E.R., et al. (2010). Fate mapping analysis reveals that adult microglia derive from primitive macrophages. Science *330*, 841–845.

Goldmann, T., Wieghofer, P., Jordão, M.J., Prutek, F., Hagemeyer, N., Frenzel, K., Amann, L., Staszewski, O., Kierdorf, K., Krueger, M., et al. (2016). Origin, fate and dynamics of macrophages at central nervous system interfaces. Nat. Immunol. *17*, 797–805.

Gomez Perdiguero, E., Klapproth, K., Schulz, C., Busch, K., Azzoni, E., Crozet, L., Garner, H., Trouillet, C., de Bruijn, M.F., Geissmann, F., and Rodewald, H.R. (2015). Tissue-resident macrophages originate from yolk-sac-derived erythro-myeloid progenitors. Nature *518*, 547–551.

Gosselin, D., Link, V.M., Romanoski, C.E., Fonseca, G.J., Eichenfield, D.Z., Spann, N.J., Stender, J.D., Chun, H.B., Garner, H., Geissmann, F., and Glass, C.K. (2014). Environment drives selection and function of enhancers controlling tissue-specific macrophage identities. Cell *159*, 1327–1340.

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol. Cell *38*, 576–589.

Huang, Y., Hoffman, C., Rajappa, P., Kim, J.H., Hu, W., Huse, J., Tang, Z., Li, X., Weksler, B., Bromberg, J., et al. (2014). Oligodendrocyte progenitor cells promote neovascularization in glioma by disrupting the blood-brain barrier. Cancer Res. *74*, 1011–1021.

Huse, J.T., Brennan, C., Hambardzumyan, D., Wee, B., Pena, J., Rouhanifard, S.H., Sohn-Lee, C., le Sage, C., Agami, R., Tuschl, T., and Holland, E.C. (2009). The PTEN-regulating microRNA miR-26a is amplified in high-grade glioma and facilitates gliomagenesis in vivo. Genes Dev. *23*, 1327–1337.

Hussain, S.F., Yang, D., Suki, D., Aldape, K., Grimm, E., and Heimberger, A.B. (2006). The role of human glioma-infiltrating microglia/macrophages in mediating antitumor immune responses. Neuro-oncol. *8*, 261–279.

Kennedy, D.W., and Abkowitz, J.L. (1997). Kinetics of central nervous system microglial and macrophage engraftment: analysis using a transgenic bone marrow transplantation model. Blood *90*, 986–993.

Kierdorf, K., Erny, D., Goldmann, T., Sander, V., Schulz, C., Perdiguero, E.G., Wieghofer, P., Heinrich, A., Riemke, P., Hölscher, C., et al. (2013a). Microglia emerge from erythromyeloid precursors via Pu.1- and Irf8-dependent pathways. Nat. Neurosci. *16*, 273–280.

Kierdorf, K., Katzmarski, N., Haas, C.A., and Prinz, M. (2013b). Bone marrow cell recruitment to the brain in the absence of irradiation or parabiosis bias. PLoS ONE *8*, e58544.

Lavin, Y., Winter, D., Blecher-Gonen, R., David, E., Keren-Shaul, H., Merad, M., Jung, S., and Amit, I. (2014). Tissue-resident macrophage enhancer landscapes are shaped by the local microenvironment. Cell *159*, 1312–1326.

Liu, W.Y., Wang, Z.B., Zhang, L.C., Wei, X., and Li, L. (2012). Tight junction in blood-brain barrier: an overview of structure, regulation, and regulator substances. CNS Neurosci. Ther. *18*, 609–615.

Locatelli, G., Wörtge, S., Buch, T., Ingold, B., Frommer, F., Sobottka, B., Krüger, M., Karram, K., Bühlmann, C., Bechmann, I., et al. (2012). Primary oligodendrocyte death does not elicit anti-CNS immunity. Nat. Neurosci. *15*, 543–550.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. *15*, 550.

Mass, E., Ballesteros, I., Farlik, M., Halbritter, F., Günther, P., Crozet, L., Jacome-Galarza, C.E., Händler, K., Klughammer, J., Kobayashi, Y., et al. (2016). Specification of tissue-resident macrophages during organogenesis. Science *353*, aaf4238.

Mildner, A., Schmidt, H., Nitsche, M., Merkler, D., Hanisch, U.K., Mack, M., Heikenwalder, M., Brück, W., Priller, J., and Prinz, M. (2007). Microglia in the adult brain arise from Ly-6ChiCCR2+ monocytes only under defined host conditions. Nat. Neurosci. *10*, 1544–1553.

Mosser, D.M., and Edwards, J.P. (2008). Exploring the full spectrum of macrophage activation. Nat. Rev. Immunol. *8*, 958–969.

Müller, A., Brandenburg, S., Turkowski, K., Müller, S., and Vajkoczy, P. (2015). Resident microglia, and not peripheral macrophages, are the main source of brain tumor mononuclear cells. Int. J. Cancer *137*, 278–288.

Murray, P.J., Allen, J.E., Biswas, S.K., Fisher, E.A., Gilroy, D.W., Goerdt, S., Gordon, S., Hamilton, J.A., Ivashkiv, L.B., Lawrence, T., et al. (2014). Macrophage activation and polarization: nomenclature and experimental guidelines. Immunity *41*, 14–20.

Noushmehr, H., Weisenberger, D.J., Diefes, K., Phillips, H.S., Pujara, K., Berman, B.P., Pan, F., Pelloski, C.E., Sulman, E.P., Bhat, K.P., et al.; Cancer Genome Atlas Research Network (2010). Identification of a CpG island methylator phenotype that defines a distinct subgroup of glioma. Cancer Cell *17*, 510–522.

Okabe, Y., and Medzhitov, R. (2016). Tissue biology perspective on macrophages. Nat. Immunol. *17*, 9–17.

Opitz, C.A., Litzenburger, U.M., Sahm, F., Ott, M., Tritschler, I., Trump, S., Schumacher, T., Jestaedt, L., Schrenk, D., Weller, M., et al. (2011). An endogenous tumour-promoting ligand of the human aryl hydrocarbon receptor. Nature *478*, 197–203.

Ostuni, R., and Natoli, G. (2011). Transcriptional control of macrophage diversity and specialization. Eur. J. Immunol. *41*, 2486–2490.

Ozawa, T., Riester, M., Cheng, Y.K., Huse, J.T., Squatrito, M., Helmy, K., Charles, N., Michor, F., and Holland, E.C. (2014). Most human non-GCIMP glioblastoma subtypes evolve from a common proneural-like precursor glioma. Cancer Cell *26*, 288–300.

Parkhurst, C.N., Yang, G., Ninan, I., Savas, J.N., Yates, J.R., 3rd, Lafaille, J.J., Hempstead, B.L., Littman, D.R., and Gan, W.B. (2013). Microglia promote learning-dependent synapse formation through brain-derived neurotrophic factor. Cell *155*, 1596–1609.

Parney, I.F., Waldron, J.S., and Parsa, A.T. (2009). Flow cytometry and in vitro analysis of human glioma-associated macrophages. Laboratory investigation. J. Neurosurg. *110*, 572–582.

Pyonteck, S.M., Akkari, L., Schuhmacher, A.J., Bowman, R.L., Sevenich, L., Quail, D.F., Olson, O.C., Quick, M.L., Huse, J.T., Teijeiro, V., et al. (2013). CSF-1R inhibition alters macrophage polarization and blocks glioma progression. Nat. Med. *19*, 1264–1272.

Quail, D.F., Bowman, R.L., Akkari, L., Quick, M.L., Schuhmacher, A.J., Huse, J.T., Holland, E.C., Sutton, J.C., and Joyce, J.A. (2016). The tumor microenvironment underlies acquired resistance to CSF-1R inhibition in gliomas. Science *352*, aad3018.

Reith, W., LeibundGut-Landmann, S., and Waldburger, J.M. (2005). Regulation of MHC class II gene expression by the class II transactivator. Nat. Rev. Immunol. *5*, 793–806.

Schulz, C., Gomez Perdiguero, E., Chorro, L., Szabo-Rogers, H., Cagnard, N., Kierdorf, K., Prinz, M., Wu, B., Jacobsen, S.E., Pollard, J.W., et al. (2012). A lineage of myeloid cells independent of Myb and hematopoietic stem cells. Science *336*, 86–90.

Sedgwick, J.D., Schwender, S., Imrich, H., Dörries, R., Butcher, G.W., and ter Meulen, V. (1991). Isolation and direct characterization of resident microglial cells from the normal and inflamed central nervous system. Proc. Natl. Acad. Sci. USA *88*, 7438–7442.

Sevenich, L., Bowman, R.L., Mason, S.D., Quail, D.F., Rapaport, F., Elie, B.T., Brogi, E., Brastianos, P.K., Hahn, W.C., Holsinger, L.J., et al. (2014). Analysis of tumour- and stroma-supplied proteolytic networks reveals a brain-metastasis-promoting role for cathepsin S. Nat. Cell Biol. *16*, 876–888.

Solga, A.C., Pong, W.W., Kim, K.Y., Cimino, P.J., Toonen, J.A., Walker, J., Wylie, T., Magrini, V., Griffith, M., Griffith, O.L., et al. (2015). RNA sequencing of tumor-associated microglia reveals Ccl5 as a stromal chemokine critical for Neurofibromatosis-1 glioma growth. Neoplasia *17*, 776–788.

Stephan, A.H., Barres, B.A., and Stevens, B. (2012). The complement system: an unexpected role in synaptic pruning during development and disease. Annu. Rev. Neurosci. *35*, 369–389.

Szulzewsky, F., Arora, S., de Witte, L., Ulas, T., Markovic, D., Schultze, J.L., Holland, E.C., Synowitz, M., Wolf, S.A., and Kettenmann, H. (2016). Human glioblastoma-associated microglia/monocytes express a distinct RNA profile compared to human control and murine samples. Glia *64*, 1416–1436.

Verhaak, R.G., Hoadley, K.A., Purdom, E., Wang, V., Qi, Y., Wilkerson, M.D., Miller, C.R., Ding, L., Golub, T., Mesirov, J.P., et al.; Cancer Genome Atlas Research Network (2010). Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR, and NF1. Cancer Cell *17*, 98–110.

Villagra, A., Cheng, F., Wang, H.W., Suarez, I., Glozak, M., Maurin, M., Nguyen, D., Wright, K.L., Atadja, P.W., Bhalla, K., et al. (2009). The histone deacetylase HDAC11 regulates the expression of interleukin 10 and immune tolerance. Nat. Immunol. *10*, 92–100.

Xue, J., Schmidt, S.V., Sander, J., Draffehn, A., Krebs, W., Quester, I., De Nardo, D., Gohel, T.D., Emde, M., Schmidleithner, L., et al. (2014). Transcriptome-based network analysis reveals a spectrum model of human macrophage activation. Immunity *40*, 274–288.

OPEN
ACCESS
CellPress

Cell Reports
Article

Formation of Chromosomal Domains by Loop Extrusion

Geoffrey Fudenberg,[1,2,6] Maxim Imakaev,[3,6] Carolyn Lu,[4] Anton Goloborodko,[3] Nezar Abdennur,[5]
and Leonid A. Mirny[1,2,3,*]
[1]Graduate Program in Biophysics, Harvard University, Cambridge, MA 01238, USA
[2]Institute for Medical Engineering and Science, Massachusetts Institute of Technology (MIT), Cambridge, MA 02139, USA
[3]Department of Physics, Massachusetts Institute of Technology (MIT), Cambridge, MA 02139, USA
[4]Program for Research in Mathematics, Engineering and Science for High School Students (PRIMES) and Undergraduate Research Opportunities Program (UROP), MIT, Cambridge, MA 02139, USA
[5]PhD Program in Computational and Systems Biology, MIT, Cambridge, MA 02139, USA
[6]Co-first author, listed alphabetically
*Correspondence: leonid@mit.edu
http://dx.doi.org/10.1016/j.celrep.2016.04.085

SUMMARY

Topologically associating domains (TADs) are fundamental structural and functional building blocks of human interphase chromosomes, yet the mechanisms of TAD formation remain unclear. Here, we propose that loop extrusion underlies TAD formation. In this process, *cis*-acting loop-extruding factors, likely cohesins, form progressively larger loops but stall at TAD boundaries due to interactions with boundary proteins, including CTCF. Using polymer simulations, we show that this model produces TADs and finer-scale features of Hi-C data. Each TAD emerges from multiple loops dynamically formed through extrusion, contrary to typical illustrations of single static loops. Loop extrusion both explains diverse experimental observations—including the preferential orientation of CTCF motifs, enrichments of architectural proteins at TAD boundaries, and boundary deletion experiments—and makes specific predictions for the depletion of CTCF versus cohesin. Finally, loop extrusion has potentially far-ranging consequences for processes such as enhancer-promoter interactions, orientation-specific chromosomal looping, and compaction of mitotic chromosomes.

INTRODUCTION

Interphase chromosome organization in three dimensions underlies critical cellular processes, including gene regulation via enhancer-promoter interactions. Mapping chromosomal interactions genome-wide has revealed that interphase chromosomes of higher eukaryotes are partitioned at a sub-megabase scale into a sequence of self-interacting regions, termed topologically associating domains (TADs; Dixon et al., 2012; Nora et al., 2012), or domains (Rao et al., 2014; Sexton et al., 2012). An increasing number of studies have found important functional roles for TADs in the control of gene expression and development (Andrey et al., 2013; Lupiáñez et al., 2015; Symmons et al., 2014).

TADs are contiguous regions of enriched contact frequency that appear as squares in a Hi-C map (Figure 1A), which are relatively insulated from neighboring regions. Many TADs have homogeneous interiors, while others have particularly enriched boundaries, or even more complex features. More recently, high-resolution maps revealed peaks of interactions between loci at the boundaries of TADs (“peak loci”; Rao et al., 2014). TADs differ from larger scale A/B compartments in that they do not necessarily form an alternating “checkerboard” pattern of enriched contact frequencies (Lajoie et al., 2015), and several TADs often reside within a single contiguous compartment (Gibcus and Dekker, 2013; Gorkin et al., 2014) (Supplemental Notes).

Although often illustrated as such, several lines of evidence indicate that TADs are not simply stable loops formed between pairs of boundary loci. First, only 50% of TADs have corner-peaks (Rao et al., 2014). Second, boundary loci do not appear to be in permanent contact either by fluorescence in situ hybridization (FISH) (Rao et al., 2014) or by their relative contact frequency (see Results). Third, while TADs are enriched in contact probability throughout the domain, polymer simulations show that simple loops display enrichment only at the loop bases, unless the loop is very short (Benedetti et al., 2014; Doyle et al., 2014). For these reasons, identifying mechanisms of how TADs are formed remains an important open question.

While polymer models have provided insight into multiple levels of chromosome organization (Baù et al., 2011; Lieberman-Aiden et al., 2009; Marko and Siggia, 1997; Naumova et al., 2013; Rosa and Everaers, 2008), relatively few have focused on TADs. Of those that have considered TADs, some have focused primarily on characterizing chromosome structure rather than the mechanisms of folding (Giorgetti et al., 2014; Hofmann and Heermann, 2015). Others (Barbieri et al., 2012; Jost et al., 2014) have considered models where monomers of the same type experience preferential pairwise attractions to produce TADs; such models, however, when generalized to the genome-wide scale, would require a separate factor to recognize and compact each TAD. With only several types of monomers, this would produce checkerboard patterns for each type, which is characteristic of compartments rather than

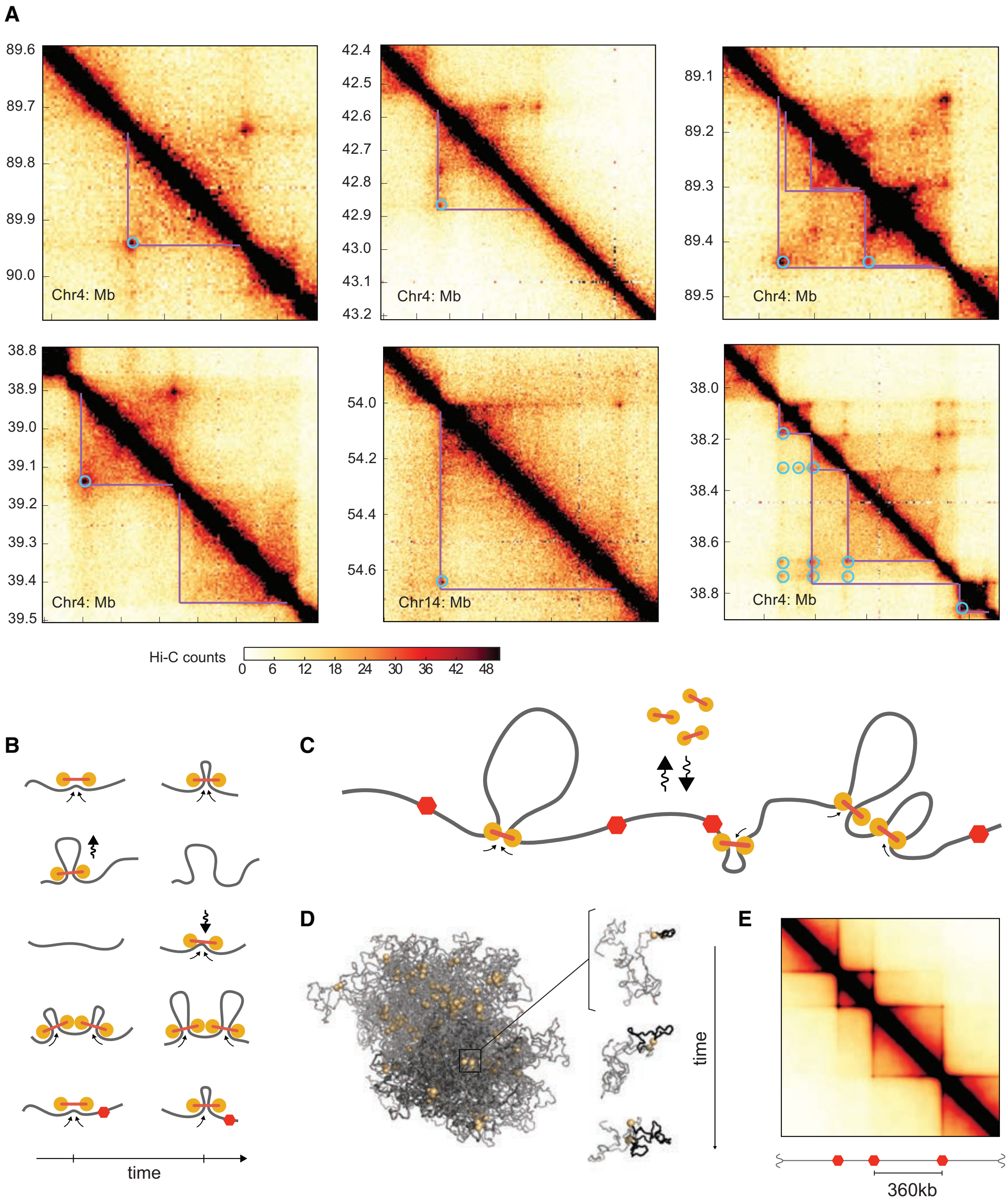


Figure 1. Loop Extrusion as a Mechanism for TAD Formation

(A) Hi-C contact maps at 5-kb resolution for six chromosomal regions (GM12878 in situ MboI) (Rao et al., 2014), highlighting TADs (purple lines) and peak loci (blue circles).

(B) Model of LEF dynamics (Figure S1A): LEFs shown as linked pairs of yellow circles, chromatin fiber shown in gray. From top to bottom: extrusion, dissociation, association, stalling upon encountering a neighboring LEF, stalling at a BE (red hexagon).

(C) Schematic of LEF dynamics (Movies S1 and S2).

(D) Conformation of a polymer subject to LEF dynamics, with processivity 120kb, separation 120kb. Left: LEFs (yellow) and chromatin (gray), for one conformation, where darker gray highlights the combined extent of three regions of sizes (180 kb, 360 kb, and 720 kb) separated by BEs. Right: the progressive extrusion of a loop (black) within a 180-kb region.

(E) Simulated contact map for processivity 120 kb and separation 120 kb.

TADs. One proposed mechanism giving good agreement to the observed TAD organization relies on supercoiling (Benedetti et al., 2014). Still, the connection between supercoiling and higher order eukaryotic chromosome organization remains unclear, since the reported agreement between supercoiling domain boundaries and TAD boundaries is roughly one in ten (Naughton et al., 2013).

Here, we propose a mechanism whereby TADs are formed by loop extrusion (Alipour and Marko, 2012; Nasmyth, 2001). In this process, *cis*-acting loop-extruding factors (LEFs; likely, cohesins) form progressively larger loops but are stalled by boundary elements (BEs), such as bound CTCF at TAD boundaries (Figures 1B and 1C). We tested this mechanism using polymer simulations of the chromatin fiber subject to the activity of LEFs. We found that it can produce TADs that quantitatively and qualitatively agree with Hi-C data. Importantly, our work provides a mechanism for preferentially forming contacts within TADs, such a mechanism is implicitly assumed in structural models of TADs formed by dynamic loops (Giorgetti et al., 2014; Hofmann and Heermann, 2015). Loop extrusion (Alipour and Marko, 2012), first introduced as processive loop enlargement by condensin (Nasmyth, 2001), has been implicated in mitotic chromosome compaction (Goloborodko et al., 2015; Naumova et al., 2013) and chromosome segregation in bacteria (Gruber, 2014; Wang et al., 2015). Importantly, however, these previous proposals did not consider any role of loop extrusion for TAD formation in interphase and did not directly test the impact of loop extrusion on 3D spatial organization or contact maps.

RESULTS

Mechanism of Loop Extrusion with BEs

To demonstrate how loop extrusion can lead to the formation of TADs, we first defined the dynamics of LEFs limited by BEs (Figures 1B and 1C; Figure S1A). Upon binding to the chromatin fiber, each LEF holds together two directly adjacent regions; then, it extrudes a loop by translocating along the chromatin fiber in both directions, holding together progressively more distant regions of a chromosome. Translocation stops when the LEF encounters an obstacle, either another LEF or a BE. If halted only on one side, LEFs continue to extrude on the other side. Throughout this process, LEFs can stochastically dissociate, releasing the extruded loop; for generality, we assume that this occurs uniformly across the genome. BEs underlie the formation of TADs by stalling LEF translocation, thus ensuring that extruded loops do not cross TAD boundaries. BEs in vivo might be formed by specifically bound architectural proteins, including CTCF, or any other impediment to LEF translocation. We note that BEs in vivo may be partially permeable because they either stochastically stall LEFs or are present in a fraction of cells.

Minimal 3D Model of Interphase Loop Extrusion

To efficiently explore how loop-extrusion dynamics spatially organize an interphase chromosome, we first studied a minimal model. In particular, we modeled a 10-Mb region of the chromatin fiber as a polymer subject to the activity of associating and dissociating LEFs limited by impermeable BEs (Figure 1C). As described previously (Naumova et al., 2013), we modeled the chromatin fiber as a polymer of 10-nm monomers (roughly three nucleosomes, or 600 bp) with excluded volume interactions and without topological constraints, subject to Langevin dynamics in OpenMM (Eastman et al., 2013). LEFs impose a system of bonds on the polymer: a bound LEF forms a bond between monomers at the two ends of an extruded loop, and the bond is re-assigned to increasingly separated pairs of monomers as a LEF translocates along the chromosome; when a LEF unbinds, this bond is removed. BEs, which halt LEF translocation, were placed at fixed positions, with sequential separations of 180 kb, 360 kb, and 720 kb through the 10-Mb region.

The dynamics of loop extrusion are determined by two independent parameters (Figure 2B; Figures S1B and S2): the average linear *separation* between bound LEFs, and the LEF *processivity*, i.e., the average size of a loop extruded by an unobstructed LEF over its lifetime (Goloborodko et al., 2015). Our model is additionally characterized by parameters governing the diffusivity of chromatin, polymer stiffness, density, and the Hi-C capture radius. For each set of parameter values, we ran polymer simulations long enough to allow $\geq$10 association/dissociation events per LEF (Movies S1 and S2). From simulations, we obtain an ensemble of chromosome conformations (Figure 1D) and compute the average contact frequency maps (“simulated Hi-C”; Figure 1E) that can be compared with experimental Hi-C data.

For many values of LEF processivity and separation, we observed the formation of TADs on a simulated Hi-C map (Figures 2C–2F). For some parameter values, we observed the formation of homogenous TADs; other simulated parameter sets led to the formation of peaks at corners of TADs, or enrichment of contacts at the boundary of TADs, seen as lines along the edge of a TAD. These simulations illustrate how the combined action of LEFs and BEs leads to enrichment of interactions within TADs and effective insulation between neighboring TADs.

TADs Are Formed by Dynamic Loops

Next, we tested the ability of our model to reproduce the Hi-C contact frequency as a function of genomic distance s ($P(s)$), used previously for quantifying polymer models (Barbieri et al., 2012; Benedetti et al., 2014; Le et al., 2013; Naumova et al., 2013; Rosa et al., 2010). We aimed to reproduce the $P(s)$ within TADs of sizes 180 kb, 360 kb, and 720 kb, as well as $P(s)$ between TADs. We determined the goodness of fit for each parameter set as the geometric standard deviation of the ratios of the four experimental and four simulated $P(s)$ curves (see Experimental Procedures). Note that $P(s)$ between TADs is $\sim$2-fold smaller and scales differently with distance (Figure 2A; Figure S2C). For each pair of values of LEF processivity and LEF separation, we quantified the best achieved goodness of fit and the number of times a pair appears among the top 100 out of 6,912 total parameter sets (Figure 2B).

We found that the best agreement with Hi-C data is achieved for LEF processivity of $\sim$120–240 kb and LEF separation of $\sim$120 kb (Figure 2B), where the resulting TADs consist of dynamically forming, growing, and dissociating loops (Figure 3A; Figures S1E–S1G). In this regime, LEFs extrude $\sim$75-kb loops relatively independently, as there are substantial gaps between LEFs (52%–69% average coverage of TADs by loops).

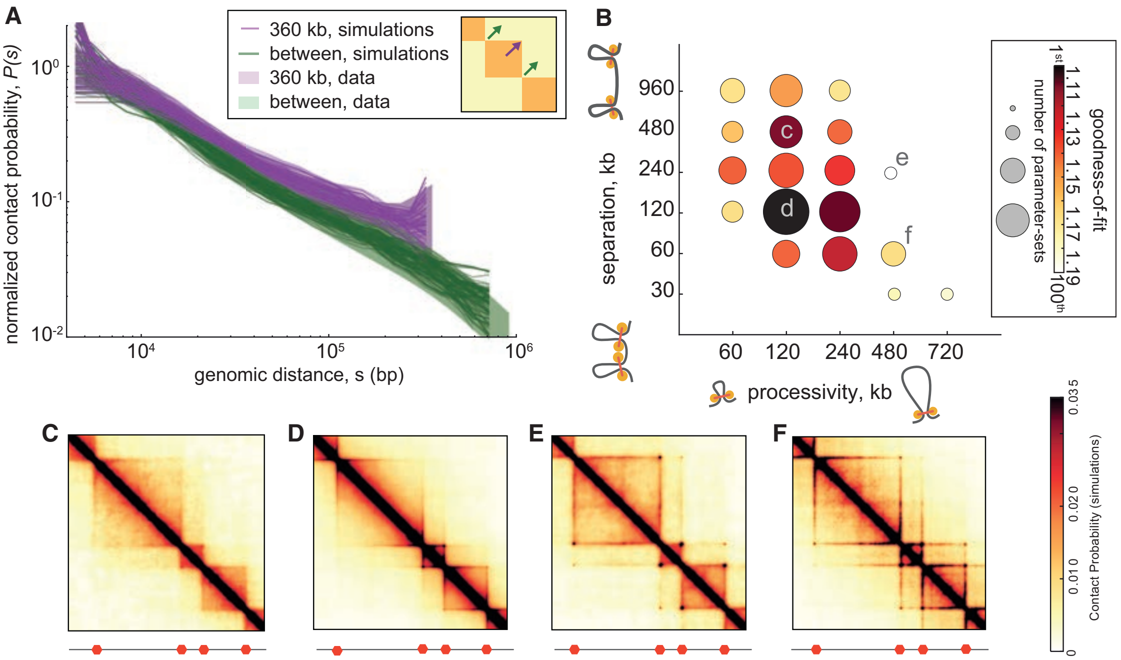


Figure 2. Quantitative Analysis of Loop Extrusion

(A) Experimental *P(s)* (shaded areas) versus simulated *P(s)* for the 100 best fitting parameter sets (lines, one per parameter set) within TADs (purple) and between TADs (green). Experimental *P(s)* calculated from 2-kb contact maps and normalized to one at 4 kb; shaded area shows 10th and 90th percentiles at each genomic distance. Simulated *P(s)* shown with vertical offsets from fitting (Experimental Procedures).

(B) Goodness of fit versus LEF processivity and separation for the 100 best fitting parameter sets (from 6,912 total parameter sets; Data S1). Circled areas represent the number of parameter sets among the top 100, while color quantifies the best fit at each processivity-separation pair; a value of 1 indicates a perfect fit.

(C–F) Simulated contact maps for the indicated processivity-separation pairs.

Notably, TADs are barely visible in simulated single-cell contact maps (Figure 3B). Moreover, only a small fraction of contacts on the map are directly mediated by LEFs. Consistently, polymer conformations display high spatial overlap between adjacent TADs, rather than appearing as segregated globules (Figure 3C; Figure S3A). Indeed, in both simulated and experimental Hi-C maps, there is only a ~2-fold depletion of contacts between neighboring TADs (Figure 2A; Figure S2C). Together, this demonstrates how TADs emerge as a population-average feature.

Loop Extrusion Recapitulates Results of Experimental TAD Boundary Deletions

Importantly, the mechanism of loop extrusion naturally recapitulates the results of TAD boundary deletion experiments (Nora et al., 2012). Upon the experimental deletion of a TAD boundary, the TAD spreads to the next boundary; this indicates that preferential interactions between loci in a TAD are not hard-wired and that BEs play crucial roles. This behavior has been confirmed with targeted disruption of CTCF-binding motifs at TAD boundaries (Guo et al., 2015; Narendra et al., 2015; Sanborn et al., 2015; de Wit et al., 2015). Consistently, in our model, deletion of a BE leads to spreading of a TAD until the next BE (Figure S1C).

TAD Corner-Peaks Are Not Permanent Loops

Many TADs appear to have peaks of interactions at their corners in Hi-C data (~50%; Rao et al., 2014). Interestingly, we found that TADs with and without peaks have similar *P(s)*, suggesting a similar underlying organizational mechanism, independent of the corner peak (Figure S2C). In agreement, our model shows that the mechanism of loop extrusion can produce both types of TADs, as increasing LEF processivity naturally strengthens peaks at TAD corners (Figures 2E and 2F; Figure S2A). Interestingly, our simulations show that TADs with visibly strong peaks do not require permanent contact between BEs, in agreement with our analyses of Hi-C data (Figures S4F and S4G).

Our simulations, together with previous polymer studies of chromatin loops (Benedetti et al., 2014; Doyle et al., 2014; Hofmann and Heermann, 2015), demonstrate that single stable loops are incapable of producing TADs. As follows, we refrain from directly identifying TAD corner peaks with loops. Indeed, we found that stable loops between BEs provide some of the worst fits to Hi-C data, with exceedingly strong corner peaks and a lack of visible TADs (Figure 4; Figure S4D). This stands in contrast with popular depictions of TADs as loops (Rao et al., 2014). Instead, our model predicts that TADs with and without corner peaks result from the collective activity of LEFs in the region between BEs.

TADs Require Long-Range Insulation

Importantly, insulation between neighboring TADs in our model does not arise from direct physical blocking of interactions by BEs. Instead, our model relies on the ability of BEs to regulate the translocation of LEFs. LEFs allow for insulation to be

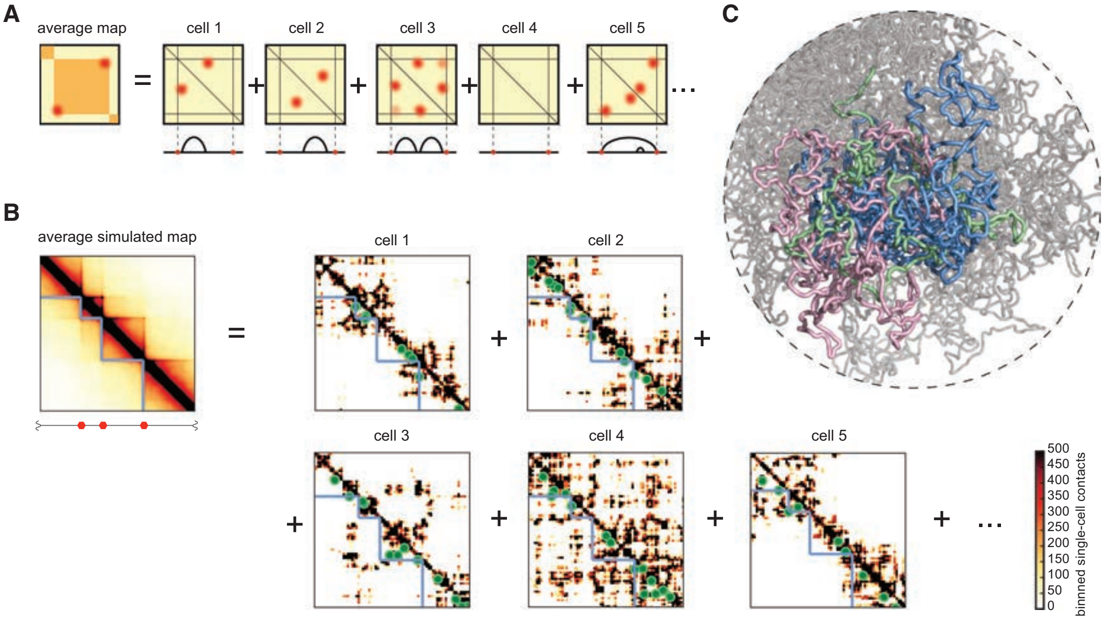


Figure 3. TADs Formed by LEFs Consist of Dynamically Forming, Growing, and Dissociating Loops

(A) Illustration of how TADs formed by loop extrusion result from averaging the dynamic positions of loop bases over many cells, including configurations with nested (cell 5) and consecutive (cell 3) loops (Figure S1).

(B) Left: the simulated contact map, as in Figure 1E, is an average of many single-cell maps. Right: simulated single-cell contact maps (18-kb resolution; green circles show LEF positions).

(C) Conformation of a polymer subject to LEF dynamics with processivity at 120 kb and separation at 120 kb. Three neighboring TADs of 180 kb, 360 kb, and 720 kb colored in green, pink, and blue, respectively. Contacts from an ensemble of such conformations are averaged together to form a contact map.

mediated over spatial and genomic distances much larger than the physical size of the BE. To rule out the possibility that a bulky BE is sufficient to insulate neighboring TADs, we performed simulations of this scenario. Indeed, in simulations where a BE is simply a bulky object, we see no long-range insulation and fail to obtain TADs (Figure 5; Figure S5). Similarly, in simulations where the chromatin fiber is locally very stiff at a BE, we again only see local insulation and fail to obtain TADs. Together, these simulations highlight the role of LEFs for imposing insulation at the scale of whole TADs.

Another important characteristic of our model is that loops extruded by LEFs act in *cis*, along the chromatin fiber, and do not impose interactions between genomically distal loci or loci on different chromosomes. Indeed, when we analyzed the interaction patterns of peak loci in Hi-C data, we found that there was no enrichment of contacts between pairs of peak loci at larger separations on the same chromosome or between different chromosomes (Figures S5G and S5H). This pattern is consistent with our model but is inconsistent with models that rely on direct interactions between BEs when such loci come into spatial proximity.

To rule out the mechanism whereby TADs are formed by direct BE-to-BE associations, we performed simulations where any two BEs would interact when they came into close spatial proximity (Figures S5I–S5L). Biologically, this represents a scenario where proteins interact to bridge cognate genomic elements (Barbieri et al., 2012; Bohn and Heermann, 2010; Brackley et al., 2015; Scolari and Cosentino Lagomarsino, 2015), for example, via interactions mediated by dimerization of bound CTCF. Our simulations confirmed that a direct BE-to-BE mechanism has no way of distinguishing between distant or proximal chromosomal regions; instead, all pairs of BEs display peaks of contact probability. Moreover, direct BE-to-BE interactions alone imposed negligible insulation between neighboring TADs, even in the case of strongly interacting BEs. Together, these results demonstrate the utility of LEFs stalled by BEs for restricting potentially interacting pairs of loci to those that are within TADs.

Molecular Roles for Cohesin and CTCF as LEFs and BEs

Next, we investigated possible molecular candidates for LEFs and BEs for TAD formation in interphase. We found that our proposed roles of cohesin as a LEF and CTCF as a BE both reconcile existing experimental results and predict outcomes of future experiments.

Multiple lines of evidence point to cohesin as a possible LEF in interphase. Like condensin, cohesin is a Structural Maintenance of Chromosomes (SMC) complex. These have been hypothesized to extrude chromatin loops (Alipour and Marko, 2012; Nasmyth, 2001), have similar molecular architectures to known motor proteins (Guacci et al., 1993; Nasmyth, 2001; Peterson, 1994), and have very recently been shown to slide along DNA (Stigler et al., 2016). Cohesins have been implicated in interphase TAD organization (Mizuguchi et al., 2014; Sofueva et al., 2013; Zuin et al., 2014) and chromatin looping (Kagey et al., 2010) beyond their role in sister chromatid cohesion. Indeed,

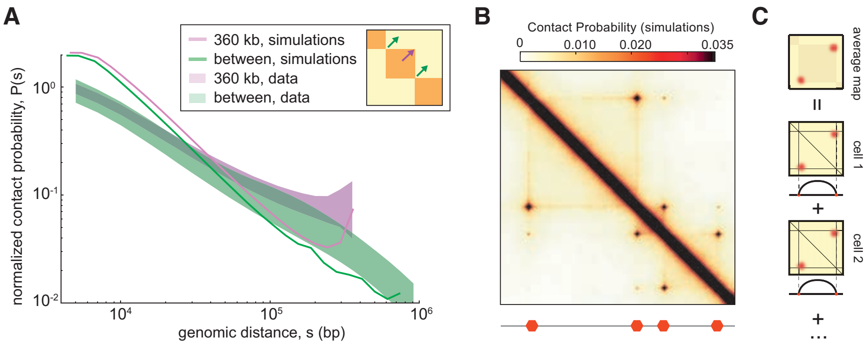


Figure 4. Simple Strong Loops Are Not TADs

(A) Experimental *P(s)* (shaded areas) versus simulated *P(s)* (solid lines) for a parameter set with a strong loop between neighboring BEs, calculated as in Figure 2A. Here, the fit is relatively poor (1.4137, rank 2,208 out of 6,912), and loops are not completely permanent, with BEs in contact 27% of the time for the 180-kb TAD and 14% of the time for the 720-kb TAD.

(B) Simulated contact map for a simple strong loop with processivity at 960 kb and separation at 960 kb.

(C) Illustration of how a single loop present in many cells leads to strong corner-peaks between neighboring BEs.

cohesins dynamically bind chromatin even before DNA replication (Gerlich et al., 2006). Finally, cohesin is enriched at interphase TAD boundaries (Dixon et al., 2012) and corner peaks (Rao et al., 2014), and its depletion makes TADs less prominent (Sofueva et al., 2013; Zuin et al., 2014).

CTCF is a similarly relevant molecular candidate for forming BEs. First, CTCF is enriched at TAD boundaries (Dixon et al., 2012), its depletion makes TADs less prominent (Zuin et al., 2014), and it has a relatively long residence time on chromatin (Nakahashi et al., 2013). Second, bound CTCF interacts with cohesin (Parelho et al., 2008) in an orientation-dependent manner (Xiao et al., 2011), similar to the interactions of shugoshin (Hara et al., 2014) and sororin (Nishiyama et al., 2010) with cohesin. Third, inward-oriented CTCF sites are enriched at TAD boundaries (Vietri Rudan et al., 2015) and TAD corner-peaks (Rao et al., 2014). We note that any impediment to LEF translocation may serve as a BE. As follows, BEs in vivo may be formed by sites with high occupancy of proteins other than CTCF (Van Bortle et al., 2014) that block LEF translocation physically rather than through a specific interaction. For example, active promoters bound by transcription-associated machinery are prominent candidates, as they are particularly bulky and are enriched at TAD boundaries (Dixon et al., 2012; Ulianov et al., 2016).

Loop Extrusion Predicts the Effects of Cohesin and CTCF Perturbations

The proposed roles of cohesin as a LEF and CTCF as a BE make predictions for patterns of cohesin and CTCF binding, genomic engineering experiments at particular loci, and global perturbations of cohesin and CTCF levels on chromosomal organization. First, analogous to LEF accumulation at BEs in our simulations (Figure 6B), we predict that cohesin accumulates at CTCF-binding sites but only when CTCF is bound at these sites (Parelho et al., 2008). Consistently, we find that cohesin chromatin immunoprecipitation sequencing (ChIP-seq) peaks are more enriched around strongly bound CTCF peaks (Figures 6C and 6D). This suggests that CTCF binding strength can translate into BE permeability.

Second, with these molecular roles, our simulations make predictions about changes in contact maps and spatial distances that would result from experimental perturbations to cohesin and CTCF (Figure S2). In particular, we predict that depletion of CTCF will result in increased permeability of BEs, thus reducing insulation between neighboring TADs, yet would have little effect on spatial distances of loci within TADs and only moderately reduce spatial distances of loci between TADs. In contrast, depletion of cohesin, modeled as an increased LEF separation, would also make TADs weaker but would be accompanied by more drastic increases of distances for loci both within and between TADs. Currently available Hi-C data support such differential, non-redundant, effects of CTCF and cohesin depletion on Hi-C maps (Zuin et al., 2014). Consistently, available imaging data support decompaction following cohesin depletion (Nolen et al., 2013; Sofueva et al., 2013; Zuin et al., 2014) and lack of decompaction following CTCF depletion (Nolen et al., 2013). Finally, our model predicts that greatly increased cohesin-binding time, modeled as greatly increased LEF processivity, would condense interphase chromosomes into a prophase-like "vermicelli" state (Figure S3), as seen upon depletion of the cohesin unloader Wapl (Tedeschi et al., 2013).

Complex TAD Architectures from Directional CTCF Boundaries

If bound CTCFs act as directional BEs, stopping LEF translocation from one side only, then the mechanism of loop extrusion can also explain the observed enrichment in convergent CTCF sites at TAD boundaries and loop bases, even at very large genomic separations (Figure 6A; Figure S6). Loop extrusion with directional BEs also provides a rationale for the results of manipulating CTCF site orientation, where flipping solely the orientation of a CTCF site can cause two neighboring TADs to merge (Guo et al., 2015; Narendra et al., 2015; Sanborn et al., 2015; de Wit et al., 2015). Interestingly, CTCF-binding sites at TAD boundaries are oriented so that the C terminus of bound CTCF (Nakahashi et al., 2013), known to interact with cohesin (Xiao et al., 2011), faces the interior of TADs. Indeed, we found that cohesin ChIP-seq peaks are enriched in this exact orientation-dependent manner around strongly bound CTCF peaks (Figures 6C–6F). Interestingly, YY1 and Znf143 were also enriched around strongly bound CTCF peaks, but only the former displayed an orientation dependence similar to cohesin. Together, these observations support a mechanism where CTCF acts as a BE that impedes loop extrusion by cohesins in an orientation-dependent manner.

To further test whether CTCF as a directional BE can recapitulate the variety of TAD domain architectures in vivo, we extended the minimal model introduced earlier to investigate a complex system of directional BEs with locus-specific

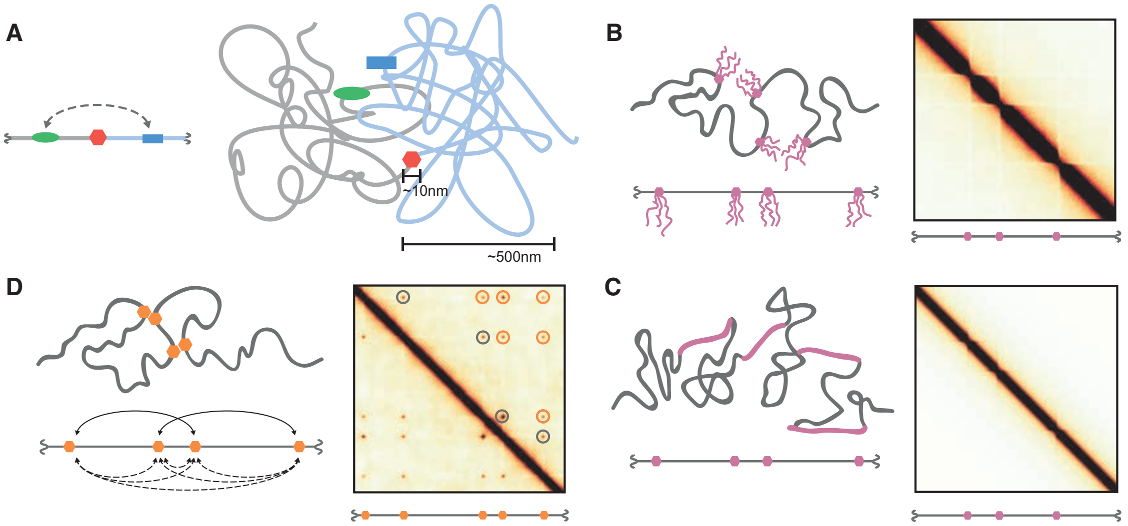


Figure 5. TADs Require Long-Range Insulation

(A) Illustration of a genomic region with an insulating element (red hexagon), a promoter (blue rectangle), and an enhancer (green oval) in 1D and 3D (Supplemental Notes).

(B) Illustration and contact map for a model of BEs as large bulky objects (e.g., bound by proteins or RNA). Each BE is bound by three polymer chains of length 10.

(C) As above, for a model of BEs as a stiff region of chromatin (ten monomers of stiffness 6).

(D) As above, for a model with direct BE-to-BE attraction (attraction strength 3). Solid arcs display interactions in this particular conformation, and dashed arcs show all possible interaction partners. Black circles indicate peaks between neighboring BEs, and orange circles indicate peaks between non-neighboring BEs.

permeability (Figure 7). For these simulations, we converted ChIP-seq data for CTCF over a 15-Mb region of human chromosome 14 (chr14) into BE permeability and directionality. First, since many CTCF peaks are not uniquely associated with a motif, we assigned directionality for ambiguous CTCF peaks using the nearest cohesin peak. Second, we transformed CTCF ChIP-seq peak heights into BE permeability using a logistic function such that more weakly bound CTCF sites formed more permeable BEs. We found that when the parameters of the best-fitting minimal model were used, this system produced contact maps with good agreement with Hi-C maps at short distances (<400 kb) but poor agreement at further distances (Figure S7). We found that better agreement at far distances can be obtained if we increased LEF processivity to 360 kb (corresponding to an increase in average loop size from ~75 kb to ~135 kb). Still, even with increased LEF processivity, agreement along the chromosome was non-uniform (Figure S7).

Upon inspecting these simulated maps (Figure 7; Figure S7), we noticed that strong CTCF sites tended to produce stronger boundaries, both in simulations and in the Hi-C data, while weak CTCF sites (e.g., 66.5 Mb) had minimal effects on both maps. Consistently, CTCF sites that are not at annotated TAD boundaries have lower CTCF occupancy (Figure S6D). In both maps, CTCF-depleted regions formed large homogenous domains (e.g., chr14:62.6–63.6 Mb). Moreover, we found that our simulations recapitulated several other features of TAD organization, including: nested TADs (e.g., 62 Mb), strong interactions between the BE and the body of the domain (“lines” at the edge, or perimeter, of the TAD, e.g., 65.9 Mb), and complex networks of interactions between several BEs (“grids” of peaks, e.g., 66 Mb). These results illustrate how not only TADs but also finer structures similar to those found in experimental Hi-C maps naturally emerged in our simulations without being directly encoded; this further supports loop extrusion as a mechanism of chromosomal organization in interphase.

DISCUSSION

In summary, our model of loop extrusion not only forms TADs in agreement with Hi-C but also agrees with a number of specific features observed experimentally. First, loop extrusion naturally produces enrichment of contacts within a domain and can create corner peaks between TAD boundaries. Second, loop extrusion naturally explains finer structures, including nested TADs, lines at TAD edges, and grids of peaks. Third, it explains why peaks of interactions are absent between distal BEs and BEs on different chromosomes. Fourth, it explains the merging of TADs seen in boundary deletion experiments. Fifth, it provides a rationale for inward-oriented CTCF motifs at TAD boundaries. Sixth, it explains a directional bias in the accumulation of cohesin peaks around bound CTCF motifs. Seventh, it is consistent with current Hi-C and microscopy results for CTCF and cohesin depletion. Finally, it allows for interaction peaks between BEs that are not simple stable loops, as necessary for consistency with available Hi-C and microscopy data (Imakaev et al., 2015). Given the ability of loop extrusion to recapitulate complex locus-specific folding patterns, we propose that TADs may be better understood as one of many emergent phenomena from loop extrusion, rather than a discrete and distinct set of genomic entities.

Loop extrusion as a model of TAD formation in interphase has recently received a significant amount of attention. In Nichols and Corces (2015), the authors proposed, but did not quantitatively

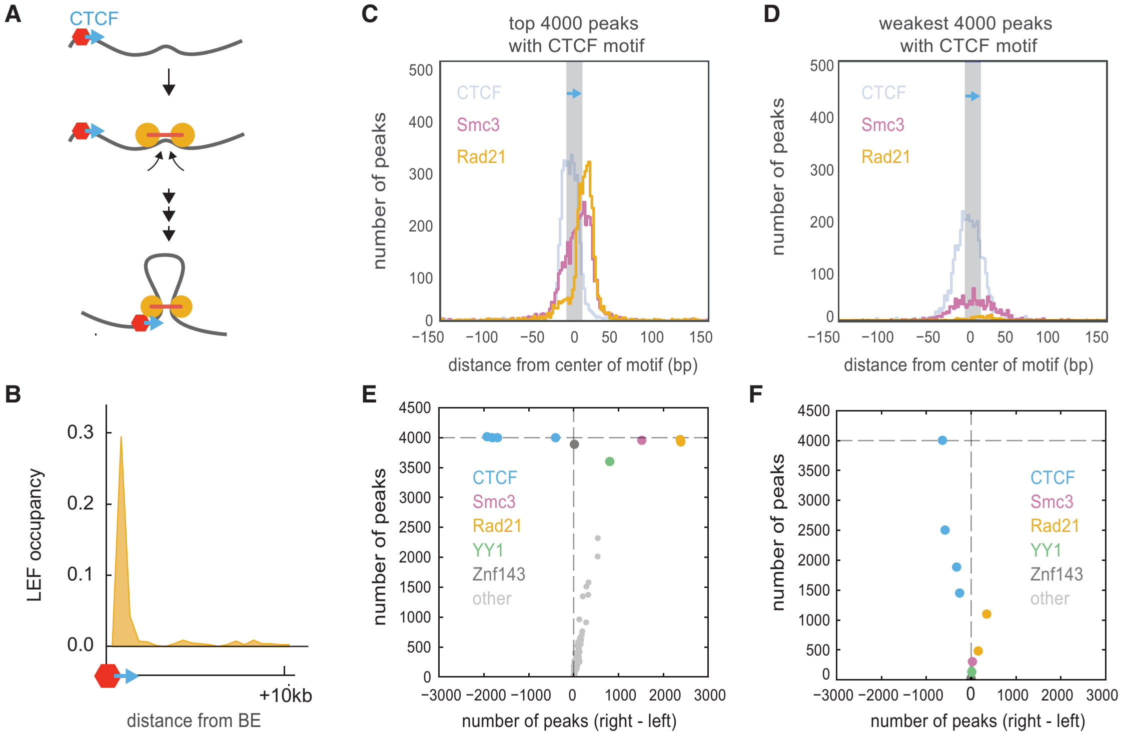


Figure 6. CTCF as a Directional BE

(A) Inward-oriented CTCF sites at TAD boundaries are consistent with loop extrusion and a directional boundary function of CTCF (Figure S6).

(B) Accumulation of LEFs at BEs for simulations with processivity at 120 kb and separation at 120 kb.

(C) Distributions of CTCF, Smc3, and Rad21 ChIP-seq peak summits in the vicinity of the 4,000 strongest motif-associated CTCF binding peaks (orientation indicated by blue arrow).

(D) Same as in (C), but for the weakest 4000 motif-associated CTCF-binding sites.

(E) Asymmetry and enrichment of factor ChIP-seq peak summits around the strongest 4,000 motif-associated CTCF sites. Each dot represents an ENCODE GM12878 ChIP-seq track. The y axis shows the number of peaks within ±200 bp of a CTCF motif. The x axis shows the difference between the number of peak summits on the right and on the left of the motif, i.e., asymmetry of the factor relative to a CTCF motif.

(F) Same, but for the weakest 4,000 motif-associated CTCF ChIP peaks.

test, that bound CTCF can load cohesin in an orientation-specific manner, holding one side of the cohesin complex while the other translocates to form an extruded loop. We note that, if the majority of cohesin is loaded in this fashion, the results of cohesin depletion would be similar to those of CTCF depletion, in contrast with current experimental results reporting non-redundant effects (Zuin et al., 2014); in the context of loop extrusion, details of LEF loading and BE function are important subjects for future study. In Sanborn et al. (2015), published while our manuscript was available as a preprint (http://biorxiv.org/content/early/2015/08/14/024620), the authors considered a tension globule model as well as a loop extrusion model similar to the one presented here. Surprisingly, for both the tension globule and the loop extrusion models, they found good agreement with experimental *P(s)* that is robust to changes in simulation parameters. In contrast, for loop extrusion we found that *P(s)* depends strongly on the parameters of LEF dynamics. This may be because we considered a wider range of LEF processivities and separations, spanning from a free-polymer regime to a permanent loop regime, to a compacted "vermicelli" regime. Also, in contrast with Sanborn et al. (2015), where 13 out of 13 genomic engineering experiments were successfully predicted, we found a non-uniform agreement as a function of genomic position between the results of our simulations and Hi-C maps. We believe that our observation of non-uniform agreement along the chromosome can reflect: additional undetermined factors underlying BEs, locus-specific details of LEF dynamics (including sites of loading and unloading), the role of higher-order active and inactive compartments (Brackley et al., 2015; Jost et al., 2014) and lamina associations (Kind et al., 2015), or locus-specific experimental details of Hi-C and ChIP-seq (Imakaev et al., 2012; Yaffe and Tanay, 2011).

Consideration of a wide range of LEF processivities and separations allowed us to make predictions regarding the effects of CTCF and cohesin perturbations. Indeed, our simulations predict distinct consequences of CTCF and cohesin depletion on Hi-C maps and spatial distances, consistent with available imaging and Hi-C data (Nolen et al., 2013; Sofueva et al., 2013; Zuin et al., 2014). Still, further validation of our predictions requires new methods for architectural protein removal, as available techniques have yet to fully disrupt TAD formation. Additionally, we note that other proteins complexes may play important roles in chromosome organization, potentially via mechanisms distinct from loop extrusion; for example, bound polycomb may be able to self-associate, thus compacting polycomb-bound

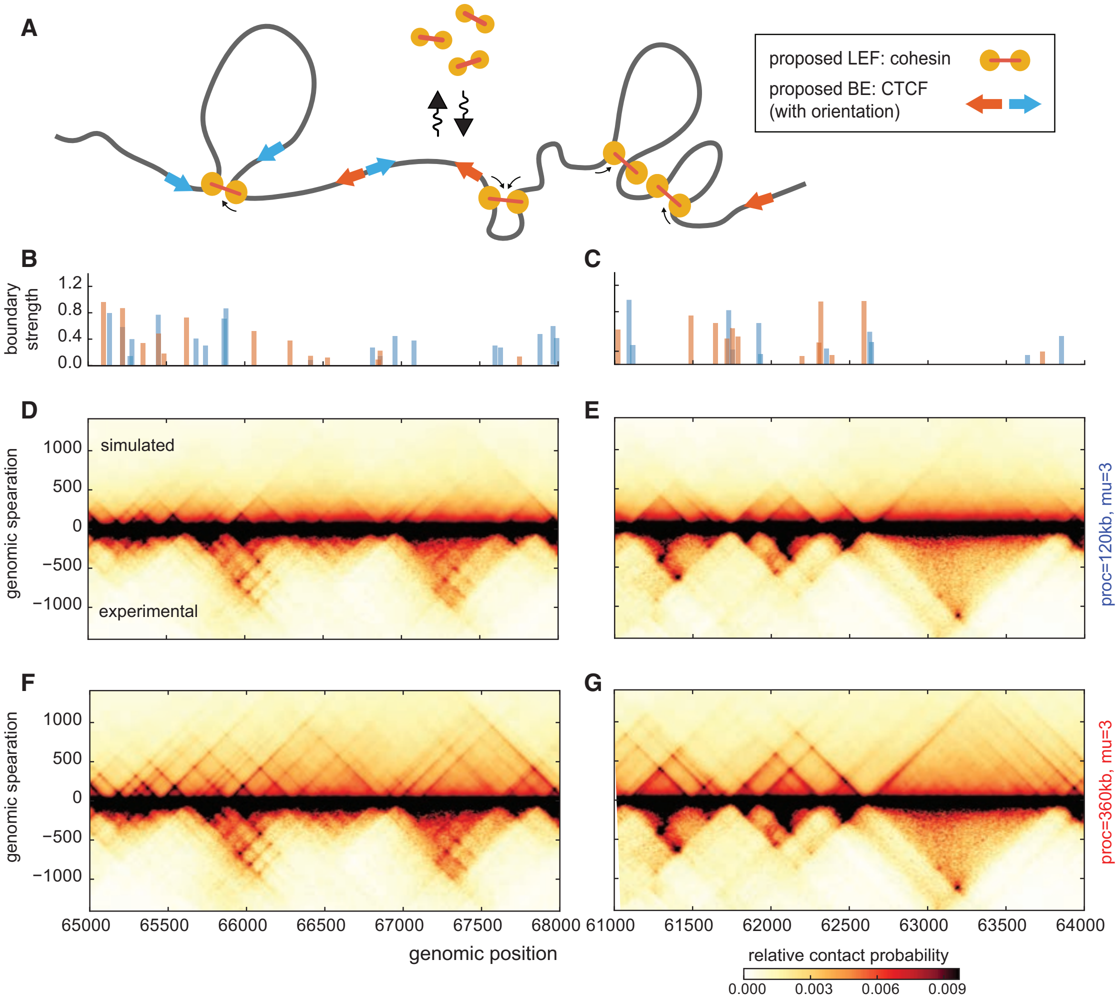


Figure 7. Complex TAD Architectures from Loop Extrusion

(A) Schematic of LEF dynamics with directional BEs.

(B and C) Directional BE strength profile (the sum of BEs occupancies within a 12-kb bin) for regions simulated in (D–G).

(D–G) Simulated contact maps for regions of human chr14, GM12878 cell type, for models with orientation-specific BEs of varying permeability. Maps are compared with experimental maps for the same regions at the same 12-kb resolution (Figure S7). LEF processivity is 120 kb (D and E) and 360 kb (F and G).

regions (Boettiger et al., 2016) and possibly contributing to the formation of TAD-like domains (Williamson et al., 2014).

The mechanism of loop extrusion in interphase has additional, potentially far-ranging, consequences for processes in the nucleus. First, enhancer-promoter pairings can be dictated by the relative placement of BEs, including CTCF (Hou et al., 2008). Second, loop extrusion may have an even stronger effect if LEFs stall at promoters, effectively turning the enhancer-promoter search process into a 1D search process and allowing for orientation-specific interactions. Third, loop extrusion may facilitate high-fidelity VDJ and class-switch recombination, as well as other processes dependent on long-range intra-chromosomal looping with specific orientations, particularly given the observed interplay between CTCF and cohesin (Alt et al., 2013; Degner et al., 2011; Dong et al., 2015; Lin et al., 2015).

Finally, the mechanism of TAD formation via loop extrusion studied here is similar to the proposed mechanism of mitotic chromosome condensation (Alipour and Marko, 2012; Goloborodko et al., 2015, 2016; Nasmyth, 2001; Naumova et al., 2013) but with the addition of BEs and many fewer, less processive, LEFs. Accordingly, increasing the number and processivity of LEFs and removing BEs could underlie the transition from interphase to mitotic chromosome organization. Conversely, upon exit from mitosis, interphase 3D chromosome organization can be re-established by restoring previous BE positions, which could potentially be epigenetically inherited bookmarks (Kadauke and Blobel, 2013).

EXPERIMENTAL PROCEDURES

LEF Dynamics with BEs

LEF translocation along a chromatin fiber was simulated on a 1D lattice, where each position was characterized by the following parameters: association (birth) probability, dissociation (death) probability, and BE occupancy (stalling probability). For the minimal model, we considered a system with uniform birth probability, constant death probability, a fixed number of LEFs, and a discrete number of completely impermeable BEs. As described previously (Goloborodko et al., 2015), we modeled LEFs as having two "heads" connected by a linker.

3D Simulations

We represent chromatin fibers as a polymer of spherical monomers connected by harmonic bonds, with stiffness and a soft-core repulsive potential. Simulations were performed with Langevin dynamics in periodic boundary conditions using OpenMM (Eastman and Pande, 2010; Eastman et al., 2013). The two monomers held by the two heads of each LEF were connected by a harmonic bond. A 3D-to-1D dynamics parameter controlled the number of 3D-simulation time steps per 1D-simulation time steps.

Minimal 3D Polymer Model

We performed simulations of a polymer chain consisting of eight groups of three TADs of 300, 600, and 1,200 monomers each, arranged sequentially (300; 600; 1,200; 300; …). Impermeable BEs were placed between neighboring TADs. For each parameter set, we collected 2,000 conformations to calculate *P(s)* and 10,000 for contact maps.

Parameter Sweep

We considered the following values for the five simulation parameters:

LEF processivity: 100; 200; 400; 800; 1,200; 1,600
LEF separation: 50; 100; 200; 400; 800; 1,600
3D-to-1D dynamics: 300/4; 1,000/4; 5,000/4
Stiffness: 0; 2; 4; 6
Density: 0.05; 0.2.

For each of these 864 separate simulations, we generated 2,000 conformations (total: 1,728,000), and calculated contact maps for the eight values of the Hi-C capture radius (the distance at which contacts between two monomers are recorded; 2, 3, 4, 5, 6, 7, 8, and 10), for a total of 6,912 parameters sets.

Experimental P(s) *and Hi-C Maps*

To calculate experimental *P(s)* within and between TADs, we used publicly available data and annotations from Rao et al. (2014). Data were processed in house using *hiclib* (Imakaev et al., 2012) for the GM12878 inSitu protocol and MboI restriction enzyme, binned at 2 kb. Displayed Hi-C maps were at a 5-kb resolution, processed similarly.

Goodness of Fit

To compare experimental and simulated *P(s)*, we averaged over experimental TADs 0.9–1.1 of the size of a simulated TAD, (180 kb, 360 kb, and 720 kb). The goodness of fit was the geometric standard deviation of the ratio of simulated to experimental *P(s)*. Since the best fitting models had diverse parameter sets, we took the first 100 best fitting models (fit values, 1.103–1.195) and assessed how frequently each pair (processivity, separation) occurs in this list and what the best fit was for each pair.

ChIP-Seq Peaks around Oriented CTCF Motifs

Motifs were assigned to ENCODE narrow-peak calls in the GM12878 cell line by interval intersection, using bedtools (Quinlan and Hall, 2010). Genome-wide CTCF motif matches were also obtained from ENCODE (Kheradpour and Kellis, 2014), using the CTCF_known1 motif. For 4,000 most and least enriched CTCF peaks, we produced histograms of the summit positions of ENCODE-called ChIP peaks for other factors.

Converting CTCF ChIP-Seq to Orientation-Specific BE Permeability

To convert ChIP-seq peak strength to the occupancy of simulated BEs, we used a logistic transformation, $f(x) = 1/(1 + \exp(-x/20 - \mu))$, where x is a sum of peak fold-change-over-input values for peaks with a given orientation in a 600-bp bin. $\mu = 3$ was used as the default, and $\mu = 2$ or $\mu = 4$ was used for higher or lower CTCF occupancy. Peak orientation was determined by the motif orientation; when unavailable, we used relative orientation of CTCF and the closest Rad21 peak.

3D Polymer Model with Orientation-Specific BEs

We modeled a 15-Mb region of human chr14, 60,000,000 to 75,000,000, using the same parameters as the best fitting minimal model. We also considered simulations with different processivity and CTCF occupancy. The simulated contact map was rescaled to 12-kb resolution and compared to GM12878 in situ data from Rao et al. (2014), corrected at 10 kb, and rescaled to 12 kb.

Models of Direct BE-to-BE Interactions, Bulky and Stiff BEs

Models had the same BE positions as in the minimal model but did not include LEF dynamics. To simulate direct BE-to-BE interactions, we used an attractive soft-core potential. Contact maps were displayed for attractive strengths of 1.5, 3, and 5, and a contact radius of 10. For bulky BEs, several polymer chains were connected at each BE (either three chains of length 10 attached one per monomer to the monomers around the BE via harmonic bonds, or five chains of length 6). For stiff BEs, the 10 monomers around the BE had an increased stiffness of 6, while other monomers had a stiffness of 1, as defined earlier.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, Supplemental Notes, seven figures, two movies, and one database file and can be found with this article online at http://dx.doi.org/10.1016/j.celrep.2016.04.085.

AUTHOR CONTRIBUTIONS

G.F., M.I., and L.A.M. conceived of the project and wrote the paper. C.L. performed initial simulations. N.A. led ChIP-seq and motif analyses. A.G. led literature curation for CTCF and cohesin. M.I. led polymer simulations. G.F. led experimental data analyses.

ACKNOWLEDGMENTS

M.I. and G.F. would like to thank the MIT PRIMES: Program for Research in Mathematics, Engineering, and Science for High School Students for the opportunity to supervise excellent students. This project began as C.L.'s project for the MIT PRIMES, supervised by M.I. and G.F., and initial results were presented at the annual PRIMES conference in May 2013 (https://math.mit.edu/research/highschool/primes/materials/2013/conf/12-2-Lu.pdf). Work including detailed comparison to experimental Hi-C data was presented at conferences in Heidelberg (May 2015) and at the Weizmann Institute (June 2015), and was available as a preprint at bioRxiv, August 2015 (http://biorxiv.org/content/early/2015/08/14/024620). The authors thank Job Dekker for many stimulating discussions. The authors also thank Edith Heard, Suzana Hadjur, Katerina Kraft, Elphege Nora, Francois Spitz, and other attendees of the Genome Regulation in 3D Conference at the Weizmann Institute for helpful discussions and feedback. We are thankful to Maxime Dahan for many productive discussions and Labex and Institut Curie for hosting L.M., G.F., and A.G. during this work. Authors are supported by grants R01 GM114190, U54 DK107980, and R01 G003143 from the National Institute of Health and 1504942 from the National Science Foundation.

Received: October 12, 2015
Revised: December 1, 2015
Accepted: April 20, 2016
Published: May 19, 2016

REFERENCES

Alipour, E., and Marko, J.F. (2012). Self-organization of domain structures by DNA-loop-extruding enzymes. Nucleic Acids Res. *40*, 11202–11212.

Alt, F.W., Zhang, Y., Meng, F.-L., Guo, C., and Schwer, B. (2013). Mechanisms of programmed DNA lesions and genomic instability in the immune system. Cell *152*, 417–429.

Andrey, G., Montavon, T., Mascrez, B., Gonzalez, F., Noordermeer, D., Leleu, M., Trono, D., Spitz, F., and Duboule, D. (2013). A switch between topological domains underlies HoxD genes collinearity in mouse limbs. Science *340*, 1234167.

Barbieri, M., Chotalia, M., Fraser, J., Lavitas, L.-M., Dostie, J., Pombo, A., and Nicodemi, M. (2012). Complexity of chromatin folding is captured by the strings and binders switch model. Proc. Natl. Acad. Sci. USA *109*, 16173–16178.

Baù, D., Sanyal, A., Lajoie, B.R., Capriotti, E., Byron, M., Lawrence, J.B., Dekker, J., and Marti-Renom, M.A. (2011). The three-dimensional folding of the α-globin gene domain reveals formation of chromatin globules. Nat. Struct. Mol. Biol. *18*, 107–114.

Benedetti, F., Dorier, J., Burnier, Y., and Stasiak, A. (2014). Models that include supercoiling of topological domains reproduce several known features of interphase chromosomes. Nucleic Acids Res. *42*, 2848–2855.

Boettiger, A.N., Bintu, B., Moffitt, J.R., Wang, S., Beliveau, B.J., Fudenberg, G., Imakaev, M., Mirny, L.A., Wu, C., and Zhuang, X. (2016). Super-resolution imaging reveals distinct chromatin folding for different epigenetic states. Nature *529*, 418–422.

Bohn, M., and Heermann, D.W. (2010). Diffusion-driven looping provides a consistent framework for chromatin organization. PLoS ONE *5*, e12218.

Brackley, C.A., Johnson, J., Kelly, S., Cook, P.R., and Marenduzzo, D. (2015). Binding of bivalent transcription factors to active and inactive regions folds human chromosomes into loops, rosettes and domains. arXiv, arXiv: 1511.01848. http://arxiv.org/abs/1511.01848.

de Wit, E., Vos, E.S.M., Holwerda, S.J.B., Valdes-Quezada, C., Verstegen, M.J.A.M., Teunissen, H., Splinter, E., Wijchers, P.J., Krijger, P.H.L., and de Laat, W. (2015). CTCF binding polarity determines chromatin looping. Mol. Cell *60*, 676–684.

Degner, S.C., Verma-Gaur, J., Wong, T.P., Bossen, C., Iverson, G.M., Torkamani, A., Vettermann, C., Lin, Y.C., Ju, Z., Schulz, D., et al. (2011). CCCTC-binding factor (CTCF) and cohesin influence the genomic architecture of the Igh locus and antisense transcription in pro-B cells. Proc. Natl. Acad. Sci. USA *108*, 9566–9571.

Dixon, J.R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J.S., and Ren, B. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature *485*, 376–380.

Dong, J., Panchakshari, R.A., Zhang, T., Zhang, Y., Hu, J., Volpi, S.A., Meyers, R.M., Ho, Y.-J., Du, Z., Robbiani, D.F., et al. (2015). Orientation-specific joining of AID-initiated DNA breaks promotes antibody class switching. Nature *525*, 134–139.

Doyle, B., Fudenberg, G., Imakaev, M., and Mirny, L.A. (2014). Chromatin loops as allosteric modulators of enhancer-promoter interactions. PLoS Comput. Biol. *10*, e1003867.

Eastman, P., and Pande, V. (2010). OpenMM: a hardware-independent framework for molecular simulations. Comput. Sci. Eng. *12*, 34–39.

Eastman, P., Friedrichs, M.S., Chodera, J.D., Radmer, R.J., Bruns, C.M., Ku, J.P., Beauchamp, K.A., Lane, T.J., Wang, L.P., Shukla, D., et al. (2013). OpenMM 4: a reusable, extensible, hardware independent library for high performance molecular simulation. J. Chem. Theory Comput. *9*, 461–469.

Gerlich, D., Koch, B., Dupeux, F., Peters, J.-M., and Ellenberg, J. (2006). Live-cell imaging reveals a stable cohesin-chromatin interaction after but not before DNA replication. Curr. Biol. *16*, 1571–1578.

Gibcus, J.H., and Dekker, J. (2013). The hierarchy of the 3D genome. Mol. Cell *49*, 773–782.

Giorgetti, L., Galupa, R., Nora, E.P., Piolot, T., Lam, F., Dekker, J., Tiana, G., and Heard, E. (2014). Predictive polymer modeling reveals coupled fluctuations in chromosome conformation and transcription. Cell *157*, 950–963.

Goloborodko, A., Marko, J.F., and Mirny, L. (2015). Chromosome compaction via active loop extrusion. Biophys J. *110*, 2162–2168.

Goloborodko, A.A., Imakaev, M., Marko, J., and Mirny, L. (2016). Compaction and segregation of sister chromatids via active loop extrusion. eLife. http://dx.doi.org/10.7554/eLife.14864.

Gorkin, D.U., Leung, D., and Ren, B. (2014). The 3D genome in transcriptional regulation and pluripotency. Cell Stem Cell *14*, 762–775.

Gruber, S. (2014). Multilayer chromosome organization through DNA bending, bridging and extrusion. Curr. Opin. Microbiol. *22*, 102–110.

Guacci, V., Yamamoto, A., Strunnikov, A., Kingsbury, J., Hogan, E., Meluh, P., and Koshland, D. (1993). Structure and function of chromosomes in mitosis of budding yeast. Cold Spring Harb. Symp. Quant. Biol. *58*, 677–685.

Guo, Y., Xu, Q., Canzio, D., Shou, J., Li, J., Gorkin, D.U., Jung, I., Wu, H., Zhai, Y., Tang, Y., et al. (2015). CRISPR inversion of CTCF sites alters genome topology and enhancer/promoter function. Cell *162*, 900–910.

Hara, K., Zheng, G., Qu, Q., Liu, H., Ouyang, Z., Chen, Z., Tomchick, D.R., and Yu, H. (2014). Structure of cohesin subcomplex pinpoints direct shugoshin-Wapl antagonism in centromeric cohesion. Nat. Struct. Mol. Biol. *21*, 864–870.

Hofmann, A., and Heermann, D.W. (2015). The role of loops on the order of eukaryotes and prokaryotes. FEBS Lett. *589* (*20 Pt A*), 2958–2965.

Hou, C., Zhao, H., Tanimoto, K., and Dean, A. (2008). CTCF-dependent enhancer-blocking by alternative chromatin loop formation. Proc. Natl. Acad. Sci. USA *105*, 20398–20403.

Imakaev, M., Fudenberg, G., McCord, R.P., Naumova, N., Goloborodko, A., Lajoie, B.R., Dekker, J., and Mirny, L.A. (2012). Iterative correction of Hi-C data reveals hallmarks of chromosome organization. Nat. Methods *9*, 999–1003.

Imakaev, M.V., Fudenberg, G., and Mirny, L.A. (2015). Modeling chromosomes: Beyond pretty pictures. FEBS Lett. *589* (*20 Pt A*), 3031–3036.

Jost, D., Carrivain, P., Cavalli, G., and Vaillant, C. (2014). Modeling epigenome folding: formation and dynamics of topologically associated chromatin domains. Nucleic Acids Res. *42*, 9553–9561.

Kadauke, S., and Blobel, G.A. (2013). Mitotic bookmarking by transcription factors. Epigenetics Chromatin *6*, 6.

Kagey, M.H., Newman, J.J., Bilodeau, S., Zhan, Y., Orlando, D.A., van Berkum, N.L., Ebmeier, C.C., Goossens, J., Rahl, P.B., Levine, S.S., et al. (2010). Mediator and cohesin connect gene expression and chromatin architecture. Nature *467*, 430–435.

Kheradpour, P., and Kellis, M. (2014). Systematic discovery and characterization of regulatory motifs in ENCODE TF binding experiments. Nucleic Acids Res. *42*, 2976–2987.

Kind, J., Pagie, L., de Vries, S.S., Nahidiazar, L., Dey, S.S., Bienko, M., Zhan, Y., Lajoie, B., de Graaf, C.A., Amendola, M., Fudenberg, G., Imakaev, M., Mirny, L.A., Jalink, K., Dekker, J., van Oudenaarden, A., and van Steensel, B. (2015). Genome-wide maps of nuclear lamina interactions in single human cells. Cell *163*, 134–147.

Lajoie, B.R., Dekker, J., and Kaplan, N. (2015). The hitchhiker's guide to Hi-C analysis: practical guidelines. Methods *72*, 65–75.

Le, T.B.K., Imakaev, M.V., Mirny, L.A., and Laub, M.T. (2013). High-resolution mapping of the spatial organization of a bacterial chromosome. Science *342*, 731–734.

Lieberman-Aiden, E., van Berkum, N.L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B.R., Sabo, P.J., Dorschner, M.O., et al. (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science *326*, 289–293.

Lin, S.G., Guo, C., Su, A., Zhang, Y., and Alt, F.W. (2015). CTCF-binding elements 1 and 2 in the Igh intergenic control region cooperatively regulate V(D)J recombination. Proc. Natl. Acad. Sci. USA *112*, 1815–1820.

Lupiáñez, D.G., Kraft, K., Heinrich, V., Krawitz, P., Brancati, F., Klopocki, E., Horn, D., Kayserili, H., Opitz, J.M., Laxova, R., et al. (2015). Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. Cell *161*, 1012–1025.

Marko, J.F., and Siggia, E.D. (1997). Polymer models of meiotic and mitotic chromosomes. Mol. Biol. Cell *8*, 2217–2231.

Mizuguchi, T., Fudenberg, G., Mehta, S., Belton, J.-M., Taneja, N., Folco, H.D., FitzGerald, P., Dekker, J., Mirny, L., Barrowman, J., and Grewal, S.I. (2014).

Cohesin-dependent globules and heterochromatin shape 3D genome architecture in S. pombe. Nature *516*, 432–435.

Nakahashi, H., Kwon, K.R., Resch, W., Vian, L., Dose, M., Stavreva, D., Hakim, O., Pruett, N., Nelson, S., Yamane, A., et al. (2013). A genome-wide map of CTCF multivalency redefines the CTCF code. Cell Rep. *3*, 1678–1689.

Narendra, V., Rocha, P.P., An, D., Raviram, R., Skok, J.A., Mazzoni, E.O., and Reinberg, D. (2015). CTCF establishes discrete functional chromatin domains at the Hox clusters during differentiation. Science *347*, 1017–1021.

Nasmyth, K. (2001). Disseminating the genome: joining, resolving, and separating sister chromatids during mitosis and meiosis. Annu. Rev. Genet. *35*, 673–745.

Naughton, C., Avlonitis, N., Corless, S., Prendergast, J.G., Mati, I.K., Eijk, P.P., Cockroft, S.L., Bradley, M., Ylstra, B., and Gilbert, N. (2013). Transcription forms and remodels supercoiling domains unfolding large-scale chromatin structures. Nat. Struct. Mol. Biol. *20*, 387–395.

Naumova, N., Imakaev, M., Fudenberg, G., Zhan, Y., Lajoie, B.R., Mirny, L.A., and Dekker, J. (2013). Organization of the mitotic chromosome. Science *342*, 948–953.

Nichols, M.H., and Corces, V.G. (2015). A CTCF code for 3D genome architecture. Cell *162*, 703–705.

Nishiyama, T., Ladurner, R., Schmitz, J., Kreidl, E., Schleiffer, A., Bhaskara, V., Bando, M., Shirahige, K., Hyman, A.A., Mechtler, K., and Peters, J.M. (2010). Sororin mediates sister chromatid cohesion by antagonizing Wapl. Cell *143*, 737–749.

Nolen, L.D., Boyle, S., Ansari, M., Pritchard, E., and Bickmore, W.A. (2013). Regional chromatin decompaction in Cornelia de Lange syndrome associated with NIPBL disruption can be uncoupled from cohesin and CTCF. Hum. Mol. Genet. *22*, 4180–4193.

Nora, E.P., Lajoie, B.R., Schulz, E.G., Giorgetti, L., Okamoto, I., Servant, N., Piolot, T., van Berkum, N.L., Meisig, J., Sedat, J., et al. (2012). Spatial partitioning of the regulatory landscape of the X-inactivation centre. Nature *485*, 381–385.

Parelho, V., Hadjur, S., Spivakov, M., Leleu, M., Sauer, S., Gregson, H.C., Jarmuz, A., Canzonetta, C., Webster, Z., Nesterova, T., et al. (2008). Cohesins functionally associate with CTCF on mammalian chromosome arms. Cell *132*, 422–433.

Peterson, C.L. (1994). The SMC family: novel motor proteins for chromosome condensation? Cell *79*, 389–392.

Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841–842.

Rao, S.S.P., Huntley, M.H., Durand, N.C., Stamenova, E.K., Bochkov, I.D., Robinson, J.T., Sanborn, A.L., Machol, I., Omer, A.D., Lander, E.S., and Aiden, E.L. (2014). A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell *159*, 1665–1680.

Rosa, A., and Everaers, R. (2008). Structure and dynamics of interphase chromosomes. PLoS Comput. Biol. *4*, e1000153.

Rosa, A., Becker, N.B., and Everaers, R. (2010). Looping probabilities in model interphase chromosomes. Biophys. J. *98*, 2410–2419.

Sanborn, A.L., Rao, S.S.P., Huang, S.-C., Durand, N.C., Huntley, M.H., Jewett, A.I., Bochkov, I.D., Chinnappan, D., Cutkosky, A., Li, J., et al. (2015). Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. Proc. Natl. Acad. Sci. USA *112*, E6456–E6465.

Scolari, V.F., and Cosentino Lagomarsino, M. (2015). Combined collapse by bridging and self-adhesion in a prototypical polymer model inspired by the bacterial nucleoid. Soft Matter *11*, 1677–1687.

Sexton, T., Yaffe, E., Kenigsberg, E., Bantignies, F., Leblanc, B., Hoichman, M., Parrinello, H., Tanay, A., and Cavalli, G. (2012). Three-dimensional folding and functional organization principles of the Drosophila genome. Cell *148*, 458–472.

Stigler, J., Çamdere, G.Ö., Koshland, D.E., and Greene, E.C. (2016). Single-molecule imaging reveals a collapsed conformational state for DNA-bound cohesin. Cell Rep. *15*, 988–998.

Sofueva, S., Yaffe, E., Chan, W.-C., Georgopoulou, D., Vietri Rudan, M., Mira-Bontenbal, H., Pollard, S.M., Schroth, G.P., Tanay, A., and Hadjur, S. (2013). Cohesin-mediated interactions organize chromosomal domain architecture. EMBO J. *32*, 3119–3129.

Symmons, O., Uslu, V.V., Tsujimura, T., Ruf, S., Nassari, S., Schwarzer, W., Ettwiller, L., and Spitz, F. (2014). Functional and topological characteristics of mammalian regulatory domains. Genome Res. *24*, 390–400.

Tedeschi, A., Wutz, G., Huet, S., Jaritz, M., Wuensche, A., Schirghuber, E., Davidson, I.F., Tang, W., Cisneros, D.A., Bhaskara, V., et al. (2013). Wapl is an essential regulator of chromatin structure and chromosome segregation. Nature *501*, 564–568.

Ulianov, S.V., Khrameeva, E.E., Gavrilov, A.A., Flyamer, I.M., Kos, P., Mikhaleva, E.A., Penin, A.A., Logacheva, M.D., Imakaev, M.V., Chertovich, A., et al. (2016). Active chromatin and transcription play a key role in chromosome partitioning into topologically associating domains. Genome Res. *26*, 70–84.

Van Bortle, K., Nichols, M.H., Li, L., Ong, C.-T., Takenaka, N., Qin, Z.S., and Corces, V.G. (2014). Insulator function and topological domain border strength scale with architectural protein occupancy. Genome Biol. *15*, R82.

Vietri Rudan, M., Barrington, C., Henderson, S., Ernst, C., Odom, D.T., Tanay, A., and Hadjur, S. (2015). Comparative Hi-C reveals that CTCF underlies evolution of chromosomal domain architecture. Cell Rep. *10*, 1297–1309.

Wang, X., Le, T.B.K., Lajoie, B.R., Dekker, J., Laub, M.T., and Rudner, D.Z. (2015). Condensin promotes the juxtaposition of DNA flanking its loading site in Bacillus subtilis. Genes Dev. *29*, 1661–1675.

Williamson, I., Berlivet, S., Eskeland, R., Boyle, S., Illingworth, R.S., Paquette, D., Dostie, J., and Bickmore, W.A. (2014). Spatial genome organization: contrasting views from chromosome conformation capture and fluorescence in situ hybridization. Genes Dev. *28*, 2778–2791.

Xiao, T., Wallace, J., and Felsenfeld, G. (2011). Specific sites in the C terminus of CTCF interact with the SA2 subunit of the cohesin complex and are required for cohesin-dependent insulation activity. Mol. Cell. Biol. *31*, 2174–2183.

Yaffe, E., and Tanay, A. (2011). Probabilistic modeling of Hi-C contact maps eliminates systematic biases to characterize global chromosomal architecture. Nat. Genet. *43*, 1059–1065.

Zuin, J., Dixon, J.R., van der Reijden, M.I.J., Ye, Z., Kolovos, P., Brouwer, R.W., van de Corput, M.P., van de Werken, H.J., Knoch, T.A., van IJcken, W.F., et al. (2014). Cohesin and CTCF differentially affect chromatin architecture and gene expression in human cells. Proc. Natl. Acad. Sci. USA *111*, 996–1001.

OPEN
ACCESS
CellPress

Cell Reports
Resource

Rapid Changes in the Translatome during the Conversion of Growth Cones to Synaptic Terminals

Kelvin Xi Zhang,[1] Liming Tan,[1] Matteo Pellegrini,[2] S. Lawrence Zipursky,[1,*] and Jason M. McEwen[1]
[1]Department of Biological Chemistry, Howard Hughes Medical Institute, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles CA 90095, USA
[2]Department of Molecular, Cell and Developmental Biology, University of California Los Angeles, P.O. Box 951606, Los Angeles, CA 90095, USA
*Correspondence: lzipursky@mednet.ucla.edu

http://dx.doi.org/10.1016/j.celrep.2015.12.102

SUMMARY

A common step in the formation of neural circuits is the conversion of growth cones to presynaptic terminals. Characterizing patterns of global gene expression during this process is problematic due to the cellular diversity of the brain and the complex temporal dynamics of development. Here, we take advantage of the synchronous conversion of *Drosophila* photoreceptor growth cones into presynaptic terminals to explore global changes in gene expression during presynaptic differentiation. Using a tandemly tagged ribosome trap (T-TRAP) and RNA sequencing (RNA-seq) at multiple developmental times, we observed dramatic changes in coding and non-coding RNAs with presynaptic differentiation. Marked changes in the mRNA encoding transmembrane and secreted proteins occurred preferentially. The 3′ UTRs of transcripts encoding synaptic proteins were preferentially lengthened, and these extended UTRs were preferentially enriched for sites recognized by RNA binding proteins. These data provide a rich resource for uncovering the regulatory logic underlying presynaptic differentiation.

INTRODUCTION

Uncovering the genetic programs regulating neural development remains a central problem in neuroscience. Genetic screens have uncovered genes regulating specific steps in neural circuit formation, including cell fate determination, axon guidance, and synapse formation (Desai et al., 1988; Newsome et al., 2000; Zallen et al., 1999; Kidd et al., 1998). Biochemical and molecular biological approaches have also led to the identification of specific genes and the proteins they encode which regulate distinct steps in circuit assembly (Serafini et al., 1994; Cheng et al., 1995; Schmucker et al., 2000). These single gene approaches have been highly effective in dissecting the genetic programs regulating specific steps in neural development.

Recent advances in high-throughput RNA sequencing (RNA-seq) technologies provide an opportunity to assess the global patterns of gene expression during neural development and to relate these to specific genetic programs. In principle, global studies would be most informative in systems where traditional genetic approaches could be applied to establish causal relationships alluded to from correlations discovered in global studies. *Drosophila* offers numerous advantages for combining these approaches. Many cell-type-specific markers enable isolation of mRNA from specific neurons and a palette of genetic approaches allow for incisive phenotypic analyses (Nern et al., 2015; Venken et al., 2011). The fly visual system is particularly well-suited to high-throughput sequencing approaches, as there are numerous different neuronal cell types, and many of these are present in relatively large numbers (Tuthill et al., 2013).

The fly visual system comprises the retina or compound eye that detects light and the optic lobe or ganglia that process visual information. The eye contains some 750 modules called ommatidia and they, in turn, each comprise eight photoreceptor (R) cells. These include R1–R6 neurons, which form synapses with target neurons in the first optic ganglion, the lamina, and the R7 and R8 neurons with distinct visual pigments and synaptic targets in the second optic ganglion, the medulla. Although early steps in R cell development occur in this asynchronous fashion (Hadjieconomou et al., 2011), the conversion of R cell growth cones to synaptic terminals occurs synchronously (Chen et al., 2014). This provides a unique opportunity to explore changes in gene expression occurring in a global fashion associated with this critical step in neuronal differentiation.

Presynaptic differentiation has been studied through genetic and biochemical approaches. It is a complex process involving multiple steps (Chia et al., 2013). These include the transport of active zone and synaptic vesicle components to the presynaptic terminus and the complex interplay between these and the actin cytoskeleton during the assembly of the presynaptic apparatus. In addition, this must be coordinated with the recognition between appropriate synaptic partners, the assembly of an adhesive synaptic complex between the pre- and postsynaptic membrane and the maintenance through secreted signals at synapses.

Here, we set out to provide a description of the changes in gene expression that occur during presynaptic differentiation at a global scale. Several different approaches have been described for isolating RNA from neurons for mRNA sequencing studies. These include using fluorescent cell tags for isolating neurons by manual picking after tissue dissociation or through fluorescence-activated cell sorting (FACs) (Abruzzi et al., 2015). The INTACT technique utilizes cell-specific expression of a nuclear envelope tag followed by affinity purification of nuclei (Henry et al., 2012; Steiner et al., 2012). The tagged ribosomal affinity purification (TRAP) technique utilizes affinity purification of ribosomes associated with mRNA from lysates prepared from tissues where an epitope-tagged ribosomal protein is selectively expressed in neurons of interest (Doyle et al., 2008; Heiman et al., 2008). Each of these techniques leads to the isolation of a specific pool of mRNA and has provided effective ways of characterizing expression of genes in specific cell types.

As we were particularly interested in understanding the interactions of developing growth cones with their environment and, hence, the proteome rather than the transcriptome, we sought a method of mRNA isolation more likely to reflect differences in protein expression than gene expression more generally. We, thus, set out to identify transcripts associated with ribosomes in R cells using a modified version of the TRAP technique, called tandemly(T)-TRAP. Here, we incorporated an additional tag for a second round of affinity purification of ribosomes from specific neurons. This two-step purification increased the specificity of mRNA isolation, thereby significantly reducing the noise from non-specific association of transcripts expressed in other cells encountered using TRAP. In this paper, we describe changes in gene expression in R cells during presynaptic differentiation. These studies uncovered prominent and dynamic changes in patterns of expression of cell surface and secreted proteins associated with the transformation of growth cones to synaptic terminals and suggest a selective role for post-transcriptional regulatory mechanisms in regulating presynaptic differentiation.

RESULTS

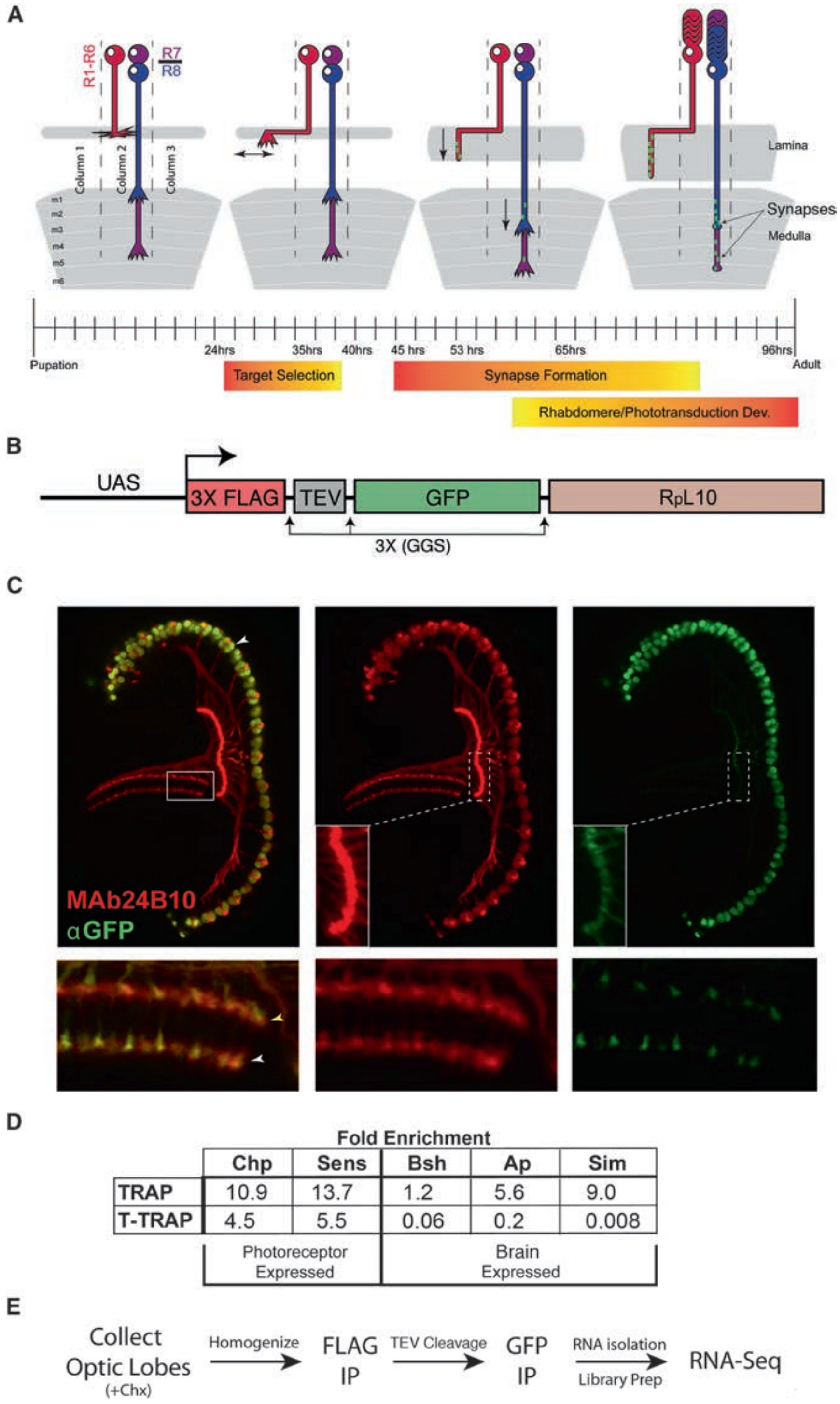
T-TRAP Purification of mRNA

To assess changes in gene expression associated with the conversion of R cell growth cones to presynaptic terminals, we purified mRNA associated with ribosomes selectively from R cells before, during, and after synapse formation (Figure 1A; see below) and analyzed these via RNA-seq. To increase the signal to noise in the TRAP method (Doyle et al., 2008; Heiman et al., 2008), we included a second tag to facilitate a second affinity purification step. We refer to this method as T-TRAP (see below and Experimental Procedures). We modified the N terminus of the *Drosophila* ribosomal protein RpL10 with two tandemly arranged epitopes, 3X FLAG and GFP, separated by the Tobacco Etch Virus (TEV) protease site and expressed this in specific cell types using the GAL4/upstream activating system (UAS) (Figure 1B). In brief, ribosomes from extracts prepared in the presence of cycloheximide were immunoprecipitated with anti-FLAG antibody containing beads, followed by cleavage with TEV protease to release the ribosome from the beads. Ribosomes were then immunoprecipitated again with antibodies to GFP, mRNA was extracted, and RNA-seq libraries were prepared (see Experimental Procedures).

To profile R cell mRNA, we used *chaoptic-Gal4* (*chp-Gal4*) to selectively drive the modified RpL10 protein under UAS control in these cells. The tagged ribosomal protein was seen in the cell bodies and, at lower levels, in axon terminals of R cells (Figure 1C). No evidence of toxicity was observed at the level of light microscopy; R cell morphology was indistinguishable from wild-type as assessed at multiple stages of development and in the adult. In addition, defects were not observed in animals in which the tag was expressed ubiquitously during development. Quantitative western blots revealed that the level of tagged Rpl10 protein normalized to the number of R cells remained constant between 40 and 65 hr; the levels at 24 and 35 hr were 33% and 160% to that observed at 40 hr. As normalized transcript levels were compared these differences do not affect comparisons of reads per kilobase of a specific mRNA per million reads (RPKM) values between samples.

To assess purification, we used qPCR to compare the enrichment of known R cell-specific genes (e.g., [*chp*] and *senseless*) to genes expressed in the optic lobe that are not expressed in R cells (e.g., *apterous*, *brain-specific homeobox*, and *single-minded*). Levels of two retina-specific and three optic lobe-specific transcripts were assessed in extracts prepared from dissected tissues including both the optic lobe and retina at 24 hr after pupae formation (APF) and compared to the levels in the complexes prepared by TRAP and T-TRAP. The levels were normalized to the ribosomal protein RpL11 in extracts and in the immunoprecipitants. For TRAP, the enrichment of retinal-specific genes over optic lobe genes was between one to ten times (Figures 1D and S2). By contrast, for T-TRAP the enrichment through the two-step purification was 25–500 times (see Experimental Procedures for calculations). This increase in enrichment is attributed to a significant depletion of optic lobe transcripts. As expected, this increase in specificity through the second round of immunoprecipitation comes at a cost of decreased yield of mRNA (~30% compared to TRAP).

cDNA libraries were prepared from mRNA associated with the affinity purified ribosomes and sequenced using an Illumina HiSeq 2000 (Figures 1E and S1; Experimental Procedures). As libraries were generated by first strand synthesis using oligo dT, there was a 3′-bias in the sequences obtained. The correlation in the distribution of RPKM values between biological replicates for each time point was excellent with an average Pearson's r of 0.99 (Figures 2A and 2B). By contrast, the further time points were away from each other, the lower the correlation score (i.e., from 0.97 to 0.77). Importantly, the correlation of RNA-seq data between biological replicates using T-TRAP was considerably improved over TRAP (0.99 versus 0.93) (Figures 2A and 2B; see Experimental Procedures for details). This is consistent with the additional affinity purification step resulting in an increase in signal-to-noise. We performed a principal component analysis on R cell expression profiles and found seven distinct subpopulations during R cell development, clearly distinguishable by the first principal component (PC1, 38% of the total variation; Figure 2C). We restricted our analysis to genes with RPKM values >1 at one or more time points (see Experimental Procedures). A total of 9,806 genes fell into this category. This



Fold Enrichment

	Chp	Sens	Bsh	Ap	Sim
TRAP	10.9	13.7	1.2	5.6	9.0
T-TRAP	4.5	5.5	0.06	0.2	0.008
	Photoreceptor Expressed		Brain Expressed		

Figure 1. T-TRAP Method for RNA-Seq from Developing R Cells

(A) Schematic of R cell differentiation during pupal development. Target selection period refers to R1–R6. Target selection for R7 and R8 overlaps with the early phase of synapse formation (i.e., post 45 hr).

(B) Schematic of the tandem-tagged ribosomal affinity purification (T-TRAP) construct. The N terminus of the *Drosophila* ribosomal protein L10 (RpL10) coding region was fused to two tandemly arranged epitopes (3xFLAG and GFP) separated by the TEV protease site. 3x (GS) refers to three tandemly arrayed tripeptide sequences (glycine, glycine, serine) used as linker segments.

(C) Tagged RpL10 was selectively expressed in R cells using Gal4 driven by a Chp enhancer element. Upper panels show low magnification, 63x images of eye and part of the optic lobes (i.e., the lamina and medulla comprising synaptic targets for R cell axons) at 24 hr APF. Tagged RpL10 protein accumulates in R cell bodies (white arrowhead), and in the growth cones of R1–R6 (insets), and R7 and R8 (lower panels: R7, yellow arrowhead; R8, white arrowhead). The tagged RpL10 was visualized with an antibody to GFP and R cells were visualized with MAb24B10.

(D) qPCR from RNA isolated from R cells using TRAP and T-TRAP. The TRAP construct comprises from the N terminus, GFP followed by RpL10. Enrichment represents the fold increase between the level of the mRNA indicated and an mRNA encoding a ribosomal subunit RpL11 (see text).

(E) Summary of T-TRAP RNA-seq. Chx, cycloheximide.

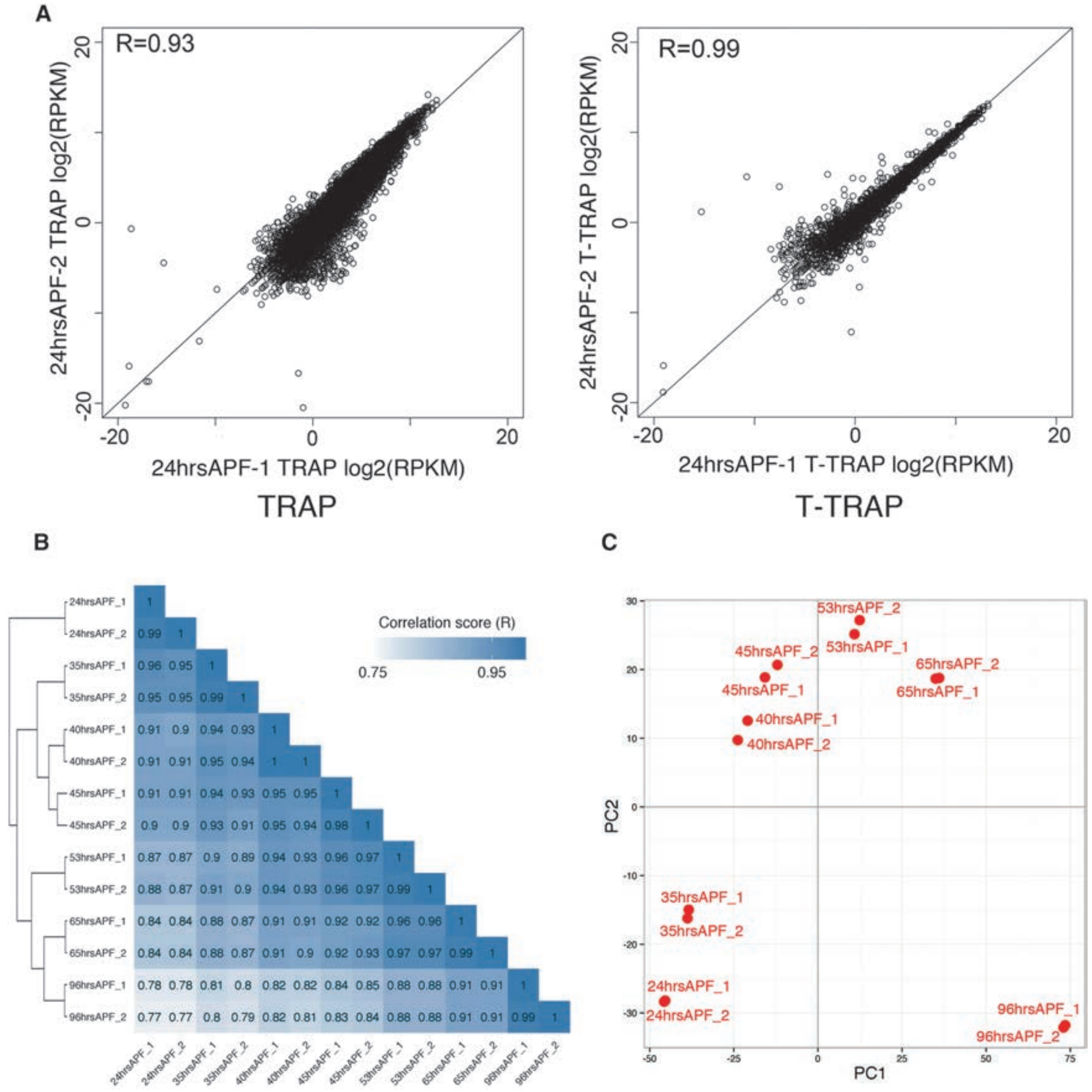
See also Figures S1 and S2.

represents 66.3% of the 14,794 genes encoding proteins annotated in *Drosophila*. In addition to mRNA encoding proteins, the expression of 578 long intergenic non-coding RNAs (lincRNAs) were also observed with RPKM values >1 at one or more time points. We describe the analysis of lincRNAs in the Supplementary Information and in Figure S3 and Table S3.

Dynamic Changes in Gene Expression during Conversion of Growth Cones to Presynaptic Terminals

Although early R cell development is asynchronous, from 24 APF onward the development of R cell terminals occurs in a largely synchronous fashion. We chose seven time points for RNA-seq analysis of R cells during pupal development corresponding to 24, 35, 40, 45, 53, 65, and 96 hr APF. These time intervals correlate with specific steps in R cell development (see Figure 1A):

1. 24–35 hr APF: R1–R6 growth cones undergo a rearrangement within the lamina plexus (Clandinin and Zipursky, 2000; Lee et al., 2003; Bate, 1993). No changes are seen in the growth cones of R7 and R8.
2. 35–40 hr APF: growth cones do not alter their positions or their morphologies.



	24hrsAPF_1	24hrsAPF_2	35hrsAPF_1	35hrsAPF_2	40hrsAPF_1	40hrsAPF_2	45hrsAPF_1	45hrsAPF_2	53hrsAPF_1	53hrsAPF_2	65hrsAPF_1	65hrsAPF_2	96hrsAPF_1	96hrsAPF_2
24hrsAPF_1	1													
24hrsAPF_2	0.99	1												
35hrsAPF_1	0.96	0.95	1											
35hrsAPF_2	0.95	0.95	0.99	1										
40hrsAPF_1	0.91	0.9	0.94	0.93	1									
40hrsAPF_2	0.91	0.91	0.95	0.94	1	1								
45hrsAPF_1	0.91	0.91	0.94	0.93	0.95	0.95	1							
45hrsAPF_2	0.9	0.9	0.93	0.91	0.95	0.94	0.98	1						
53hrsAPF_1	0.87	0.87	0.9	0.89	0.94	0.93	0.96	0.97	1					
53hrsAPF_2	0.88	0.87	0.91	0.9	0.94	0.93	0.96	0.97	0.99	1				
65hrsAPF_1	0.84	0.84	0.88	0.87	0.91	0.91	0.92	0.92	0.96	0.96	1			
65hrsAPF_2	0.84	0.84	0.88	0.87	0.91	0.9	0.92	0.93	0.97	0.97	0.99	1		
96hrsAPF_1	0.78	0.78	0.81	0.8	0.82	0.82	0.84	0.85	0.88	0.88	0.91	0.91	1	
96hrsAPF_2	0.77	0.77	0.8	0.79	0.82	0.81	0.83	0.84	0.88	0.88	0.91	0.91	0.99	1

Figure 2. Correlation of RNA-Seq Data from Different Libraries

(A) Scatterplot comparison of biological replicates using the single purification approach (TRAP) and the double purification approach (T-TRAP).

(B) Correlograms showing the correlation score matrix across all libraries at all time points.

(C) Principal component analysis (PCA) of expression in R cells for seven time points. PC1, first principle component; PC2, second principle component. R, Pearson correlation coefficient.

3. 40–45 hr APF: this interval marks the onset of the morphological transformation of R cell growth cones of each class into presynaptic terminals (Chen et al., 2014).
4. 45–53 hr APF: conversion of bulbous growth cones to elongated shape of synaptic terminals is complete. Bruchpilot (Brp) puncta, markers for the presynaptic active zone, begin to accumulate (Chen et al., 2014).
5. Post 53 hr APF: postsynaptic differentiation in target neurons occurs (Chen et al., 2014). Maturation of synaptic terminals overlaps with the development of the rhabdomere, the photosensitive organelle containing rhodopsin (Kumar and Ready, 1995).

In summary, the conversion of R cell growth cones to differentiating presynaptic structures takes place between 35 and 53 hr APF, spanning four time points in our time series.

We sought to assess how many genes exhibited marked changes in their levels of expression between different time points (Figure 3). Here, we performed the differential expression analysis between two successive time points with an adjusted p value < 0.001 and plotted the number of genes, with RPKM values >1 at any given time point, that changed five times or more (Figure 3A). As expected, there is a substantial change during the 30 hr time interval between 65–96 hr APF corresponding to the massive change in morphology and the acquisition of phototransduction properties in these cells. Importantly, changes in the expression of many genes were seen in two narrow time intervals (i.e., 5 hr, 35–40 hr, and 40–45 hr), corresponding to a period prior to and during the onset of presynaptic differentiation, respectively. Between 35–40 hr APF, just before the onset of presynaptic differentiation, ∼450 genes were up or downregulated by greater than five times with more genes downregulated during this interval. A similar number of genes were regulated by greater than five times between 40–45 hr, corresponding to the onset of presynaptic differentiation. By contrast, considerably more genes were upregulated than downregulated during this second interval. Interestingly, the changes during the subsequent two intervals (45–53 hr APF and 53–65 hr APF) were less pronounced. The rate of addition of presynaptic active zones in all three classes of photoreceptor neurons during this period occurs in a largely linear fashion (Chen et al., 2014). Thus, marked changes in gene expression were seen just prior to synapse formation (i.e., 35–40 hr APF) and coincident with the onset of the conversion of a growth cone to a synaptic terminal (i.e., 40–45 hr).

Changes in mRNA Encoding Cell Surface and Secreted Proteins Associated with Ribosomes during Conversion of Growth Cones to Presynaptic Terminals

To obtain a global view of the biological processes represented in these differentially expressed genes, we carried out Gene Ontology (GO) term enrichment analysis using a p value of 0.05 (Figure 3B). There was an increase in genes expressed during later stages (from 65–96 hr APF) involved in photosensitivity and neuronal function, consistent with the terminal differentiation of these cells into photoreceptor neurons. In addition, a marked change in gene expression was observed between 40–45 hr APF coincident with the onset of presynaptic differentiation. GO enrichment analysis revealed preferential increases in the expression of genes during this interval involved in "cell surface receptor-linked signal transduction," "cell adhesion," "extracellular structure organization," "regulation of membrane potential," "ion transport," and "neurological system process." The marked changes in chitin metabolism seen in this time interval were also observed in T-TRAP of lamina neurons (data not shown) and thus this is likely to represent background due to high levels of expression associated with the formation of the lens and other chitinous tissues. The enrichment in "cell surface-linked signal transduction" and "cell adhesion" and "extracellular structure organization" are consistent with the cellular processes associated with the transformation of the growth cone to a presynaptic terminal.

Interestingly, in contrast to the increase in "cell surface receptor-linked signal transduction" and "cell adhesion" gene categories observed between 40–45 hr APF, there was a substantial decrease in the levels of this class of transcripts between 35–40 hr APF (see below). This corresponds to a stage in which growth cones largely remain "quiescent," where there are no obvious changes in growth cone dynamics or position. Thus, a marked re-programming of gene expression regulating intercellular communication occurs during two sequential time intervals corresponding to a period following axon guidance and the onset of the conversion of R cell growth cones to presynaptic terminals.

We sought to explore the expression pattern of genes encoding cell surface and secreted proteins in more detail. Examination of GO terms in FlyBase revealed that some proteins with known functions in axon guidance and cell adhesion were not included in the GO terms "cell adhesion" and "cell surface receptor-linked signal transduction." To provide a more comprehensive assessment of the expression patterns, we used the set of 968 genes identified by Kurusu et al. (2008) as putative cell surface and secreted proteins encoded in the fly genome and generated additional "hand" curated lists for an additional 15 categories (Table S1).

We focused on three time intervals marking the transition from growth cones to differentiating presynaptic terminals, including a stage prior to overt changes in growth cone morphology (35–40 hr), the onset of morphological transformation to a presynaptic terminal (40–45 hr), and the appearance of presynaptic active zones (45–53 hr APF), respectively (Figure 3C). Changes in selective expression of cell surface proteins for all three time intervals were observed. Between 35–40 hr APF, 69 transcripts encoding cell surface/secreted proteins changed by >5-fold. Approximately one-third of these transcripts increased and the remaining two-thirds decreased. By contrast, many more transcripts encoding cell surface proteins increased during the 40–45 hr APF interval, with 62 transcripts changing five times or more, of which 56 increased and the remaining six decreased. In addition, between 35–40 and 45–53 hr APF there were selective changes in genes associated with photoreceptor function, consistent with the role of these cells in the adult. Although genes in other categories, including those encoding synaptic proteins (see below), were differentially expressed, the number of such genes was not different from that expected by chance.

Many of the genes encoding cell surface and secreted proteins that were upregulated between 40–45 hr APF, were downregulated in the developmental window just prior to this interval (i.e., 35–40 hr APF). The transcripts for some 25 cell surface proteins changed during 45–53 hr APF, with ∼50% increasing and the remaining decreasing (Figures 3C and 3D). In summary, in the union set of transcripts changing between 40–45, 45–53, and 40–53 hr APF, corresponding to the onset of the final targeting step and onset of presynaptic differentiation, 125 cell surface proteins were up or downregulated by greater than five times.

We categorized the patterns of expression of genes in the union set that go up greater than five times into three major classes (Figures 4A and 4B): class I genes were downregulated before being upregulated between 40–53 hr APF; class II genes were unchanged between 35–40 hr APF and were subsequently upregulated between 40–45 hr APF; and class III genes were upregulated between 45–53 hr APF. These fall into different protein families (Figure 4C; see Table S8 for a complete list of expression of protein families at all time points). It should be noted that the five times cut-off is arbitrary and, indeed, many other cell surface proteins are present at each time point and may change modestly (e.g., 2- to 4-fold) or not at all (see below). Indeed, several genes known to contribute to wiring during these developmental steps are expressed by a change less than five times between different developmental time points within the broader interval from 35–53 hr APF (data not shown). Nevertheless, the changes in expression of many genes corresponding to discrete

Biololgical category	35-40hrs APF			40-45hrs APF			45-53hrs APF			Total in fly genome
	Up (5X)	Down (5X)	P-value	Up (5X)	Down (5X)	P-value	Up (5X)	Down (5X)	P-value	
cell_surface	21	48	3.00E-07	56	6	1.31E-08	18	12	1.15E-04	950
phototransduction	8	1	1.52E-06	0	1	4.02E-01	10	0	1.40E-09	43
splicing_regulator	1	0	5.18E-02	0	0	9.84E-01	0	0	1.27E-01	150
gap	0	0	1.80E-01	0	0	2.43E-01	0	0	3.06E-01	21
synapse	1	0	3.55E-01	1	1	6.08E-01	0	0	3.25E-01	44
kinase_activity	10	1	4.65E-01	5	2	1.43E-01	0	1	9.64E-01	373
TF	12	12	4.75E-01	15	5	3.35E-01	3	3	9.07E-02	751
actin_binding	1	3	5.25E-01	3	1	6.02E-01	1	0	3.03E-01	120
ecdysone_inducible	1	3	6.66E-01	2	1	4.23E-01	0	2	7.59E-01	109
phosphatase_activity	2	5	7.30E-01	2	4	6.39E-01	1	0	2.00E-01	178
translation_regulation	4	1	8.41E-01	1	3	7.72E-01	0	0	2.63E-01	78
RNA_binding	8	7	9.55E-01	8	11	3.42E-01	0	1	9.98E-01	741
gef	2	1	9.76E-01	1	0	5.82E-01	0	0	3.25E-01	24
membrane_transporter	8	27	9.77E-01	30	3	9.83E-01	4	2	6.55E-02	765
chromatin_assembly	0	0	9.90E-01	1	1	1.63E-01	0	0	9.61E-02	138
microtubule	4	2	9.96E-01	0	1	2.58E-01	0	0	2.16E-01	63

(legend on next page)

development steps in wiring raise the possibility that they play roles at these developmental steps or alternatively that signaling in growth cones changes the spectrum of proteins loaded onto ribosomes for use at later stages.

Comparison between FACS and T-TRAP Generated Transcriptomes

We next sought to assess whether the changes observed in T-TRAP reflected changes in transcription or in post-transcriptional mechanisms (i.e., association of mRNAs with ribosomes). To do this, we compared patterns of gene expression from FACS-sorted cells, using the GFP in the T-TRAP-labeled RpL10 protein as a marker for cell sorting and mRNA isolated from flies of the same genotype using T-TRAP at two time points, 40 hr and 53 hr APF. We then assessed whether changes in gene expression observed between these two time points reflected changes in transcription or in post-transcriptional mechanisms. There were 669 and 308 transcripts that changed by more than five times between these two time points in T-TRAP and FACS, respectively (Table S2). Of the overlap set of 142 genes differentially expressed between these time points in both the T-TRAP and FACS analysis, 138 were regulated in a similar manner in both FACS and T-TRAP. For 531 differentially expressed genes in T-TRAP (79.3%), however, there was either no change in FACS or the change was in the opposite direction. Thus, although changes in gene expression associated with presynaptic differentiation occur at the level of transcription, the changes in ribosome association suggests that post-transcriptional regulatory mechanisms play a prominent role in presynaptic differentiation.

Changes in Presynaptic 3′ UTR Regulation during Synapse Formation

We next sought to assess the relationship between the expression of genes encoding presynaptic proteins and the transformation of R cell growth cones to presynaptic terminals. Despite the abrupt change in morphology correlated with the transition from growth cones to synaptic terminals and the dramatic changes in the expression of cell surface and secreted proteins associated with this transition, a similar change in the expression of presynaptic proteins, as a class, was not observed (Figure 5A; compare to Figure 4A).

Most genes encoding presynaptic proteins were expressed at moderate levels as early as 24 hr APF, well before synapse formation and these generally increased over time (Table S1). Furthermore, as a class, these genes were no more likely to be upregulated than expected by chance (see Figure 3C). Only five genes encoding known presynaptic proteins exhibited substantial changes in levels between 40 and 53 hr APF. Two synaptic vesicle proteins, synapsin (Syn) and synaptogyrin (Syngr), increase seven times and five times, respectively, during this period. Syngr increases another 32 times between 53–96 hr APF, consistent with its requirement for normal synaptic vesicle biogenesis (Stevens et al., 2012). Adult R cells are densely packed with synaptic vesicles and Syngr expression appears linked to this accumulation. Other transcripts encoding proteins involved with synaptic vesicle function, fusion and recycling (i.e., n-synaptobrevin [n-Syb], synaptosomal-associated protein 25 kDa [SNAP-25], and endophilinA [EndoA]) increased modestly between 53–95 hr APF (two to three times) (Bhattacharya et al., 2002; Guichet et al., 2002; Vilinsky et al., 2002). Transcripts encoding active zone proteins Cacophony (Cac), Straight-jacket (Stj), and Brp increased two to seven times between 24 and 53 hr; the patterns of expression exhibited different kinetics and the level of these transcripts generally decreased by two to ten times by 96 hr APF (Ly et al., 2008; Smith et al., 1996). The level of *highwire* (*hiw*) mRNA, encoding a ubiquitin E3 ligase that acts as a negative regulator of active zones (Wan et al., 2000), was very low at all times points with only one RPKM value >1 at 40 hr APF. Together, these observations are consistent with a role for post-transcriptional regulation during presynaptic differentiation.

Given the implication of the importance of post-transcriptional regulation during synapse formation, we examined the 3′ UTRs of mRNAs encoding presynaptic proteins. We observed that 3′ UTRs of 55% of these genes (45 genes analyzed) changed between 24–96 hr APF (Figure 5; Table S4). By contrast, only 26% of five randomly selected sets of genes (45 genes in each set) during this time period showed changes in their 3′ UTRs. This difference is highly significant (p = 1.71e-07, z test). In other systems, enzymes regulating 3′ UTR cleavage and polyadenylation site selection are downregulated during development (Di Giammartino et al., 2011; Ji et al., 2009), thereby resulting in longer 3′ UTRs. Consistent with this, we observed a universal downregulation of enzymes regulating the 3′ UTRs during development. These included the cleavage and polyadenylation specificity factor (CPSF), cleavage stimulation factor (CSTF), and cleavage factor IM (CFIM) ranging from 3.5 times to 44 times from 24–96 hr APF (Table S5). In addition, regulatory factors of the Elav family of RNA binding proteins including Elav, Rpb9, and Fne show dynamic changes in expression (Samson and Chalvet, 2003) (Table S5). Elav protein has been shown to promote the formation of longer 3′ UTRs in the developing fly embryonic nervous system (Hilgers et al., 2012). In developing R cells,

Figure 3. Differential Gene Expression during R Cell Development

(A) The number of upregulated and downregulated genes during R cell development. RPKM values are greater than one (see text).

(B) Enriched biological processes across different stages during R cell development using GO term enrichment analysis. The differentially expressed genes were classified into two categories: upregulated (red) and downregulated (blue).

(C) Biological categories (see text) of differentially regulated genes (five times upregulated and five times downregulated) between 35–40 hr APF, 40–45 hr APF, and 45–53 hr APF. The p value represents the likelihood that the number of genes expressed greater than five times between the indicated time points is significantly different from that expected from a random distribution in each biological category (see Experimental Procedures).

(D) The number cell surface and secreted molecules upregulated and downregulated greater than five times. For both (C) and (D), RPKM for a transcript at one time point must be greater than one and levels between time points are greater than five times.

See also Tables S1 and S2.

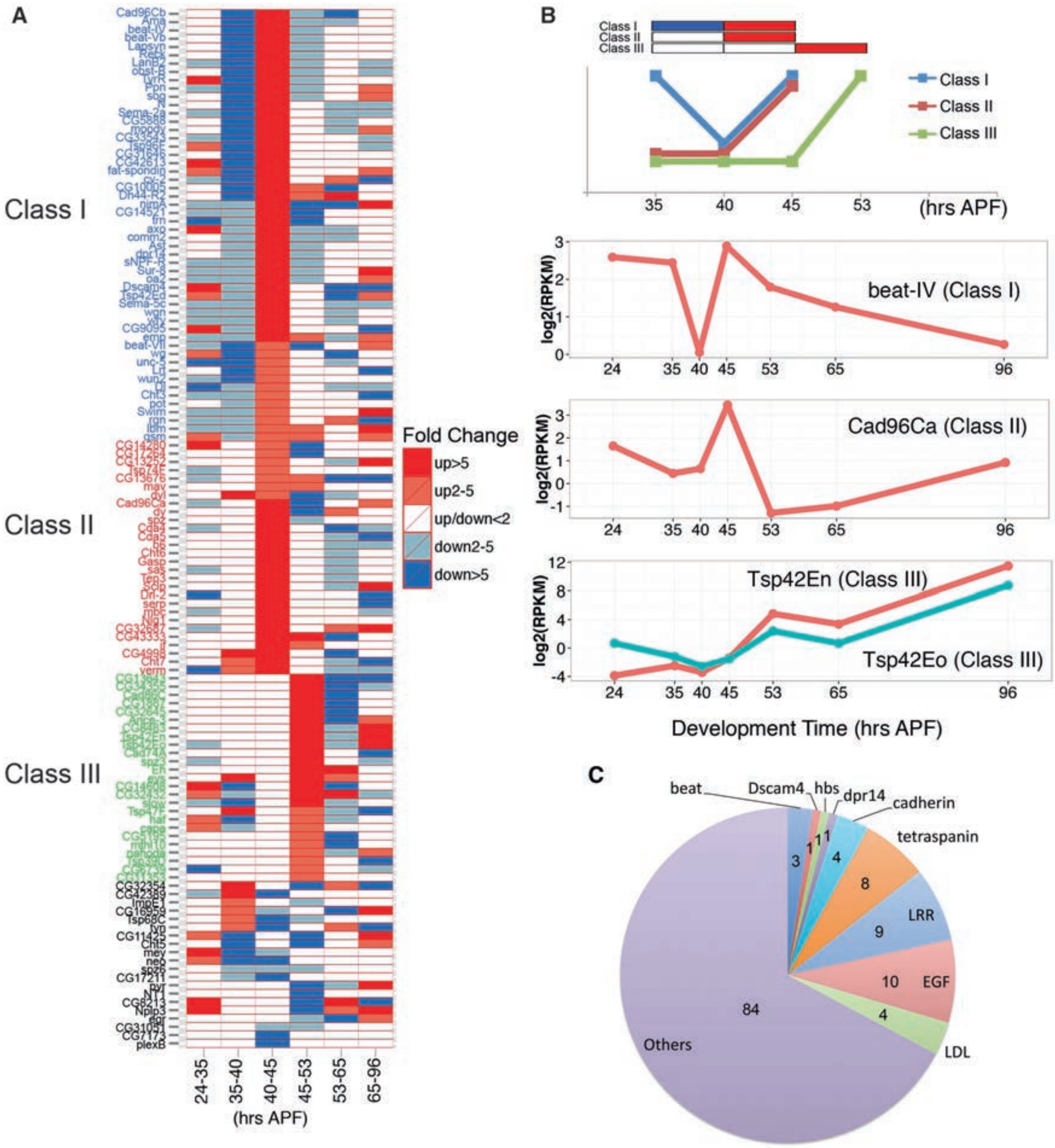


Figure 4. Dynamic Regulation of Cell Surface and Secreted Molecules Expressed during Presynaptic Differentiation

(A) Heatmap showing differential regulation of 125 cell surface or secreted molecules changing greater than 5 times in the union set of differentially expressed genes between 40–45 hr APF, 45–53 hr APF, and 40–53 hr APF. Color code as indicated. Expression for each of transcript is shown across the entire time series. The fold change refers to the significant change in level between two neighboring time points. Note that these are relative changes not absolute values and boxes showing genes with RPKMs less than one in both time points are labeled as "no change" (white).

(B) Schematic presentation of gene expression patterns of a group of genes as indicated. Gene expression patterns of four cell surface or secreted molecules are shown (see text).

(C) Pie chart showing the numbers of 125 differentially expressed genes in different gene families during presynaptic development. There are six Ig superfamily proteins: three Beat paralogs, Dscam4, Hbs, and Dpr14.

See also Table S8.

elav mRNA expression peaks at 40 hr APF and gradually falls to low levels at 96 hr APF. By contrast, Rpb9 increases continuously to high levels at 96 hr APF and Fne shows a complex pattern of regulation with peak levels at both 35 and 45 hr APF before decreasing gradually to low levels by 96 hr APF. Thus, although there is a general trend toward lengthening the 3′ UTR in all genes in R cells as development proceeds, there is a selective increase in 3′ UTR length for genes encoding presynaptic components and correlated changes in mRNAs encoding proteins and paralogs of them, which regulate processing of 3′ UTRs.

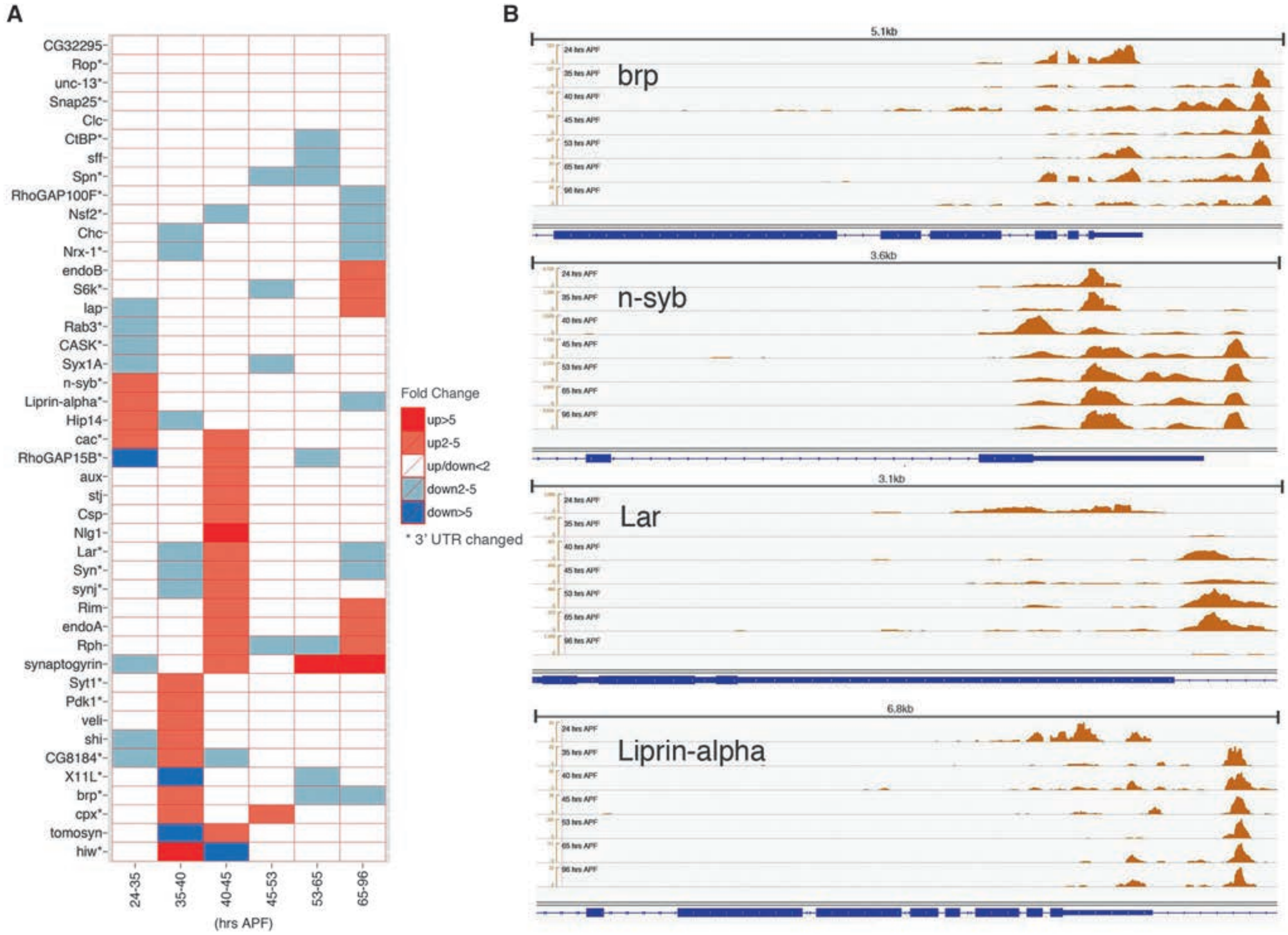


Figure 5. Identification of Alternative 3′ UTRs in Presynaptic Proteins

(A) The heatmap showing gene expression level changes of 45 presynaptic proteins between two adjacent time points during development.

(B) Four examples of transcripts (*brp*, *n-syb*, *Lar*, and *Liprin-alpha*) demonstrating the extended 3′ UTRs during development.

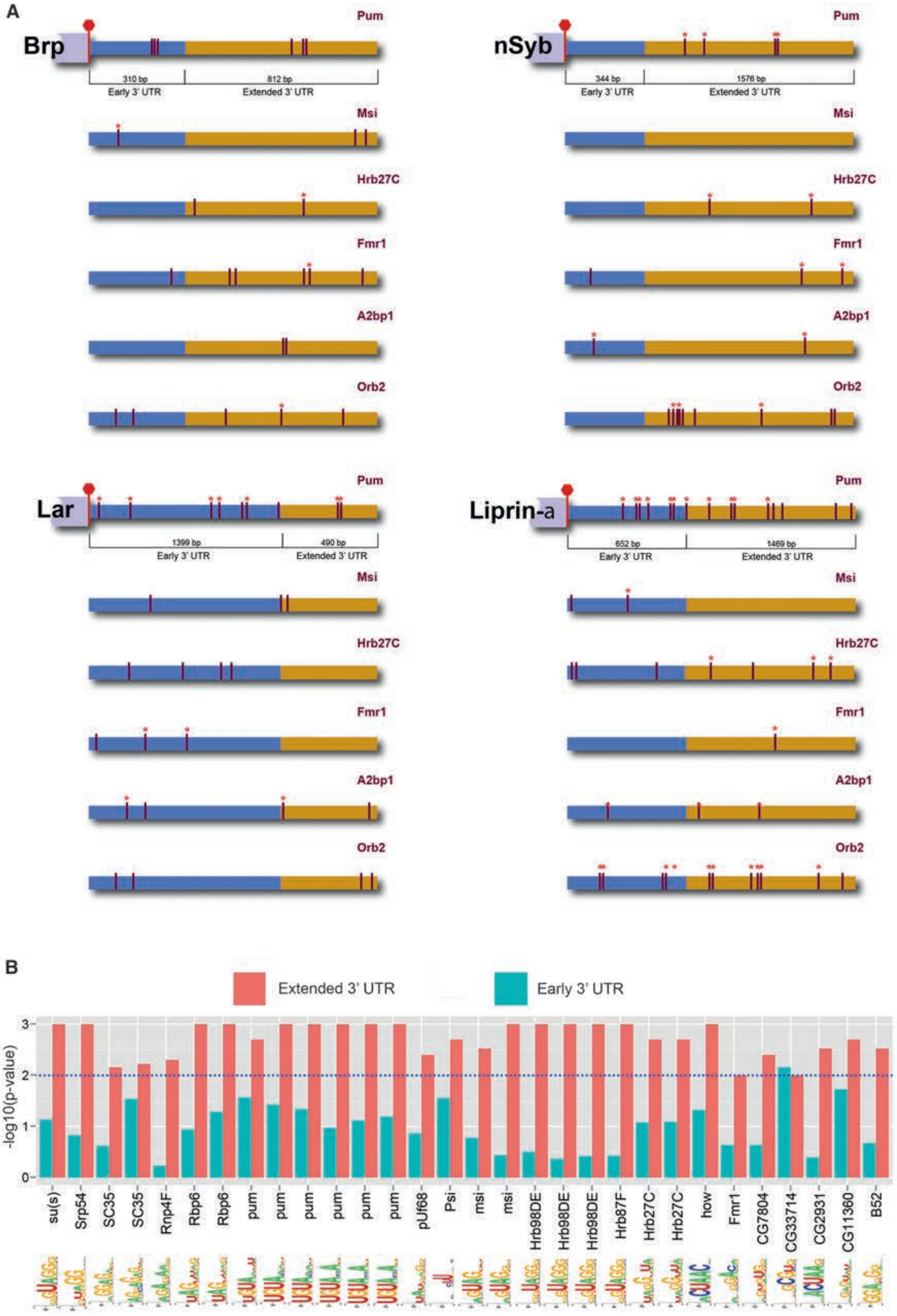
See also Table S6.

Identification of Enriched RNA Binding Proteins Binding Motifs in Synaptic Transcripts

The correlation between presynaptic differentiation and changes in the 3′ UTR of many transcripts encoding synaptic proteins raised the possibility that these regions are sites of translational regulation through microRNAs, RNA binding proteins or both. To explore this possibility, we compared the sequences of synaptic transcripts with extended 3′ UTRs, to those lacking the 3′ UTR extension for binding sites for both microRNA and RNA binding proteins. This comparison included 18 transcripts with 3′ UTR extensions, on average 1.3 kb in length, to 20 transcripts that were not extended, with 3′ UTRs with an average length of 1.1 kb (Table S6).

To test for enrichment of microRNA binding sites within 3′ UTRs, we searched for potential binding sites for 466 known fly microRNAs in the miRBase database (Kozomara and Griffiths-Jones, 2014). The numbers of predicted binding sites found in both classes of 3′ UTRs were very similar ($R = 0.84$, p value $< 2.2e-16$, Pearson correlation coefficient). This suggests that the transcripts with extended 3′ UTRs in R cells are no more likely to be regulated by microRNAs than their unextended counterparts. Furthermore, the 3′ UTRs were no more likely to contain miRNA binding sites than randomly selected 3′ UTRs (p value $= 0.057$, z test). Thus, our analysis does not support the notion that the increased length of transcripts reflects an increased propensity of transcripts encoding presynaptic proteins to be regulated by miRNAs.

We next assessed whether extended 3′ UTRs are more likely to be sites of regulation through RNA binding proteins than their counterparts that were not extended. We first performed the RNA binding proteins (RBP) binding site prediction analysis using a list of known position weight matrices (PWMs) of 224 fly RNA binding proteins (Ray et al., 2013). This revealed that RNA binding motifs were over-represented in mRNAs encoding synaptic proteins when compared to random sequences of genomic DNA regardless of whether the 3′ UTRs were extended. Furthermore, there were marked differences between the 3′ UTRs. Thirty significant motifs were found in genes encoding synaptic



(legend on next page)

proteins with extended 3′ UTR, whereas only one motif was significantly enriched in genes lacking an extended 3′ UTR (i.e., with a p value of 0.01) (Figure 6B). Four examples are shown in Figure 6A. Sequence comparison of all the RBP binding sites between the genomes of 14 insect species revealed that 88.9% (16 out of 18) of the presynaptic genes with extended 3′ UTRs have at least one highly conserved RBP binding site (e.g., asterisks in Figure 6A and Table S7). These findings suggest that post-transcriptional mechanisms contribute significantly to synaptic differentiation.

Interestingly, several RNA binding proteins known to regulate mRNA localization and translation (e.g., 4EHP, Hrb27c, and Imp), were also dynamically regulated during this period. These findings are consistent with the notion that post-transcriptional regulatory mechanisms may play a prominent role in regulating the expression of synaptic components. Several RBPs that exhibited changes in expression during R cell synapse formation and for which binding sites in mRNAs encoding synaptic proteins are enriched have been shown to regulate synaptic strength at the fly neuromuscular junction (Menon et al., 2004) and learning and memory (Keleman et al., 2007). Thus, together these data support the view that post-transcriptional mechanisms play an important role in regulating presynaptic differentiation.

DISCUSSION

Characterization of Gene Expression during Synapse Formation Using T-TRAP

The conversion of a growth cone to a presynaptic terminal is a common step in the neuronal differentiation. As this step occurs synchronously in R cells, these neurons provide a unique opportunity to characterize global patterns of gene expression during this process. We isolated mRNA associated with ribosomes labeled with two tandemly arranged tags (T-TRAP), rather than a single one, separated by a protease recognition site. This facilitated a two-step affinity purification of mRNA associated with ribosomes thereby increasing signal-to-noise and reproducibility between biological replicates over that achieved via TRAP (e.g., from 0.93 to 0.99). This comes at a cost of a significant, though acceptable, decrease in yield (~30%). No developmental or morphological abnormalities resulted from overexpression of the tandem-tagged ribosomal subunit in various cell types, in contrast to overexpression of other tags in the same cells used for mRNA purification, such as the poly A binding protein (J.M.M. and S.L.Z, unpublished data). We have recently modified the transgenes encoding T-TRAP to further reduce background by increasing expression via the inclusion of non-coding sequences enhancing translation (Pfeiffer et al., 2012) and by mitigating the effects of leaky expression of the UAS construct by inserting a transcriptional stop sequences flanked by FRT-recombination sites (unpublished data). Targeting recombinase expression to the cells of interest and coupling this with cell-type-specific expression of GAL4 further increases the cell-type specificity of this method (Golic and Lindquist, 1989). Given the cellular diversity of neural tissue we anticipate that the T-TRAP technique will provide effective ways for isolating transcripts from increasingly smaller subsets of cells in flies and perhaps in mouse as well.

Selective Changes in Transcripts Encoding Cell Surface Proteins with the Onset of Presynaptic Differentiation

The detailed time course of gene expression in R cells during the transformation of R cell growth cones to presynaptic terminals revealed substantial changes in the expression of many mRNAs encoding cell surface proteins, including those implicated in cellular recognition and synapse formation. The changes in expression were highly selective for cell surface and secreted proteins. The only other category with selective expression within this interval was “phototransduction” reflecting a precocious upregulation of a subset of these genes prior to the onset of opsin expression and, hence, photosensitivity. Indeed, just prior to the onset of morphological changes in R cell growth cones preceding their transformation into presynaptic terminals (i.e., 35–40 hr APF), many more mRNAs encoding cell surface components (approximately three times) were downregulated than upregulated. By contrast, during the subsequent time interval correlating with the first morphological manifestation of presynaptic differentiation (i.e., 40–45 hr APF), many more cell surface proteins (approximately ten times) were upregulated than downregulated. The changes in the association of these mRNAs with ribosomes during these time intervals suggests that a massive restructuring of the cell surface plays an important role in the transition of the growth cone to a synaptic terminal. Comparison of expression differences between FACS and T-TRAP indicate that many of these changes occur preferentially at the post-transcriptional level. The complementary and alternating expression of the translational inhibitors, 4EHP and 4E-BP (or Thor) at each time point (i.e., when one is high the other is low), is consistent with expression being highly regulated at the level of translation just prior to and during synapse formation. The changes in the cell surface proteome, however, also reflect the stability of different membrane proteins. As the half-lives of proteins vary over many orders of magnitude, the changes in the ribosome association of mRNAs of proteins encoding different cell surface proteins only represent part of the story. Measuring these changes in axon terminals for proteins in vivo, even for one protein, remains an experimental challenge.

The transcripts changing greater than five times during these intervals included those encoding members of large families of cell surface molecules previously implicated, or shown to be required for, neural circuit formation including immunoglobulin-containing proteins (e.g., Dscam4 and beat-IV) (Fambrough

Figure 6. RNA Binding Proteins Binding Motifs in 3′ UTRs of Presynaptic Proteins

(A) Four transcripts encoding synaptic proteins (*brp*, *n-syb*, *Lar*, and *Liprin-α*) with binding sites for the indicated RNA binding proteins (RBPs) in both early 3′ UTR (blue) and extended 3′ UTR (gold). RBP binding sites highly conserved across species are labeled with asterisks (see text).

(B) Bar plot showing 30 significant RBP binding motifs were found in genes encoding synaptic proteins with extended 3′ UTR (red bar), whereas only one motif was significantly enriched in genes without an extended 3′ UTR (blue bar). The blue dash line indicates a threshold of p value of 0.01.

See also Figure S4 and Tables S6 and S7.

and Goodman, 1996; Millard et al., 2010; Yamagata and Sanes, 2008), leucine-rich repeat containing proteins (e.g., lapsyn) (Guan et al., 2011), cadherins (e.g., Cad96Ca), integrin family members (e.g., *if*), and tetraspanins (Hoang and Chiba, 1998; Kopczynski et al., 1996; Lee et al., 2001). In addition, other proteins that changed considerably during this time, including components of the extracellular matrix (e.g., laminin and various proteases), guidance receptors (e.g., Drl-2), secreted signals (e.g., wingless, semaphorin2a), antagonists of BMP signaling and spatzle family members, regulate various steps in development, including synapse formation, terminal branch morphology, and cell survival (García-Alonso et al., 1996; Liebl et al., 2008; Zhu et al., 2008; Zlatic et al., 2009). These changes are consistent with dynamic intercellular interactions occurring within the neuropil as complex patterns of connections between R cell terminals and their targets are assembled (Rivera-Alba et al., 2011; Takemura et al., 2013).

Post-transcriptional Regulation through 3′ UTRs

Although marked changes in cell surface proteins were readily observed between closely spaced time points prior to and during synapse formation, changes in the levels of transcripts encoding synaptic proteins were modest. A doubling in the length of the 3′ UTRs encoding presynaptic proteins was correlated with an increase in the number of binding sites for RNA binding proteins implicated in regulating mRNA localization, stability and translation. These binding proteins were expressed in R cells, some at constant levels (e.g., Fmr) through the time course examined, whereas others changed considerably (e.g., Hrb27c). Importantly, in each of these extended transcripts at least one RBP binding site is conserved with orthologous transcripts in other *Drosophila* species supporting the notion that these sites are functionally significant. Functional studies in *Caenorhabditis elegans* also support a role for increased 3′ UTR length in regulating presynaptic differentiation, with mutations in a negative regulator of polyadenylation (i.e., favoring shortened 3′ UTR) leading to defects in synapse formation (Van Epps et al., 2010). Together these data suggest that increased 3′ UTR length is an important regulator of presynaptic differentiation and that increased length promotes post-transcriptional regulation through interactions with RNA binding proteins.

Regulation through the 3′ UTRs may also play a more widespread role in regulating presynaptic differentiation (Chia et al., 2014). For instance, actin assembly is required for presynaptic differentiation and actin regulatory proteins are regulated post-transcriptionally via the evolutionarily conserved protein Imp. Imp binds selectively to sequences within the 3′ UTRs of actin regulatory proteins (Medioni et al., 2014). Indeed, biochemical studies in *Drosophila* S2 cells in culture identified 40 actin regulatory proteins with Imp binding sites (Hansen et al., 2015). All of these are expressed in R cells, many at very high levels, during synapse formation. Importantly, in our dataset, Imp transcript is upregulated some 40 times prior to the onset of presynaptic differentiation.

A recent study suggests that different 3′ UTRs may also regulate the transport of membrane proteins to the cell surface (Berkovits and Mayr, 2015). Here, human HuR proteins bind to 3′ UTRs of mRNAs encoding cell surface proteins as they are being translated at the ER. HuR proteins then recruit a protein complex that selectively associates with the C terminus of the nascent polypeptide and this complex, in turn, promotes transport of the newly synthesized protein to the cell surface. There are three *Drosophila* homologs of HuR (Elav, Fne and Rpb9). These three proteins have been implicated in synapse formation in the *Drosophila* neuromuscular junction (Zaharieva et al., 2015), and, as we discussed above, they are also implicated in lengthening of 3′ UTRs. Interestingly, these proteins are expressed in very different and dynamic ways in R cells during synaptic development (Table S5). Rbp9 expression increases continuously between 24 and 53 hr APF with a change of nearly 20 times. By contrast, Elav and Fne are expressed in a largely complementary fashion with peak levels of Elav at 24 and 40 hr APF, whereas Fne shows high levels of expression at 35 and 45 hr APF, with reduced levels at 40 hr APF.

Together, these data suggest that diverse post-transcriptional mechanisms mediated through 3′ UTR sequences play an important role in regulating synapse formation in R cells.

Concluding Remarks

Decades of genetic analysis have provided an extensive tool kit for exploring neural circuit assembly in the fly visual system. Here, progress has been made in large part based on the ability to genetically manipulate specific cell types, single cells, or small subsets of them to assess their phenotypes in an otherwise normal or wild-type background. We recently demonstrated that all R cells progress through synapse formation synchronously. This synchrony within a single fly and the ability to synchronize large numbers of animals at the onset of pupal development, provided a unique opportunity to explore gene expression at a global level during the conversion of growth cones to presynaptic terminals. Coupling this global approach to genetic and functional analysis at the level of single genes provide a unique opportunity to dissect the molecular program underlying synaptic development.

EXPERIMENTAL PROCEDURES

T-TRAP Flies

TRAP and T-TRAP constructs were generated by insertion of dRpL10a with an N-terminal Not-I site following the ATG inserted into pUAST using EcoRI/XbaI. Then a cassette containing either Not-I flanked GFP (TRAP) or Not-I flanked 3xFLAG-TEV-GFP (T-TRAP) was inserted. Flies were injected with either construct and F2 progeny were screened for transgenic animals exhibiting the lowest UAS expression in the absence of GAL4 as assessed by western blot against GFP.

T-TRAP Purification

Brain dissections were performed in dissection buffer (1× HBSS, 2.5 mM HEPES [pH 7.4], 35 mM glucose, 4 mM $NaHCO_3$) containing 100 μg/ml of cycloheximide (CHX). The central brain was discarded and optic lobes and retina were frozen on dry ice and maintained at −80 in Lysis buffer (20 mM HEPES [pH 7.4], 150 mM KCl, 5 mM MgCl + protease inhibitors) containing 100 μg/ml CHX until enough optic lobes were accumulated. For each replicate with the Chp-GAL4 driver, 40 optic lobes were used.

Optic lobes were thawed and homogenized using a motorized Teflon pestle in 500 μl of lysis buffer with RNase inhibitors; one-ninth the volume of Igepal CA-360 was added and the tube is gently inverted, followed by addition of one-ninth the volume of 1,2-diheptanoyl-*sn*-glycero-3-phosphocholine. The tube containing this mixture was inverted and placed on ice for 5 min to lyse the cells. Lysates were centrifuged at 16,000 rpm for 15 min at 4°C. The supernatant was added to tubes containing anti-FLAG M2 affinity magnetic beads (Sigma) and incubated at 4°C for 1 hr with gentle rotation. Beads were collected and washed three times with Wash Buffer #1 (20 mM HEPES [pH 7.4], 150 mM KCl, 5 mM MgCl, 1% Igepal CA-360) and 500 µl of IP buffer #2 (20 mM HEPES [pH 7.4], 150 mM KCl, 5 mM MgCl, 1% Igepal CA-360, + protease and RNase inhibitors) was added to the beads. Five microliters of TEV protease (10 U/µl) was added (Life Technologies) and in some cases 3 µl Turbo DNase (2 U/µl) (Ambion), and mixtures were incubated at room temperature for 1 hr with gentle rotation. The supernatant was collected and added to protein-G beads (Life Technologies) pre-bound with 50 µg each of anti-GFP monoclonal antibodies HtzGFP-19C8 and HtzGFP-19F7 (Monoclonal Antibody Core Facility, Sloan-Kettering Cancer Center) and incubated at 4°C for 30 min. Beads were collected and washed with Wash Buffer #2 (20 mM HEPES [pH 7.4], 300 mM KCl, 5 mM MgCl, 1% Igepal CA-360) three times. RNA was eluted from the beads by addition of 350 µl of RLT buffer (QIAGEN) and incubated for 5 min with occasional inversion. RNA is then isolated using the RNA-min elute kit from QIAGEN. RNA was then amplified in a linear fashion using Arcturus RiboAmp HS kit (Life Technologies). cDNA was then generated for quality assessment and then paired-end Illumina, multiplexed sequencing libraries were prepared.

Library Preparation from Cells Isolated by FACS

Cell dissociation procedures, FACS purification of cells from 40 and 53 hr APF, RNA purification, and library constructions were done as described in Tan et al. (2015).

Additional experimental procedures are included in the Supplemental Information.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, five figures, and eight tables and can be found with this article online at http://dx.doi.org/10.1016/j.celrep.2015.12.102.

AUTHOR CONTRIBUTIONS

K.X.Z., J.M., and S.L.Z. designed and conceived research. J.M.M. generated T-TRAP constructs and flies, performed dissections and biochemical RNA isolation, and prepared libraries. L.T. performed FACS experiments. K.X.Z. collected Illumina sequencing data, developed data analysis pipelines, and analyzed data. M.P. designed data analysis pipelines. K.X.Z., J.M.M, and S.L.Z. interpreted data and wrote the paper.

ACKNOWLEDGMENTS

We would like to thank Douglas Black, Kelsey Martin, William Yang, Kathrin Plath, Giovanni Coppola, and members of the S.L.Z. laboratory for critical reading of the manuscript. K.X.Z. is supported by the Canadian Institute of Health Research Fellowship. J.M.M. was supported by the NIH-Cellular Neurobiology Training Program (5T32NS07101:29-30). S.L.Z. is an investigator of the Howard Hughes Medical Institute.

Received: April 30, 2014
Revised: December 4, 2015
Accepted: December 21, 2015
Published: January 28, 2016

REFERENCES

Abruzzi, K., Chen, X., Nagoshi, E., Zadina, A., and Rosbash, M. (2015). RNA-seq profiling of small numbers of Drosophila neurons. Methods Enzymol. *551*, 369–386.

Bate, M. (1993). The Development of Drosophila Melanogaster. M. Bate and A. Martinez-Arias, eds. (Cold Spring Harbor Press), pp. 1363–1491.

Berkovits, B.D., and Mayr, C. (2015). Alternative 3′ UTRs act as scaffolds to regulate membrane protein localization. Nature *522*, 363–367.

Bhattacharya, S., Stewart, B.A., Niemeyer, B.A., Burgess, R.W., McCabe, B.D., Lin, P., Boulianne, G., O'Kane, C.J., and Schwarz, T.L. (2002). Members of the synaptobrevin/vesicle-associated membrane protein (VAMP) family in Drosophila are functionally interchangeable in vivo for neurotransmitter release and cell viability. Proc. Natl. Acad. Sci. USA *99*, 13867–13872.

Chen, Y., Akin, O., Nern, A., Tsui, C.Y., Pecot, M.Y., and Zipursky, S.L. (2014). Cell-type-specific labeling of synapses in vivo through synaptic tagging with recombination. Neuron *81*, 280–293.

Cheng, H.J., Nakamoto, M., Bergemann, A.D., and Flanagan, J.G. (1995). Complementary gradients in expression and binding of ELF-1 and Mek4 in development of the topographic retinotectal projection map. Cell *82*, 371–381.

Chia, P.H., Li, P., and Shen, K. (2013). Cell biology in neuroscience: cellular and molecular mechanisms underlying presynapse formation. J. Cell Biol. *203*, 11–22.

Chia, P.H., Chen, B., Li, P., Rosen, M.K., and Shen, K. (2014). Local F-actin network links synapse formation and axon branching. Cell *156*, 208–220.

Clandinin, T.R., and Zipursky, S.L. (2000). Afferent growth cone interactions control synaptic specificity in the Drosophila visual system. Neuron *28*, 427–436.

Desai, C., Garriga, G., McIntire, S.L., and Horvitz, H.R. (1988). A genetic pathway for the development of the Caenorhabditis elegans HSN motor neurons. Nature *336*, 638–646.

Di Giammartino, D.C., Nishida, K., and Manley, J.L. (2011). Mechanisms and consequences of alternative polyadenylation. Mol. Cell *43*, 853–866.

Doyle, J.P., Dougherty, J.D., Heiman, M., Schmidt, E.F., Stevens, T.R., Ma, G., Bupp, S., Shrestha, P., Shah, R.D., Doughty, M.L., et al. (2008). Application of a translational profiling approach for the comparative analysis of CNS cell types. Cell *135*, 749–762.

Fambrough, D., and Goodman, C.S. (1996). The Drosophila beaten path gene encodes a novel secreted protein that regulates defasciculation at motor axon choice points. Cell *87*, 1049–1058.

García-Alonso, L., Fetter, R.D., and Goodman, C.S. (1996). Genetic analysis of Laminin A in Drosophila: extracellular matrix containing laminin A is required for ocellar axon pathfinding. Development *122*, 2611–2621.

Golic, K.G., and Lindquist, S. (1989). The FLP recombinase of yeast catalyzes site-specific recombination in the Drosophila genome. Cell *59*, 499–509.

Guan, Z., Buhl, L.K., Quinn, W.G., and Littleton, J.T. (2011). Altered gene regulation and synaptic morphology in Drosophila learning and memory mutants. Learn. Mem. *18*, 191–206.

Guichet, A., Wucherpfennig, T., Dudu, V., Etter, S., Wilsch-Bräuniger, M., Hellwig, A., González-Gaitán, M., Huttner, W.B., and Schmidt, A.A. (2002). Essential role of endophilin A in synaptic vesicle budding at the Drosophila neuromuscular junction. EMBO J. *21*, 1661–1672.

Hadjieconomou, D., Timofeev, K., and Salecker, I. (2011). A step-by-step guide to visual circuit assembly in Drosophila. Curr. Opin. Neurobiol. *21*, 76–84.

Hansen, H.T., Rasmussen, S.H., Adolph, S.K., Plass, M., Krogh, A., Sanford, J., Nielsen, F.C., and Christiansen, J. (2015). Drosophila Imp iCLIP identifies an RNA assemblage coordinating F-actin formation. Genome Biol. *16*, 123.

Heiman, M., Schaefer, A., Gong, S., Peterson, J.D., Day, M., Ramsey, K.E., Suárez-Fariñas, M., Schwarz, C., Stephan, D.A., Surmeier, D.J., et al. (2008). A translational profiling approach for the molecular characterization of CNS cell types. Cell *135*, 738–748.

Henry, G.L., Davis, F.P., Picard, S., and Eddy, S.R. (2012). Cell type-specific genomics of Drosophila neurons. Nucleic Acids Res. *40*, 9691–9704.

Hilgers, V., Lemke, S.B., and Levine, M. (2012). ELAV mediates 3′ UTR extension in the Drosophila nervous system. Genes Dev. *26*, 2259–2264.

Hoang, B., and Chiba, A. (1998). Genetic analysis on the role of integrin during axon guidance in Drosophila. J. Neurosci. *18*, 7847–7855.

Ji, Z., Lee, J.Y., Pan, Z., Jiang, B., and Tian, B. (2009). Progressive lengthening of 3′ untranslated regions of mRNAs by alternative polyadenylation during mouse embryonic development. Proc. Natl. Acad. Sci. USA *106*, 7028–7033.

Keleman, K., Krüttner, S., Alenius, M., and Dickson, B.J. (2007). Function of the Drosophila CPEB protein Orb2 in long-term courtship memory. Nat. Neurosci. *10*, 1587–1593.

Kidd, T., Brose, K., Mitchell, K.J., Fetter, R.D., Tessier-Lavigne, M., Goodman, C.S., and Tear, G. (1998). Roundabout controls axon crossing of the CNS midline and defines a novel subfamily of evolutionarily conserved guidance receptors. Cell *92*, 205–215.

Kopczynski, C.C., Davis, G.W., and Goodman, C.S. (1996). A neural tetraspanin, encoded by late bloomer, that facilitates synapse formation. Science *271*, 1867–1870.

Kozomara, A., and Griffiths-Jones, S. (2014). miRBase: annotating high confidence microRNAs using deep sequencing data. Nucleic Acids Res. *42*, D68–D73.

Kumar, J.P., and Ready, D.F. (1995). Rhodopsin plays an essential structural role in Drosophila photoreceptor development. Development *121*, 4359–4370.

Kurusu, M., Cording, A., Taniguchi, M., Menon, K., Suzuki, E., and Zinn, K. (2008). A screen of cell-surface molecules identifies leucine-rich repeat proteins as key mediators of synaptic target selection. Neuron *59*, 972–985.

Lee, C.H., Herman, T., Clandinin, T.R., Lee, R., and Zipursky, S.L. (2001). N-cadherin regulates target specificity in the Drosophila visual system. Neuron *30*, 437–450.

Lee, R.C., Clandinin, T.R., Lee, C.H., Chen, P.L., Meinertzhagen, I.A., and Zipursky, S.L. (2003). The protocadherin Flamingo is required for axon target selection in the Drosophila visual system. Nat. Neurosci. *6*, 557–563.

Liebl, F.L., Wu, Y., Featherstone, D.E., Noordermeer, J.N., Fradkin, L., and Hing, H. (2008). Derailed regulates development of the Drosophila neuromuscular junction. Dev. Neurobiol. *68*, 152–165.

Ly, C.V., Yao, C.K., Verstreken, P., Ohyama, T., and Bellen, H.J. (2008). straightjacket is required for the synaptic stabilization of cacophony, a voltage-gated calcium channel alpha1 subunit. J. Cell Biol. *181*, 157–170.

Medioni, C., Ramialison, M., Ephrussi, A., and Besse, F. (2014). Imp promotes axonal remodeling by regulating profilin mRNA during brain development. Curr. Biol. *24*, 793–800.

Menon, K.P., Sanyal, S., Habara, Y., Sanchez, R., Wharton, R.P., Ramaswami, M., and Zinn, K. (2004). The translational repressor Pumilio regulates presynaptic morphology and controls postsynaptic accumulation of translation factor eIF-4E. Neuron *44*, 663–676.

Millard, S.S., Lu, Z., Zipursky, S.L., and Meinertzhagen, I.A. (2010). Drosophila dscam proteins regulate postsynaptic specificity at multiple-contact synapses. Neuron *67*, 761–768.

Nern, A., Pfeiffer, B.D., and Rubin, G.M. (2015). Optimized tools for multicolor stochastic labeling reveal diverse stereotyped cell arrangements in the fly visual system. Proc. Natl. Acad. Sci. USA *112*, E2967–E2976.

Newsome, T.P., Asling, B., and Dickson, B.J. (2000). Analysis of Drosophila photoreceptor axon guidance in eye-specific mosaics. Development *127*, 851–860.

Pfeiffer, B.D., Truman, J.W., and Rubin, G.M. (2012). Using translational enhancers to increase transgene expression in Drosophila. Proc. Natl. Acad. Sci. USA *109*, 6626–6631.

Ray, D., Kazan, H., Cook, K.B., Weirauch, M.T., Najafabadi, H.S., Li, X., Gueroussov, S., Albu, M., Zheng, H., Yang, A., et al. (2013). A compendium of RNA-binding motifs for decoding gene regulation. Nature *499*, 172–177.

Rivera-Alba, M., Vitaladevuni, S.N., Mishchenko, Y., Lu, Z., Takemura, S.Y., Scheffer, L., Meinertzhagen, I.A., Chklovskii, D.B., and de Polavieja, G.G. (2011). Wiring economy and volume exclusion determine neuronal placement in the Drosophila brain. Curr. Biol. *21*, 2000–2005.

Samson, M.L., and Chalvet, F. (2003). found in neurons, a third member of the Drosophila elav gene family, encodes a neuronal protein and interacts with elav. Mech. Dev. *120*, 373–383.

Schmucker, D., Clemens, J.C., Shu, H., Worby, C.A., Xiao, J., Muda, M., Dixon, J.E., and Zipursky, S.L. (2000). Drosophila Dscam is an axon guidance receptor exhibiting extraordinary molecular diversity. Cell *101*, 671–684.

Serafini, T., Kennedy, T.E., Galko, M.J., Mirzayan, C., Jessell, T.M., and Tessier-Lavigne, M. (1994). The netrins define a family of axon outgrowth-promoting proteins homologous to C. elegans UNC-6. Cell *78*, 409–424.

Smith, L.A., Wang, X., Peixoto, A.A., Neumann, E.K., Hall, L.M., and Hall, J.C. (1996). A Drosophila calcium channel alpha1 subunit gene maps to a genetic locus associated with behavioral and visual defects. J. Neurosci. *16*, 7868–7879.

Steiner, F.A., Talbert, P.B., Kasinathan, S., Deal, R.B., and Henikoff, S. (2012). Cell-type-specific nuclei purification from whole animals for genome-wide expression and chromatin profiling. Genome Res. *22*, 766–777.

Stevens, R.J., Akbergenova, Y., Jorquera, R.A., and Littleton, J.T. (2012). Abnormal synaptic vesicle biogenesis in Drosophila synaptogyrin mutants. J. Neurosci. *32*, 18054–18067.

Takemura, S.Y., Bharioke, A., Lu, Z., Nern, A., Vitaladevuni, S., Rivlin, P.K., Katz, W.T., Olbris, D.J., Plaza, S.M., Winston, P., et al. (2013). A visual motion detection circuit suggested by Drosophila connectomics. Nature *500*, 175–181.

Tan, L., Zhang, K.X., Pecot, M.Y., Nagarkar-Jaiswal, S., Lee, P.T., Takemura, S.Y., McEwen, J.M., Nern, A., Xu, S., Tadros, W., et al. (2015). Ig Superfamily ligand and receptor pairs expressed in synaptic partners in Drosophila. Cell *163*, 1756–1769.

Tuthill, J.C., Nern, A., Holtz, S.L., Rubin, G.M., and Reiser, M.B. (2013). Contributions of the 12 neuron classes in the fly lamina to motion vision. Neuron *79*, 128–140.

Van Epps, H., Dai, Y., Qi, Y., Goncharov, A., and Jin, Y. (2010). Nuclear pre-mRNA 3′-end processing regulates synapse and axon development in C. elegans. Development *137*, 2237–2250.

Venken, K.J., Simpson, J.H., and Bellen, H.J. (2011). Genetic manipulation of genes and cells in the nervous system of the fruit fly. Neuron *72*, 202–230.

Vilinsky, I., Stewart, B.A., Drummond, J., Robinson, I., and Deitcher, D.L. (2002). A Drosophila SNAP-25 null mutant reveals context-dependent redundancy with SNAP-24 in neurotransmission. Genetics *162*, 259–271.

Wan, H.I., DiAntonio, A., Fetter, R.D., Bergstrom, K., Strauss, R., and Goodman, C.S. (2000). Highwire regulates synaptic growth in Drosophila. Neuron *26*, 313–329.

Yamagata, M., and Sanes, J.R. (2008). Dscam and Sidekick proteins direct lamina-specific synaptic connections in vertebrate retina. Nature *451*, 465–469.

Zaharieva, E., Haussmann, I.U., Bräuer, U., and Soller, M. (2015). Concentration and localization of coexpressed ELAV/Hu proteins control specificity of mRNA processing. Mol. Cell. Biol. *35*, 3104–3115.

Zallen, J.A., Kirch, S.A., and Bargmann, C.I. (1999). Genes required for axon pathfinding and extension in the C. elegans nerve ring. Development *126*, 3679–3692.

Zhu, B., Pennack, J.A., McQuilton, P., Forero, M.G., Mizuguchi, K., Sutcliffe, B., Gu, C.J., Fenton, J.C., and Hidalgo, A. (2008). Drosophila neurotrophins reveal a common mechanism for nervous system formation. PLoS Biol. *6*, e284.

Zlatic, M., Li, F., Strigini, M., Grueber, W., and Bate, M. (2009). Positional cues in the Drosophila nerve cord: semaphorins pattern the dorso-ventral axis. PLoS Biol. *7*, e1000135.

OPEN
ACCESS
CellPress

Cell Reports
Article

Hypofunctional TrkA Accounts for the Absence of Pain Sensitization in the African Naked Mole-Rat

Damir Omerbašić,[1,5,7] Ewan St. J. Smith,[1,2,7] Mirko Moroni,[1] Johanna Homfeld,[1] Ole Eigenbrod,[1] Nigel C. Bennett,[3]
Jane Reznick,[1] Chris G. Faulkes,[4] Matthias Selbach,[5] and Gary R. Lewin[1,6,8,*]
[1]Molecular Physiology of Somatic Sensation, Max Delbrück Center for Molecular Medicine, 13125 Berlin, Germany
[2]Department of Pharmacology, University of Cambridge, Cambridge CB2 1PD, UK
[3]Department of Zoology and Entomology, University of Pretoria, Pretoria, Hatfield 0028, Republic of South Africa
[4]School of Biological and Chemical Sciences, Queen Mary University of London, Mile End Road, London E1 4NS, UK
[5]Proteome Dynamics Group, Max Delbrück Center for Molecular Medicine, 13125 Berlin, Germany
[6]Excellence Cluster Neurocure, Charité Universitätsmedizin Berlin, 10117 Berlin, Germany
[7]Co-first author
[8]Lead Contact
*Correspondence: glewin@mdc-berlin.de

http://dx.doi.org/10.1016/j.celrep.2016.09.035

SUMMARY

The naked mole-rat is a subterranean rodent lacking several pain behaviors found in humans, rats, and mice. For example, nerve growth factor (NGF), an important mediator of pain sensitization, fails to produce thermal hyperalgesia in naked mole-rats. The sensitization of capsaicin-sensitive TRPV1 ion channels is necessary for NGF-induced hyperalgesia, but naked mole-rats have fully functional TRPV1 channels. We show that exposing isolated naked mole-rat nociceptors to NGF does not sensitize TRPV1. However, the naked mole-rat NGF receptor TrkA displays a reduced ability to engage signal transduction pathways that sensitize TRPV1. Between one- and three-amino-acid substitutions in the kinase domain of the naked mole-rat TrkA are sufficient to render the receptor hypofunctional, and this is associated with the absence of heat hyperalgesia. Our data suggest that evolution has selected for a TrkA variant that abolishes a robust nociceptive behavior in this species but is still compatible with species fitness.

INTRODUCTION

Inflammation and tissue injury cause hypersensitivity of the affected tissue so that mild mechanical and thermal stimuli become painful. This phenomenon is called hyperalgesia (Lewin et al., 2014; Smith and Lewin, 2009). A critical endogenous mediator of inflammatory thermal and mechanical hyperalgesia, both in rodents and in humans, is nerve growth factor (NGF). Early studies in rodents and humans revealed that a single local dose of exogenous recombinant NGF can produce profound and long-lasting thermal and mechanical hyperalgesia (Dyck et al., 1997; Lewin and Mendell, 1993; Lewin et al., 1993; Petty et al., 1994). Furthermore, loss of function mutations in the *NGF* gene or *NTRK1*, which encodes the high-affinity NGF receptor TrkA, cause a range of congenital pain insensitivity syndromes in humans (Carvalho et al., 2011; Einarsdottir et al., 2004; Indo et al., 1996). The NGF/TrkA signaling system is critical for the genesis and maintenance of hypersensitivity states in mammals (Lewin et al., 1994, 2014; Woolf et al., 1994). The importance of increased NGF signaling during pain has recently been reinforced by the fact that blocking NGF signaling appears to be highly effective in treating pain in humans on the basis of phase 2 clinical trial data (Katz et al., 2011; Lane et al., 2010).

The naked mole-rat (*Heterocephalus glaber*) is a eusocial African rodent that displays a range of extreme physiological characteristics from cancer resistance and extreme longevity to complete insensitivity to acid (Liang et al., 2010; O'Connor et al., 2002; Park et al., 2008; Smith et al., 2011; Schuhmacher et al., 2015). We discovered that this species completely lacks behavioral heat hyperalgesia when challenged with NGF and the pro-inflammatory agents capsaicin and complete Freund's adjuvant (Park et al., 2008). The polymodal, capsaicin-gated ion channel TRPV1 is also required in mice for the development of NGF-induced heat hyperalgesia (Chuang et al., 2001). However, our studies have shown that although naked mole-rats are behaviorally insensitive to capsaicin, they have sensory neurons that express a TRPV1 channel with ligand sensitivity and biophysical properties indistinguishable from that found in mice or humans (Smith et al., 2011). Here, we investigated how heat hyperalgesia has been disabled in the naked mole-rat over the course of evolution. We addressed this question using molecular and cellular approaches to dissect out at which stage of the sensitization pathway heat sensitization fails. A cellular model of heat hyperalgesia is the rapid and potent sensitization of TRPV1 currents that has been studied in isolated sensory neurons (Shu and Mendell, 2001). We show that rapid sensitization of TRPV1-mediated currents is absent in sensory neurons from naked mole-rats. However, the naked mole-rat TRPV1 protein can be sensitized when expressed in mouse sensory neurons. We show that the cloned naked mole-rat TrkA receptor is less efficient at engaging signal transduction pathways leading to TRPV1 sensitization. Furthermore, we demonstrate that unique amino acid variants in the kinase domain of the naked mole-rat

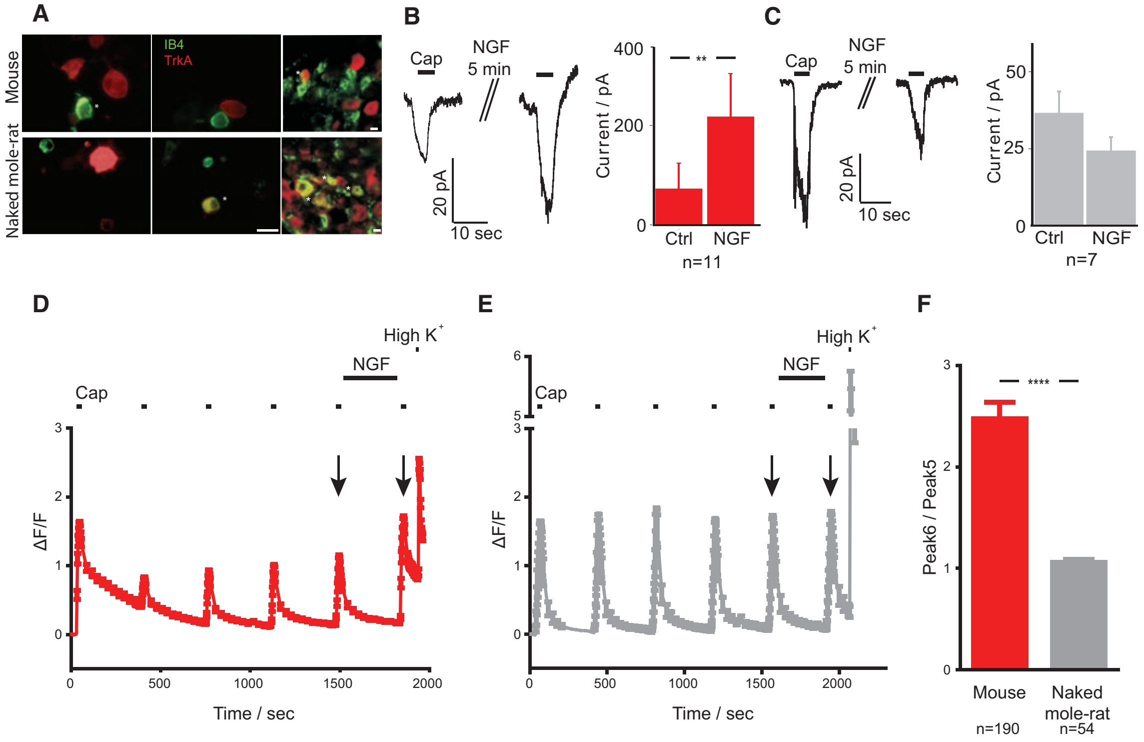


Figure 1. Naked Mole-Rat Dorsal Root Ganglia Neurons Are Not Sensitized by NGF

(A) IB4 (green) and TrkA (red) label largely different populations of mouse DRG neurons, greater co-labeling occurs in naked mole-rat TrkA-positive DRG neurons in culture (left panels) and sections (right panels). Asterisks denote double-labeled neurons; scale bar, 20 μm.

(B and C) NGF potentiates capsaicin-gated currents in mouse DRG neurons (B) but has no effect in naked mole-rat neurons (C).

(D and E) NGF superfusion causes sensitization of mouse DRG neurons, observed as increase in calcium influx (D), but not in naked mole-rat DRG neurons (E); high-potassium solution (40 mM KCl) was used to verify cell viability.

(F) Ratios of sixth and fifth capsaicin response from (D) and (E), as labeled by arrows.

Mann-Whitney *U* test was used in (B), (C), and (F) (**p < 0.01; ****p < 0.0001). Data are presented as mean ± SEM.

TrkA receptor likely render the receptor hypofunctional. Thus, millions of years of evolution appear to have led to an efficient and possibly single-molecule change that disables heat hyperalgesia.

RESULTS

TRPV1 Is Not Sensitized by NGF in Naked Mole-Rat Nociceptors

We made whole-cell patch-clamp recordings from isolated mouse and naked mole-rat sensory neurons live-labeled with fluorescently tagged isolectin B4 (IB4). IB4 predominantly binds to non-peptidergic small-diameter sensory neurons in mice, while TrkA immunoreactivity is specific to peptidergic sensory neurons that do not bind IB4 (Averill et al., 1995). Immunohistochemistry confirmed that IB4-negative sensory neurons are TrkA positive in mouse and naked mole-rat, but as in the rat (Price and Flores, 2007), some naked mole-rat TrkA-positive neurons were IB4 positive (Figure 1A). We also immunostained cultured naked mole-rat sensory neurons and found that 50% (34/68 cells) were TrkA positive while 35% (24/68) were IB4 positive, and only a small proportion of TrkA-positive cells were IB4 positive (15% [5/34]). We thus focused our analysis on IB4-negative neurons to increase the likelihood of recording from naked mole-rat sensory neurons that possess TrkA receptors. The ability of NGF to rapidly sensitize TRPV1 was measured by comparing capsaicin-evoked current amplitudes before and after a 5-min NGF superfusion (100 ng/mL). As in rat sensory neurons (Shu and Mendell, 1999), there was a substantial increase in the average size of the capsaicin-evoked current (>2-fold) after acute NGF treatment of IB4-negative mouse sensory neurons (Figure 1B). However, in naked mole-rat IB4-negative sensory neurons, NGF never sensitized TRPV1 currents (Figure 1C). In order to confirm these results we also performed calcium imaging on isolated mouse and naked mole-rat sensory neurons. Given that both capsaicin and calcium influx cause desensitization and tachyphylaxis of TRPV1 (Koplas et al., 1997; Lishko et al., 2007), we applied five consecutive pulses (100 nM capsaicin, 30 s pulse) in order to obtain stable calcium signals before exposing the neurons to NGF (Hanack et al., 2015), followed by the sixth capsaicin pulse (Figures 1D and 1E). In mouse sensory neurons, NGF caused robust sensitization of capsaicin responses, but no increase in calcium influx was observed in naked mole-rat sensory neurons (Figure 1F). Thus, the absence of behavioral signs of NGF-induced heat hyperalgesia in the naked mole-rat (Park et al., 2008) can be accounted for by molecular changes, intrinsic

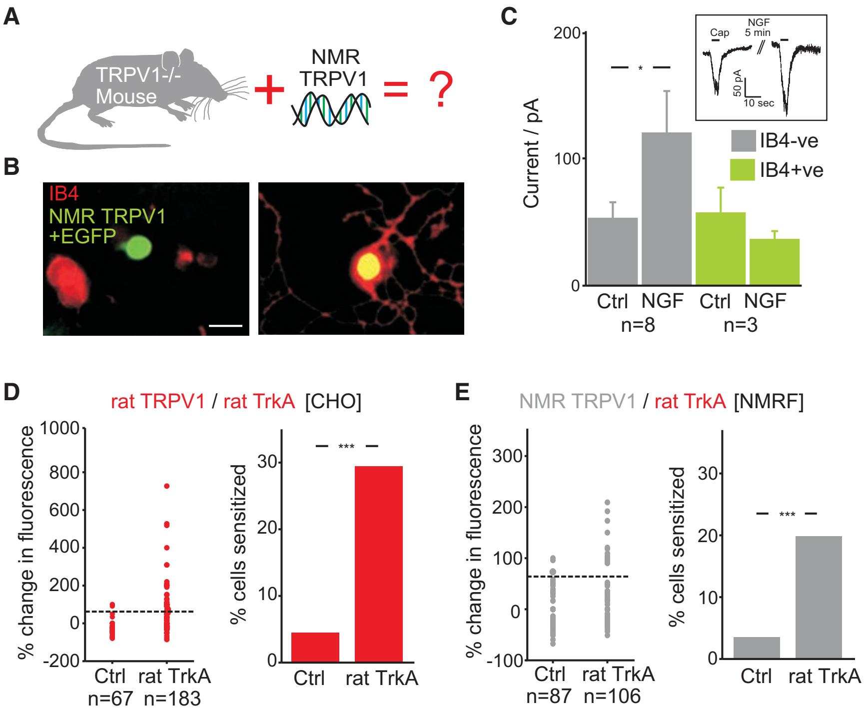


Figure 2. Naked Mole-Rat TRPV1 Currents Can Be Sensitized by NGF

(A) Naked mole-rat *Trpv1* cDNA was transfected into DRG neurons originating from *Trpv1*$^{-/-}$ mice.

(B) *Trpv1*$^{-/-}$ DRG neurons expressing naked mole-rat TRPV1 channels were identified by co-transfection with *EGFP*; IB4-568 labeling allowed targeting of TrkA-positive neurons. Scale bar, 50 μm.

(C) NGF potentiates naked mole-rat TRPV1-mediated capsaicin currents in IB4-negative, but not IB4-positive, *Trpv1*$^{-/-}$ DRG neurons.

(D) In CHO cells co-expressing rat *Trpv1*/rat *TrkA*, NGF sensitized capsaicin responses, unlike in control cells.

(E) Naked mole-rat fibroblast cells expressing naked mole-rat *Trpv1*/rat*TrkA* were sensitized by NGF when compared to controls.

Sensitization in (D) and (E) was scored if change in $[Ca^{2+}]_i$ intensity > (mean + 2 SD) of controls (dotted lines in lefthand panels). Mann-Whitney *U* test was used in (C) and chi-square test in (D) and (E) (*$p < 0.05$; ***$p < 0.001$). Data in (C) are presented as mean ± SEM.

to naked mole-rat nociceptors, that have disabled NGF sensitization of TRPV1.

The cloned naked mole-rat *TRPV1* receptor (*nmrTrpv1*) displays biophysical properties similar to its mouse counterpart with respect to proton, capsaicin, and heat gating (Smith et al., 2011). It is, however, possible that the naked mole-rat TRPV1 protein cannot be phosphorylated on critical residues that are required for full sensitization. Several conserved amino acid residues that can be phosphorylated within the TRPV1 molecule have been shown to be important for sensitization (Bhave et al., 2003; Chuang et al., 2001; Prescott and Julius, 2003; Zhang et al., 2005); however, all but one of these residues were conserved in *nmrTrpv1* (Figure S1A). Thus, Ser502 (numbering for *ratTrpv1*), a normally conserved residue involved in protein kinase C epsilon type (PKCε)-mediated sensitization (Numazaki et al., 2002), was substituted by a threonine in the naked mole-rat protein. By using the phorbol-12-myristate-13-acetate ester (PMA) to activate PKCε in cells transfected with *ratTrpv1*, we observed robust sensitization using calcium imaging as the readout (Figure S1B). To measure PKCε sensitization of *nmrTRPV1*, we used a new naked mole-rat fibroblast cell line (Figure S1C) to enable recording of PKCε-mediated sensitization of *nmrTRPV1* in its native environment, which was robustly observed (Figure S1D). We also generated a naked mole-rat *TRPV1*^{T502S} mutant that was also sensitized by PMA in naked mole-rat fibroblast cell lines (Figure S1E) and conclude that in terms of TRPV1 sensitization, threonine is functionally equivalent to serine at position 502.

To demonstrate more directly that naked mole-rat TRPV1 is fully capable of being sensitized, we expressed it in mouse sensory neurons from *Trpv1*$^{-/-}$ mice. Using an Alexa-Fluor-568-conjugated IB4, mouse *Trpv1*$^{-/-}$ IB4-negative sensory neurons were selected by their green fluorescence after transfection with plasmids encoding EGFP and the naked mole-rat *Trpv1* cDNA (Figures 2A and 2B). We used whole-cell patch-clamp electrophysiology to demonstrate that capsaicin-evoked currents are present in transfected *Trpv1*$^{-/-}$ sensory neurons and that these currents could be sensitized by NGF (Figure 2C). In contrast, no sensitization of the capsaicin current in mouse IB4-positive sensory neurons was observed presumably because of the absence of TrkA in these cells (Figures 2C and 1A). Transfected *Trpv1*$^{-/-}$ sensory neurons had heat-gated currents with an activation threshold of 44.4°C ± 0.7°C (n = 5) and pH-gated currents sensitive to ruthenium red (Figure S1F). Heat-activated currents are reported to be otherwise rare in *Trpv1*$^{-/-}$ sensory neurons (Caterina et al., 2000). Thus, the naked mole-rat TRPV1 protein can rescue capsaicin and heat sensitivity in *Trpv1*$^{-/-}$ sensory neurons with a heat-activation threshold concomitant with the heat-activation threshold of nmrTRPV1 (Smith et al., 2011) and is fully capable of NGF initiated sensitization in the mouse cellular context.

Rapid TRPV1 sensitization via NGF activation of TrkA receptors can be reconstituted in human and animal-derived cell lines as well as *Xenopus laevis* oocytes (Bonnington and McNaughton, 2003; Prescott and Julius, 2003; Zhang et al., 2005). It is possible that TRPV1 sensitization by NGF cannot take place in the naked mole-rat cellular context. We used calcium imaging to measure TRPV1 sensitization after Chinese hamster ovary (CHO) cells were transfected with *ratTrpv1*/*ratTrkA*. In control experiments, *ratTrkA* was either not transfected or buffer instead of NGF was superfused between the capsaicin pulses. NGF produced a robust sensitization with 29.5% of capsaicin-responsive CHO cells showing increased calcium signals post-NGF compared to just 4.5% in controls (Figure 2D). Similarly, using naked mole-rat fibroblast cells co-transfected with *nmrTrpv1/ratTrkA*, 19.8% of capsaicin-responsive naked mole-rat fibroblast cells showed increased calcium signals post-NGF compared to just 3.5% in controls (Figure 2E). These data indicate that naked mole-rat cells possess the necessary signaling components for TRPV1 sensitization.

Naked Mole-Rat TrkA Is Hypofunctional

We cloned the naked mole-rat *TrkA* cDNA from mRNA isolated from sensory neurons (*nmrTrkA*). The *nmrTrkA* sequence was identical to that predicted from the naked mole-rat genome assembly (Keane et al., 2014; Kim et al., 2011). The predicted naked mole-rat TrkA peptide sequence was aligned with orthologous sequences from 26 other mammalian species (Figure S2). There was significant sequence divergence in the extracellular TrkA domains, including the juxtamembrane NGF-binding domain; however, the intracellular sequences within the kinase domain were highly conserved (Figure S2B). All tyrosine residues important for receptor activation were conserved in all the species, including the naked mole-rat. We reasoned that at least some of the amino acid variants in the kinase domain of *nmrTrkA* may be common variants found in African mole-rats (family Bathyergidae). In order to screen for such variants, we obtained TrkA sequences from five further African mole-rat species: the Damaraland mole-rat (*Fukomys damarensis*), the Mashona mole-rat (*Fukomys darlingi*), the giant mole-rat (*Fukomys mechowii*), the Natal mole-rat (*Cryptomys hottentotus natalensis*), and Emin's mole-rat (*Heliophobius emini*) (Figure S3A). We used genomic DNA from these species to PCR amplify the exonic regions of the TrkA gene, guided by variants found in nmrTrkA. However, we also assembled TrkA transcripts from published RNA sequencing (RNA-seq) data from African mole-rat species (Davies et al., 2015). In addition, we obtained RNA from the brains of three Mashona mole-rats and performed RNA-seq followed by de novo transcriptome assembly (Table S1). An African mole-rat phylogeny was constructed including the new transcriptome data from the Mashona mole-rat (Figure S3C), and this was in close agreement with previous analyses that had not included this species (Davies et al., 2015). Alignment of the available predicted TrkA amino acid sequences from African mole-rats revealed that the nmrTrkA kinase domain has accumulated at least three amino acid variants that are either absent or rare in the animal kingdom, including African mole-rats (Figure S3B). There was just one amino acid change that appeared to be unique to naked mole-rat, which was a leucine (rat) to cysteine substitution at position 774 (Figure S2B). The accumulation of amino acid variants in the *nmrTrkA* kinase domain encouraged us to carry out a functional analysis of the ability of this receptor to participate in nociceptor sensitization. To do this, we tested the ability of the naked mole-rat TrkA receptor to sensitize TRPV1 using electrophysiology with *X. laevis* oocytes as the heterologous expression system. Oocytes were injected with a *ratTrpv1* cRNA and cRNAs coding for either *ratTrkA* or *nmrTrkA*. We observed that 1 μM capsaicin causes substantial and long-lasting desensitization of TRPV1 currents in oocytes and thus decided to record proton-gated TRPV1 currents to quantify NGF sensitization, as others have done (Zhang et al., 2005). Using a two-electrode voltage clamp, we showed that an acidic stimulus (pH 5.8) produced robust inward currents in TRPV1-expressing oocytes that were absent in non-injected oocytes (data not shown). In oocytes injected with *ratTrkA* and *ratTrpv1* cRNA, superfusion of NGF (100 ng/mL, 5 min) caused a robust sensitization of acid-gated currents (Figure 3B). However, the same NGF concentration produced a significantly smaller sensitization of TRPV1 currents in oocytes injected with *nmrTrkA* and *ratTrpv1* cRNA (Figures 3B and 3E). Comparable amounts of rat and naked mole-rat TrkA protein were present in membranes isolated from *X. laevis* oocytes (Figure 3C), indicating that differences in TrkA protein levels was unlikely to account for the reduced TRPV1 sensitization. We next varied NGF concentration (1–1,000 ng/mL) but kept the superfusion time constant (5 min). TrkA is a high-affinity NGF receptor with a dissociation constant K_d of less than 10^{-9} M (Kaplan et al., 1991; Klein et al., 1991). When oocytes were stimulated with 1,000 ng/mL NGF, activation of the naked mole-rat TrkA receptor produced a degree of sensitization similar to that observed with rat TrkA (Figure 3C). These results strongly suggest that the naked mole-rat TrkA molecule is less efficient at initiating sensitization with NGF concentrations of ~100 ng/mL, which was shown to be saturating in adult rat sensory neurons (Shu and Mendell, 1999). It is conceivable that recombinant human NGF used in this study (rhNGF) displays stronger binding affinity to rat TrkA than to the naked-mole-rat TrkA. To test this idea, we cloned chimeric TrkA receptors containing the N-terminal, extracellular part of the receptor from rat TrkA together with the transmembrane domain and entire intracellular kinase domain from the naked mole-rat molecule (Figures 3D and 3E). HEK293 cells were transiently transfected with either rat or chimeric *TrkA* construct to assess NGF-stimulated TrkA activation (Figure S4). An antibody raised against extracellular rat TrkA domain was used to measure the total level of TrkA protein in cell lysates (total TrkA), and two antibodies that recognize phosphorylated tyrosine residues in the TrkA kinase domain were employed to study receptor activation. Anti-phospho-TrkA (Tyr674/675; numbering for human TrkA) was used to measure the phosphorylation levels of the activation loop tyrosines (Segal and Greenberg, 1996; Segal et al., 1996), and an anti-phospho-TrkA (Tyr490) was used that recognizes the activated putative Shc binding site (Obermeier et al., 1993a). NGF stimulation triggered rapid phosphorylation of Tyr674/675 in rat TrkA, but not in chimeric TrkA (Figures S4A and S4B). In contrast to rat TrkA, NGF treatment did not have any effect on activation of Tyr674/675 in the chimeric TrkA receptor. However, the Tyr674/675 residues in both chimeric TrkA and rat TrkA displayed strong basal receptor phosphorylation in the absence of NGF, probably triggered by receptor dimerization events due to overexpression. This observation is in agreement with previous findings that an antibody against the TrkA extracellular domain can itself crosslink two receptors, causing their activation in PC12 cells (Clary et al., 1994; Hempstead et al., 1992). NGF triggered increased phosphorylation of the Tyr490 residue in the rat TrkA molecule after 1 min but did not have any apparent effect on the phosphorylation level of the chimeric TrkA Tyr490 residue (Figures S4C and S4D). Next, we tested chimeric TrkA in the context of NGF-mediated TRPV1 sensitization. Proton acid-gated TRPV1 currents in *X. laevis* oocytes co-expressing chimeric TrkA could only be

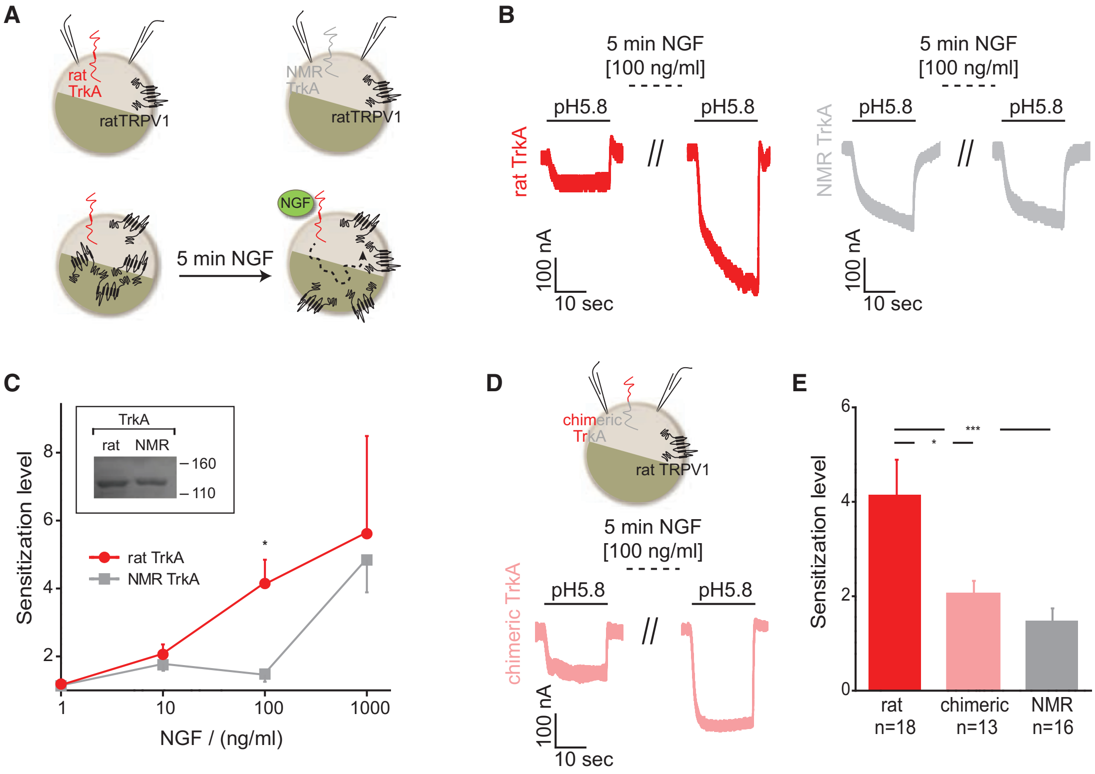


Figure 3. Naked Mole-Rat TrkA Is Impaired in TRPV1 Current Potentiation

(A) Schematic representation of the transfection and recording conditions used.

(B) NGF causes substantial sensitization of proton-gated rat TRPV1 currents when signaling through rat TrkA, but this effect is reduced in the oocytes co-expressing the naked mole-rat TrkA receptor.

(C) Sensitization levels across NGF concentrations were calculated as the ratio of a current immediately after and before NGF superfusion. High NGF concentration rescues the sensitization through NGF TrkA. For all the measurements, at least three pH stimuli were applied before the NGF superfusion in order to obtain the stable current responses, while at least two acid-gated currents were recorded post-NGF application. Between 2 and 18 oocytes were recorded for every NGF concentration. Inset: ten oocytes, injected with equal concentration of rat or naked mole-rat*TrkA* cRNA, were lysed, and pelleted membranes were subjected to western blotting. Prior to blotting, the protein concentration was measured to ensure equal sample loading. TrkA is expressed as a 140-kDa protein.

(D and E) 100 ng/mL NGF potentiated acid-gated TRPV1 currents recorded in *X. laevis* oocytes via chimeric TrkA (D), but the sensitization level was significantly smaller than for rat TrkA and not different from the NMR TrkA receptor quantified in (E).

Two-way ANOVA with Sidak's multiple comparison was used in (C), and one-way ANOVA with Bonferroni's multiple comparison test was used in (E) (*$p < 0.05$; ***$p < 0.001$). Data are presented as mean ± SEM, except in (C) for NMR TrkA at 1 ng/mL NGF (only mean current plotted), where two oocytes were recorded.

moderately sensitized with 100 ng/mL NGF; indeed, the mean level of sensitization observed was not significantly larger than that found with the full-length *nmrTrkA* (Figure 3E). In contrast, sensitization of TRPV1 proton currents by NGF-stimulated oocytes co-expressing *ratTrkA* was at least twice as large as with full-length *nmrTrkA* or chimeric receptors. These results strongly suggest that a hypo-functional naked mole-rat TrkA kinase domain underlies the lack of TRPV1 sensitization in this species.

Quantitative Proteomics Reveals Hypofunctional Downstream Signaling of the nmrTrkA Intracellular Domain

A quantitative proteomics approach was used that combined HEK293 cell stable isotope labeling by amino acids in cell culture (SILAC) (Ong et al., 2002) with high-resolution liquid chromatography coupled with tandem mass spectrometry (LC-MS/MS). Heavy-stable isotope (Lys-8 and Arg-10)-labeled HEK cells transiently expressing rat or chimeric TrkA were stimulated with murine NGF (100 ng/mL) for 10 min and light-stable isotope (Lys-0 and Arg-0) cells were left untreated (Figure 4A) (Olsen et al., 2006). Following stimulation, cells were lysed, and equal amounts of protein were used for each SILAC pair (Figure 4A). Titanium dioxide (TiO_2) chromatography was used for phosphopeptide enrichment, and peptides were analyzed with LC-MS/MS. Typically, phosphopeptides were sequenced several times in different forms (such as oxidized methionine or missed tryptic cleavage), and overall, a similar number of phosphopeptide sites were quantified and identified in cells transfected with rat TrkA or chimeric TrkA (Figure 4B), indicating a similar overall number of phosphorylated proteins in each condition. Class I phosphopeptide sites comprise those residues with the highest localization probability for the phospho-group (>0.75); that is, the sum of probabilities of other potential sites is less than 0.25. From more than 2,000 identified and quantified class I phospho-sites,

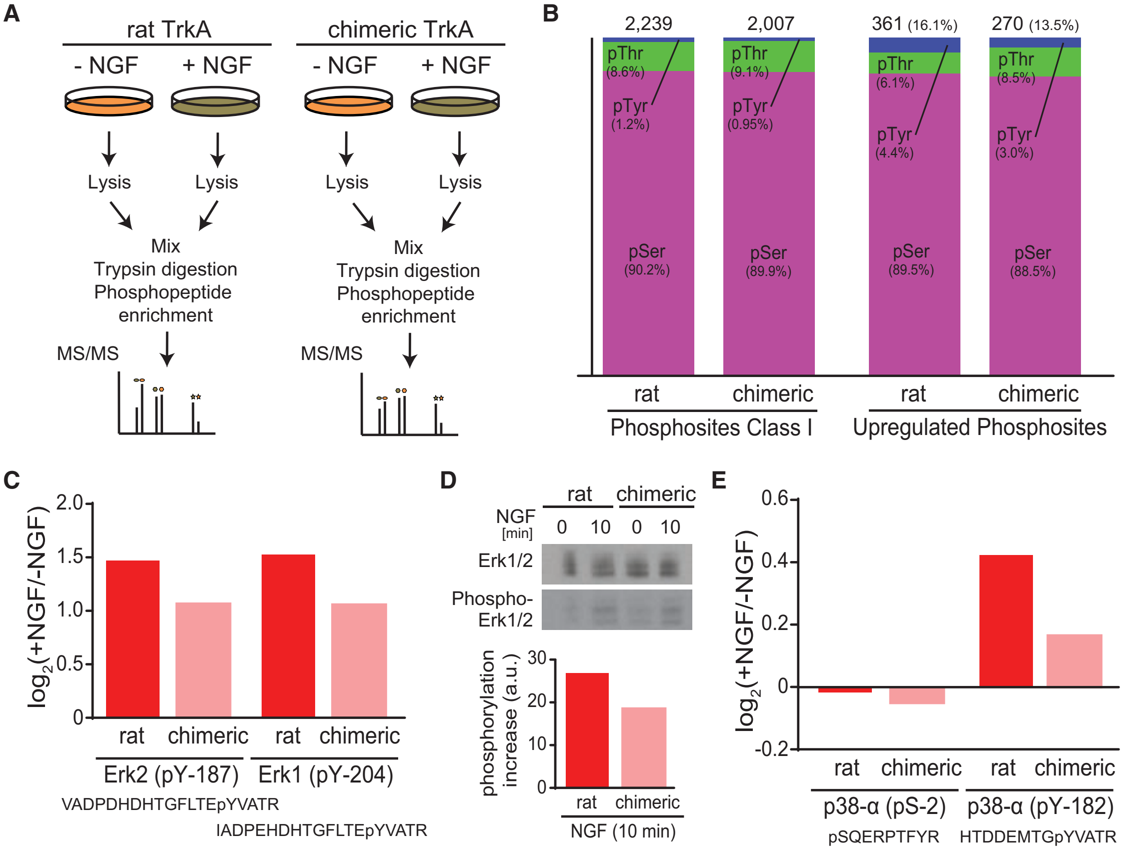


Figure 4. Quantitative Proteomics Reveals Hypofunctional Signaling of the nmrTrkA Intracellular Domain

(A) Overview of the SILAC experiment, performed as a biological duplicate for both receptors. Following the trypsin digestion, samples were enriched for phosphopeptides on TiO_2 columns.

(B) Similar number of phosphopeptides were identified and quantified in both rat and chimeric TrkA samples, but significantly more peptides were upregulated when signaling through rat TrkA than through chimeric TrkA (Fisher's exact test, two-tailed p = 0.0155).

(C and D) Activation of rat TrkA triggers stronger phosphorylation of Erk1/2 key regulatory residues than activation of chimeric TrkA.

(E) Regulatory tyrosine-182 residue of p38-α kinase shows stronger phosphorylation when signaling through rat TrkA, but NGF activation does not change phosphorylation levels of the serine-2 residue.

Data in (C)–(E) are presented as mean values from two experiments.

the distribution of phosphoserine (pSer), phosphothreonine (pThr), and phosphotyrosine (pTyr) sites observed in cells transfected with rat or chimeric TrkA was similar to distributions reported previously with cells stimulated with NGF or epidermal growth factor (EGF) (Emdal et al., 2015; Olsen et al., 2006). We next examined the NGF-upregulated phospho-sites and found that significantly more phosphopeptides were upregulated in NGF-treated cells with rat TrkA (361/2,239 [16.8%]) compared to cells with chimeric TrkA (270/2,007 [13.5%]; Figure 4B).

Analysis of upregulated phospho-site sequence motifs was used to extract over-represented and enriched sequence patterns (Chou and Schwartz, 2011; Schwartz and Gygi, 2005). NGF stimulation was associated with the upregulation of a similar pattern of sequence motifs surrounding pSer residues in cells with rat or chimeric TrkA (Figure S5). However, stimulation of the rat TrkA receptor was associated with a more substantial enrichment of proline-containing motifs compared to chimeric TrkA, which is an indicator of stronger activation of MAPK/CDK protein families (Amanchy et al., 2007).

An additional quantitative proteomics experiment revealed that there was a stronger activation of specific phosphopeptides from Erk2 (MAPK1, pTyr-187) and Erk1 (MAPK3, pTyr-204) (Critton et al., 2008; Sacco et al., 2009) after stimulation of the cells expressing rat TrkA compared to chimeric TrkA (Figure 4C). In addition, western blotting for phosphorylated Erk in HEK293 cells transfected with rat or chimeric TrkA after NGF stimulation revealed reduced levels of phospho-Erk protein after stimulation of the chimeric receptor (Figure 4D). We could also identify and quantify the changes in phosphopeptides from p38-α (MAPK14), a kinase involved in TRPV1 regulation in sensory neurons (Ji et al., 2002; Raingeaud et al., 1995). We observed a stronger increase in the abundance of p38-α derived-phosphopeptides containing the pTyr182 residue after stimulation of rat TrkA compared to chimeric TrkA (Figure 4E). A phosphoserine residue on the same protein (pSer2) (Olsen et al., 2010) did not

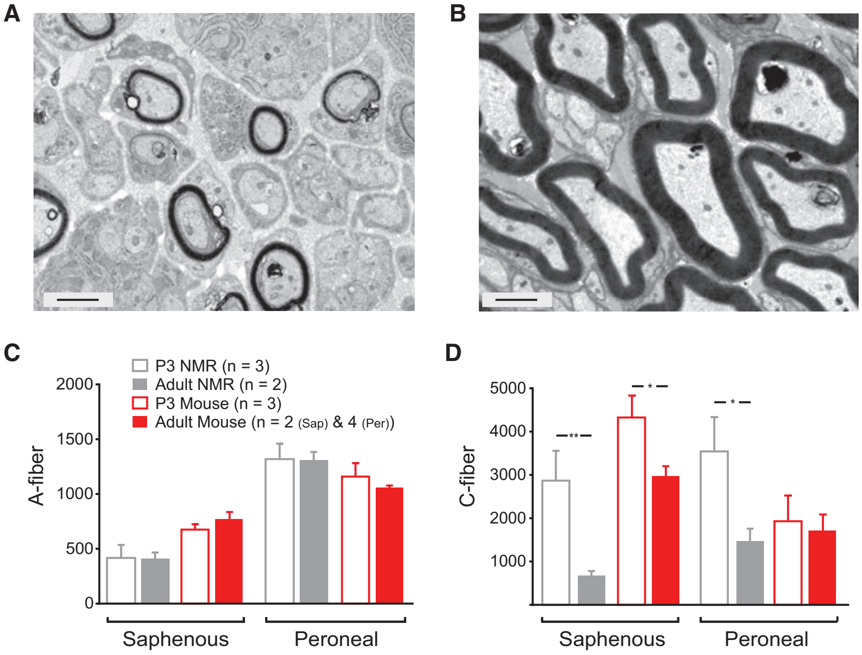


Figure 5. Naked Mole-Rat Pups Have More C-Fibers in Peripheral Nerves than Adults

(A and B) Example electron micrograph of the saphenous nerve of an NMR P3 pup (A) and an adult animal (B). Different myelination stages of single A-fibers and C-fibers within Remak bundles are visible; scale bar, 1 μm.

(C) Numbers of fibers with detectable myelination were comparable for neonatal and adult nerves in both naked mole-rat and mouse.

(D) Quantification of C-fiber number for the saphenous and peroneal nerve in the pup compared to the adult nerve from naked mole-rat and mouse. For comparison, naked mole-rat adult data were taken from St. John Smith et al. (2012) and mouse adult data were taken from Moshourab et al. (2013) and Robertson and Sima (1980) for mouse saphenous and common peroneal nerve, respectively.

Numbers in (C) and (D) indicate the number of animals used for quantification (two nerves per animal). Mann-Whitney *U* test was used (*$p < 0.05$; **$p < 0.01$). Data are presented as mean ± SEM.

show any significant change in either condition after NGF stimulation indicating specificity of NGF-mediated activation (Figure 4E).

Developmental Consequences of Hypofunctional TrkA in the Naked Mole-Rat

NGF-TrkA signaling is essential for the survival of embryonic sensory neurons (Lallemend and Ernfors, 2012; Lewin and Barde, 1996). Adult naked mole-rats have a striking paucity of C-fibers in cutaneous nerves (St John Smith et al., 2012), a feature that is reminiscent of NGF/TrkA loss of function in humans and mice (Crowley et al., 1994; Indo et al., 1996). We thus used transmission electron microscopy to quantify the numbers of myelinated and unmyelinated fibers in peripheral nerves of postnatal day 3 (P3) naked mole-rats and mice (Figures 5A and 5B). We compared the numbers of myelinated (or myelinating) axons in the saphenous and common peroneal nerves in neonates with the published values for adult mice and naked mole-rats using identical methods. We found that the number of unmyelinated C-fibers counted in cross-sections from the purely cutaneous saphenous nerve and the mixed common peroneal nerve from naked mole-rats was between 2- and 3.5-fold higher than the number observed in adult nerves (Figure 5D). However, the number of unmyelinated fibers found in the mouse common peroneal nerve did not change between P3 and adult mole-rats, although there was a small attrition of C-fibers from the saphenous nerve (Figure 5D). In contrast, although the peripheral nerves of P3 naked mole-rats and mice are still undergoing myelination (Figures 5A and 5B), the number of fibers with a myelin sheath (A-fibers) was not different between nerves from the neonate and adult (Figure 5C). These data suggested that there is substantial loss of unmyelinated axons from cutaneous and mixed peripheral nerves of naked mole-rats between P3 and adulthood.

DISCUSSION

We dissected the molecular mechanism that underlies the absence of thermal hyperalgesia in the African naked mole-rat (*H. glaber*) (Park et al., 2008). NGF is central player in the generation of thermal hyperalgesia and acts via its receptor TrkA to initiate hyperalgesia in a TRPV1-dependent manner (Bonnington and McNaughton, 2003; Chuang et al., 2001; Lewin et al., 2014). We have shown that lack of heat hyperalgesia in the naked mole-rat is associated with absence of NGF-induced TRPV1 sensitization in sensory neurons. Our data indicate that the key molecular change in the signal transduction pathway from NGF to hyperalgesia is a unique but minimal sequence change in the naked mole-rat TrkA molecule. We provide evidence that between one and three unique amino acid substitutions within the kinase domain make the naked mole-rat TrkA receptor less efficient at engaging downstream signal transduction, including members of the MAPK family of effectors. Efficient NGF signaling is also a prerequisite for the survival and terminal branching of embryonic sensory neurons in the mouse (Crowley et al., 1994; Patel et al., 2000). Interestingly, a hypofunctional TrkA receptor in the naked mole-rat is associated with a striking paucity of unmyelinated C-fibers in adult peripheral nerves (St John Smith et al., 2012). A comparative anatomical study of six other African mole-rat species (for which TrkA sequences were obtained here) indicated that the C-fiber deficit appears to be unique to naked mole-rats (St John Smith et al., 2012). Even though the Mashona mole-rat (*F. darlingi*) shares two out of the three unique amino acid variants found in TrkA kinase domain of the naked mole-rat receptor (Figure S3B), it does not lack C-fibers (St John Smith et al., 2012). We thus postulate that hypofunctional TrkA signaling in vivo may lead to a loss of C-fibers in naked mole-rats. However, newborn naked mole-rats were found to have many more C-fibers in peripheral nerves than adults. This finding suggests that C-fibers in the naked mole-rat are lost between P3 and adulthood, perhaps as a consequence of hypofunctional TrkA signaling.

Among vertebrate receptors, the TrkA receptor displays the strictest conservation in the intracellular kinase domain (Figure S2). Using chimeric TrkA receptors (rat extracellular/naked mole-rat intracellular), we could show directly that the reduced ability of the naked mole-rat TrkA receptor to sensitize TRPV1 currents is likely localized to the kinase domain (Figure 3). Indeed, biochemical experiments demonstrated a striking reduction in signaling capacity in terms of ligand-dependent tyrosine phosphorylation (Figure S4); however, all the important tyrosine residues in the kinase domain are conserved in the naked mole-rat TrkA molecule (Figure S2). We speculate that insertion of a cysteine for a leucine at position 774 in the naked mole-rat TrkA receptor may alter the efficiency of phosphorylation or recognition of the flanking tyrosine's Tyr751 and Tyr785. Tyrosine 751 has been implicated in binding of the p85 subunit of phosphoinositide 3-kinase (Obermeier et al., 1993b), and Tyr785 serves as a major and selective interaction site for phosphoinositide phospholipase C-γ (Obermeier et al., 1993a). It is of course also possible that accumulated effects of the other variants that are not specific to naked mole-rat TrkA (Figure S3B) contribute to the reduction in receptor signaling we have observed.

Our data strongly suggest that molecular changes in naked mole-rat TrkA molecule alter signal transduction efficiency. Ligand concentration of 100 ng/mL produced almost maximal sensitization of TRPV1 in our oocyte expression system, a similar dose dependence to that found for capsaicin current sensitization in rat sensory neurons (Shu and Mendell, 1999). In contrast, NGF stimulation of chimeric TrkA receptor produced little sensitization of TRPV1 currents at 100 ng/mL but normal sensitization at 1,000 ng/mL (Figure 3C). The maintained efficacy of the naked mole-rat TrkA receptor at very high NGF concentrations is consistent with our previous observation that NGF (500 ng/mL) promotes neurite outgrowth of both mouse and naked mole-rat sensory neurons in culture (Park et al., 2008). However, it is well known that orders-of-magnitude lower concentrations of NGF (<1 ng/mL) are capable of promoting maximal neuronal survival or neurite outgrowth in developing neurons (Davies et al., 1993; Vaillant et al., 2002; Ye et al., 2003). Using a high-resolution quantitative proteomics approach, we found that 10 min after NGF stimulation with 100 ng/mL, there were subtle but significant differences in upregulated phosphopeptides between rat TrkA and a chimeric TrkA containing the naked mole-rat intracellular domain. We obtained evidence of reduced pTyr on peptides belonging to MAPK proteins, including p38α, which has been directly implicated in the sensitization of TRPV1 (Ji et al., 2002).

Surprisingly, naked mole-rat pups do not show the deficit in C-fibers that we had observed in adult animals (Figure 5). It thus appears that the signaling capacity of the naked mole-rat TrkA is sufficient to support the survival and functional development of sensory neurons during embryonic development (Crowley et al., 1994; Lechner et al., 2009). NGF is functionally important for the maintenance of mature sensory neurons (Lewin et al., 2014), but rodents exposed to NGF-function blocking antibodies exhibit death of sympathetic neurons, but probably not sensory neurons (Gorin and Johnson, 1980; Lewin et al., 1992; Ruberti et al., 2000). The concentrations of NGF that robustly sensitize TRPV1 in adult neurons are clearly much higher than those needed to support embryonic survival (see above). It is thus conceivable that the molecular changes in the naked mole-rat TrkA receptor that we describe are more relevant to physiological processes that follow strong receptor stimulation. In this context it is interesting to note that NGF signaling in adult naked mole-rat is still capable of producing mechanical hyperalgesia, a process that does not involve TRPV1 (Lewin et al., 2014). It is possible that hypofunctional TrkA signaling leads to the loss of sensory neurons in naked mole-rats after birth. Naked mole-rats have an extraordinarily long gestation period of ~70 days and can live for up to 32 years (Jarvis, 1991; Sanchez et al., 2015). It is thus feasible that developmental events that occur just after birth, like the loss of TrkA expression in approximately half of the nociceptors (Bennett et al., 1998; Molliver et al., 1997), occur over a more protracted period in the naked mole-rat. In this context, it is important to note that some nociceptors in rats and mice are still dependent on NGF for survival for a few days after birth (Crowley et al., 1994; Lewin et al., 1992). We find that the paucity of C-fibers in cutaneous nerves is correlated with molecular changes in the TrkA receptor associated with reduced signaling (Figure S4). However, it is still possible that effects of other as-yet-unknown gene variants in the naked mole-rat potentiate the effects of the TrkA variants to promote postnatal nociceptor loss.

In summary, we provide evidence that evolution has selected for a single-molecule change in the naked mole-rat NGF receptor TrkA that is sufficient to abolish heat hyperalgesia in this species. Mutations in the *trkA* gene are highly detrimental in humans, but here we show that evolution has selected for sequence change(s) in the naked mole-rat gene that are not only functionally powerful but also compatible with species survival and continued fitness. We speculate that heat hyperalgesia is not an essential phenotypic attribute for the naked mole-rat that is adapted to a subterranean habitat in equatorial East Africa, where temperatures have remained constant for millions of years. Other African mole-rat species have apparently not dispensed with efficient TrkA signaling, and we speculate that one reason for this is that the naked mole-rat is probably the most energetically challenged species in this family (Bennett and Faulkes, 2000). Thus, naked mole-rats can make do with a stripped-down nociceptive system, equipped with fewer C-fibers, that requires less energy but is sufficient for acute nociception and mechanical hyperalgesia following injury (Park et al., 2008). Our study illustrates how evolution can select for mechanistically novel single-molecule changes that exert dramatic phenotypic effects but are compatible with the maintenance of species fitness.

EXPERIMENTAL PROCEDURES

DRG Neuron and Cell Culture

Animal housing, care, and protocols for euthanasia were approved by German federal authorities (State of Berlin). Dorsal root ganglia (DRG) neurons were prepared from both naked mole-rat and mouse as described previously (Park et al., 2008) and plated onto glass coverslips plated with poly-L-lysine (PLL; 200 mg/mL) and laminin (20 μg/mL). CHO and naked mole-rat fibroblast cells were cultured in F12-Ham medium (Life Technologies) and incubated at 37°C in 5% CO_2. For electrophysiology experiments, cells were plated onto PLL-coated plastic dishes and the following day transfected with Lipofectamine (Invitrogen).

Electrophysiology

Recordings from DRG neurons took place after a 10- to 20-min incubation with either IB4-Alexa 488 or IB4-Alexa Fluor 568 (Invitrogen). Whole-cell recordings were made using pipettes (3–6 MΩ resistance) pulled with a Flaming-Brown puller (Sutter Instruments). Extracellular solution contained 140 mM NaCl, 1 mM $MgCl_2$, 2 mM $CaCl_2$, 4 mM KCl, 4 mM glucose, and 10 mM HEPES (pH 7.4) with NaOH. Electrodes were filled with 110 mM KCl, 10 mM NaCl, 1 mM $MgCl_2$, 1 mM EGTA, and 10 mM HEPES (pH 7.3). Solutions were applied and heated using a gravity-driven multi-barrel perfusion system (WAS-02) (Dittert et al., 2006).

X. laevis defoliculated oocytes (stage V or VI) were purchased from EcoCyte Bioscience. Each oocyte was injected in Barth solution using the Nanoject II Auto-Nanoliter Injector (Drummond) with 32.2 nL cRNA mix. Two-electrode voltage-clamp recordings were performed at room temperature 3–5 days after injection using a GeneClamp500B Amplifier, Digidata 1322A, and pClamp 8.0 Software (Axon Instruments). Additional details are available in Supplemental Experimental Procedures.

Molecular Biology and RNA Sequencing

Cloning of naked mole-rat Trpv1 was described before (Smith et al., 2011). In order to clone naked mole-rat TrkA, total RNA was isolated from DRGs with TRIzol (Life Technologies) and dissolved in 30 μL RNase-free water. 1–3 μg total RNA and oligo(dT) and random hexamers (BioTeZ) were used for cDNA synthesis using SuperScript III Reverse Transcriptase (Life Technologies).

To sequence the coding DNA sequence (CDS) for the TrkA intracellular kinase domain of other African mole-rat species, primers specific for NMR TrkA were used to amplify exons 12–17 from species' genomic DNA. Five mole-rats representative of the Bathyergidae family were used: Giant (*F. mechowii*), Damaraland (*Fukomys damarensis*), Mashona (*Fukomys darlingi*), Natal (*Cryptomys hottentotus natalensis*), and Emin's (*Heliophobius emini*) mole-rats. RNA-seq data for *F. darlingi* were generated from three brain samples using paired-end, strand-specific (dUTP) libraries that were sequenced on an Illumina HiSeq2000 platform. The accession number for the annotated transcriptome and the sequencing reads from the Mashona mole-rat reported in this paper is NCBI: PRJNA303968. Additional details are available in Supplemental Experimental Procedures.

Immunocytochemistry and Calcium Imaging

Standard immunohistochemistry and immunocytochemistry protocols on NMR and mouse DRGs were used using an anti-TrkA antibody (kind gift from L.F. Reichardt, UCSF) and IB4-488. Immunofluorescent images were examined with a Leica DM 5000B microscope and MetaVue software (Visitron).

Calcium imaging was conducted as described previously (Milenkovic et al., 2007). Standard Fura-2 ratiometric calcium imaging was conducted to measure responses to capsaicin in CHO and naked mole-rat fibroblast cells transfected with rat TRPV1and NMR TRPV1, respectively, with or without rTrkA. An inverted microscope (Zeiss Observer A1) equipped with the MetaFluor photonics imaging system, including Polychromator V or DG4 (Sutter Instruments), and a CoolSNAP ES camera (Visitron) was used for cell imaging. Additional details are available in Supplemental Experimental Procedures.

MS-Based Protein Quantification Using SILAC

SILAC-labeled HEK293 cells were transfected with TrkA constructs (rat or chimeric) and pEGFP plasmid (5:1) with polyethylenimine. 24 hr after transfection, cells were serum starved and stimulated for 10 min with 100 ng/mL NGF (murine 2.5S, Promega) or left untreated. Equal amounts of protein from each SILAC pair were mixed together. Protein mixtures were reduced with DTT, alkylated with iodoacetamide, pre-digested with Lysyl endopeptidase (LysC, Wako), and subjected to trypsin digestion overnight. Peptides were purified from stop-and-go extraction (STAGE) tips. Phosphopeptide enrichment was performed on 0.5 mg TiO_2 beads. Phosphopeptides were separated on a monolithic column (100 μm inner diameter × 2,000 mm, MonoCap C18 High Resolution 2000 [GL Sciences]; kindly provided by Dr. Yasushi Ishihama [Kyoto University]). The Q Exactive instrument (Thermo Fisher Scientific) was operated in the data-dependent mode, and MaxQuant software was used to identify and quantify proteins. MS/MS spectra were searched using the Andromeda search engine. Additional details are available in Supplemental Experimental Procedures.

ACCESSION NUMBERS

The accession number for the annotated transcriptome and the sequencing reads from the Mashona mole-rat reported in this paper is NCBI: PRJNA303968.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, five figures, and one table and can be found with this article online at http://dx.doi.org/10.1016/j.celrep.2016.09.035.

AUTHOR CONTRIBUTIONS

D.O., E.S.J.S., M.M., and G.R.L. designed experiments; D.O., E.S.J.S., and M.M. performed cloning, electrophysiology, and imaging experiments. J.H. and D.O. performed biochemical studies. D.O. and M.S. designed, analyzed, and performed quantitative proteomic studies of TrkA signaling. O.E. and J.R. isolated mRNA and performed transcriptome assemblies and bioinformatics analysis. D.O., J.R., N.B., and G.R.L. collected Mashona mole-rat samples. C.G.F. donated DNA samples from African mole-rats for TrkA sequence analysis. D.O., E.S.J.S., and G.R.L. wrote the manuscript with input from all authors.

ACKNOWLEDGMENTS

We thank members of the G.R.L. lab for their comments, T.J. Park for helpful discussions, and Djordje Vasiljevic and Koshi Imami for help with mass spectrometry experiments. Anja Wegner and Heike Thränhardt provided excellent technical assistance. We are grateful to Bettina Purfürst for transmission electron microscopy. This work was supported by a European Research Council grant (grant 294678 Extremeophile Mammal) to G.R.L. E.S.J.S. acknowledges support from the Alexander von Humboldt foundation.

Received: January 16, 2016
Revised: June 23, 2016
Accepted: September 13, 2016
Published: October 11, 2016

REFERENCES

Amanchy, R., Periaswamy, B., Mathivanan, S., Reddy, R., Tattikota, S.G., and Pandey, A. (2007). A curated compendium of phosphorylation motifs. Nat. Biotechnol. *25*, 285–286.

Averill, S., McMahon, S.B., Clary, D.O., Reichardt, L.F., and Priestley, J.V. (1995). Immunocytochemical localization of trkA receptors in chemically identified subgroups of adult rat sensory neurons. Eur. J. Neurosci. *7*, 1484–1494.

Bennett, N.C., and Faulkes, C.G. (2000). African Mole-Rats: Ecology and Eusociality (Cambridge University Press).

Bennett, D.L., Michael, G.J., Ramachandran, N., Munson, J.B., Averill, S., Yan, Q., McMahon, S.B., and Priestley, J.V. (1998). A distinct subgroup of small DRG cells express GDNF receptor components and GDNF is protective for these neurons after nerve injury. J. Neurosci. *18*, 3059–3072.

Bhave, G., Hu, H.-J., Glauner, K.S., Zhu, W., Wang, H., Brasier, D.J., Oxford, G.S., and Gereau, R.W., 4th. (2003). Protein kinase C phosphorylation sensitizes but does not activate the capsaicin receptor transient receptor potential vanilloid 1 (TRPV1). Proc. Natl. Acad. Sci. USA *100*, 12480–12485.

Bonnington, J.K., and McNaughton, P.A. (2003). Signalling pathways involved in the sensitisation of mouse nociceptive neurones by nerve growth factor. J. Physiol. *551*, 433–446.

Carvalho, O.P., Thornton, G.K., Hertecant, J., Houlden, H., Nicholas, A.K., Cox, J.J., Rielly, M., Al-Gazali, L., and Woods, C.G. (2011). A novel NGF mutation clarifies the molecular mechanism and extends the phenotypic spectrum of the HSAN5 neuropathy. J. Med. Genet. *48*, 131–135.

Caterina, M.J., Leffler, A., Malmberg, A.B., Martin, W.J., Trafton, J., Petersen-Zeitz, K.R., Koltzenburg, M., Basbaum, A.I., and Julius, D. (2000). Impaired nociception and pain sensation in mice lacking the capsaicin receptor. Science *288*, 306–313.

Chou, M.F., and Schwartz, D. (2011). Biological sequence motif discovery using motif-x. Curr. Protoc. Bioinformatics *Chapter 13*, Unit 13.15–24.

Chuang, H.H., Prescott, E.D., Kong, H., Shields, S., Jordt, S.E., Basbaum, A.I., Chao, M.V., and Julius, D. (2001). Bradykinin and nerve growth factor release the capsaicin receptor from PtdIns(4,5)P2-mediated inhibition. Nature *411*, 957–962.

Clary, D.O., Weskamp, G., Austin, L.R., and Reichardt, L.F. (1994). TrkA cross-linking mimics neuronal responses to nerve growth factor. Mol. Biol. Cell *5*, 549–563.

Critton, D.A., Tortajada, A., Stetson, G., Peti, W., and Page, R. (2008). Structural basis of substrate recognition by hematopoietic tyrosine phosphatase. Biochemistry *47*, 13336–13345.

Crowley, C., Spencer, S.D., Nishimura, M.C., Chen, K.S., Pitts-Meek, S., Armanini, M.P., Ling, L.H., McMahon, S.B., Shelton, D.L., Levinson, A.D., et al. (1994). Mice lacking nerve growth factor display perinatal loss of sensory and sympathetic neurons yet develop basal forebrain cholinergic neurons. Cell *76*, 1001–1011.

Davies, A.M., Lee, K.F., and Jaenisch, R. (1993). p75-deficient trigeminal sensory neurons have an altered response to NGF but not to other neurotrophins. Neuron *11*, 565–574.

Davies, K.T.J., Bennett, N.C., Tsagkogeorga, G., Rossiter, S.J., and Faulkes, C.G. (2015). Family Wide Molecular Adaptations to Underground Life in African Mole-Rats Revealed by Phylogenomic Analysis. Mol. Biol. Evol. *32*, 3089–3107. http://dx.doi.org/10.1093/molbev/msv175.

Dittert, I., Benedikt, J., Vyklický, L., Zimmermann, K., Reeh, P.W., and Vlachová, V. (2006). Improved superfusion technique for rapid cooling or heating of cultured cells under patch-clamp conditions. J. Neurosci. Methods *151*, 178–185.

Dyck, P.J., Peroutka, S., Rask, C., Burton, E., Baker, M.K., Lehman, K.A., Gillen, D.A., Hokanson, J.L., and O'Brien, P.C. (1997). Intradermal recombinant human nerve growth factor induces pressure allodynia and lowered heat-pain threshold in humans. Neurology *48*, 501–505.

Einarsdottir, E., Carlsson, A., Minde, J., Toolanen, G., Svensson, O., Solders, G., Holmgren, G., Holmberg, D., and Holmberg, M. (2004). A mutation in the nerve growth factor beta gene (NGFB) causes loss of pain perception. Hum. Mol. Genet. *13*, 799–805.

Emdal, K.B., Pedersen, A.-K., Bekker-Jensen, D.B., Tsafou, K.P., Horn, H., Lindner, S., Schulte, J.H., Eggert, A., Jensen, L.J., Francavilla, C., and Olsen, J.V. (2015). Temporal proteomics of NGF-TrkA signaling identifies an inhibitory role for the E3 ligase Cbl-b in neuroblastoma cell differentiation. Sci. Signal. *8*, ra40.

Gorin, P.D., and Johnson, E.M.J., Jr. (1980). Effects of exposure to nerve growth factor antibodies on the developing nervous system of the rat: an experimental autoimmune approach. Dev. Biol. *80*, 313–323.

Hanack, C., Moroni, M., Lima, W.C., Wende, H., Kirchner, M., Adelfinger, L., Schrenk-Siemens, K., Tappe-Theodor, A., Wetzel, C., Kuich, P.H., et al. (2015). GABA blocks pathological but not acute TRPV1 pain signals. Cell *160*, 759–770.

Hempstead, B.L., Rabin, S.J., Kaplan, L., Reid, S., Parada, L.F., and Kaplan, D.R. (1992). Overexpression of the trk tyrosine kinase rapidly accelerates nerve growth factor-induced differentiation. Neuron *9*, 883–896.

Indo, Y., Tsuruta, M., Hayashida, Y., Karim, M.A., Ohta, K., Kawano, T., Mitsubuchi, H., Tonoki, H., Awaya, Y., and Matsuda, I. (1996). Mutations in the TRKA/NGF receptor gene in patients with congenital insensitivity to pain with anhidrosis. Nat. Genet. *13*, 485–488.

Jarvis, J. (1991). Reproduction of naked mole-rats. In The Biology of the Naked Mole-Rat, P.W. Sherman, J.U.M. Jarvis, and R.D. Alexander, eds. (Princeton University Press), pp. 384–425.

Ji, R.-R., Samad, T.A., Jin, S.-X., Schmoll, R., and Woolf, C.J. (2002). p38 MAPK activation by NGF in primary sensory neurons after inflammation increases TRPV1 levels and maintains heat hyperalgesia. Neuron *36*, 57–68.

Kaplan, D.R., Hempstead, B.L., Martin-Zanca, D., Chao, M.V., and Parada, L.F. (1991). The trk proto-oncogene product: a signal transducing receptor for nerve growth factor. Science *252*, 554–558.

Katz, N., Borenstein, D.G., Birbara, C., Bramson, C., Nemeth, M.A., Smith, M.D., and Brown, M.T. (2011). Efficacy and safety of tanezumab in the treatment of chronic low back pain. Pain *152*, 2248–2258.

Keane, M., Craig, T., Alföldi, J., Berlin, A.M., Johnson, J., Seluanov, A., Gorbunova, V., Di Palma, F., Lindblad-Toh, K., Church, G.M., and de Magalhães, J.P. (2014). The Naked Mole Rat Genome Resource: facilitating analyses of cancer and longevity-related adaptations. Bioinformatics *30*, 3558–3560.

Kim, E.B., Fang, X., Fushan, A.A., Huang, Z., Lobanov, A.V., Han, L., Marino, S.M., Sun, X., Turanov, A.A., Yang, P., et al. (2011). Genome sequencing reveals insights into physiology and longevity of the naked mole rat. Nature *479*, 223–227.

Klein, R., Jing, S.Q., Nanduri, V., O'Rourke, E., and Barbacid, M. (1991). The trk proto-oncogene encodes a receptor for nerve growth factor. Cell *65*, 189–197.

Koplas, P.A., Rosenberg, R.L., and Oxford, G.S. (1997). The role of calcium in the desensitization of capsaicin responses in rat dorsal root ganglion neurons. J. Neurosci. *17*, 3525–3537.

Lallemend, F., and Ernfors, P. (2012). Molecular interactions underlying the specification of sensory neurons. Trends Neurosci. *35*, 373–381.

Lane, N.E., Schnitzer, T.J., Birbara, C.A., Mokhtarani, M., Shelton, D.L., Smith, M.D., and Brown, M.T. (2010). Tanezumab for the treatment of pain from osteoarthritis of the knee. N. Engl. J. Med. *363*, 1521–1531.

Lechner, S.G., Frenzel, H., Wang, R., and Lewin, G.R. (2009). Developmental waves of mechanosensitivity acquisition in sensory neuron subtypes during embryonic development. EMBO J. *28*, 1479–1491.

Lewin, G.R., and Barde, Y.A. (1996). Physiology of the neurotrophins. Annu. Rev. Neurosci. *19*, 289–317.

Lewin, G.R., and Mendell, L.M. (1993). Nerve growth factor and nociception. Trends Neurosci. *16*, 353–359.

Lewin, G.R., Ritter, A.M., and Mendell, L.M. (1992). On the role of nerve growth factor in the development of myelinated nociceptors. J. Neurosci. *12*, 1896–1905.

Lewin, G.R., Ritter, A.M., and Mendell, L.M. (1993). Nerve growth factor-induced hyperalgesia in the neonatal and adult rat. J. Neurosci. *13*, 2136–2148.

Lewin, G.R., Rueff, A., and Mendell, L.M. (1994). Peripheral and central mechanisms of NGF-induced hyperalgesia. Eur. J. Neurosci. *6*, 1903–1912.

Lewin, G.R., Lechner, S.G., and Smith, E.S.J. (2014). Nerve growth factor and nociception: from experimental embryology to new analgesic therapy. Handbook Exp. Pharmacol. *220*, 251–282.

Liang, S., Mele, J., Wu, Y., Buffenstein, R., and Hornsby, P.J. (2010). Resistance to experimental tumorigenesis in cells of a long-lived mammal, the naked mole-rat (Heterocephalus glaber). Aging Cell *9*, 626–635.

Lishko, P.V., Procko, E., Jin, X., Phelps, C.B., and Gaudet, R. (2007). The ankyrin repeats of TRPV1 bind multiple ligands and modulate channel sensitivity. Neuron *54*, 905–918.

Milenkovic, N., Frahm, C., Gassmann, M., Griffel, C., Erdmann, B., Birchmeier, C., Lewin, G.R., and Garratt, A.N. (2007). Nociceptive tuning by stem cell factor/c-Kit signaling. Neuron *56*, 893–906.

Molliver, D.C., Wright, D.E., Leitner, M.L., Parsadanian, A.S., Doster, K., Wen, D., Yan, Q., and Snider, W.D. (1997). IB4-binding DRG neurons switch from NGF to GDNF dependence in early postnatal life. Neuron *19*, 849–861.

Moshourab, R.A., Wetzel, C., Martinez-Salgado, C., and Lewin, G.R. (2013). Stomatin-domain protein interactions with acid-sensing ion channels modulate nociceptor mechanosensitivity. J. Physiol. *591*, 5555–5574.

Numazaki, M., Tominaga, T., Toyooka, H., and Tominaga, M. (2002). Direct phosphorylation of capsaicin receptor VR1 by protein kinase Cepsilon and identification of two target serine residues. J. Biol. Chem. *277*, 13375–13378.

O'Connor, T.P., Lee, A., Jarvis, J.U.M., and Buffenstein, R. (2002). Prolonged longevity in naked mole-rats: age-related changes in metabolism, body composition and gastrointestinal function. Comp. Biochem. Physiol. A Mol. Integr. Physiol. *133*, 835–842.

Obermeier, A., Lammers, R., Wiesmüller, K.H., Jung, G., Schlessinger, J., and Ullrich, A. (1993a). Identification of Trk binding sites for SHC and phosphatidylinositol 3′-kinase and formation of a multimeric signaling complex. J. Biol. Chem. *268*, 22963–22966.

Obermeier, A., Halfter, H., Wiesmüller, K.H., Jung, G., Schlessinger, J., and Ullrich, A. (1993b). Tyrosine 785 is a major determinant of Trk—substrate interaction. EMBO J *3*, 933–941.

Olsen, J.V., Blagoev, B., Gnad, F., Macek, B., Kumar, C., Mortensen, P., and Mann, M. (2006). Global, in vivo, and site-specific phosphorylation dynamics in signaling networks. Cell *127*, 635–648.

Olsen, J.V., Vermeulen, M., Santamaria, A., Kumar, C., Miller, M.L., Jensen, L.J., Gnad, F., Cox, J., Jensen, T.S., Nigg, E.A., et al. (2010). Quantitative phosphoproteomics reveals widespread full phosphorylation site occupancy during mitosis. Sci. Signal. *3*, ra3.

Ong, S.-E., Blagoev, B., Kratchmarova, I., Kristensen, D.B., Steen, H., Pandey, A., and Mann, M. (2002). Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. Mol. Cell. Proteomics *1*, 376–386.

Park, T.J., Lu, Y., Jüttner, R., Smith, E.S.J., Hu, J., Brand, A., Wetzel, C., Milenkovic, N., Erdmann, B., Heppenstall, P.A., et al. (2008). Selective inflammatory pain insensitivity in the African naked mole-rat (Heterocephalus glaber). PLoS Biol. *6*, e13.

Patel, T.D., Jackman, A., Rice, F.L., Kucera, J., and Snider, W.D. (2000). Development of sensory neurons in the absence of NGF/TrkA signaling in vivo. Neuron *25*, 345–357.

Petty, B.G., Cornblath, D.R., Adornato, B.T., Chaudhry, V., Flexner, C., Wachsman, M., Sinicropi, D., Burton, L.E., and Peroutka, S.J. (1994). The effect of systemically administered recombinant human nerve growth factor in healthy human subjects. Ann. Neurol. *36*, 244–246.

Prescott, E.D., and Julius, D. (2003). A modular PIP2 binding site as a determinant of capsaicin receptor sensitivity. Science *300*, 1284–1288.

Price, T.J., and Flores, C.M. (2007). Critical evaluation of the colocalization between calcitonin gene-related peptide, substance P, transient receptor potential vanilloid subfamily type 1 immunoreactivities, and isolectin B4 binding in primary afferent neurons of the rat and mouse. J. Pain *8*, 263–272.

Raingeaud, J., Gupta, S., Rogers, J.S., Dickens, M., Han, J., Ulevitch, R.J., and Davis, R.J. (1995). Pro-inflammatory cytokines and environmental stress cause p38 mitogen-activated protein kinase activation by dual phosphorylation on tyrosine and threonine. J. Biol. Chem. *270*, 7420–7426.

Robertson, D.M., and Sima, A.A. (1980). Diabetic neuropathy in the mutant mouse [C57BL/ks(db/db)]: a morphometric study. Diabetes *29*, 60–67.

Ruberti, F., Capsoni, S., Comparini, A., Di Daniel, E., Franzot, J., Gonfloni, S., Rossi, G., Berardi, N., and Cattaneo, A. (2000). Phenotypic knockout of nerve growth factor in adult transgenic mice reveals severe deficits in basal forebrain cholinergic neurons, cell death in the spleen, and skeletal muscle dystrophy. J. Neurosci. *20*, 2589–2601.

Sacco, F., Tinti, M., Palma, A., Ferrari, E., Nardozza, A.P., Hooft van Huijsduijnen, R., Takahashi, T., Castagnoli, L., Cesareni, G., van Huijsduijnen, R.H., et al. (2009). Tumor suppressor density-enhanced phosphatase-1 (DEP-1) inhibits the RAS pathway by direct dephosphorylation of ERK1/2 kinases. J Biol Chem. *284*, 22048–22058.

Sanchez, J.R., Milton, S.L., Corbit, K.C., and Buffenstein, R. (2015). Multifactorial processes to slowing the biological clock: Insights from a comparative approach. Exp. Gerontol. *71*, 27–37.

Schuhmacher, L.N., Husson, Z., and Smith, E.S. (2015). The naked mole-rat as an animal model in biomedical research: current perspectives. Open Access Anim. Physiol. *7*, 137.

Schwartz, D., and Gygi, S.P. (2005). An iterative statistical approach to the identification of protein phosphorylation motifs from large-scale data sets. Nat. Biotechnol. *23*, 1391–1398.

Segal, R.A., and Greenberg, M.E. (1996). Intracellular signaling pathways activated by neurotrophic factors. Annu. Rev. Neurosci. *19*, 463–489.

Segal, R.A., Bhattacharyya, A., Rua, L.A., Alberta, J.A., Stephens, R.M., Kaplan, D.R., and Stiles, C.D. (1996). Differential utilization of Trk autophosphorylation sites. J. Biol. Chem. *271*, 20175–20181.

Shu, X., and Mendell, L.M. (1999). Nerve growth factor acutely sensitizes the response of adult rat sensory neurons to capsaicin. Neurosci. Lett. *274*, 159–162.

Shu, X., and Mendell, L.M. (2001). Acute sensitization by NGF of the response of small-diameter sensory neurons to capsaicin. J. Neurophysiol. *86*, 2931–2938.

Smith, E.S.J., and Lewin, G.R. (2009). Nociceptors: a phylogenetic view. J. Comp. Physiol. A Neuroethol. Sens. Neural Behav. Physiol. *195*, 1089–1106.

Smith, E.S.J., Omerbašić, D., Lechner, S.G., Anirudhan, G., Lapatsina, L., and Lewin, G.R. (2011). The molecular basis of acid insensitivity in the African naked mole-rat. Science *334*, 1557–1560.

St John Smith, E., Purfürst, B., Grigoryan, T., Park, T.J., Bennett, N.C., and Lewin, G.R. (2012). Specific paucity of unmyelinated C-fibers in cutaneous peripheral nerves of the African naked-mole rat: comparative analysis using six species of Bathyergidae. J. Comp. Neurol. *520*, 2785–2803.

Vaillant, A.R., Zanassi, P., Walsh, G.S., Aumont, A., Alonso, A., and Miller, F.D. (2002). Signaling mechanisms underlying reversible, activity-dependent dendrite formation. Neuron *34*, 985–998.

Woolf, C.J., Safieh-Garabedian, B., Ma, Q.-P., Crilly, P., and Winter, J. (1994). Nerve growth factor contributes to the generation of inflammatory sensory hypersensitivity. Neuroscience *62*, 327–331.

Ye, H., Kuruvilla, R., Zweifel, L.S., and Ginty, D.D. (2003). Evidence in support of signaling endosome-based retrograde survival of sympathetic neurons. Neuron *39*, 57–68.

Zhang, X., Huang, J., and McNaughton, P.A. (2005). NGF rapidly increases membrane expression of TRPV1 heat-gated ion channels. EMBO J. *24*, 4211–4223.

F · S · T®

FINE SCIENCE TOOLS

Creating a Selection to Fit Your Needs

Scissors · Retractors · Magnifiers · Probes & Hooks · Bone Instruments · Animal Identification
Hemostats · Forceps · Surgical & Laboratory Equipment · Feeding Needles · Spatulae & Spoons
Wound Closure · Surgical Plates · Instrument Care & Sterilization · Rongeurs · Scalpels & Knives
Clamps · Pins & Holders · Needles & Needle Holders · Student Quality Instruments & Much More

FINE SURGICAL INSTRUMENTS FOR RESEARCH™
Visit us at **finescience.com** or call **800 521 2109**